



Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

PHD THESIS

TRINITY COLLEGE DUBLIN

THE UNIVERSITY OF DUBLIN

SCHOOL OF COMPUTER SCIENCE AND STATISTICS

Topological and Perceptual Representations for Reference Frame Retrieval in Hand-Drawn Animation

Author:

Rachael Elizabeth Schwartz

Supervisor:

Dr. Rachel McDonnell

Dr. John Dingliana

Submitted
July 22, 2026

A dissertation submitted in partial fulfillment of the degree of
Doctor in Philosophy

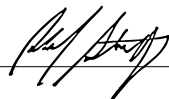
Declaration of Work

I hereby declare that this dissertation is entirely my own work and that it has not been submitted as an exercise for a degree at this or any other university.

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at <http://www.tcd.ie/calendar>.

I have completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at <http://tcd-ie.libguides.com/plagiarism/ready-steady-write>.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Signed:  _____

Date: July 22, 2026

Abstract

Hand-drawn animation is a resolutely human art form, with its unique artistic and commercial appeal emerging directly from the emotion and soul captured within each of the 100,000+ hand-crafted frames of a feature-length film. Remarkably, watching hand-drawn animation invokes the “illusion of life” within the human mind, whereby highly stylized characters formed from collections of simple 2D shapes are somehow perceived as real 3D entities, despite their penchant for physically impossible motion and bodily deformations. Although this perceptual phenomenon has not been widely empirically studied, the small yet vibrant field of hand-drawn animation technology research aims to computationally model its key elements, such that frames’ material collections of 2D lines and shapes are automatically interpreted with respect to the “living” characters they depict.

Despite the extensive and complex technical challenges hand-drawn studios face during animated film production, many of the technical solutions produced by hand-drawn animation technology researchers are rarely used by high-level animation studios. This fact reveals a disconnect between technology researchers’ and professional animators’ respective understandings of technology’s role within the hand-drawn animation process. Our research seeks to gain insights into and begin to resolve this disconnect.

This thesis begins by evaluating the present state of hand-drawn animation technology research through a novel analytical lens. We consider the field subject to its deep theoretical roots in computational topology, aided by artistic and practical insights from hand-drawn animation industry experts. Our analysis indicates that technologies’ artistic advantages and limitations arise directly from their chosen topological approaches, with topological sophistication generally improving artistic functionality. Crucially, advanced topological approaches to hand-drawn animation technology appear to emulate cognitive processes likely underpinning hand-drawn animation perception.

With this in mind, our research pursues the first application of Topological Data Analysis (TDA) to hand-drawn animation technology, focusing on advanced topological frameworks’ potential to model human visual perception of hand-drawn animation. Advised by hand-drawn animation industry experts, we focus on the specific problem of computationally modeling animators’ Reference Frame Quality (RFQ) perception – that is, how good one hand-drawn frame is as drawing reference for another. To this end, we conduct perceptual experiments on professional animators at the hand-drawn studio Cartoon Saloon, each exploring how animators’ specialized visual perception skills practically impact different aspects of RFQ perception.

From our perceptual findings, we construct a novel general-purpose TDA framework for complex system analysis, rigorously defining its mathematical, conceptual, computational, and data analytical facets. To demonstrate our framework’s efficacy, we detail its application to perceptually-based RFQ analysis, illustrating its suitability for modeling hand-drawn frame perception. In future, our framework will underpin a reference frame retrieval system for use within hand-drawn animation studios.

Acknowledgements

Enormous thanks to my primary academic supervisor, Rachel McDonnell, for her patience, true support, kindness, clarity of thought, and willingness to let me take risks. Equally massive thanks to my industry supervisor, Mark Mullery, whose passion, friendship, intelligence, and knowledge guided this project through seemingly impassible hurdles. I am very grateful to my secondary academic supervisor, John Dingliana, for his informed insights and always thoughtful advice, and to Faraz Ahmad who, alongside being a wonderful research collaborator, brought indispensable depth, expertise, and rigor to this project. Finally, endless thanks to this project's first supervisor, Robert Bosch, whose belief in me and my work has always meant the world to me.

This thesis is tremendously indebted to the truly extraordinary support and brilliance of so many people at Cartoon Saloon. Thank you to Catherine Roycroft, Gerry Shirren, Tomm Moore, Nora Twomey, and Paul Young, without whom this project would never have happened. I am deeply grateful for the assistance and skill of Gary Chadwick and Alex O'Toole, who have both done a great deal for this project. Helga Bjarnadóttir, Bonnie Mier, Sandra Andersen, John Walsh, Beth Witchalls, Cal McLoughlin, Ferg Brennan, Steven McClurg, and Anne Bolger have all provided essential knowledge, advice, and support throughout this project, and I am exceptionally thankful for their care and willingness to help. Likewise, I owe so much to every animator who devoted time, effort, and skill towards developing this research. Many thanks to Glen Keane for his contribution to this work, and to Ted Schilowitz for his wonderful guidance, expertise, and friendship over the years. I would also like to thank the Irish Research Council for funding this project in partnership with Cartoon Saloon.

Finally, I am infinitely grateful to Jad, Helen, Paul, Kirsten, Grace, Ngaio, Gina, Sue, Lily, Claire, and, above all else, Mom and Dad.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Research Questions	4
1.3	Methodology	4
1.4	Scope & Limitations	6
1.5	Contributions	7
1.6	Thesis Overview	8
2	Background	9
2.1	Hand-drawn Animation	9
2.1.1	Hand-drawn Animation Production Pipeline	10
2.1.2	Hand-drawn Animation Frames	12
2.2	Hand-drawn Animation Technology	14
2.2.1	Frame Deformation	15
2.2.2	Frame Correspondence	22
2.2.3	Volumetric Modeling	28
2.3	Human Pose Estimation	36
2.4	Visual Perception	39
2.4.1	Human Visual Perception	40
2.4.2	Deep Learning & Visual Perception	42
2.5	Topological Data Analysis	44
2.6	Discussion	46
3	Perceptual Experiments	50
3.1	Perceptual Experiment 1: Local Frame Isometry	51
3.1.1	Variable & Data Selection	52
3.1.2	Method	53
3.1.3	Results	54
3.1.4	Discussion	55
3.2	Perceptual Experiment 2: 3D Body Pose Perception	57

3.2.1	Variable & Data Selection	58
3.2.2	Method	59
3.2.3	Results	61
3.2.4	Discussion	66
3.3	Conclusions	68
4	GENEO Sheaves	70
4.1	Sheaves & Group Equivariant Non-expansive Operators	72
4.1.1	Sheaves	72
4.1.2	Group Equivariant Non-expansive Operators (GENEOs)	74
4.2	GENEO Sheaves	75
4.3	GENEO Sheaf Data Fusion	77
4.3.1	The GENEO Sheaf Laplacian	78
4.3.2	Harmonic Extension for Data Fusion	79
4.4	Heterogeneous Multiresolution Equivariance	81
4.4.1	GENEO Sheaf Morphisms	81
4.4.2	Stalk Symmetries	82
4.4.3	Stalk Permutations	85
4.5	Pseudometrics on GENEO Sheaves	87
4.5.1	Obstruction Analysis	88
4.5.2	Tracking Subsystem Influence	91
4.6	GENEO Sheaves for Reference Frame Quality Analysis	92
4.6.1	GENEO Sheaf Model of RFQ Perception	92
4.6.2	RFQ GENEO Sheaves for Hand-drawn Character Pose Estimation	96
4.6.3	Proposed Industry Pipeline	101
4.7	Discussion	103
4.8	Conclusions	105
5	Conclusions	107
5.1	Contributions	107
5.2	Future Work & Limitations	110
5.3	Conclusions	115
	Bibliography	117
A	RFQ GENEO Sheaf Implementation Results	142
B	Interviews with Hand-drawn Animation Experts	145
B.1	Glen Keane Interview Transcript	145
B.2	Mark Mullery Interview Transcript	150

List of Figures

1.1	Examples of model sheets from Cartoon Saloon’s hand-drawn feature film <i>Wolfwalkers</i> . . .	3
1.2	Multiresolution visual data integration on fine vs. coarse data topologies (left and right, respectively). Figure adapted from [271].	4
1.4	Left: Heterogeneous multiresolution symmetries on hand-drawn animation frames: local rotation (top, darker blue) and near-global reflection (bottom, paler blue). Right: Conceptual diagram of multiresolution data consistency across local-to-global subsystems within a complex system. Figure adapted from [1].	5
1.3	Conceptual illustration of a sheaf, with blue arrows depicting consistency constraints. Figure adapted from [154].	5
2.1	Color model sheets from <i>Wolfwalkers</i> for Digital Ink & Paint.	10
2.2	Hand-drawn animation production pipeline stages for a single Cartoon Saloon <i>Wolfwalkers</i> frame: (a) storyboard, (b) posing, (c) rough animation, (d) clean-up animation, (e) DIP, and (f) compositing. The diagrams on the rough frame provide directions for clean-up animators; the colored lines on the clean-up frame specify technical information for DIP and compositing.	11
2.3	(a) Occluding contours (highlighted red) on a 3D polygon mesh, from [193]. (b-c) Standard line drawing examples from [78, 158].	12
2.4	Examples of consecutive clean-up keyframes frames from <i>Wolfwalkers</i> illustrating complex stroke topology change arising from a physically small articulate movement.	14
2.5	As-Rigid-As-Possible lattice deformation from Sýkora et al. [302]: (a) identifying target vertex location (boxed red), (b) lattice square isometry, (c) centroid of corners (circled red).	15
2.6	Example of transformation lattices assigned to each body part in a walk cycle from Chen et al. [64]. Finer resolution lattices animate both hands.	16
2.7	Example of semi-automatic inbetweening method from Carvalho et al. [51]. Hand-drawn keyframes and color-coded corresponding gesture strokes shown in boxes, with intermediate frames between them.	17
2.8	Example of stroke groups with separate lattices (green and red) merging into a single stroke group with different lattice topology (blue) through the lattice registration method from Even et al. [107]. Registration transforming green strokes into blue strokes is depicted; red strokes undergo an equivalent process.	17

2.9	Mesh deformation approach of Brodt & Bessmelstev [40]: (a) hand-drawn source and target keyframes with 2D skeletons; (b) 2.5D mesh cuts matching articulate self-occlusion discontinuities from the projected smooth 3D deformation field; (c) piecewise-smooth 2.5D deformation field; (d) an output intermediate frame.	18
2.10	Compatible frame interpolation results from Yang et al. [344]: compatible meshes given by strokes (a) and original strokes (b) from input keyframes and a generated intermediate frame; (c) mesh triangle collapse for stroke topology change.	19
2.11	Mesh deformation techniques from Zhu et al. [367]: (a) topological cut operation simulated by vertices and edges initially occupying the same position; (b) edge flip, split, and collapse operations for changing mesh topology.	19
2.12	Examples of Principles of Animation and other hand-drawn animation techniques that are challenging to model computationally: exaggerated motion creating physically impossible poses (circled red), squash-and-stretch (circled green), stylized visual perspective due to preference for symbolic over realistic depictions of character pose (circled blue).	20
2.13	Intermediate frame generation method from Dalstein et al. [87] demonstrating stroke and region collapse across frames (top) and stroke complexes (bottom).	21
2.14	Drawing unseen strokes when using stroke visibility toggles. Figure from Yang et al. [346].	23
2.15	Stroke graph diffusion (indicated by red arrows) from Xing et al. [336] comparing respective source and target stroke neighborhoods (red circles) s'_8 and s_{11} . Diffusion compares increasingly distant stroke neighborhoods (blue circles), capturing multiresolution stroke topology on topologically distinct stroke graphs.	23
2.16	Stroke correspondence (color-coded) failure case of CACAI [344] due to lack of global topology awareness. Insets (outlined blue) show stroke topology imprecisions handled by α -topology but not by methods enforcing exact stroke connectivity like Whited et al. [330]. Figure from Yang et al. [346].	24
2.17	Frame analysis pipeline and stroke correspondence results (5-6, blue arrows) from Yu et al. [353].	25
2.18	Stroke vertex matching between synthetic source and target frames and a 3D character mesh posed to match the source frame from Siyao et al. [292].	25
2.19	Results from Mo et al. [225] illustrating stability of strokes vs. regions over articulate motion. Top: input keyframes with corresponding strokes numbered and differences indicated by arrows. Bottom: color-coded many-to-one stroke and one-to-one region correspondence, with dotted lines showing strokes missing (indicated by arrows) from the output target frame.	26
2.20	RAG and correspondence graph construction from Liu et al. [198].	27
2.21	Approach of Zhu et al. [365]. Top: network flow on correspondence graph. Bottom: consecutive frames matching the network flow.	27
2.22	Results of Countinho et al. [82] positioned according to a pitch-yaw parametric viewpoint space adapted from Rivers et al. [269].	29
2.23	Example of view-dependent geometry in a hand-drawn animated character design, sourced from <i>Wolfwalkers</i> . The helmet face guard has a center bar in the side profile design (right), but not in the front (left) and partial turn (center) designs.	29
2.24	Stroke depth processing pipeline of Entem et al. [105] recursively partitioning strokes into depth-equivalent stroke groups via medial axis and Gestalt cue comprehension.	30

2.25	(a) Half-edge graph modeling self-overlap without internal hole [105]. Half-edge graph failure cases (b-d): (b) cyclically-overlapping regions [105], (c) region (soda can label) whose visibility changes with view angle [169], (d) self-overlap with internal hole, successfully modeled by subdividing strokes (blue edge) [86].	31
2.26	3D modeling results by Bessmeltsev et al. [24]: (a) hand-drawn 2D skeleton; (b) sketched and (c) colored 2D frame; (d) input 3D skeleton and strokes extracted from colored 2D frame; (e) output 3D mesh character model.	32
2.27	Inflation with different types of stroke graph depth comprehension. Top: Topologically “stitching” together components related by depth (in)equality from Dvorožňák et al. [98]. Bottom: Fusing components modeled with standard stroke depth graphs by manipulating component shape from Cordier et al. [81].	33
2.28	Results from Ma et al. [208]: input 2D character drawings (red boxes), 2D component meshes and skeletons embedded in 3D space (blue boxes), and output 3D meshes and skeletons (green boxes).	34
2.29	Different types of construction curves: (a) sparse cross-hatching and modeling results from Li et al. [189]; (b) example of construction curves (left) vs. detail (right) from Noris et al. [239]; (c) construction curves in professional hand-drawn animated character design, presented with permission from Cartoon Saloon.	35
2.30	Construction curves from Zhang et al. [356]. Top: input user-drawn construction curves. Bottom: local component symmetry planes for the input pose (left) and a novel generated pose (right).	36
2.31	DNN-based inbetweening results from (a) [292], (b) [162], and (c) [366].	37
2.32	Dynamic topology-based body model from [262], capturing ground-truth body topology (left) and superficial topology changes over articulate motion and viewpoint change (right).	38
2.33	Human pose estimation methods for hand-drawn animation frames: (a) a method designed to automatically animate children’s drawings [293], (b) Sketch2Pose [39] with accompanying SMPL body models [204], (c) HRNet [300] trained on Human-Art [174], (d) Chen & Zwicker’s transfer learning framework [67]. Tests performed in collaboration with Mohamed Fazil Tajudeen.	39
2.34	Depiction of an impossible object, demonstrating how impossibility perception arises from multiresolution feature integration. Local patches of material features (color-coded) depict physically possible object components, becoming impossible only when perceived collectively.	40
2.35	Figures from [11]: Top: original shape; middle: globally scrambled shape, impairing human recognition but not DNN object classification; bottom: local contour perturbations, impairing DNN classification but not human recognition.	42
2.36	Stimuli from [166]. Left: All figures share material features, but due to 3D shape comprehension, humans perceive the figures in (a) as less similar than the respective figures in (b) or (c); DNNs treated figures (a), (b), and (c) as equally similar. Right: Humans perceive figures depicting consistent surface-pattern curvatures (e) as more similar than figures with inconsistent surface-pattern curvatures (d); DNNs do not appear to exhibit this property, indicating that they cannot detect surface-pattern curvature inconsistency.	43
2.37	Illustration of the boundary and coboundary operators across topological spaces of descending and ascending dimension, respectively.	45

3.1	Experiment frame pairs demonstrating how each transformation variable affects geometric and topological cues, with RAGs superimposed for instructional purposes (not shown in experiment). Top = (high NIT, low Reflection, low Rotation). Middle = (low NIT, high Reflection, low Rotation). Bottom = (low NIT, low Reflection, high Rotation). RAG topology is preserved only in Middle and Bottom cases.	52
3.2	Experiment frame pairs with superimposed skeleton poses (in blue), demonstrating that physically similar poses can provide poor drawing reference (top; average RFQ rating = 2.09) and physically different poses can provide good drawing reference (bottom; average RFQ rating = 5.0).	53
3.3	Example experiment question (average RFQ rating = 4.5).	54
3.4	Significant Interactions between (left) NIT and Reflection, (center) NIT and Rotation, and (right) Reflection and Rotation. The Y-axis shows average RFQ ratings across participants on a 1-7 scale slider, with 1 being “Not at all helpful” and 7 being “Extremely helpful.” . .	55
3.5	Posable wooden maquette used for drawing reference by professional animators.	59
3.6	Blender instruction sheet given to participants. Instructions are superimposed over the Blender file containing the skeleton rig and viewport setup used during the experiment. .	60
3.7	Average per-frame Difficulty ratings (top left), Distortion ratings (top right), 3D pose Deviation (bottom left), and 2D pose Deviation (bottom right). Error bars: 95% CI. . . .	62
3.8	Experiment frames (top row) and participants’ pose estimates for each frame, shown from front-view (2 nd row from top), side-view (3 rd row from top), and aerial-view (bottom row). Pose estimates are color-coded by participant, with average pose shown in black. Pose estimates’ heads, necks, hands, and feet are not shown since participants were not required to pose these bones. Continued in Figure 3.9.	64
3.9	Experiment frames (top row) and participants’ pose estimates for each frame, shown from front-view (2 nd row from top), side-view (3 rd row from top), and aerial-view (bottom row). Pose estimates are color-coded by participant, with average pose shown in black. Pose estimates’ heads, necks, hands, and feet are not shown since participants were not required to pose these bones. Continued from Figure 3.8.	65
4.1	Conceptual diagram of heterogeneous multiresolution symmetries (top) acting over multi-scale heterogeneous subsystems (red, green, blue) vs. homogeneous symmetries (bottom) exclusively acting either locally (purple-outlined vertices) or globally (purple overlay). . .	71
4.2	Illustration of a GENEIO sheaf $\mathbf{F}$ on graph Γ (Definition 4.2.1) at adjacent $v, u \in V(\Gamma)$ sharing edge $e \in E(\Gamma)$. Mathematical notations are given alongside conceptual descriptions for expository clarity.	76
4.3	Illustration of the GENEIO sheaf Laplacian (Definition 4.3.1) at vertex $v \in V(\Gamma)$, showing the cycle of data exchange and integration across vertex and edge stalks at each diffusion step.	79

4.4	GENEO sheaf data fusion. (a) Initial admissible 0-assignment comprising input raw, real-world sensor observations (blue nodes) and model sensor states (red nodes). (b) $\mathbf{F}$ during data fusion, with model sensors (pink and purple nodes) evolving to match fixed real-world sensor observations φ_A over subgraph A (blue nodes). (c) $\mathbf{F}$ after data fusion, showing the $H_{\mathbf{F}}$ -harmonic extension of real-world sensor observations (nodes outlined dark blue) to the model system. Obstructions to global consistency are recorded in the model system (lavender nodes).	80
4.5	Illustration of Theorem 4.4.8, stating that GENE0 sheaf data fusion is stalk symmetry-equivariant. Note that $\varphi'_v = \varphi_v \circ g_{v,e} \circ a_{v,e}$ in the theorem statement.	85
4.6	Illustration of Theorem 4.4.12, stating that GENE0 sheaf data fusion is stalk permutation-equivariant.	86
4.7	Bidirectional visual data exchange between lower-level (purple) and higher-level (blue and pink) brain areas. Figure from [358].	93
4.8	Multiresolution integration and material feature composability over increasingly coarse resolutions of visual field. Figure from [268].	94
4.9	RFQ GENE0 sheaf left and right arm pose estimates (bottom) for a hand-drawn frame (top).	97
4.10	Diagram of our RFQ GENE0 sheaf implementation, addressing the problem of 3D (arm) pose estimation on 2D hand-drawn animation frames.	98
4.11	Example 3D arm pose estimates from our RFQ GENE0 sheaf implementation, shown alongside input pairs of synthetic upper and lower arm regions, as well as the <i>Wolfwalkers</i> frames on which they are based. Complete results are available in Appendix A.	99
4.12	Kanizsa triangle illusion.	100
4.13	Results of our RFQ GENE0 sheaf implementation indicating that stalk symmetries (left column) and stalk permutations (right column) computationally behave as anticipated by Theorems 4.4.8 and 4.4.12.	100

Chapter 1

Introduction

1.1 Motivation

Transformative technological advancements of recent decades are deeply connected to progress in animated filmmaking. Seminal graphics and human-computer interaction technologies underpinning our modern digital world emanated from the genesis of 3D animated filmmaking in the late 20th century, and synergy between animation technology research and broader technological development has persisted ever since. Animation technology research investigates fundamental questions at the intersection of art, psychology, mathematics, computer science, and our understanding of the natural world, often yielding cutting-edge advancements in both artistic expression and scientific knowledge. The 3D animation industry in particular has gained immense fiscal and cultural prosperity by proffering unprecedented artistic spectacles grounded in exciting new technologies. 3D animation’s technical foundations fluently meld with lucrative emerging entertainment sectors like gaming, virtual production, and augmented reality, securing 3D studios’ continued prosperity in coming decades.

Despite these many successes arising from reformulating hand-drawn animation for the digital age, hand-drawn animation itself has not benefited from 3D animation’s extraordinary technological progress. Hand-drawn films and TV shows are made with a semi-digitized version of the original Disney production pipeline devised in the 1930s, lacking the technological developments integral to 3D animation’s present eminence. While hand-drawn animation technology research is an active field, many of its technological contributions to the art form are not used in industry settings. This sharply contrasts the present state of 3D animation, where studios have prominent in-house research departments developing proprietary software tailored to animators’ artistic and technical needs. These departments also routinely collaborate with technology giants like Google, Adobe, and NVIDIA on research projects with broader applications. Audiences’ continued interest in hand-drawn animation is demonstrated by immense successes of 2D animated TV shows like *Avatar: The Last Airbender*, *Bob’s Burgers*, and *Rick and Morty*, alongside those of 3D animation incorporating hand-drawn techniques like *Spiderman: Into the Spider-Verse*, *Arcane*, and *KPop Demon Hunters*. This raises the possibility that 3D animation’s present dominance over hand-drawn animation is linked to the industries’ divergent approaches to technological advancement. Most concretely, hand-drawn studios’ technological brittleness will likely hinder synergy with profitable emerging entertainment technology markets in coming decades.

An essential technological difference between 3D and hand-drawn animation is that 3D animated characters are material digital entities, whereas hand-drawn characters exclusively exist as collections of pixels or vector strokes. This disparity has profound ramifications for both industries. 3D models and

rigs (i.e., digital skeletons used to animate characters) hard-code 3D characters’ visual and functional qualities, enabling large animation teams to work from the same ground-truth designs. These qualities can also be accessed and altered programmatically, supporting extensive artistic and technical customization of 3D animation pipelines. Meanwhile, hand-drawn characters wholly emerge from animation teams’ collective understanding and consistent renderings of characters’ visual and functional qualities – in other words, characters are primarily designed as the film is made. Consequentially, animators and film directors solely interact with their characters by watching the film or manually sifting through image files comprising the film’s 100,000+ frames, contrasting the dynamic interactions possible in 3D pipelines. Hand-drawn characters’ lack of ground-truth representations precludes 3D-style programmatic interventions and dynamic interactions. However, it is the linchpin of their unique artistic appeal and cannot be jettisoned in favor of technological ease. For hand-drawn studios to enable such interventions and interactions, they must adopt computer vision strategies for extracting and evaluating visual and functional qualities of raw hand-drawn frames.

Computer vision has made phenomenal progress in recent years, largely due to advances in deep neural networks (DNNs). Yet such techniques rely heavily on the dense color and texture data inherent to real-world photo and video inputs, often faltering on data-sparse inputs like uncolored hand-drawn animation frames. DNNs also appear insensitive to visual features governing human perception of hand-drawn frames, limiting their ability to provide hand-drawn studios with viable, perceptually-based computer vision frameworks and artistic technologies. Although extensive and highly-specialized training may resolve these issues, most hand-drawn studios lack the in-house technical knowledge needed to train DNNs. Moreover, training may be needed for each new production to address DNNs’ poor generalization across hand-drawn films’ diverse artistic styles. This could render DNN-based hand-drawn animation technologies unusable until late into production, when sufficiently many training frames have been created.

Novel computer vision approaches are therefore critical for advancing hand-drawn studios’ technical infrastructures. Such approaches should aim to optimally leverage hand-drawn frames’ sparse visual data in ways compatible with professional animators’ frame perception, enabling dynamic and customizable access to hand-drawn production data. Then, understanding hand-drawn frames from both technical and perceptual perspectives is essential to devising such approaches. Despite hand-drawn animation technology research’s limited industry uptake, the field offers profound insights into hand-drawn frames’ technical qualities. Strikingly, computational topology pervades hand-drawn animation technology research, with the vast majority of methods employing some form of topological approach. This is no accident – computational topology and hand-drawn animation share deep parallels, with topological techniques frequently capturing fundamental properties of and relationships between hand-drawn frames. Crucially, research indicates that topological techniques are also uniquely adept at capturing frame features vital to hand-drawn frame perception. Although existing techniques’ topological simplicity and consequential artistic limitations can deter industry usage, their mathematical foundations are profoundly enhanced by the emerging field of Topological Data Analysis (TDA). Adapting previously pure topological constructs to accessible computational settings, TDA offers sophisticated mathematical approaches to data modeling and analysis. These advanced topological ideas explainably capture data relationships inaccessible by other analytical approaches, presenting an ideal setting for developing hand-drawn animation technologies tailored to studios’ artistic and technical needs.

This thesis takes initial steps towards integrating TDA computer vision techniques into hand-drawn animation technology research. Specifically, we address the problem of reference frame retrieval during hand-drawn animated film production. Because hand-drawn characters lack ground-truth body representations, animators continually *resolve* poses as they work – that is, they determine how to best

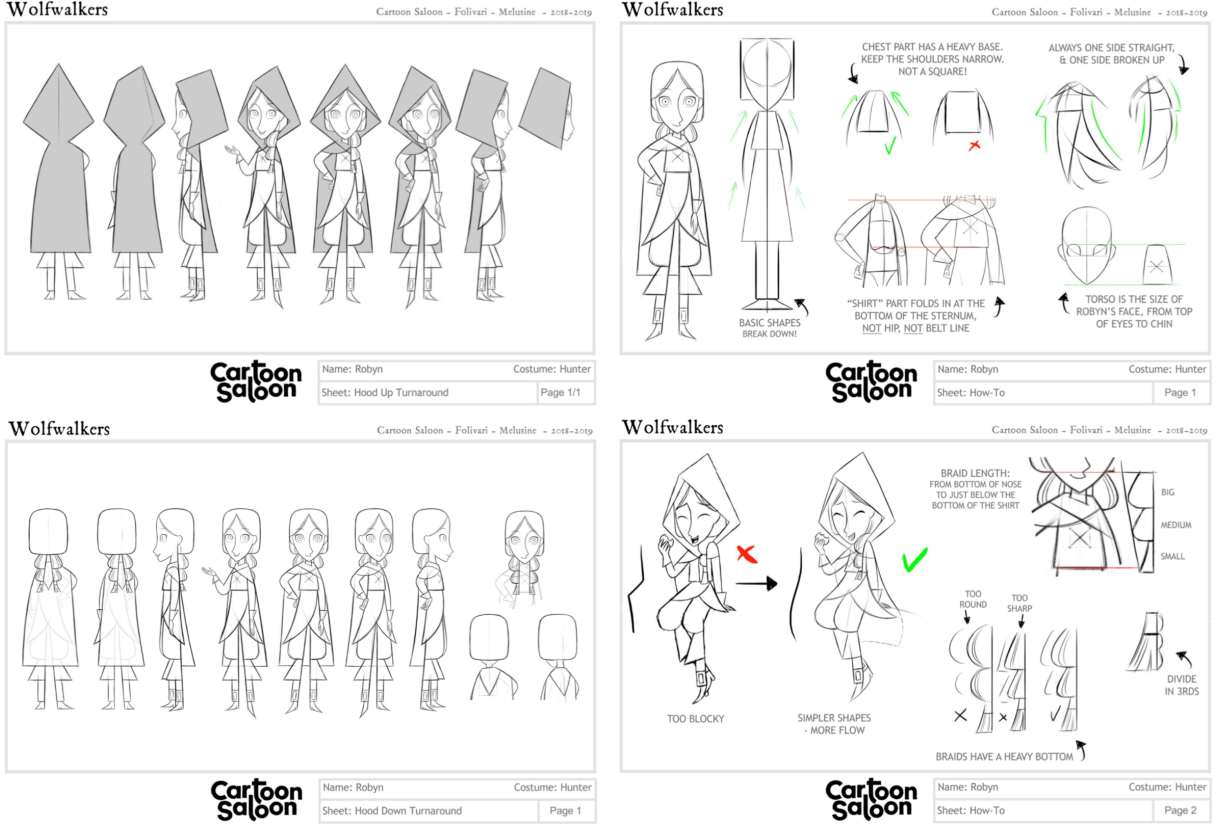


Figure 1.1: Examples of model sheets from Cartoon Saloon’s hand-drawn feature film *Wolfwalkers*.

represent the 3D pose they aim to draw within the 2D character design’s artistic and technical confines. Incorrectly resolved poses must be redrawn, costing valuable production time. To avoid this, animators reference technical character design guidelines called *model sheets* (Figure 1.1), though these depict only a small portion of the many character poses appearing in the film. Animators therefore seek out existing director-approved frames for drawing reference, but must remember exactly where in the film their desired frame appears and search through dozens of image files to find it. To support future construction of an automated reference frame retrieval system for hand-drawn studios, our research explores *reference frame quality* (RFQ) – i.e., how helpful one frame is as drawing reference for another – from both perceptual and topological perspectives.

Though such approaches are hallmarks of hand-drawn animation technology research, several aspects of this thesis diverge from prior work. By devising our RFQ analysis framework alongside professional animators at the acclaimed hand-drawn animation studio Cartoon Saloon, we bypass common pitfalls impeding industry adoption of new technologies. Our perceptual studies of Cartoon Saloon animators reveal specialized, artistic task-specific attributes of animators’ hand-drawn frame perception, informing RFQ framework design. Our framework offers a novel 3D pose estimation approach for hand-drawn characters leveraging perceptual factors found to impact animators’ RFQ perception. This entails topologically integrating frames’ salient, heterogeneous visual information across multiple scales and abstraction levels, emulating visual perception within a fully explainable TDA computer vision framework. We model diverse facets of animators’ hand-drawn frame perception, including *multiresolution* (local-to-global and global-to-local) visual data integration (Figure 1.2), local (body part-wise) symmetry invariance, visual data (in)consistency detection, and abstract functional relationship comprehension. In future, this will endow our reference frame retrieval system with advantageous functionalities like usability early in production,

pose estimation for both partial-body and full-body frames, pose resolution strategy retrieval up to multiresolution symmetries, and fluent accommodation of hand-drawn characters’ bodily deformations. Our work extends beyond hand-drawn animation technology, underpinning a novel TDA theory and analytical framework for complex system analysis across manifold diverse applications.

1.2 Research Questions

This thesis addresses the following research questions:

1. *What are the artistic and technical barriers to industry adoption of hand-drawn animation technology research, and how can they be resolved?*
2. *Which perceptual factors influence professional animators’ judgments of reference frame quality, and how can they be empirically characterized?*
3. *How can these perceptual factors be computationally modeled to systematically compare hand-drawn animation frames according to animators’ perception of reference frame quality?*

1.3 Methodology

Hand-drawn animation technology research is deeply grounded in computational topology, illuminating profound connections between topological structures and hand-drawn frames. However, topological structures presently employed in hand-drawn animation technology research are mathematically and conceptually basic, comprising mainly graphs and cell complexes. Although these structures fluently model contour connectivity, stroke junctions, medial axes, and other local frame features pivotal to line drawing perception, they cannot emulate the human brain’s multiresolution integration of local strokes and shapes into globally coherent scene understandings. Methods thus often require significant user interventions, which can interfere with animators’ workflows and artistic freedom. According to hand-drawn animation experts consulted throughout our research, these limitations may inhibit industry adoption.

In this thesis, we explore how advanced computational topology can elevate hand-drawn animation technology research by addressing pervasive barriers to industry uptake. After surveying the research field and offering topological and industry-based recommendations for improving capabilities and desirability of future work, we illustrate a practical application of this ethos to the problem of reference frame retrieval. To identify the optimal topological approach to RFQ analysis, we conduct perceptual studies on Cartoon Saloon animators, investigating how established traits of artists’ visual perception manifest in their work. Our findings imply that animators’ RFQ perception is moderately multiresolution isometry-invariant (Figure 1.4) and sensitive to both visual and pose features; moreover, frames with inconsistent visual features may engender different pose perceptions between animators. To topologically model these perceptual traits, we employ *sheaves* – graphs with local data spaces attached to vertices and edges, “stitched” together by data consistency constraints (Figure 1.3) – and *Group Equivariant Non-Expansive Operators (GENEOs)*, a theoretical formalization of equivariant data analysis. We synthesize these theories

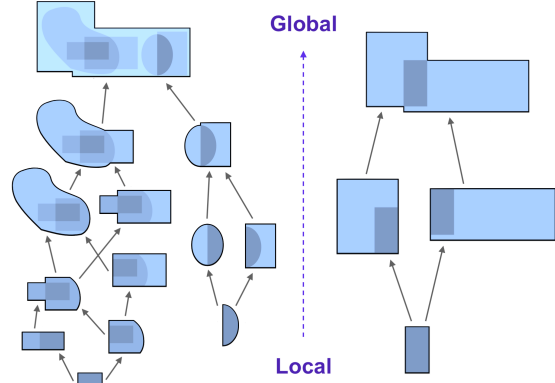


Figure 1.2: Multiresolution visual data integration on fine vs. coarse data topologies (left and right, respectively). Figure adapted from [271].

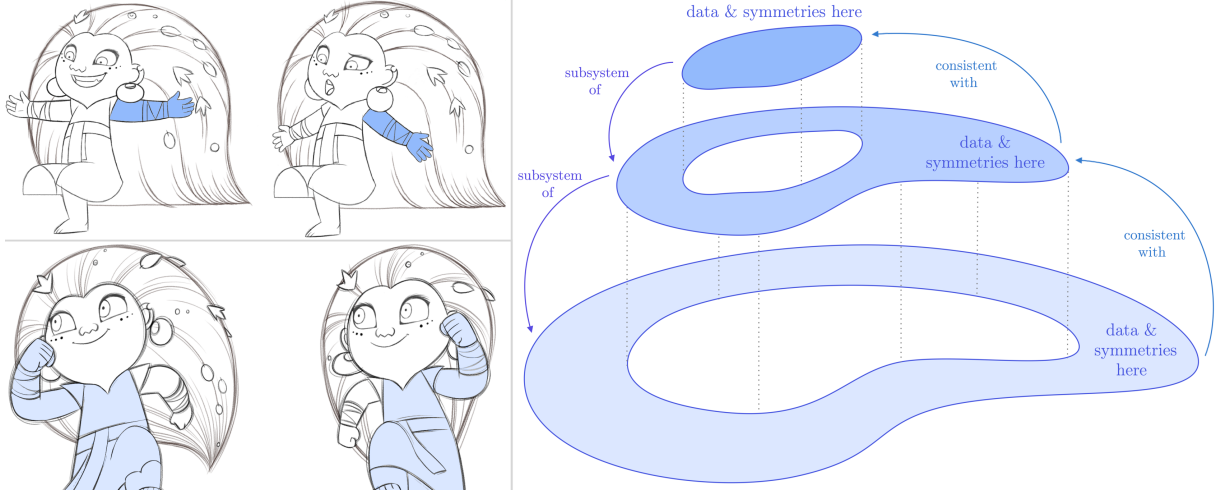


Figure 1.4: Left: Heterogeneous multiresolution symmetries on hand-drawn animation frames: local rotation (top, darker blue) and near-global reflection (bottom, paler blue). Right: Conceptual diagram of multiresolution data consistency across local-to-global subsystems within a complex system. Figure adapted from [1].

to create *GENEO sheaves*, a novel TDA framework which, when applied to RFQ analysis, explainably professional animators’ hand-drawn frame perception. RFQ GENEO sheaves are designed to integrate easily into studio pipelines, readily accommodating hand-drawn characters’ physically impossible poses, data-sparse depictions, and heightened stylization. Unlike many DNN body pose estimators, GENEO sheaves’ local-to-global awareness handles both whole-body and partial-body depictions of hand-drawn characters, both of which are common in animated films.

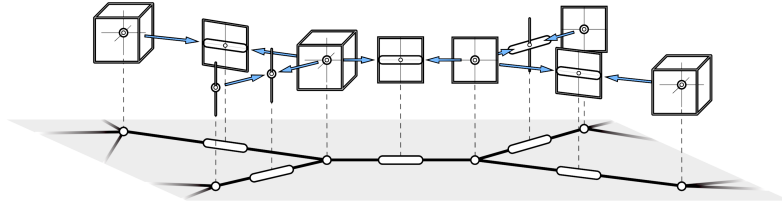


Figure 1.3: Conceptual illustration of a sheaf, with blue arrows depicting consistency constraints. Figure adapted from [154].

GENEO sheaves inhabit the intersection of heterogeneous data fusion and equivariant data analysis, addressing key research gaps in these increasingly popular disciplines. Sheaves’ sophisticated data representation and multiresolution consistency evaluation features are rapidly gaining traction in network analysis, especially for heteroge-

neous network data fusion and enhancing graph neural networks (GNNs). However, these features are designed for vector data, inhibiting applicability to tasks like image processing. Equivariant neural networks (EqNNs) are hugely popular for controllably modeling regularities like rotation and translation invariance in real-world data, but only when regularities act globally over a system. Yet many complex systems comprise heterogeneous regularities acting locally on different subsystems (Figure 1.4). GENEO sheaves address these research gaps through a novel theory of *heterogeneous multiresolution equivariance*, modeling complex systems’ heterogeneous subsystem data and regularities while maintaining system-wide data consistency.

1.4 Scope & Limitations

Due to the interdisciplinary nature of this thesis, we pursue a limited research scope in order delve deeply into each facet of our work, thereby establishing novel bonds between and insights into our varied fields of interest. We focus exclusively on uncolored hand-drawn animation frames depicting human characters with *articulate* bodies – i.e., comprising multiple local, independent, yet interrelated parts able to assume complex physical and visual pose configurations – drawn by professional animators for high-level animated feature films. These character designs comprise sets of *regions* (closed simple 2D curves) formed by *strokes* (lines drawn by animators), matching the vast majority of hand-drawn animated character designs throughout history, and ensuring the relevance of our work to industry settings (Figure 1.1). As discussed in Chapter 2, these qualities set hand-drawn frames apart from typical line drawings, distinguishing our work from the broader field of sketch processing. Accordingly, our survey of hand-drawn animation technology strictly reviews topological approaches to processing professional-quality frames over articulate motion, excluding topics such as sketch vectorization, frame segmentation, line drawing auto-coloring, and sketch-based object recognition.

Our perceptual experiments endeavor to establish basic tenets of professional animators’ hand-drawn frame perception with respect to their studio working environments and artistic tasks routinely encountered in their work. This experimental scope deviates from prior perception research, which studies professional artists’ performance on general visual tasks in generic experimental settings. It was therefore imperative that we exclusively test animators’ perception of real, raw hand-drawn frames, since edited or synthetically generated frames would have intractably confounded our experimental results. Consequently, our experimental frame data was limited to frames drawn for real Cartoon Saloon feature films. These did not always contain numerous examples of every variable case we aimed to test, limiting our experiments to small frame data sets. Moreover, the formidable subject of empirical perception-based metrics for comparing artistic data far exceeds the scope of this thesis, leading us to classify and compare experiment frames by eye. Lastly, because this is an industry-based thesis pursued in partnership with Cartoon Saloon, our experiments were conducted on Cartoon Saloon animators and frames only, precluding generalization of our perceptual results to other hand-drawn studios’ diverse animation styles.

Unlike the vast majority of hand-drawn animation technology research, this thesis’ primary aim is not to create a new frame processing algorithm, but to construct a wholly novel perceptually-based mathematical and computational framework for understanding hand-drawn frames across numerous interconnected data analytical dimensions. As such, we focus on carefully crafting a rigorous mathematical theory with deep philosophical and neuroscientific motivations, rather than chiefly outputting software and user trials; however, we computationally verify our key mathematical findings via proof-of-concept code. Accordingly, while the work presented in this thesis will underpin a reference frame retrieval system for Cartoon Saloon and other hand-drawn studios in future, development of this system exceeds our present scope. Yet prioritizing mathematical development alone ultimately magnifies our data analytical scope far beyond hand-drawn frame processing, yielding a novel, general-purpose data analytical framework unifying the increasingly prominent fields of heterogeneous data fusion and equivariant data analysis. Although hand-drawn frames remain our primary objects of study, our mathematical contributions position our research firmly within the theoretical foundations of TDA and Topological Signal Processing. Our work is thus prospectively applicable to knowledge modeling, protein analysis, scene reconstruction, and many other diverse fields. While we do not contribute to deep learning directly, GENEIO sheaves’ strong compatibility with DNNs are discussed in this thesis and will be cultivated in future work.

1.5 Contributions

- **Hand-drawn Animation Technology Through the Lens of Computational Topology**

Our first main contribution is the first state-of-the-art review of hand-drawn animation technology research subject to the field’s mathematical foundations in computational topology, illuminating how methods’ artistic advantages and limitations arise directly from their topological approaches. Combining our analysis with insights from hand-drawn animation experts Glen Keane (Disney) and Mark Mullery (Cartoon Saloon), we diagnose practical barriers to industry adoption of the methods surveyed, highlight overarching research trends, and offer informed guidance for future work.

Associated Publication: Schwartz, R., Mullery, M., Dingliana, J., & McDonnell, R. (2025). Computational topology for hand-drawn animation technology: A survey. *Computers & Graphics*, 129, 104225. doi:[10.1016/j.cag.2025.104225](https://doi.org/10.1016/j.cag.2025.104225)

- **Controlled Evaluation of Professional Animators’ Reference Frame Quality Perception**

We contribute the first controlled studies on professional animators’ RFQ perception, addressing the research gap regarding how artists’ specialized perceptual traits practically impact their work. Our first novel finding is that animators’ RFQ perception is local isometry-invariant, indicating that RFQ perception prioritizes pose resolution strategy preservation over depicted pose similarity. Secondly, we contribute evidence that animators’ perceptions of hand-drawn characters’ 3D body poses grow more divergent as frames’ visual information grows less consistent, meaning multiple animators may differently interpret the same frame. These findings motivate our RFQ analysis framework.

Associated Publications: Schwartz, R., Mullery, M., Dingliana, J., & McDonnell, R. (2025). Perception of Drawing Reference Quality among Professional Hand-drawn Animators. In *ACM/EG Expressive Symposium – WICED: Eurographics Workshop on Intelligent Cinematography and Editing – Artworks, Posters, Demos*. doi:[10.2312/exw.20251068](https://doi.org/10.2312/exw.20251068)

Schwartz, R. & McDonnell, R. (2026). Visual Data Consistency in Cartoon Poses Governs Pose Interpretation in Professional Animators. In *ACM Symposium on Applied Perception 2026 (SAP ’26)*. doi:[10.1145/3821409.3821478](https://doi.org/10.1145/3821409.3821478)

- **GENEO Sheaves & Heterogeneous Multiresolution Equivariance**

Our second main contribution is GENEIO sheaves, a novel TDA framework for modeling complex systems comprising heterogeneous subsystems and subsystem interactions across multiple resolutions, which are increasingly prominent across data analysis applications. GENEIO sheaves underpin our novel theory of heterogeneous multiresolution equivariance which, unlike prior equivariant analysis techniques, heterogeneously modulates subsystems locally while preserving global data consistency. We contribute numerous GENEIO sheaf data analysis features and evaluation metrics, all of which have strong potential for theoretical expansion and diverse applications.

Associated Peer Review: An early version of this work was submitted to the International Symposium on Computational Geometry 2025. Though the work was too preliminary for publication, the expert reviewer praised its “deep and novel mathematical theory,” “rigorous formulation of a soft problem,” and “lots of potential future directions for research,” calling it “one of the most fun papers I’ve read in a long while.” We have since addressed the reviewer’s critiques.

- **Reference Frame Quality Analysis Framework**

We contribute a new GENEIO sheaf-based RFQ analysis framework tailored to professional animators’ perceptual processes and hand-drawn studios’ practical needs. Heterogeneous multiresolution equivariance undergirds the first TDA 3D pose estimation technique for hand-drawn characters, designed to handle partial-body shots, physically impossible poses, body part occlusion, and diverse artistic styles. GENEIO sheaf frame comparison metrics model animators’ local isometry invariance, pose perception subjectivity, and complex perceptual relationships between frames’ visual and pose similarities, supporting highly customizable reference frame retrieval.

1.6 Thesis Overview

The research presented in this thesis is organized as follows:

- **Chapter 2** overviews the interdisciplinary research fields foundational to this thesis, including hand-drawn animated film production, hand-drawn animation technology, human pose estimation, human visual perception and its DNN-based computational models, and Topological Data Analysis. This chapter contains our review and analysis of hand-drawn animation technology research.
- **Chapter 3** presents our perceptual studies on professional hand-drawn animators, investigating their visual understandings of hand-drawn animation frames. We study animators’ frame perception subject to their typical hand-drawn animation studio working environments, and discuss how our results further our understanding of animators’ RFQ perception.
- **Chapter 4** introduces GENEIO sheaves, our novel general-purpose data analytical framework for modeling and evaluating complex real-world systems. After briefly reviewing the relevant mathematical background, we rigorously catalog GENEIO sheaves’ foundational definitions, theorems, metrics, and data analytical tools. We present our GENEIO sheaf-based conceptual and computational model of animators’ RFQ perception, demonstrating its efficacy through a proof-of-concept implementation estimating 3D arm poses depicted in hand-drawn frames.
- **Chapter 5** reviews this thesis’ contributions, conclusions, and directions for future work.

Chapter 2

Background

This thesis spans a wide range of specialized topics: hand-drawn animation as an art form and an industry, hand-drawn animation technology research, visual perception research, computer vision, and advanced mathematics including topology, group theory, and category theory. Our interdisciplinary scope warrants provision of considerable background information on each of these subjects, given in this chapter. In particular, a major contribution of this chapter is our state-of-the-art review on hand-drawn animation technology (Section 2.2) through the novel, dual lens of the field’s mathematical foundations in computational topology alongside artistic and practical insights from hand-drawn animation industry experts (interview transcripts in Appendix B). Our review advances our understanding of hand-drawn animation technology research across artistic, mathematical, and practical dimensions, enabling diagnosis of reasons why hand-drawn studios may be reluctant to adopt many of the field’s technological outputs. Using this analysis, we provide informed guidance for future work.

The insights gained from this survey guide our research approach to hand-drawn animation technology throughout this thesis, thereby also determining the additional topics covered in this chapter. Section 2.1 overviews the standard hand-drawn animation production pipeline through which modern hand-drawn films are made, also discussing the nature of hand-drawn frames as data entities. Section 2.2 presents our state-of-the-art review, throughout which we provide in-depth technical and artistic analysis of each approach discussed. Section 2.3 briefly overviews the expansive computer vision subfield of human pose estimation research, focusing on techniques presently or potentially applicable to hand-drawn frames. In Section 2.4, we review research on human visual cognition, highlighting findings relevant to hand-drawn frame perception and critically evaluating DNNs’ emulation of human perceptual processes. Section 2.5 summarizes results in Topological Data Analysis (TDA), and begins establishing the deep connections between TDA and hand-drawn animation technology fundamental to this thesis. This chapter concludes with Section 2.6, where we knit together these many disparate themes and explore what they collectively reveal about how hand-drawn animation technology’s present limitations can best be addressed, alongside exciting opportunities for the research field’s future.

2.1 Hand-drawn Animation

This section provides an introduction to hand-drawn animation from artistic, procedural, and data analytical viewpoints. First, we overview the production pipeline through which each hand-drawn frame passes as an animated film is made, from initial conception to final product. We discuss the order in which frames pass through various production departments, as well as each department’s artistic and

technical responsibilities. Subsequently, we detail properties of uncolored hand-drawn animation frames distinguishing them from other types of line drawings. We discuss frames as material data entities upon which frame processing and computer vision tasks can be performed. Frames are also described as products of animators’ artistic expression, drawing on insights from hand-drawn animation experts (Appendix B).

2.1.1 Hand-drawn Animation Production Pipeline

Hand-drawn animation production pipelines can vary between studios, and even across different productions within the same studio. Each production presents a unique set of artistic, practical, and technical challenges, which determine if and how particular softwares are used, mechanisms of asset passage between studio departments, and many other intricacies of procedure, practice, communication, and data management. However, the general steps involved in hand-drawn animation production have remained fixed since their inception nearly a century ago; the most seismic pipeline change since then has been studios’ migration to drawing on computers rather than paper. To equip readers with the technical knowledge necessary to understand this thesis, we review the core stages of hand-drawn animation production pipelines specifically involved in character animation; scene design and non-character animation stages are not covered. We focus primarily on the artistic pipeline; technical pipelines comprise many interconnected softwares for creation and transfer of artistic assets between production departments. For detailed information on hand-drawn animation production, please consult [311, 333, 171].

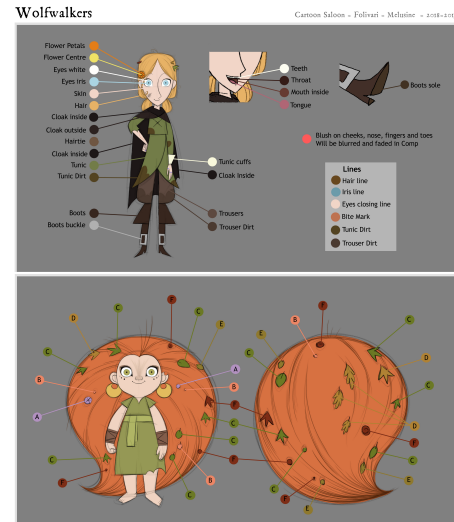


Figure 2.1: Color model sheets from *Wolfwalkers* for Digital Ink & Paint.

Storyboard Storyboarding is a process also used in live-action filmmaking in which the entire film is sketched shot-for-shot, resembling a comic strip. Because this process involves more high-level creative decision-making than artistic expertise, storyboard artists are not necessarily animators, although many industry professionals possess both skill sets. Aside from playing a pivotal role in visually crafting the film’s narrative, storyboard lays the groundwork for every frame animators will draw. Storyboarding defines characters’ emotional beats, physical interactions, prop involvement, and basic movement choreography throughout the entire film, defining the artistic, emotive, and technical challenges animators will need to meet to bring the director’s vision to life. Storyboard panels and audio recordings of corresponding screenplay dialogue are sequentially compiled into the *animatic*, the most primitive version of the film, from which all subsequent production departments work.

Character Design Character designers are often high-level artists and animators tasked with defining characters’ appearances and developing exacting artistic and technical drawing guidance for animators. With this guidance, animators are better equipped to draw characters matching the film director’s desired look. When challenges like view-dependent character geometry arise (Figure 2.23), designers must devise strategies by which animators can retain characters’ visual cohesion over articulate motion without violating complex or conflicting design constraints. Designers ultimately produce model sheets (Figure 1.1), which are used by animators as drawing reference throughout the entire animation process. Color versions of model sheets (Figure 2.1) also carry detailed information for DIP and compositing (see below), which must be replicated exactly across each individual frame to ensure visual continuity.

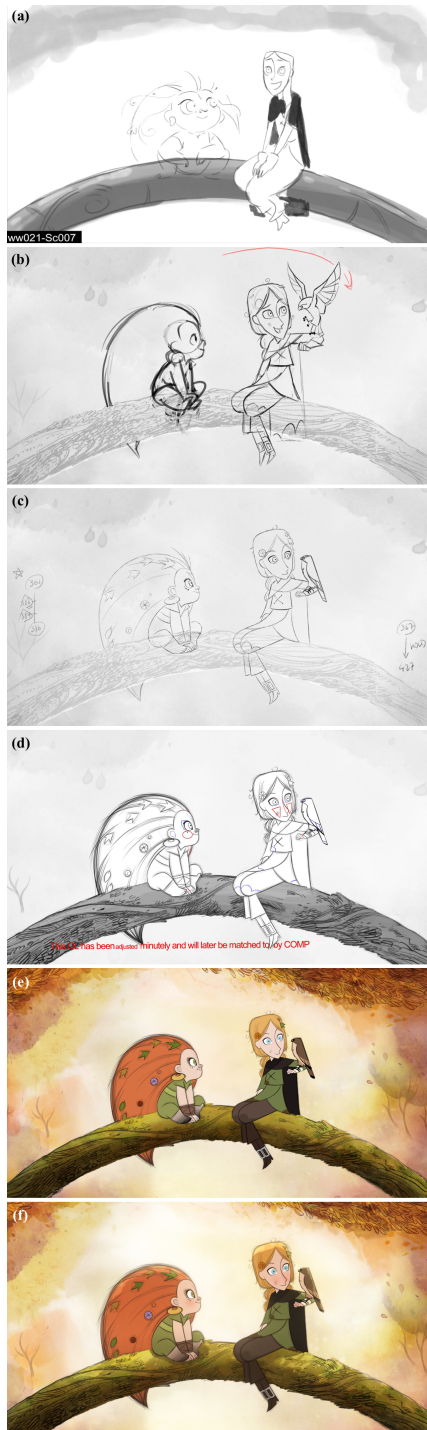


Figure 2.2: Hand-drawn animation production pipeline stages for a single Cartoon Saloon *Wolfwalkers* frame: (a) storyboard, (b) posing, (c) rough animation, (d) clean-up animation, (e) DIP, and (f) compositing. The diagrams on the rough frame provide directions for clean-up animators; the colored lines on the clean-up frame specify technical information for DIP and compositing.

Posing Posing serves an analogous purpose to blocking in a stage play. Posing artists work primarily from storyboards and the animatic in order to choreograph characters’ physical motions, positions, and interactions across each scene of the film. This is often the stage at which the emotional beats set forth in storyboards begin to assume visual and physical manifestations, with each pose expressing the character’s state of mind at a moment in time. Posing thus serves as animators’ emotional and physical roadmap for characters’ internal and external progressions throughout each scene. Moreover, it bridges the gap between characters’ emotional vs. visual appeal, beginning the process of capturing gesture and emotion within model sheets’ design constraints.

Rough Animation Rough animation begins the process through which consecutive poses are converted into fluid, articulate character motion. Working approximately yet not always stringently from model sheets, rough animators produce several drafts of *keyframes* depicting source and target poses differing by a smooth character motion. Upon finalizing keyframes, rough animators begin the process of drafting intermediate or *inbetween* frames depicting the motion itself, through which they dictate the exact physicality and quality of character motion throughout the film. Because motion quality is the essence of hand-drawn animation’s artistry, rough animators make many of the film’s pivotal artistic decisions and are thus generally senior to clean-up animators. Rough animators dictate when and how Principles of Animation (Section 2.1.2) are deployed – particularly timing, exaggeration, anticipation, follow-through, motion arcs, and appeal. They also further develop pose resolution strategies comports with the director’s artistic vision, providing artistic direction for clean-up animators to follow.

Clean-up Animation Clean-up animators draw the frames which get colored in and, ultimately, appear in the completed film. Their job is to preserve the articulate motion quality given by rough animation while satisfying model sheets’ strict artistic design guidelines. Clean-up fills in intermediate frames not yet drawn by rough animators, as well as refining and finalizing rough animators’ pose resolution strategies. Clean-up animators also ensure that each frame’s line quality meets the requisite technical standards to be colored in by Digital Ink & Paint (see below), balancing rough animation’s line and emotion quality, character design consistency, character recognizably, depicted pose clarity, and line technicalities essential for subsequent production stages. Since finalized clean-up frames are fully director-approved and offer the best-quality artistic and technical visual information, they typically provide the most helpful drawing references for both rough and clean-up animators.

Digital Ink & Paint (DIP) DIP colors in hand-drawn frames after they are animated. Because DIP is an independent production stage typically beginning well into animation, this thesis exclusively studies uncolored animation frames, as they are most abundant at the time in production when animators are seeking drawing references. DIP artists reference color versions of model sheets when coloring each frame’s regions; depending on the film’s style, both regions and strokes may need to be colored for each frame. Because hand-drawn characters lack ground-truth body representations, it may be unclear to DIP artists how certain regions in clean-up frames correspond to the regions defined in model sheets, sometimes requiring them to track regions across multiple frames to determine their representational meanings. DIP involves highly technical artistic procedures contributing significantly to the film’s final style – on *Wolfwalkers*, for instance, all region colors were slightly offset to emulate woodblock printing, and strokes were softened to simulate watercolor.

Compositing Compositing is the final animation stage in which scene design, character animation, and non-character animation are visually integrated. This not only involves receiving, managing, and assembling assets from numerous production departments into cohesive scenes, but also adding lighting effects, color gradation, and shading (although sometimes shading is partially completed by a separate department). Prior to compositing, the various components of an animated scene may appear visually flat or discombobulated; compositing artists are thereby tasked with making these assets visually harmonious, ultimately delivering the final version of each frame in the film.

2.1.2 Hand-drawn Animation Frames

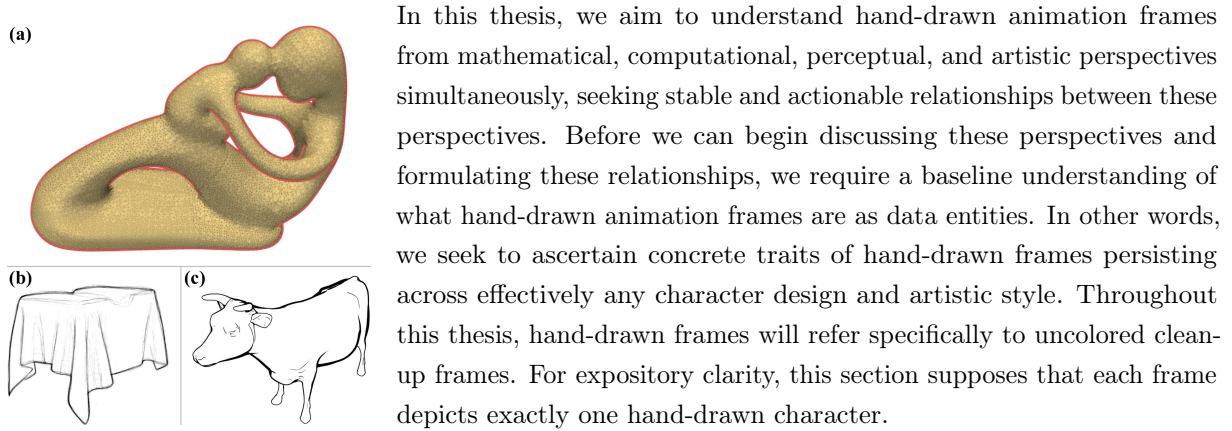


Figure 2.3: (a) Occluding contours (highlighted red) on a 3D polygon mesh, from [193]. (b-c) Standard line drawing examples from [78, 158].

In this thesis, we aim to understand hand-drawn animation frames from mathematical, computational, perceptual, and artistic perspectives simultaneously, seeking stable and actionable relationships between these perspectives. Before we can begin discussing these perspectives and formulating these relationships, we require a baseline understanding of what hand-drawn animation frames are as data entities. In other words, we seek to ascertain concrete traits of hand-drawn frames persisting across effectively any character design and artistic style. Throughout this thesis, hand-drawn frames will refer specifically to uncolored clean-up frames. For expository clarity, this section supposes that each frame depicts exactly one hand-drawn character.

While standard line drawings and hand-drawn frames both comprise strokes and regions, they present and employ these features very differently. Line drawings tend to place strokes along *occluding contours*, or edges bounding the visible portion of a surface at a given viewpoint (Figure 2.3) [158], imbuing line drawings with depth and volume information [289]. Though hand-drawn frames also use occluding contours, strokes often delineate adjacency between distinct yet depth-equivalent design elements (e.g., Figure 2.1, bottom, cuff and arm). Line drawings may also contain open curves (Figure 2.3b-c), whereas hand-drawn frames typically only comprise the closed simple 2D curves bounding regions. This is because DIP uses region closure to color frames consistency and efficiently, a practice established at Disney during the art form’s inception. Consequently, image processing methods designed for standard line drawings may not be suitable for hand-drawn frames, although frames’ unique visual features are leveraged to great effect throughout hand-drawn animation technology research.

Hand-drawn frames contain two types of visual information: *material features* and *depicted features*. Material features comprise a connected set of regions formed by a collection of boundary strokes, with region *geometry* (shape) and *topology* (connectivity) dictated by the character’s model sheet (Figure 1.1). Depicted features include the character’s depicted volumetric body geometry, body topology (i.e., ground-truth connectivity between depicted body parts), body pose, physical motion (e.g., walking, running, jumping), and abstract functional relationships between body parts (e.g., joint motion constraints, body parts’ inability to self-intersect). Here, *volumetric* refers to animators’ mental conceptions of the body poses they aim to draw or, relatedly, perceived physical poses of hand-drawn characters. Throughout this thesis, the term denotes mental or computational representations of hand-drawn characters as non-planar entities, including but not limited to 2.5D and 3D modeling. As discussed throughout this chapter, mental conceptions of volumetric pose are far more flexible and abstract than the material geometry of 3D character models. Hand-drawn frame perception fluently accommodates physically impossible volumetric poses (Figure 2.12), view-dependent geometries (Figure 2.23), and other visual irregularities.

Understanding relationships between frames’ material and depicted features is vital to understanding animators’ artistic processes, human visual perception of hand-drawn frames, and computational frame processing methods. As discussed throughout this thesis, these relationships are highly complex and not fully understood; however, we outline some basic premises here. Material and depicted topology and geometry are heavily related but not synonymous. Depicted topology is fixed across articulate motion because adjacent body parts’ corresponding regions remain materially connected; e.g., in Figure 1.1, sleeve and cuff regions remain connected (unless their shared edge is occluded) to preserve depicted connectivity between forearm and wrist. Meanwhile, material topology and geometry change continually across frames in order to depict volumetric body motion, although changes are subject to strict constraints. Material topology changes primarily convey self-contacts between functionally unrelated body parts (e.g., hand on hip), foreground-background occlusion relationships, and body part-wise perspective change. The last of these typically accompanies material geometry change (e.g., Figure 2.4, forearm) preserving the body part’s depicted geometry. That is, body part thickness and length should not appear to drastically change as characters move (although exceptions exist, discussed below). Besides these types of material topology changes, model sheet region topology is otherwise preserved over articulate motion, creating the illusion of hand-drawn characters as coherent physical entities and ensuring character recognizability across frames.

It should now be clear that material feature changes are governed by depicted features: strokes and regions are entirely subject to the volumetric pose they collectively aim to convey. Since hand-drawn frames are so data-sparse, material feature degradation compromises depicted feature clarity: e.g., missing, misplaced, or deformed strokes can make character pose difficult to interpret. Yet because hand-drawn characters have no ground-truth body representations, it follows that depicted features are an emergent property of interactions between material features and the human visual system. Even more remarkably, perception of hand-drawn characters as coherent physical entities is extraordinarily stable under unnatural material feature changes. Although most frames depict poses reasonably consistent with viewers’ abstract topological, geometric, and functional understandings of character body structure (e.g., human legs do not bend along multiple joints), *Principles of Animation* frequently engender perception of physical impossible poses and bodily deformations (Figure 2.12). While human vision easily identifies such deformations, they generally do not hinder our ability to coherently perceive frames’ volumetric features by visually processing their material features. Principles of Animation underpin the unique visual qualities of hand-drawn characters’ motion, causing bodies to appear elastic, motion to be artistically timed and exaggerated, and characters to look and move with immense visual appeal [311].

Hand-drawn characters are artistically *stylized* in myriad ways, fluidly varying between realistic and

symbolic depictions of their physical form, sometimes across different body parts within a single frame. *Symbolic* depictions are visual denotations of physical entities, varying enormously across films’ artistic styles. They are not limited to objects, also denoting physical states. For instance, a character’s hand on their hip may be symbolized by the topological hole enclosed by their arm and torso, instantly recognizable to viewers even when the arm shape conveys bone foreshortening physically impossible for the symbolized pose. This interplay between realistic and symbolic dominates the technical process of pose resolution, which is balanced against animators’ paramount goal of *finding the character*, or conveying a character’s emotive and artistic essence even at the expense of their design and physical motion constraints. The related idea of *finding the line* involves visually sculpting characters’ volumetric form from beautiful and carefully chosen hand-drawn strokes. As such, material frame features contain visual data far beyond that contained in standard line drawings. As discussed by hand-drawn animation experts in Appendix B, material features convey emotion and artistry indispensable to frames’ artistic quality.

2.2 Hand-drawn Animation Technology

Hand-drawn animation technology explores deep problems at the intersection of art and computation, facing formidable and often ill-defined technical and practical challenges. Hand-drawn frames’ data sparsity, stylization, frequent deformations, variation between visual realism and symbolism render them inherently difficult to computationally analyze and manipulate. However, this is not hand-drawn animation technology’s only obstacle to success. Technologies must integrate seamlessly with hand-drawn studios’ artistic and technical pipelines; most essentially, they should respect animators’ artistic processes and agency while relieving some of the tediousness and fastidiousness of their work. Although the vast majority of hand-drawn animation technology research endeavors to meet these goals, hand-drawn animation studios are often hesitant to adopt new technologies created by researchers. This indicates a schism between hand-drawn studios’ technological needs and researchers’ understandings of those needs, with the latter determining the technologies researchers choose to develop.

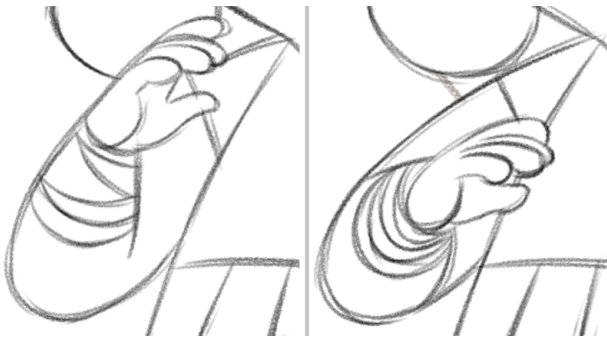


Figure 2.4: Examples of consecutive clean-up keyframes from *Wolfwalkers* illustrating complex stroke topology change arising from a physically small articulate movement.

This section comprehensively evaluates the present state of animation technology research in an effort to understand the disconnect between hand-drawn animation researchers and industry professionals, ultimately aiming to diagnose barriers to industry adoption of hand-drawn animation technology. Our review employs a novel analytical approach, examining hand-drawn animation technology subject to the field’s mathematical foundations in computational topology while incorporating insights from interviews conducted with hand-drawn animation industry experts. Connections between topology and hand-drawn animation are detailed throughout our survey, guiding researchers towards topological techniques best suited for their

computational objectives. We describe how methods’ artistic advantages and limitations arise explicitly from their chosen topological approaches, informed by our interviews with Glen Keane, master Disney animator, executive producer, and director (credits include *The Little Mermaid*, *Beauty and the Beast*, *Aladdin*, and *Tangled*), and Mark Mullery, Cartoon Saloon Chief Technology Officer and former assistant

director (credits include *The Breadwinner* and *Wolfwalkers*). Full transcripts of interviews with Keane and Mullery are available in Appendix B.

While related surveys include image processing techniques for cartoon drawings [361, 309], sketch-based modeling [34, 338], and topological techniques for various sketch analysis tasks [192], our work focuses exclusively on methods specialized for hand-drawn frame processing over complex articulate character motion (Figure 2.4), which is often beyond the scope of sketch-based modeling methods. We review three main research subfields: frame deformation, frame feature correspondence, and volumetric modeling of articulate hand-drawn characters. By incorporating knowledge from industry veterans, we aim to better inform researchers about how their work may be received and implemented in industry settings. Because we seek to diagnose present barriers to industry adoption of hand-drawn animation technology, we focus primarily on the past 15 years of research output. All methods reviewed are underpinned by computational topology and are designed to process hand-drawn clean-up frames of the data format described in the previous section.

2.2.1 Frame Deformation

Frame Deformation is a major realm of hand-drawn animation technology research. By automatically deforming or interpolating between keyframes, these methods seek to relieve animators of manually drawing intermediate frames. As several intermediate frames typically separate each pair of consecutive keyframes, this can drastically reduce animators’ drawing time during film production. Intermediate frames can be generated by attaching (parts of) frames to deformable topological structures – like lattices (Section 2.2.1) or meshes (Section 2.2.1) – and deforming the structure to produce character motion. Alternatively, pairs of corresponding source and target keyframe features – usually strokes (Section 2.2.1) – can be interpolated. Some methods, such as mesh-based region interpolation and deformable stroke topologies, straddle multiple categories. This section explores the efficacy of various frame deformation strategies, examining the advantages and limitations of their topological techniques and their usability for professional animators.

Lattices

Lattice deformation assigns a lattice to the image plane, dividing a frame into square pixel patches (Figure 2.6). When the lattice is deformed, each frame patch takes on the local deformation of its lattice square. Lattices topologize the image plane rather than frame features, making them lightweight regardless of feature intricacy. Without direct stroke or region tracking, their only feature preservation mechanism is topology-preserving deformation: if lattice squares cannot collapse, their image patches cannot degenerate. Preservation of animators’ line quality, faithfulness to character design, and sound visual communication with viewers therefore entirely depends on lattice topology preservation during deformation.

Animation arising from these confines strongly preserves stroke geometry, particularly suiting articulate *planar motion* (parallel to view plane) of characters with rigid skeletons (Figure 2.6). Planar

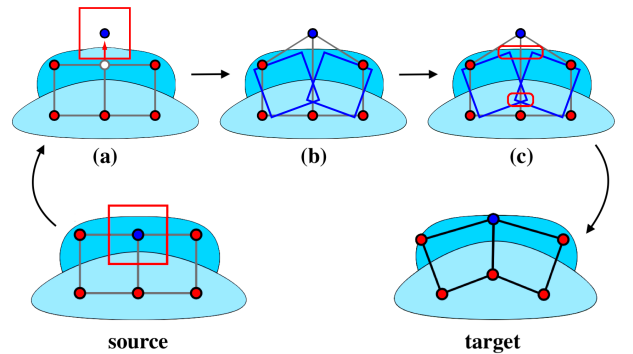


Figure 2.5: As-Rigid-As-Possible lattice deformation from Sýkora et al. [302]: (a) identifying target vertex location (boxed red), (b) lattice square isometry, (c) centroid of corners (circled red).

motion is well-approximated by As-Rigid-As-Possible (ARAP) lattice deformation, demonstrated by Sýkora et al.’s influential technique [302]. Each source keyframe lattice vertex moves to the target keyframe location most visually similar to its source location. Each lattice square isometrically transforms to optimally align with its repositioned corner vertices, which move to the centroid of their transformed squares’ corners, restoring lattice topology (Figure 2.5). While robust geometry preservation maintains character design, it lacks hand-drawn animation’s expressiveness. Kazi et al. [179] aim to improve lattice deformation’s expressiveness by defining seven lattice deformation modes, each modeling one of the 12 Principles of Animation. Though this achieves good results on small motions, larger motions are noticeably synthetic. Many lattice methods suffer from similar issues, arising from their treatment of hand-drawn strokes as textures on a deformable surface – i.e., the image plane. Strokes are then easily distorted beyond the physical arcs of hand-drawn lines, degrading their hand-crafted appearance. To address this, Chen et al. [64] generate novel strokes at intermediate frames by applying small transformations to keyframe strokes, preserving each stroke’s hand-drawn shape. Keyframe stroke contours and density maps are then ARAP lattice interpolated to guide generated stroke placement for each intermediate frame, yielding natural looking sketchy animation (Figure 2.6).

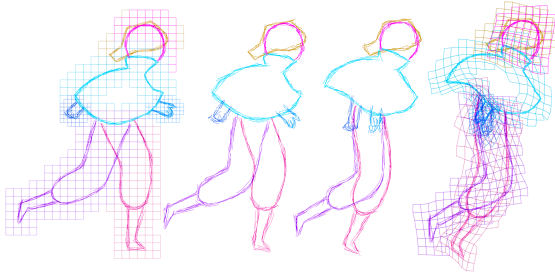


Figure 2.6: Example of transformation lattices assigned to each body part in a walk cycle from Chen et al. [64]. Finer resolution lattices animate both hands.

Choice of lattice resolution heavily impacts visual deformation quality. Finer lattices deform fewer pixels per square, reducing artifacting but becoming incomputable when squares get too small [302]. Raster frames deformed by coarser, computable lattices are post-processed with linear interpolation smoothing operations to restore image quality [118, 179], sometimes alongside shape preservation handling [302]. However, automatic line reconstruction may degrade line naturalness. Moreover, visually imperceptible artifacts often persist after smoothing and can cause color-fill issues for DIP [171, 306]. Finer resolution also supports articulation by reducing likelihood of trapping nearby but

semantically unrelated strokes within the same square and forcing them to move together during animation. To address resolution issues without sacrificing computability, Noris et al. [238] capture strokes’ topological and geometric relationships by applying a distance transform, then iteratively employing Sýkora et al.’s ARAP algorithm at increasingly fine lattice resolutions. Stroke deformation data is transferred across resolutions to speed up extraction of stroke-wise motion, offering computationally inexpensive lattice deformation at fine resolution.

Since unwanted distortion of nearby body parts necessarily arises when a single lattice globally topologizes an entire frame, many methods define local layered lattices for each body part [179, 107, 64, 108, 97, 114] (Figure 2.6). These work for coarse articulate motion, but cannot animate body part articulation within each local lattice. Carvalho et al. [51] resolve this issue by individually lattice-interpolating hand-drawn gesture strokes, each representing several detailed strokes along the gesture’s contour (Figure 2.7). Corresponding keyframe gesture strokes are optimally aligned by local isometry followed by thin-plate spline warp. The thin-plate spline warp non-linearly interpolates points sampled along pairs of corresponding source and target strokes, using lattice interpolation derived from sample point interpolation to align keyframe strokes. Hand-drawn gesture strokes relieve the need for larger lattices to maintain strokes’ relative spatial relationships and, alongside a hierarchical stroke layering scheme, enrich articulation opportunities. Mo et al. [225] automatically detect and parse unrelated nearby strokes using

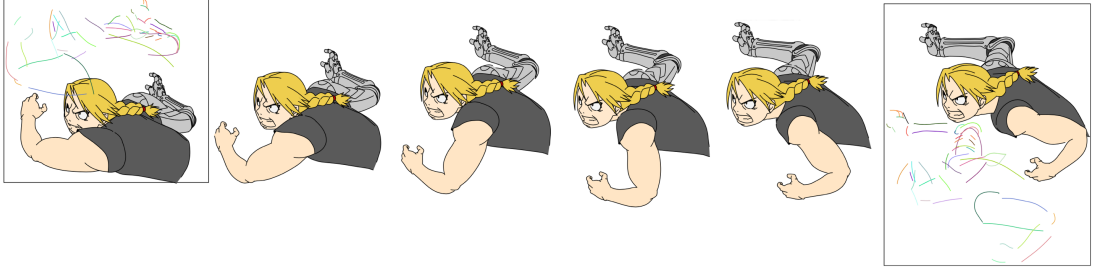


Figure 2.7: Example of semi-automatic inbetweening method from Carvalho et al. [51]. Hand-drawn keyframes and color-coded corresponding gesture strokes shown in boxes, with intermediate frames between them.

perceptual cues derived from stroke curvature and geometry at shared endpoints (Section 2.2.2). For corresponding keyframe stroke groups, a learning-based Adaptive Spatial Transformation Module predicts the affine transformation interpolating their local lattices over intermediate frames. The module learns to homogenize similar transformations on neighboring lattices and negotiate transformation disparities between overlapping lattices to depict smooth articulate rigid motion. Resulting animation is high-quality but needs user intervention to appear hand-drawn, which learning-based methods unfortunately preclude.

Complex stroke topology changes are inherent and ubiquitous to hand-drawn articulate motion (Figure 2.4), yet unattainable from standard lattice deformation. Because lattices topologize the image plane, they cannot undergo square collapse necessary for modeling stroke topology change without permanently degenerating collapsed squares’ feature data, distorting nearby features, and destroying lattice structure. Further, cutting along lattice edges or independently repositioning individual squares leaves holes in the image plane. While many of these operations are possible in mesh- and stroke graph-based frame interpolation, lattices remain popular for their computational efficiency, usability, and robust preservation of relative feature positions, sizes, shapes, and angles.

To animate stroke topology change while retaining lattices’ advantages, Even et al. [107] introduce a technique for interpolating frames with topologically unrelated lattice structures (Figure 2.8). An invertible ARAP lattice warp optimally aligns source to target keyframe strokes, registering the source lattice to the target lattice despite their differing stroke topologies. Warped source strokes are cross-faded with invertibly-warped target strokes to generate intermediate frames, simulating topological changes like stroke splitting and merging. The target frame receives new appropriate lattice topology upon becoming the new source frame, and

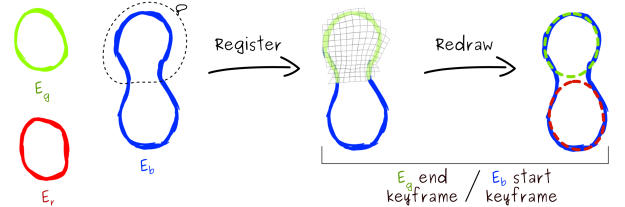


Figure 2.8: Example of stroke groups with separate lattices (green and red) merging into a single stroke group with different lattice topology (blue) through the lattice registration method from Even et al. [107]. Registration transforming green strokes into blue strokes is depicted; red strokes undergo an equivalent process.

Bezier lattice trajectories offer smooth interpolation over topology change. Even et al. [108] extend this technique by automatically layering stroke group lattices and cross-fading layers undergoing depth-order change during articulate motion, handling stroke disappearance/reappearance over view angle change. Fukusato et al. [128] achieve even more freedom by storing frame features in both lattices and meshes, providing backup records of artistic data when lattice topology breaks. To restore broken lattice topology, *graph Laplacian* diffusion [363] finds each square’s ARAP transformation optimally matching the given frame deformation, with mesh vertex position weights expressing feature change over deformation. The graph Laplacian is a matrix representation of a graph that can be used to discretely diffuse data over

weighted graphs. Here, the graph Laplacian’s usage reveals that the relationship between feature and lattice topology change is so intimate that the former can be used to reconstruct the latter.

Meshes

Meshes topologize shapes rather than the image plane, naturally relating their topology and geometry to frame features. Whereas lattice topology change breaks image plane topology and permanently damages feature data, mesh topology can alter substantially without disturbing frames’ visual information. Topology change is straightforward because meshes comprise triangles rather than uniformly sized squares; triangles can collapse and emerge without degrading overall mesh shape. Mesh topology change is also more localizable than lattice topology change since nearby triangles can vary in size and be subdivided or merged if needed.

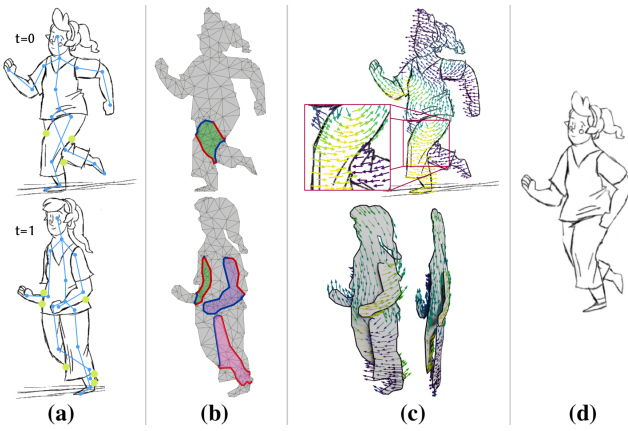


Figure 2.9: Mesh deformation approach of Brodt & Bessmeltsev [40]: (a) hand-drawn source and target keyframes with 2D skeletons; (b) 2.5D mesh cuts matching articulate self-occlusion discontinuities from the projected smooth 3D deformation field; (c) piecewise-smooth 2.5D deformation field; (d) an output intermediate frame.

try over deformation [347]. Since Laplacian influence on meshes depends on the graph path distance rather than proximity in the image plane, single-layer meshes can move articulately without undesirable motion of nearby body parts. However, these methods confine mesh deformation and skeleton motion to the plane, consequently struggling to model complex volumetric body part articulation. Brodt & Bessmeltsev [40] improve articulation by framing the problem inversely: they search for a mesh deformation field matching a given 2D skeleton motion. They observe that the smooth 3D deformation field of a hand-drawn character’s depicted volumetric motion projects to a piecewise-smooth 2.5D deformation field of frame feature motion, with discontinuities arising from articulate self-occlusion (Figure 2.9). Deep learning and graph optimization combine to construct the 3D deformation field from input 2D skeleton motion and find 2.5D mesh cuts along keyframe strokes. Cuts match the projected piecewise-smooth deformation field while minimizing stroke distortion. The method succeeds on local out-of-plane articulate motion but may degenerate strokes, a problem pervading learning-based hand-drawn frame deformation [66].

Mesh geometry can be defined in myriad ways depending on the problem at hand. While intentional triangle collapse can help generate expressive articulate animation, unintentional triangle collapse during deformation can alter frame features undesirably. To mitigate such occurrences, nearly all mesh methods

Many methods leverage deformation locality for articulate motion by equipping meshes with 2D character skeletons. Skeletons are associated to meshes by imposing each bone’s transformations on its intersecting mesh edges [325], intersecting mesh triangles [324, 60], or vertices bounding intersecting triangles [347]. When bones move and transform their intersecting mesh elements, the graph Laplacian propagates those transformations throughout the mesh. Harmonic weights associated with the Laplacian refine transformation locality [325] and shape preservation [324, 60]. For transformations propagated over triangles, ARAP lattice deformation [302] is adapted to isometrically transform each triangle independently before restoring mesh topology [324, 60]. Local rigidity can also be enforced vertex-wise by optimally preserving each mesh vertex’s nearest-neighbor star graph geometry

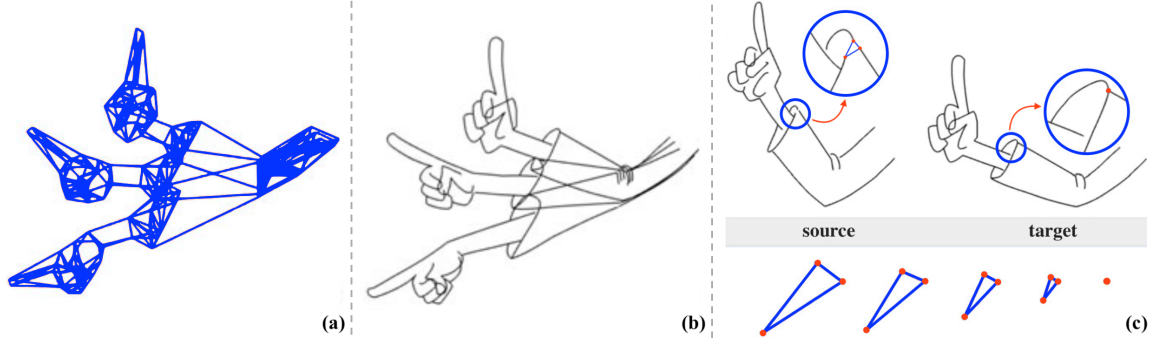


Figure 2.10: Compatible frame interpolation results from Yang et al. [344]: compatible meshes given by strokes (a) and original strokes (b) from input keyframes and a generated intermediate frame; (c) mesh triangle collapse for stroke topology change.

construct *Delaunay triangulations* maximizing every mesh triangle’s internal angles, as skinny triangles easily collapse (Figure 2.10c, bottom). Delaunay triangulations and their dual *Voronoi diagrams* are widely used in hand-drawn animation technology for partitioning and grouping strokes [25, 349, 234], interpolating frames in parametric pose space [269, 172, 82], and modeling natural phenomena like mountain ranges [90] and water caustics [168]. Their relationship to convex hulls of 3D point sets [90] likely compliments their use in volumetric modeling (Section 2.2.3). To further enhance feature preservation, Delaunay triangulations can be constructed as *compatible triangulations*: graph-isomorphic triangulations on ordered keyframe vertices such that corresponding source and target triangles have matching vertex orientations. Compatible triangulations are guaranteed to prevent mesh degeneration over deformation, and any topology-preserving mesh deformation defines compatible source and target triangulations. Applications to hand-drawn frame interpolation often relax compatibility requirements, balancing robust feature preservation with artistic expressivity. A suite of methods use Euclidean Minimum Spanning Trees to compatibly triangulate piecewise-linear stroke approximations of input keyframes [345, 344, 346]. Since directly interpolating compatibly triangulated frames precludes the stroke topology changes requisite for articulate motion, compatible interpolation is approximated by edge-wise ARAP interpolation. Triangle collapse enables small stroke topology change while optimally preserving larger visual features (Figure 2.10). Individual pairs of corresponding keyframe strokes can also be compatibly triangulated and interpolated [127]. Still, like lattices, treatment of compatible meshes as planar stroke embeddings hinders larger topology changes and occlusion handling. These limitations are somewhat remedied by depth-ordering triangles in compatible meshes [327, 202].

Zhu et al. [367] further temper compatible meshes’ mathematical rigor with artistic expressivity, devising a technique for compatibly interpolating arbitrary shapes. Topological operations like cutting, gluing, and self-overlap are dynamically simulated on a single mesh by permitting multiple vertices and edges to simultaneously occupy the

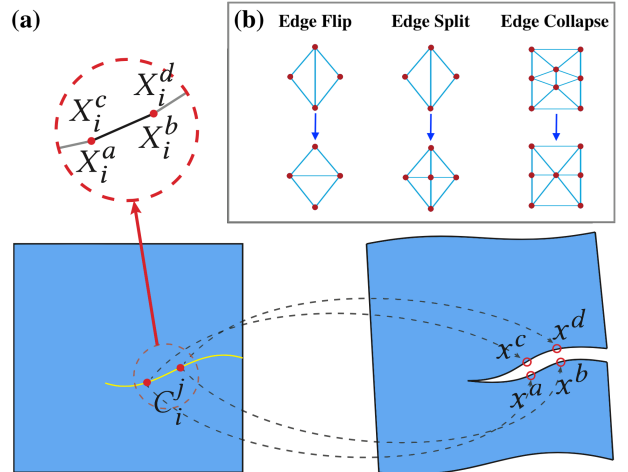


Figure 2.11: Mesh deformation techniques from Zhu et al. [367]: (a) topological cut operation simulated by vertices and edges initially occupying the same position; (b) edge flip, split, and collapse operations for changing mesh topology.

same position (Figure 2.11). Every triangulation of a point set is reachable by a sequence of edge flips [90] which, alongside shape preservation optimization, are used reestablish Delaunay meshes after extreme deformation. Because deformed meshes are compatible and simultaneous edge flips on compatible meshes preserve compatibility [301], flips can be propagated to all other deformations of the given mesh. Mesh compatibility is thus maintained across all deformations while the current deformation is retopologized. Although this method could model hand-drawn region topology change during articulate motion, animators would need to manually specify these intricate changes and fine-tune deformation rules to ensure aesthetically pleasing animation. Mullery reports that hand-drawn studios may eschew technologies demanding such effort, which is more time-consuming and draining for animators than simply drawing frames. Bai et al. [10] offer an approach that, though lacking support for extreme deformation and topology change, allows animators to specify region correspondences between frames with different region topologies. Interpolations map between disjoint subspaces of isomorphically-meshed region shapes to interpolate mesh topologies, creating one of the most self-contained and user-friendly interpolation workflows to date. Combining Bai et al.’s workflow with Zhu et al.’s shape deformation capabilities could provide robust region deformation and reduce demands for manual mesh edits and deformation rules, promoting adoption in industry settings.

Stroke Complexes

Hand-drawn animation technology’s central data structure is the *cell complex*: a topological space of connected vertices, edges, faces, and their higher-dimensional counterparts. Lattices, meshes, and graphs are all examples of cell complexes. *Stroke complexes* – cell complexes of strokes – enjoy unparalleled popularity among frame interpolation methods for their many artistically expressive and computationally appealing incarnations. These methods often indicate strong connections between stroke complexes, visual perception, and animators’ drawing process.

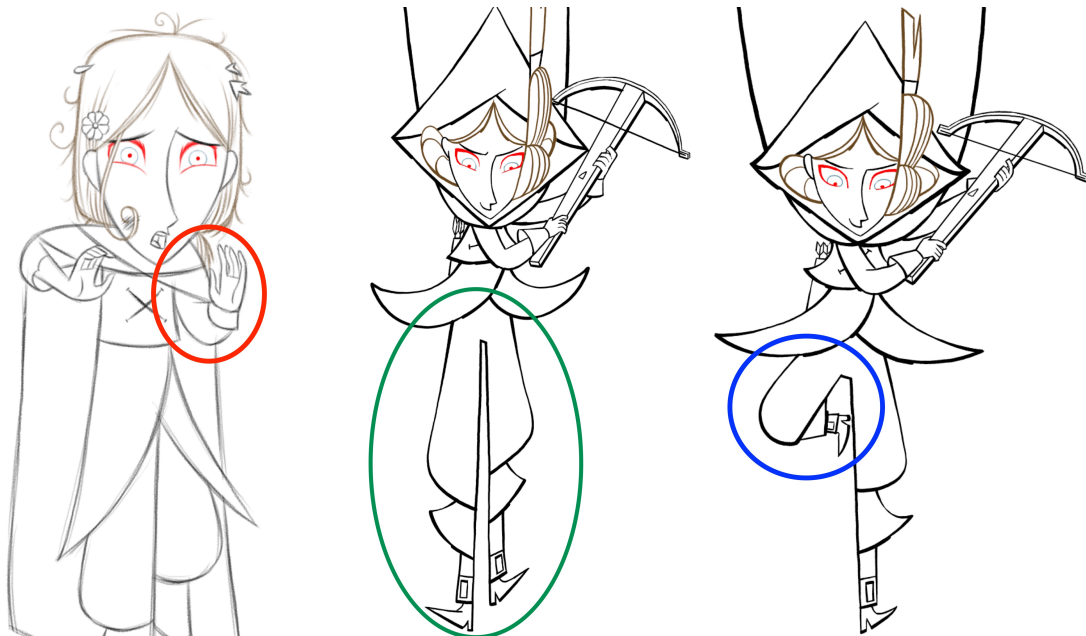


Figure 2.12: Examples of Principles of Animation and other hand-drawn animation techniques that are challenging to model computationally: exaggerated motion creating physically impossible poses (circled red), squash-and-stretch (circled green), stylized visual perspective due to preference for symbolic over realistic depictions of character pose (circled blue).

Stroke complexes most often appear as *stroke graphs* of stroke edges connecting vertices at strokes’ shared endpoints (Figure 2.13, bottom), or vice versa. After defining isomorphic or highly similar stroke graphs for input keyframes (Section 2.2.2), many methods attach a local stroke data space to each edge. Intermediate frames are generated by interpolating each corresponding key stroke pair within their common edge’s local data space, modeling animators’ practice of tracking individual stroke shape and motion across consecutive frames. These methods typically employ stroke-wise ARAP interpolation [330, 345, 336, 344, 346], although least squares interpolation on bipartite-matched key stroke sample points may offer smoother motion via bidirectionally optimal matching and deformation continuity [51]. Stroke tangent consistency is checked and smoothed at strokes’ shared endpoints, ensuring curvature continuity across intermediate frames. Consistency checking edge-wise transformations helps identify articulate body parts, which often manifest as connected stroke subgraphs undergoing near-uniform isometries. This is a powerful example of topological frame analysis extracting sophisticated information from simple features, indicating deep synergy between topology and hand-drawn animation. Through the simple construction of local data spaces attached to a stroke graph, key elements of animators’ drawing process can be emulated and complex articulate body structure becomes discernible. Once a body part is identified, its strokes’ transformations are homogenized for motion uniformity and given a logarithmic spiral trajectory [330, 344, 346], modeling the motion arcs foundational to hand-drawn character motion [311]. However, animators move and deform a body part’s strokes at slightly different rates and shapes to convey weight, express emotion, and heighten appeal. Their deference to core animation principles – e.g., squash-and-stretch, follow-through, and exaggeration [311] – and penchant for symbolic over realistic depiction greatly complicates automatic stroke interpolation. Homogenized motion can look mechanical, lacking stroke naturalness and artistry essential to professional hand-drawn animation (Figure 2.12). Animators also rely heavily on their internal proprioception and volumetric mental models of characters when choosing stroke shape and motion across frames [160, 183]; in the words of animator Norman McLaren, “animation is not the art of drawings that move but the art of movements that are drawn” [294]. Neither ARAP nor bipartite matching interpolation incorporates volumetric knowledge, often yielding noticeably synthetic line quality. These methods also cannot modify stroke graph topology, relying on stroke visibility toggles to simulate limited feature topology change (Section 2.2.2).

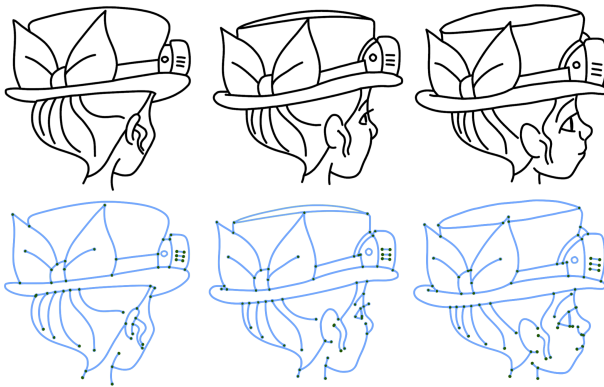


Figure 2.13: Intermediate frame generation method from Dalstein et al. [87] demonstrating stroke and region collapse across frames (top) and stroke complexes (bottom).

A competing stroke complex formulation offers superior topological flexibility by exploiting parallels between hand-drawn stroke topology changes and topological cell complex operations. By treating each frame as a formal topological cell complex, Dalstein et al. [86] supports operations like cutting, gluing, and cell collapse on stroke complexes, similarly to Zhu et al.’s mesh deformation scheme [367]. A sequence of stroke topology changes over consecutive frames is modeled as a space-time cell complex, which can be “sliced” at any point in time to yield frames of valid stroke complexes [87]. These space-time complexes are distinct from 3D stroke networks, simulating rather than explicitly computing out-of-plane motion via planar stroke deformation. Animators’ concurrent

consideration of individual strokes and the regions they form is modeled by the *boundary operator*, a topological cell complex operator exchanging data between faces (regions) and their surrounding edges

(strokes) (see Section 2.5 and Figure 2.37 for formal details). Whereas most frame deformation methods force animators to work exclusively with either strokes or regions, the boundary operator enables joint exploitation of strokes’ artistic freedom and regions’ consistency over articulate motion. Unlike stroke graph methods, body parts can be identified as regions rather than by detecting uniform stroke group transformations, leaving animators free to unsynchronize individual stroke shapes and motions for artistic effect without potentially inhibiting body part identification. Because regions and strokes are explicitly encoded as topological cells, they can alter topologically and undergo cell collapse to depict feature topology change and collapse without need for stroke visibility toggles (Figure 2.13).

Although topological stroke operations and hand-drawn stroke manipulation share many similarities, they are not equivalent. Some visually simple stroke edits are difficult or unattainable via topological stroke operations, costing animators time and energy better spent translating their artistry onto the page. Given Dalstein et al.’s user feedback [86], Mullery notes that animators’ frustrations may be exacerbated if they fail to grasp the topological logic needed for stroke editing workarounds, a pitfall pervading methods modeling frames as formal mathematical spaces. Clearly, each frame deformation method’s choice of cell complex and use of accompanying mathematical devices – local data attachment to cells, graph diffusion, topological operations – profoundly impact its advantages, limitations, and aesthetic results.

2.2.2 Frame Correspondence

Before keyframes are interpolated, correspondences between their features must be established. Although many researchers approach fame deformation and correspondence simultaneously, the problems are distinct: correspondence refers to establishing relationships between features appearing in multiple (typically consecutive) keyframes, while deformation refers to strategies for transforming source frame features into target frame features. Correspondence concerns recognition of the same stroke or region as a character changes pose and view angle, rather than deformation of said features to animate pose and view change. In practice, stroke correspondence methods (Sections 2.2.2 and 2.2.2) typically provide the correspondence relationships needed to interpolate keyframe strokes, while region correspondence methods (Section 2.2.2) serve to automatically propagate region colorization across sequences or sets of frames. This section reviews the wide variety of stroke and region correspondence techniques across topological, combinatorial, and spectral domains.

Stroke Graph Isomorphism

To interpolate stroke data over isomorphic keyframe stroke graph topology as described in Section 2.2.1, a stroke graph isomorphism between keyframes must be found. As with topology-preserving lattice deformation and compatible triangulation, fixing isomorphic stroke topology across a frame sequence guarantees preservation of hand-drawn keyframe features throughout intermediate frames.

Isomorphisms are often defined by simultaneously tracing isomorphic paths along input keyframes’ stroke graphs until all edges have been reached. Whited et al.[330] begin at a user-defined pair of corresponding key strokes, halt upon encountering non-isomorphic 1-hop subgraph topology or a visually dissimilar stroke pair, and resume once the user provides another stroke correspondence. Without the ability to detect and parse stroke connectivity arising from contact or occlusion between disjoint body parts, this method is best suited to interpolating frames with highly similar stroke topology, which can be tedious for animators to draw. However, hand-drawn animation is most often *on 2s*, with new drawings on every other frame of the standard 24 frames-per-second rate. Consecutive frames therefore likely differ by articulate motion, frequently involving complex stroke topology changes precluding isomorphism. Mo et al. [225] bypass stroke T-junctions during tracing to parse superficial stroke connectivity, handling

self-contacts and simple occlusion with their learning-based local transformation technique (Section 2.2.1). Unfortunately, strokes are missing or erroneously added when keyframes lack bijective stroke correspondence (Figure 2.19). Both [330, 225] exclusively trace connected strokes, which can cause oversensitivity to natural artistic variation in exact stroke connectivity across hand-drawn frames.

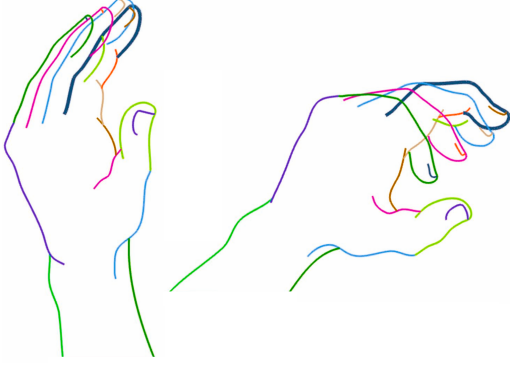


Figure 2.14: Drawing unseen strokes when using stroke visibility toggles. Figure from Yang et al. [346].

at coarser resolutions (Figure 2.16). Still, global topology preservation is not guaranteed and, like all tracing methods, user intervention is needed to correct correspondence errors. To prevent nearby but semantically distinct strokes from moving together, α -close strokes not touching in one or both keyframes are α -disconnected if their proximity differs excessively between keyframes. Though this improves on previous tracing methods’ articulation handling, users may still need to remove superficial connectivity between unrelated body parts. Tracing methods’ imposition of bijective key stroke correspondence also creates extra work for animators in practice. Although animators track individual strokes across frames, natural key stroke relationships are rarely bijective and can require exhaustive fine-tuning to appear as such (Figure 2.19). A prevalent example of this are stroke visibility toggles, used to ensure bijectivity and simulate certain (but not all) stroke topology changes by compelling animators to draw unseen strokes (Figure 2.14). Visibility is automatically modulated over intermediate frames at T-junctions intersecting seen and unseen strokes [330, 344], sometimes alongside user-defined stroke splitting/merging [346], to create immersion-like functionality without explicit layering.

Despite tracing methods reporting substantial time saved for animators in user studies [330, 344, 346], drawing freely costs animators far less energy than intricate interventions like manual stroke matching and drawing unseen strokes. According to Mullery, talented animators are rarely talented machinists, and hand-drawn studios will always choose the former over the latter. Energy spent on anything besides finding the character, maintaining their design, and conveying emotion is likely not energy that hand-drawn studios are willing to spend. Even drawing interventions like unseen strokes may be extrinsic to animators’ artis-

Other tracing methods define α -topology stroke graphs connecting disjoint yet visually cohesive strokes that should move together during animation (e.g., head and facial features) (Figure 2.16, red box at right). CACAI [344] defines k -nearest neighbor α -topology and greedily matches strokes by geometric and α -topological features during tracing. However, matchings do not guarantee global stroke topology preservation and the fixed k may be too small for high-complexity stroke neighborhoods and too large for low-complexity neighborhoods. Addressing these issues, Yang et al. [346] introduce dynamically optimized proximity-based α -topology. They strengthen CACAI’s stroke matching algorithm by requiring local α -topology isomorphism between corresponding key strokes, improving topology preservation

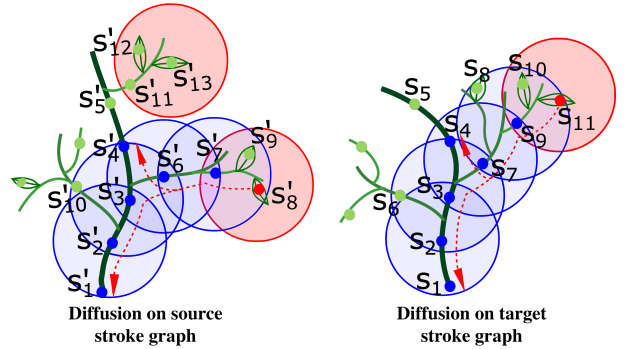


Figure 2.15: Stroke graph diffusion (indicated by red arrows) from Xing et al. [336] comparing respective source and target stroke neighborhoods (red circles) s'_8 and s_{11} . Diffusion compares increasingly distant stroke neighborhoods (blue circles), capturing multiresolution stroke topology on topologically distinct stroke graphs.

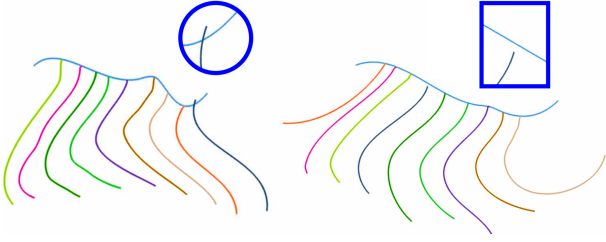


Figure 2.16: Stroke correspondence (color-coded) failure case of CACAI [344] due to lack of global topology awareness. Insets (outlined blue) show stroke topology imprecisions handled by α -topology but not by methods enforcing exact stroke connectivity like Whited et al. [330]. Figure from Yang et al. [346].

[234] use stroke proximity, connectivity, and visual contiguity reasoning to embed a frame’s strokes in a tight fitting α -topological mesh. Differential operators ill-suited to stroke graphs can be stably applied to the mesh, and minimizing the commutativity of the Laplacian for such operators yields correspondences between frames with arbitrary stroke topology and cardinality differences. This method’s main limitations are its run time and its inability to generate intermediate frames. In Xing et al.’s [336] tracing approach, stroke graph vertices represent groups of neighboring strokes, and edges indicate neighborhoods sharing strokes. Tracing diffuses over keyframe stroke graphs to contextualize stroke neighborhoods within global stroke topology, tracking *multiresolution* (local-to-global) geometry and topology similarities and differences between frames (Figure 2.15). Stroke neighborhoods are compared without defining bijections between them, enabling many-to-many stroke correspondence across topologically distinct local neighborhoods. Large-scale topology change is also possible since stroke graph isomorphism is optimally approximated but not required. Recently, Chen et al. [68] use optical flow to simulate rather than explicitly compute stroke graph isomorphism between raster keyframes. To generate intermediate frames, vectorized source strokes are linearly interpolated according to their local optical flows, optimally aligning with their corresponding target strokes. User interaction simply involves redrawing local stroke inconsistencies arising from stroke topology change, a much more intuitive and natural task for animators than interventions used in stricter stroke graph isomorphism methods.

Stroke Graph Matching

Seeking improved correspondence flexibility, several methods eschew formal topological approaches to stroke graphs. Instead, they favor combinatorial and analytical techniques such as graph labeling, graph fusion, and spectral graph theory. These approaches fluently bypass topological constraints like bijective stroke matching and local isomorphism, showcasing the diverse graph processing approaches applicable to keyframe stroke correspondence.

A major advantage of graph matching methods is their ability to output and then optimize many-to-many stroke matchings. Isomorphic stroke graph tracing methods do not accommodate many-to-many matchings at all. To enable one-to-many and many-to-many matchings, Carvalho et al. [51] conducts maximum bipartite matching on keyframes’ guideline strokes and regions. However, this technique performs best on one-to-one matchings, becoming increasingly unreliable in more complex cases. Yu et al. [353] and Song et al. [295] support many-to-many correspondence by respectively utilizing graph learning and graph statistics within similar pipelines. Both methods sample stroke feature points, label points with local rotation- and scale-invariant stroke shape descriptors, generate a many-to-many labeling, and

tic process and slow their workflow. To alleviate need for such interventions, several methods adapt stroke graph tracing and α -topology to accommodate more flexible keyframe correspondences. As with compatible triangulations – another type of isomorphism – the most successful approaches soften isomorphism constraints enough to substantially enhance artistic freedom, but not enough to sacrifice mathematical feature preservation guarantees. Some use graph Laplacian diffusion to process coarser-resolution stroke topology, finding larger stroke group correspondences impossible to detect from 1-hop stroke neighborhoods. Myronova et al.

iteratively prune graph edges to reach optimal correspondence. For optimal labeling, Yu et al. alternates graph transductive labeling – a semi-supervised learning-based graph labeling technique – and distance metric learning; Song et al. alternates maximum a posteriori estimation modeled by Markov random fields and solved by stroke graph cut minimization with maximum likelihood estimation, such that minimizing graph cut minimizes correspondence energy. Alternation continues until the correspondence function converges to an optimal many-to-many labeling. Both methods outperform Hungarian bipartite matching, although neither is evaluated by animators so their efficacy in practice is unknown.

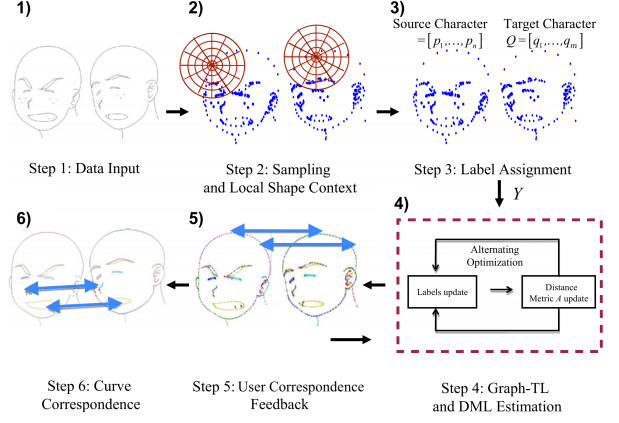


Figure 2.17: Frame analysis pipeline and stroke correspondence results (5-6, blue arrows) from Yu et al. [353].

Spectral graph theory is becoming increasingly popular in hand-drawn animation technology for its flexible yet powerful global topology awareness. For instance, spectral analysis can reconstruct the Delaunay triangulation of a point set from graph matrix data alone [296]. A suite of methods called Patch Alignment Frameworks marry spectral graph theory with manifold learning to compute stroke graph correspondence. Keyframes’ stroke feature vectors given by local shape context descriptors are mapped into a single high-dimensional feature space, where either k -nearest neighbor clustering [352, 195] or sparse coding [322] is used to construct patches of similar source and target stroke feature vectors. Semi-supervised patch construction allows users to add correspondence constraints [352]. Patches are then aligned via spectral analysis and projected into lower-dimensional clusters with respect to shape and proximity data, where they are matched within each cluster to find stroke correspondence. Nearest-neighbor matching naturally accommodates many-to-many correspondence [352]; Hungarian bipartite matching requires reformulating many-to-many matching as one-to-one matching to achieve comparable flexibility [322, 195]. This technique offers substantially relaxed stroke correspondence constraints compared to tracing methods, yet provides far fewer opportunities for manual correction and falters on exaggerated motion and occlusion.

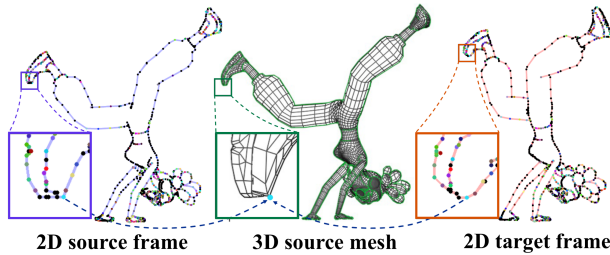


Figure 2.18: Stroke vertex matching between synthetic source and target frames and a 3D character mesh posed to match the source frame from Siyao et al. [292].

To better handle occlusion, Siyao et al. [292] train a Vertex Correspondence Transformer on synthetic line drawings generated from 3D character meshes (Figure 2.18). The Transformer processes pixel data, stroke vertex position, and spectral stroke graph topology. Vertices with high-confidence correspondences between keyframes propagate their articulate motion vectors to nearby unmatched vertices, completing vertex correspondence by leveraging spectral methods’ efficacy on incomplete graph data. An occlusion-handling visibility mask predictor tracks 3D mesh vertices alongside their projected 2D counterparts, yielding correspondence robustness unattainable from 2D data alone. This method performs competitively against other learning-based frame interpolation methods like Chen et al. [66] and outperforms all comparison methods in perceptual tests. However, authors note that strokes often degrade in intermediate frames, a common issue for learning-based frame interpolation methods

[66]. Despite this limitation, many aspects of this approach are compelling and will likely inspire future methods combining topological and learning-based correspondence techniques.

Region Adjacency Graphs

Despite the intimate relationship between strokes and regions (Section 2.2.1), regions are the more stable frame feature (Figure 2.19). Regions always explicitly correspond to ground-truth character design, whereas strokes more often indicate body contours appearing at particular poses and view angles. Stroke methods’ reliance on isomorphism likely arises from inherent challenges predicting and semantically interpreting stroke emergence and evolution over articulate motion; contrastingly, the clear relationship from volumetric body structure to region shape and connectivity often makes regions easier to identify. Hand-drawn characters comprise a ground-truth set of regions, each typically symbolizing a body part. Ground-truth region connectivity then strongly correlates to skeleton structure. Region shape change, connectivity change, and collapse respectively indicate volumetric motion, partial or full occlusion, and viewpoint change. Therefore a character’s *Region Adjacency Graph* (RAG) – comprising region vertices and edges indicating region adjacency – deeply relates to their depicted volumetric pose. This connection parallels Brodt & Bessmeltsev’s association between 2.5D and 3D deformation fields (Section 2.2.1). Still, the relationship may not be straightforward in practice. According to Keane, hand-drawn characters’ volumetric forms are frequently subservient to their symbolic depictions. Region shape and RAG topology can thus deviate unpredictably from projected 3D character pose (Figure 2.12).

Early RAG correspondence methods exclusively exploit local topology and geometry. Nascimento et al. [235] compare functions encoding 1-hop RAG neighborhood topology and geometry to recursively track regions over frames, handling mild occlusions but failing to correctly track similarly shaped regions. Sato et al. [279] formulate correspondence as quadratic programming graph matching, comparing graph vertices by their corresponding regions’ length, area, and midline angle similarity. Despite its simplicity, this method achieves good results on frames with highly similar RAGs, though fails on RAG topology change. Recent methods’ integration of intermediate and global topology awareness improves results. Similarly to stroke graph tracing methods (Section 2.2.2), Miyauchi et al. [223] greedily trace keyframes’ RAGs simultaneously to match regions by shape, local topology, and depth cues. Depth cues provide intermediate topology awareness, as they are a consequence of global 3D skeleton pose. Comparative tests of different cue combinations show shape alone performs far worse than shape and topology together, with depth cues only aiding region matching when correct. The authors extend their existing approach to incorporate 2D skeleton estimation, which they use to improve depth cue accuracy and provide global topology awareness by relating skeleton and RAG topology [224]. Better intermediate and global topology awareness improve their prior results. Liu et al. [197] also employ greedy RAG tracing, attempting to

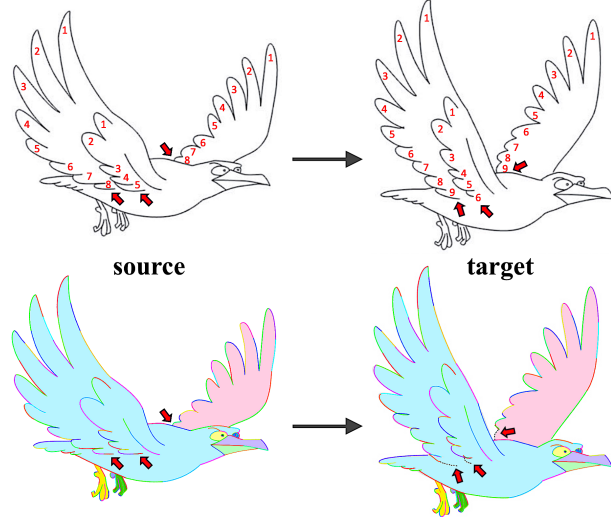


Figure 2.19: Results from Mo et al. [225] illustrating stability of strokes vs. regions over articulate motion. Top: input keyframes with corresponding strokes numbered and differences indicated by arrows. Bottom: color-coded many-to-one stroke and one-to-one region correspondence, with dotted lines showing strokes missing (indicated by arrows) from the output target frame.

reinforce their local region geometry- and topology-matching scheme with global RAG topology awareness. Their matching scheme involves bipartite matching on keyframe correspondence graphs – one for regions and another for 1-hop region neighborhoods – edge-weighted with Kendall shape similarity measures. Kendall measures are isometry-invariant for robust shape matching over articulate motion, but computable only for shapes with equal numbers of sample points, causing shape detail loss for larger regions. As in stroke correspondence, bipartite matching also precludes one-to-many and many-to-many correspondence.

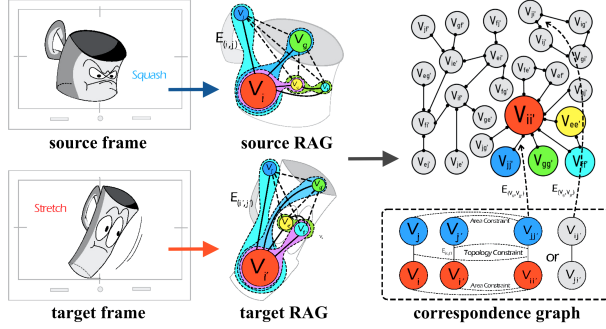


Figure 2.20: RAG and correspondence graph construction from Liu et al. [198].

this graph outputs region correspondence maximizing shape and topology similarity. Comparison testing between both of Liu et al.’s methods [197, 198] indicates global topology comprehension via spectral analysis outperforms greedy tracing in correspondence accuracy. However, both methods’ focus on 1-hop RAG topology causes difficulties when correctly matching corresponding regions with geometrically and topologically distinct local neighborhoods, which frequently arise in articulate motion (Figure 2.4). Matching labeled source and unlabeled target regions, Chen et al. [70] minimize harmonic energy on a connected graph of keyframe region vertices with shape similarity edge weights. Optimal correspondence is reached when each source region’s label equals the vertex-wise graph Laplacian, which is the average of the adjacent target regions’ labels with respect to incident edges’ shape weights. To aid shape- and symmetry-aware global topology comprehension, users label semantically equivalent regions (e.g., left and right feet). Mixed integer quadratic programming optimizes correspondence between topologically similar but geometrically distorted RAG subgraphs, supplying additional global topology awareness. Overall, spectral RAG methods appear more robust than spectral stroke methods (Section 2.2.2). This phenomenon may arise from regions being more stable than strokes during articulate motion, alluding to the possibility that spectral methods capture part of the relationship between RAG topology and volumetric body structure.

In addition to spatial multiresolution analysis, temporal multiresolution analysis also proves useful for region correspondence. Such methods match regions over sets of more than two frames, leveraging this increased data availability to handle complex region topology changes. For a set of consecutive frames, Zhu et al. [365] define a correspondence graph with vertices for potentially

Given greedy tracing methods’ instability from lack of multiresolution topology awareness (Section 2.2.2), recent RAG methods turn to spectral RAG analysis for better results. Continuing ideas from [197], Liu et al. [198] define a graph with a vertex for each source and target region pair, weighted by Kendall shape similarity. Edges connect vertices whose source and target regions are adjacent in source and target RAGs, respectively, only if source and target regions’ respective shared boundaries are of the same connectivity type (Figure 2.20). Edges are weighted with Kendall shared boundary similarity. Spectral decomposition on

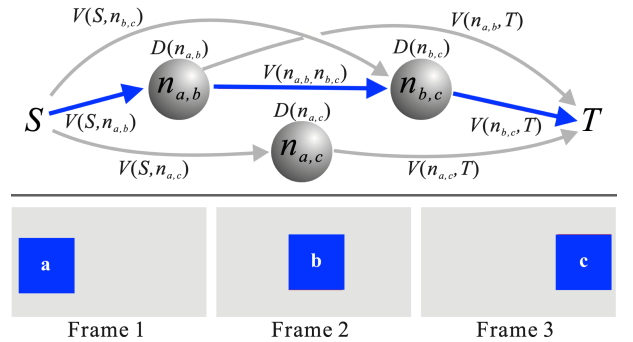


Figure 2.21: Approach of Zhu et al. [365]. Top: network flow on correspondence graph. Bottom: consecutive frames matching the network flow.

corresponding region pairs in source and target frames, and edges between vertices when one vertex’s source region equals another vertex’s target region. Regions correspond when a globally optimal network flow passes through their representative vertex (Figure 2.21). Network flow splitting and merging effectively models one-to-many and many-to-many correspondences. Graph vertices are weighted with isometry-invariant shape similarity, best suiting this method to rigid region transformations. Analyzing several frames at once provides superior occlusion handling than a similar approach analyzing only frame pairs [199]. Although region topology is not considered, it could potentially improve this technique’s region tracking over long-range occlusions by adding ground-truth body structure awareness. Maejima et al. [210] is the only method to date marrying spatial and temporal multiresolution analysis, spectrally matching target and multiple source frames’ RAGs. This approach addresses region data loss arising from occlusion, and correspondence ambiguity arising from articulate RAG topology change. Results demonstrate that as keyframe RAG topologies increasingly differ, more reference frames are needed for correct correspondence. Future spectral methods may be able to avoid complex correspondence graph constructions by simply considering more frames during region matching.

2.2.3 Volumetric Modeling

Volumetric modeling methods leverage stroke and region topology and geometry cues to build 2.5D or 3D representations of hand-drawn characters. Such methods echo human visual perception of cohesive volumetric body pose exclusively from the uncolored strokes and regions comprising hand-drawn frames. Many methods therefore use psychological *Gestalt cues* – especially stroke continuity, proximity, and closure – to assess depth relationships between features. Cues are used to detect stroke junctions indicating depth changes and complete open curves defining region boundaries. Volumetric modeling can be approached in many different ways, from depth-ordering strokes (Section 2.2.3) to generating 3D character models and poses (Section 2.2.3) to “sculpting” volumetric forms from hand-drawn strokes (Section 2.2.3). Such techniques serve to generate depictions of characters at novel view angles, automatically layer depth-inequivalent features, build 3D character models, and give animators volumetric understanding of frames as they draw. This section examines topological structures and techniques utilized by each of these approaches and discuss challenges inherent to volumetrically modeling hand-drawn 2D characters.

Stroke Depth Ordering

Besides manual layering and simulating occlusion with visibility toggles, the stroke methods reviewed thus far exhibit virtually no depth awareness. Yet stroke depth ordering is an active research field, producing techniques for automatically depicting hand-drawn characters at novel view angles, informing 3D character model generation, and automatically structuring input frames into layers. Stroke depth ordering is a challenging task, with some approaches involving automatic recognition and processing of Gestalt depth cues thought to play pivotal roles in human visual perception. These necessitate stroke graph formulations topologically distinct from those used for deformation and correspondence. Other stroke depth approaches, particularly viewpoint generation methods, parametrize stroke depth relationships across viewpoints within topological spaces.

Rivers et al.’s [269] influential viewpoint generation approach leverages relative stroke positions and user-defined partial stroke depth orderings from select key view angles. Novel views then are generated by interpolating stroke shape, position, and depth within parametric viewpoint space. Despite its lack of stroke merging and region overlap handling, this simple approach to viewpoint generation inspired many similar methods [136, 130, 181, 125], some incorporating 3D mesh awareness [170, 200]. Fukusato et al.

[126] extend Rivers et al.’s framework to generate novel viewpoints for view-dependent character designs, compatibly triangulating each region across its view-dependent depictions to ensure their smooth interpolation. A roll dimension is added to Rivers et al.’s original pitch and yaw parametric viewpoint space (Figure 2.22) to improve results. View-dependent character designs are exceedingly common in hand-drawn animation because artistic stylization increases likelihood that geometries drawn in certain poses may not be possible in others. Even if they are possible, their visual meaning may be unclear to the viewer, prompting pose-specific redesigns (Figure 2.23). Because frames themselves are discrete, most view-dependent design changes are discrete rather than continuous and often cannot be captured by simple stroke or shape interpolation. Moreover, these methods only permit topological events resolvable by stroke depth ordering and therefore may not accommodate symbolic view-dependent geometries, as these are rarely expressible as consistent 3D geometries (Figure 2.23). To model discrete view-dependent geometry changes, Coutinho et al. [82] define a hierarchical 2.5D skeleton and associate to each bone the set of hand-drawn depictions of its corresponding body part across various local view angles. Posing the skeleton causes each bone’s strokes to deform towards the depiction best matching its current local pose, mix-and-matching discrete depictions of local body parts to generate novel poses and views (Figure 2.22). Still, articulation is limited since multiple depth values cannot be assigned to the same stroke.

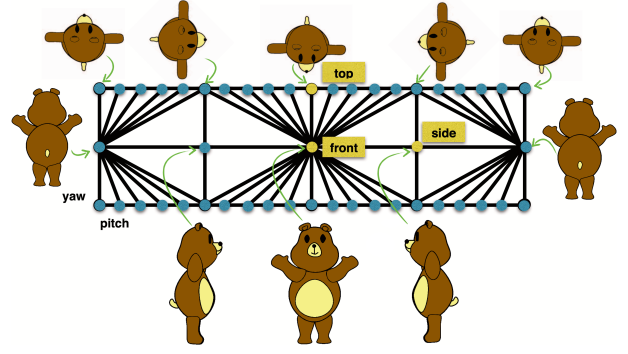


Figure 2.22: Results of Coutinho et al. [82] positioned according to a pitch-yaw parametric viewpoint space adapted from Rivers et al. [269].

Many methods seek to define stroke graph data structures capable of representing depth. A critical goal of such techniques is to model strokes as planar immersions – rather than embeddings, as in Sections 2.2.1 and 2.2.2 – so that strokes bounding depth-inequivalent adjacent regions can be depth-equivalent along the regions’ shared boundary (Figure 2.27, top). Immersion greatly benefits articulate pose representation: for instance, an arm can be in front of the torso while connected to the torso at the shoulder. Early attempts to model such poses with standard stroke graphs are unsuccessful. Sýkora et al. [303] estimate stroke depth by segmenting regions with LazyBrush [306], with user-defined partial depth orderings on regions serving as color labels. Topological sorting gives a consistent depth order from the resulting depth graph, which can be propagated across consecutive frames via ARAP registration of moving components [302]. However, it is unclear how character motion inducing depth order change might be handled. Liu

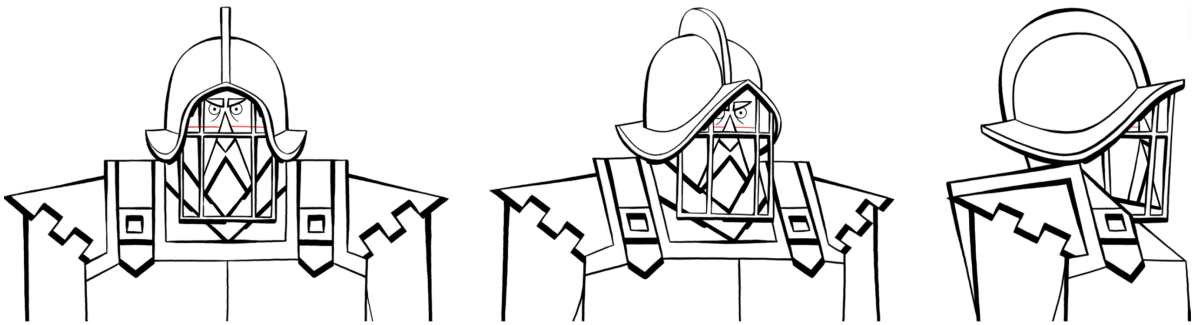


Figure 2.23: Example of view-dependent geometry in a hand-drawn animated character design, sourced from *Wolfwalkers*. The helmet face guard has a center bar in the side profile design (right), but not in the front (left) and partial turn (center) designs.

et al. [199] similarly utilize topological sorting on a depth inequality graph, incorporating T-junction stroke depth estimation and network flow-based region correspondence between frame pairs to improve performance. Cordier et al. [81] estimate stroke depth by combining Gestalt cues with topological depth cues devised for modeling impossible figures, an approach potentially suited to modeling symbolic view-dependent geometry. However, their graph of depth-equivalent stroke groups limits 3D model generation capabilities (Section 2.2.3). Both methods require depth-equivalent region interiors and boundary strokes, precluding smooth connections between depth-inequivalent components.

Later methods overwhelmingly employ half-edge stroke graphs to model articulate stroke immersions. Originally devised for planar embeddings of polyhedra, half-edge graphs replace each stroke edge with a pair of oppositely-oriented *half-edges* which can assume different depth values (Figure 2.25a). Half-edge graphs’ boundary operator on interior half-edges describes region depth; their boundary operator on exterior half-edges describes adjacent regions’ pairwise depth relationships. Unlike stroke layering schemes, which are inherently depth-discrete, half-edge graphs easily model articulate stroke immersions, allowing continuous stroke tracing across depth levels. Methods automatically constructing half-edge graphs from frames ubiquitously rely on Gestalt stroke depth cues. Some detect Gestalt junctions and cusps by applying Laplacians on strokes, with stroke-wise diffusion rates signifying local continuity change [303, 349]. Others combine Gestalt cues with stroke graph tracing, sometimes incorporating awareness of region containment relationships for better detecting foreground-background relationships [104, 265]. Jiang et al. [169] attach occluding surfaces to foreground half-edges, simulating occlusion-aware stroke topology change while keeping stroke graph topology static. Despite this unique and important decoupling of stroke and stroke graph topology over articulate motion, users must re-define strokes’ intricate occlusion relationships for each frame with novel stroke graph structure, likely dissuading industry use.

More automatic half-edge approaches jointly detect stroke depth with Gestalt cues and partition line drawings into their depth-equivalent components, either by graph-cut on Delaunay stroke triangulation [349] or medial axis region detection [105]. These techniques are deeply related, as Voronoi diagrams are bounded by medial axes [90]. By topologically sorting a graph of stroke junctions, Yeh et al. [349] globally optimizes region depth order consistency over local stroke groups’ intersection angles and respective areas. Entem et al. [105] recursively subdivide half-edge stroke graphs of input frame components until every subgraph exclusively comprises depth-equivalent strokes (Figure 2.24). Subgraphs’ recursive processing order is decided by matching their stroke geometries to Gestalt depth cue cases. Although both methods output half-edge complexes of depth-equivalent regions, Yeh et al. [349] is designed specifically for

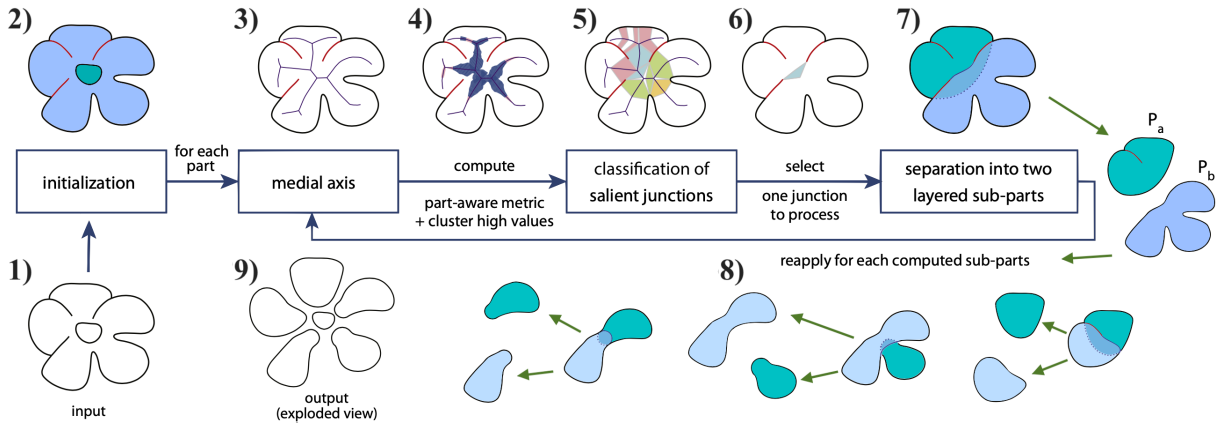


Figure 2.24: Stroke depth processing pipeline of Entem et al. [105] recursively partitioning strokes into depth-equivalent stroke groups via medial axis and Gestalt cue comprehension.

animating hair strands, while Entem et al. [105] processes arbitrary line drawings.

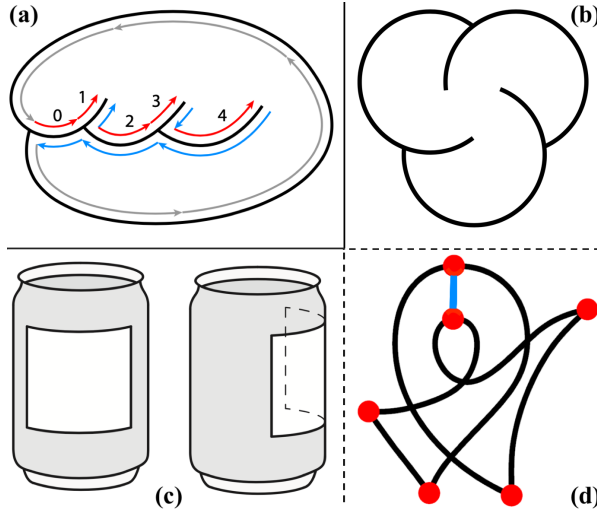


Figure 2.25: (a) Half-edge graph modeling self-overlap without internal hole [105]. Half-edge graph failure cases (b-d): (b) cyclically-overlapping regions [105], (c) region (soda can label) whose visibility changes with view angle [169], (d) self-overlap with internal hole, successfully modeled by subdividing strokes (blue edge) [86].

Similarly to Dalstein et al.’s stroke complexes (Section 2.2.1), half-edge stroke graph methods’ limitations arise from exchanging artistic expressivity for topological and computational simplicity. Methods searching for globally consistent depth orderings expect adjacent regions to have simple foreground-background relationships, failing to accommodate complex depth orderings like intertwined regions [349], view-dependent region visibility changes [169], and cyclically-overlapping regions [105] (Figure 2.25). Gestalt cues’ exclusive sensitivity to local depth relations and obliviousness complex depth orderings emerging at intermediate and global scales may compound these challenges. Extending Gestalt comprehension to coarser stroke resolutions may help resolve some of these failure cases. Although Entem et al. can model self-overlapping regions without internal holes (Figure 2.25a), half-edge methods cannot model self-overlapping regions with internal holes [86, 169, 105] (Figure 2.25d).

Strokes can be subdivided and assigned different

depths as a work-around, but inconsistencies may present during animation [86, 169]. Such limitations arise because half-edge methods assume that depth-inequalities exclusively occur at adjacent regions’ shared boundaries, trading complex depth processing for algorithmic ease and straightforward Gestalt comprehension. Still, there is no apparent reason why half-edge graphs and Gestalt cues could not be differently combined to handle complex cases. Future work in this vein would greatly enhance usability in industry settings, since the distinction between cases like Figures 2.25a and 2.25d is near-meaningless to animators while drawing and should not require complex work-arounds. As with certain topological operations on stroke complexes, these cases highlight the confluence and tension between computational topology, computability, and animators’ drawing processes.

3D Character Modeling

Hand-drawn characters’ symbolic and view-dependent geometry, as well as their tendency to adopt physically impossible poses, render them inherently resistant 3D modeling. Despite its challenges, the many applications of 3D character modeling – from estimating 3D character pose to creating digital versions of character reference sculptures used by 20th century Disney animators – make it a vibrant research area. Only methods using computational topology data structures and techniques to generate 3D models from uncolored line drawings are discussed in this subsection, covering a small sliver of an expansive field. Most methods reviewed explicitly generate 3D character meshes from input frames; others estimate frames’ depicted poses and use them to pose existing 3D models.

Several methods generate 3D models by fitting an existing character skeleton into an input frame, defining a skeleton graph planar embedding with respect to stroke Gestalt cues. For an input frame and a correspondingly posed 3D character skeleton, Bessmeltsev et al. [24] first segment the drawing into strokes overlaying each body part sub-skeleton. Strokes are then further segmented into individual bones’ occluding

contours by defining a Voronoi diagram for the sub-skeleton graph’s planar embedding. Self-occluding strokes are handled by Gestalt continuity consistency checks. A 3D surface-of-revolution mesh is generated around each curved medial skeleton bone, preventing unnatural cylindrical body part modeling. Users can refine the generated 3D model by drawing the input 3D pose at other view angles (Figure 2.26). A similar technique, enriched with strict bone length foreshortening cues, automatically poses existing 3D character models from input gesture drawings [25]. By fixing bone lengths, 3D joint positions are uniquely determined by 3D bones’ projected lengths. This enormously simplifies joint location space, which in turn computationally facilitates the respective 2D skeleton embedding and 3D pose estimation optimizations. Here, the trade-off between computability and artistic expressivity is significant. Animators constantly break realistic foreshortening cues to better preserve character design, express emotion, heighten visual appeal, and improve visual communication with viewers, hindering this method’s applicability. Often, underlying pose is more reliably indicated by character body topology than by bone length (Figure 2.12). Favoring body topology over foreshortening, Sketch2Pose [39] adds self-contact awareness to a learning-based 3D pose estimation scheme. To find regions involved in self-contacts, SMPL [204] body models are aligned with input frames before computing intersecting convex hulls of non-adjacent skeleton subgraphs. Because 2D self-contacts do not always indicate 3D self-contacts (e.g., a hand in front of, but not on, the hip), flexible foreshortening is estimated by statistically correlating bone angle distributions from the training frame data set and a motion capture dataset. Despite its many benefits, this method may fail on highly stylized hand-drawn characters. It particularly struggles on poses drawn more symbolically than realistically, which may interfere with the initial SMPL alignment phase. Jain et al. [167] offer an even more topological approach with three increasingly abstract pose representation modes: full 3D models, generic 3D bone cylinders, and graphs encoding connectivity relationships between entities depicted in hand-drawn frames. By focusing on abstract topological data and only including geometric data when feasible, this approach can model highly symbolic and volumetrically inconsistent hand-drawn characters. Its major drawback is intensive manual intervention during skeleton fitting: users must annotate each frame’s skeleton and body parts, then search a motion capture database for motion resembling the input animation. This extent of user intervention would likely deter usage in industry settings.

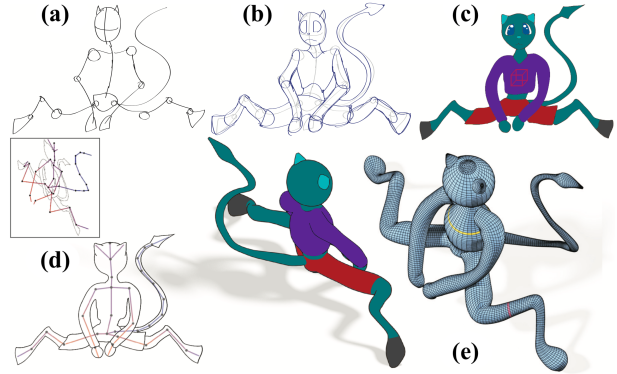


Figure 2.26: 3D modeling results by Bessmeltsev et al. [24]: (a) hand-drawn 2D skeleton; (b) sketched and (c) colored 2D frame; (d) input 3D skeleton and strokes extracted from colored 2D frame; (e) output 3D mesh character model.

A common 3D model generation technique is inflation, where 2D region meshes are deformed into 3D volumes and fused into character models. Inflation methods typically employ stroke depth comprehension to handle articulate body structure. Generally, more sophisticated depth comprehension yields better modeling results. Buchanan et al. [44] compute a medial axis-like skeleton for a character silhouette, then use a simple and effective polygonal surface-of-revolution estimator to generate a 3D mesh. Unfortunately, their lack of stroke depth comprehension wholly prevents articulate body part modeling, placing strong restrictions on the types of drawings producing acceptable outputs. While methods utilizing standard stroke depth graphs fare better, their limited depth representation capabilities cannot fully resolve all modeling challenges. Cordier et al. [81] use a standard stroke graph of depth-equivalent stroke groups to

generate 3D models of mirror-symmetric shapes from line drawings. Stroke groups’ component models are created separately, then fused together. While this technique models a variety of complex articulate shapes, foreground components must be flattened to intersect adjacent background shapes’ exterior contours, creating undesirable component mesh shapes (Figure 2.27, bottom). Olsen et al. [241] use edge-weighted RAGs to encode pairwise region depth relationships estimated from Gestalt cues, then employ Delaunay-based inflation to generate models. Because they encode region and stroke depth relationships in standard – rather than half-edge – stroke graphs, articulate body part modeling may require user intervention (e.g., Figure 2.27, top). Parakkat et al. [248] similarly exploit the relationship between Delaunay triangulations and convex hulls [90] to create a simple yet effective inflation technique intended for automatically shading hand-drawn characters. Once components are modeled separately, they are attached with respect to user-defined standard stroke graph depth ordering. Again, the familiar impediments to modeling of articulate poses with complex inter-component depth relationships arise. Borosan et al. [35] use medial axis partitioning to segment silhouettes into cylindrical component shapes with their own internal skeletons. Components are joined by manually connecting their local skeletons, enabling construction of complex rigged character meshes. Still, lack of stroke depth comprehension causes all drawn components to be depth-equivalent, with articulation made possible only by posing the generated rig after modeling.

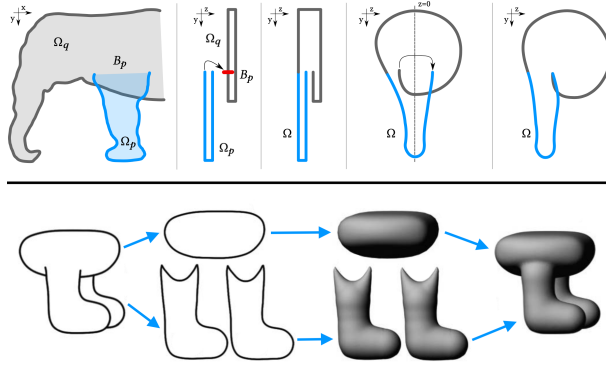


Figure 2.27: Inflation with different types of stroke graph depth comprehension. Top: Topologically “stitching” together components related by depth (in)equality from Dvorožňák et al. [98]. Bottom: Fusing components modeled with standard stroke depth graphs by manipulating component shape from Cordier et al. [81].

Inflation methods utilizing half-edge graphs or similar stroke depth (in)equality comprehension exhibit markedly better performance. Some use stroke graph tracing and Gestalt cues to automatically construct half-edge graphs from input frames as described in Section 2.2.3 [104, 265]. Others semi-automatically construct half-edge graphs from stroke annotations [98] or stroke drawing order [104]. The latter approach assumes strokes are drawn in depth order, constraining users’ artistic process. Once a half-edge graph is constructed, inflation can be approached in various ways. Ramos et al. [265] compute 2D normal vectors at points sampled along each component’s contour, then assess component geometry to convert these into a 3D normal field. Component volumes are connected by interpolating surface normals with respect to

half-edge depths, yielding smooth depth-equivalent junctions connecting depth-inequivalent components. Other methods approach modeling as a global non-linear optimization problem, offering better results than modeling components independently before fusing them together [99]. From stroke depth-annotated input drawings, Dvorožňák et al. [99] use Ink-and-Ray [305] to generate layered component meshes topologically connected by stroke-wise depth (in)equality constraints. Jointly optimizing inflation and shifting functions globally inflates, reshapes, and repositions all component meshes to slot together with minimum discontinuity while respecting (in)equality constraints. However, only bias relief meshes are generated rather than full volumetric models. Extending this framework, Dvorožňák et al. [98] inflate double-sided component meshes into full 3D models while topologically “stitching” together adjacent meshes (Figure 2.27, bottom). Finally, 3D components are ARAP repositioned to optimally align their external contours with their input stroke groups. Optimization on stroke-wise depth (in)equality and contour alignment constraints recomputes when users animate the model, maintaining mesh topology and

component depth order over character movement. Consequently, component depth order cannot change during animation. Since only side profile drawings are accepted as input, articulation is severely limited. Users also report that generated animation is sometimes not what they envision, which is suitable for exploratory artistic work, but could easily frustrate the exacting technical artistic demands of industry settings.

Other methods incorporate both stroke depth awareness and medial axis skeletons into their inflation approach. Entem et al. [104] find the medial axis of each depth-equivalent half-edge stroke group, then inflate and fuse body parts according to their local medial axis shape and global half-edge depth order. This generates smooth surfaces rather than rigged articulate characters, and is limited to modeling side-profile drawings of animals. Ma et al. [208] construct complex rigged articulate character meshes via stroke depth comprehension without employing half-edge graphs. A rapid shape intersection test

estimates pairwise component depth relationships as users draw. To generate component skeletons, each depth-equivalent stroke group is simplified to piecewise-linear strokes whose edges are propagated inwards, adding skeleton junctions where edges topologically split or merge during propagation. The resulting skeleton – a piecewise-linear version of the medial axis – is then simplified and used to further segment the component if necessary, similarly to [105, 104, 35]. Extending Borosan et al.’s approach of cylindrically approximating components and connecting their skeletons, Ma et al. combine skeleton and stroke depth data to automatically determine how to best connect articulate depth-inequivalent components. Prior algorithms including Dvorožňák et al. [98] are used to generate and smooth the final 3D mesh. Ma et al. comparison test their method against Borosan et al. and Dvorožňák et al.’s full 3D modeling approach, finding that they outperform both methods on users’ modeling speed and accuracy. This is likely because Ma et al. use pose-specific stroke depth data to infer pose-invariant articulate 3D skeletons. Contrastingly, Dvorožňák et al. constrain 3D articulation with respect to initial half-edge graphs, while Borosan et al. fail to accommodate complex stroke depth relationships. Although half-edge stroke graphs are not explicitly used, Ma et al.’s success reveals their potential suitability for constructing underlying 3D character skeletons. Alongside their proven efficacy for informing component surface connectivity [265, 99, 98], half-edge stroke depth graphs are powerful data structures for 3D modeling hand-drawn uncolored frames.

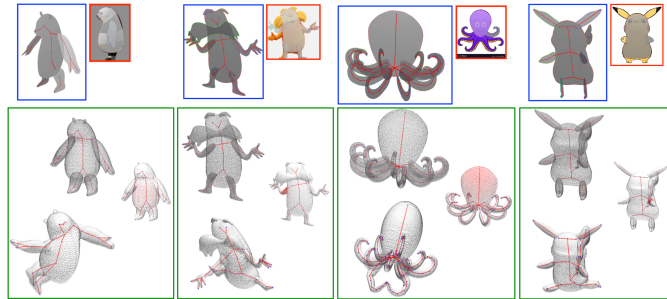


Figure 2.28: Results from Ma et al. [208]: input 2D character drawings (red boxes), 2D component meshes and skeletons embedded in 3D space (blue boxes), and output 3D meshes and skeletons (green boxes).

Sculptural Drawing

Sculptural drawing techniques generate volumetric understandings of hand-drawn characters as users draw, with strokes acting to refine volumetric form. Hand-drawn animators often lack expertise with standard 3D modeling software and techniques, so sculptural drawing offers 3D modeling capabilities within their existing skill set. While the closely-related field of sketch-based modeling is extremely well-developed for non-articulate objects, the nascent field of sculptural drawing focuses on modeling articulate characters.

Sculptural drawing methods heavily rely on user-drawn *construction curves* denoting 3D surface curvature within the 2D drawing plane (Figure 2.29). Projecting 3D curves to the plane creates false

intersections between them, making depth ambiguity a central challenge of sculptural drawing. Andre et al. [6] requires user annotations to parse curve depths at intersection points, while other methods embed intersecting construction curves in different view planes to circumvent the need for annotations [222, 323, 356]. Both Andre et al. and Miao et al. [222] generate 3D surfaces from construction curves by sweeping along one curve’s length while modulating shape with respect to other curves’ contours. Alternatively, Zhang et al. [356] optimally inflate Delaunay-meshed construction curves to satisfy curve shape and bilateral symmetry constraints while preserving mesh topology. As a computational bonus, topology preservation enables the given optimization matrix to be reused to update output when users modify their sketches. Wang et al. [323] update 3D models as users sketch, using construction curves along X, Y, Z -axis cross-section planes to both enhance users’ volumetric understanding of their model as they draw and to generate the model itself. Construction curves along cross-section planes are extruded until they intersect, forming a topologically connected mesh. The mesh is then smoothed and optimally aligned with users’ strokes. Li et al. [189] opt for sparse cross-hatch curves rather than traditional construction curves, solving a stroke graph labeling problem to determine whether each input stroke indicates surface convexity or concavity (Figure 2.29a). Similarly to Ramos et al.’s [265] inflation approach, a 3D mesh is generated via diffusion over a discrete surface curvature field, computed from labeled input strokes. Similar approaches are commonly used throughout sketch-based object modeling [337, 141], but have yet to be widely applied to articulate characters.

Sculptural drawing is not yet a well-developed field, although this may soon change given the recent explosion of interest and success in sketch-based modeling. For now, sculptural drawing methods exhibit several limitations. Because construction curves depict occluding contours of volumetric forms, they are associated with a particular view angle. Sculptural drawing methods exploit this relationship to generate 3D models, often constraining functionality in the process. For instance, Miao et al. rely heavily on view planes, designing their method for characters exclusively comprising bilaterally symmetric body parts. Characters must be drawn in side profile so that all component symmetry planes are parallel. Stroke depth ordering is obtained by tracking users’ stroke drawing order, adding additional usability constraints. Users indicate when components have mirror-symmetric occluded counterparts (e.g., right leg occluded by the left leg), so visible construction curves can be duplicated, reflected, and depth-ordered, a technique also used by Sýkora et al. [98]. Contrastingly, Zhang et al. [356] let users draw and reposition local component-wise bilateral symmetry planes in 3D, supporting characters drawn in myriad poses and ensuring proper component depth regardless of drawing order (Figure 2.30). This method outperforms Borosan et al. [35] on usability and output quality in user trials, requires fewer annotations than Andre et al., and supports more diverse articulate poses than Sýkora et al. Still, it only works on non-frontal view drawings and cannot model non-bilaterally symmetric or intertwined components.

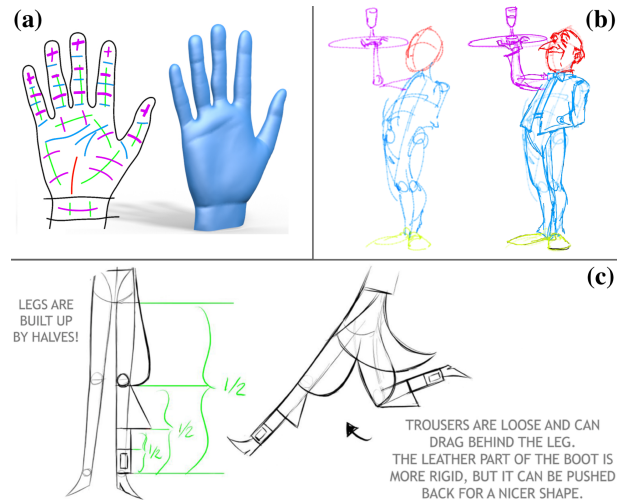


Figure 2.29: Different types of construction curves: (a) sparse cross-hatching and modeling results from Li et al. [189]; (b) example of construction curves (left) vs. detail (right) from Noris et al. [239]; (c) construction curves in professional hand-drawn animated character design, presented with permission from Cartoon Saloon.

Perhaps sculptural drawing’s primary limitation is that hand-drawn characters often cannot be faithfully translated into consistent 3D models. As discussed throughout this review, phenomena including view-dependent geometry (Figure 2.23), subservience of 3D form to 2D symbol, and physically impossible character motion (Figure 2.12) create immense difficulty in volumetrically modeling hand-drawn characters. At times, even hand-drawn animators themselves must confront unresolvable volumetric inconsistencies in character designs as they draw (Figure 2.23). Despite these challenges, sculptural drawing is worthy



Figure 2.30: Construction curves from Zhang et al. [356]. Top: input user-drawn construction curves. Bottom: local component symmetry planes for the input pose (left) and a novel generated pose (right).

of future exploration. Volumetric comprehension of hand-drawn characters could help studios with their colossal frame data management tasks. Some animators, including Glen Keane, hope to integrate volumetric knowledge directly into animation software. Keane believes this would improve his focus on finding the character and expressing himself through line, leaving accessories like clothing and props to be added automatically. Both Keane and Mullery conjecture that sculptural drawing in virtual reality may prove useful to animators in future [274, 209], although existing studio workflows are not yet ready to adopt such technology. To achieve these aims, researchers will have to think carefully about the relationships between hand-drawn frames and the volumetric body poses they depict, especially since such relationships have yet to be fully captured by topological modeling alone.

2.3 Human Pose Estimation

To computationally model professional animators’ RFQ perception, we must be able to assess similarities and differences between frames depicting a hand-drawn character in different body poses. Human pose estimation (HPE) is therefore relevant to this thesis’ research objectives, particularly Research Questions 2 and 3 (Section 1.2). HPE is an expansive research field, a thorough survey of which [59, 362] far surpasses the scope of this thesis. Overwhelmingly, it is dominated by learning-based computer vision techniques, especially and increasingly those employing DNNs.

The vast majority of DNN HPE methods – and DNN computer vision methods generally – comprise layers of multiscale pixel patch processing. Deeper networks with more pixel patch processing layers generally outperform shallower networks, implying that extensive, iterative pixel patch processing underpins these methods’ success [59]. However, this paradigm’s efficacy depends upon input images’ pixel data density and variability at small scales, likely engendering DNNs’ intractable preference for highly local texture-based image features [37]. Real-world photo and video data generally contain considerable pixel-wise variability, offering DNNs abundant local texture data. Contrastingly, hand-drawn frames’ mostly all white pixel patches preclude DNNs’ usual modes of image feature extraction [5, 66], as illustrated by results of pixel patch-based DNN inbetweening methods for uncolored hand-drawn frames, which exhibit poor stroke quality and connectivity preservation [292, 162, 366] (Figure 2.31). In HPE, this issue persists for other types of data-sparse inputs like low-resolution photographs [190], highlighting DNNs’ difficulties leveraging image features essential for human visual cognition (Section 2.4.2). Without these features, DNN HPE methods are prone to overfitting and poor visual abstraction on other computer vision

tasks [354], prospectively hindering their handling of diverse artistic styles across hand-drawn animated films. We therefore constrain our review to HPE methods which, at least in theory, can accommodate hand-drawn frames’ data sparsity and stylistic diversity, focusing on multiresolution data integration and body structure-awareness. These techniques have markedly advanced HPE in recent years, likely because they imbue DNNs with key traits of human vision.

Most learning-based HPE models are either top-down (global-to-local) like AlphaPose [115] or bottom-up (local-to-global) like OpenPose [49]. The former localize skeleton joint keypoints within whole-body bounding boxes, whereas the latter detect and label small pixel patches likely containing keypoints before globally connecting them according to human skeletal structure. Contrastingly, multiresolution data integration supports bidirectional passage and synthesis of visual information between local and global scales, mitigating local pixel patch biases and better emulating visual perception [75, 351, 300]. An early multiresolution integration effort obtained local pixel patches for each body part, before contextualizing them within global body structure and learning body part appearance with a convolutional neural network (CNN) [113]. Stacked Hourglass Networks [236, 103] pioneered repeated layers of top-down and bottom-up processing, outperforming other state-of-the-art models and cementing the importance of multiresolution data integration in HPE research. Tang et al. [308] learn a CNN-based hierarchical pose model by transmitting and synthesizing feature data bidirectionally across local and global hierarchical levels, outperforming state-of-the-art methods on multiple benchmarks. To optimize body part descriptors, Cai et al.’s [47] model learns the scales at which each body part’s salient features appear, then synthesize visual data exclusively across these scales. Scale-selectivity improves efficiency and noise reduction, with this model winning the 2019 COCO Keypoint Challenge. HRNet [300], a particularly influential example of multiresolution data integration, implements parallel subnetworks each processing data at a different resolution. Iterative multiresolution data exchange and synthesis across these parallel networks, outperforming numerous state-of-the-art pose estimators.

As well as integrating data across multiple visual resolutions, propagating local, body part-wise data globally across skeleton models also proves a successful HPE strategy. Global (whole-body) diffusion of local (body part-wise) data helps contextualize body parts within global body structure, constrain physically feasible poses, and identify occluded body parts, emphasizing multiscale visual structure over local pixel features [76, 359, 196]. An initial attempt imbues each joint’s feature map with additional information from nearby joints, exceeding state-of-the-art performance at the time [74]. Many recent efforts employ graph convolutional networks (GCNs), a version of graph neural networks convolving subgraphs of graph-structured data. GCNs’ underlying graphs model various aspects of human skeleton structure, proficiently capturing and propagating dependency relationships between body parts and joints. GCNs can learn to dynamically modulate the direction and path-length of information passage between body parts, adaptively finding optimal information propagation patterns for the given input [357]. Introduced by Qiu et al. [262], Dynamic GCNs synthesize results from joint data diffusion across both nearest-neighbor and fully-connected keypoint graphs [71, 148]. The former encodes keypoint relationships closer to ground-truth skeletal structure, while the latter encodes any kind of keypoint relationship (Figure

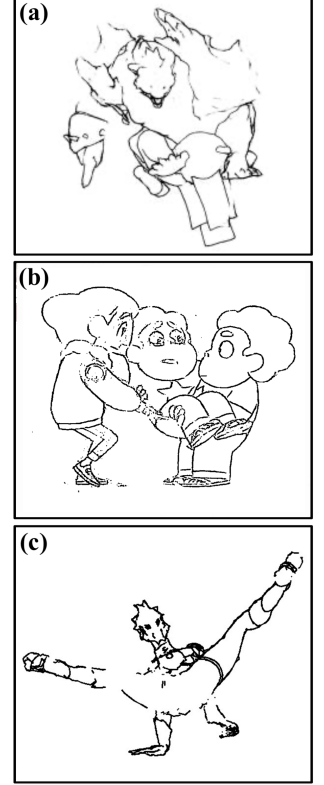


Figure 2.31: DNN-based inbetweening results from (a) [292], (b) [162], and (c) [366].

2.32). Human body structure is learned as both graphs vary over pose and view angle changes, excelling at complex pose and occlusion handling.

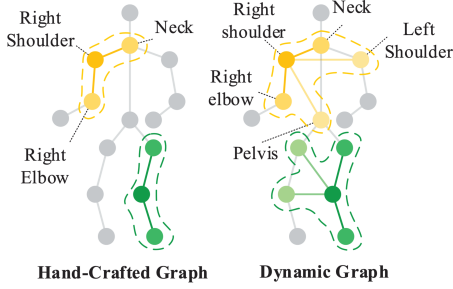


Figure 2.32: Dynamic topology-based body model from [262], capturing ground-truth body topology (left) and superficial topology changes over articulate motion and viewpoint change (right).

Since graphs are topological body models, GCNs dominate this line of inquiry [362]. Qiu et al. [260] combine a heatmap joint locator with a GCN pose refiner to label joints as either visible or occluded, outperforming state-of-the-art methods on complex poses by leveraging this joint (in)visibility information. More ambitiously, Xu et al. [339] attempt to conceptually unify multiresolution data integration, local information propagation, and topological body structure within a single GCN model, learning multiscale skeleton structure and dynamics by propagating visual feature data across distinct local and global levels of a highly deformable skeleton graph.

Unfortunately, as seen with stroke complexes and RAGs throughout Section 2.2, graphs simply lack the topological sophistication needed to fluidly transition between local and global scales within a single model, causing Xu et al. and other GCN HPE methods to employ multiple, separate tactics when attempting multiresolution integration. Moreover, GCNs struggle with heterogeneous graphs, often precluding body part-wise shape comprehension [348], and are restricted to ground-truth skeleton structure, requiring additional functionality to model semantic relationships between body parts [340]. In Chapter 4, we discuss ways to mitigate these limitations while preserving GCNs’ desirable functionalities.

Hand-drawn characters Despite the litany of research on volumetric modeling for hand-drawn characters (Section 2.2.3) and HPE for real-world photos and videos, pose estimation methods for hand-drawn characters are scarce [264, 192]. Research presently falls into two categories, neither of which is suitable for industry-quality hand-drawn animation frames. The first category comprises learning-based methods reliant on color data from fully colored (and sometimes also shaded) characters, which seldom handle uncolored frames [293, 218, 207, 69]. The second category includes methods designed for generic character depictions like silhouette outlines and stick figures, which cannot accommodate hand-drawn films’ complex region-based character designs [25, 321, 144, 315]. The only true contenders for hand-drawn frame pose estimation are Sketch2Pose [39] (Section 2.2.3), HRNet trained on MSCOCO [191] and Human-Art [174] datasets, Brodt & Bessmeltsev’s 2.5D-to-3D mesh topology method [40] (Section 2.2.1), and Chen & Zwicker’s transfer learning framework [67]. Of these, Sketch2Pose is the only fully 3D pose estimator.

While Sketch2Pose is trained on many different kinds of line drawings, our preliminary tests show poor performance on uncolored Cartoon Saloon frames. Sketch2Pose struggles to identify character body parts, handle partial-body frames, and match SMPLs model to character geometry, yielding substantial



Figure 2.33: Human pose estimation methods for hand-drawn animation frames: (a) a method designed to automatically animate children’s drawings [293], (b) Sketch2Pose [39] with accompanying SMPL body models [204], (c) HRNet [300] trained on Human-Art [174], (d) Chen & Zwicker’s transfer learning framework [67]. Tests performed in collaboration with Mohamed Fazil Tajudeen.

pose degeneration (Figure 2.33b). In our preliminary tests, HRNet trained on Human-Art outperformed Chen & Zwicker’s transfer learning framework (Figure 2.33c,d). Though we cannot establish the precise reasons for the methods’ differing results, Human-Art’s size and artistic diversity may help HRNet learn general features of stylized artistic representations. Still, Human-Art-trained HRNet cannot accommodate partial-body frames, which seems to be a general weakness of many DNN-based HPE methods (Figure 2.33a,c). Since partial-body frames are extremely common in hand-drawn animated films, this is a considerable barrier to hand-drawn character HPE methods’ usability in industry settings.

2.4 Visual Perception

As this thesis endeavors to design perceptually-based hand-drawn animation technology, we briefly overview relevant research on human visual perception. Because hand-drawn frames are so data-sparse, we focus on aspects of visual cognition likely vital to hand-drawn frame perception, including lower-level processes like contour and shape comprehension alongside higher-level processes like multiresolution data integration and abstract visual reasoning. Subsequently, we motivate our novel TDA-based approach to perceptual computing by critically evaluating DNN-based models of visual cognition, which are presently the gold standard for perceptual computing. We compare DNNs’ visual data processing capabilities with those constituting human vision, considering whether DNNs are suitable for hand-drawn frame processing, or whether alternative approaches are warranted.

2.4.1 Human Visual Perception

Our discussion of human visual perception research focuses on cognitive processes particularly relevant to line drawing perception, as well as overviews of perceptual studies performed on professional artists. Numerous studies confirm that humans process photographs and uncolored line drawings along nearly identical neural pathways, indicating that many generalized features of visual cognition are equally applicable to line drawings and real-world scenes [289, 201, 319]. These findings extend to children and infants, implying that line drawing perception is innate rather than learned [280, 92, 203]. Studies also show that visual perception of surfaces is both illumination invariant [332] and insensitive to illumination inconsistencies [243]. Researchers conclude that when encountering complex or inconsistent illumination, the visual system assumes that the light source is co-located with the viewer [332], corroborating Hertzmann’s theory of why line drawings are so fluently perceived. Hertzmann proposes that line drawings are perceived analogously to 3D scene renderings in which the light source is co-located with the camera, casting shadows exclusively along surfaces’ occluding contours [158] (Figure 2.3c).

Contours are indeed highly salient for human visual perception. Unlike illumination, contour inconsistencies – including misalignment [4], distortion [318], decomposition [331, 319], and disconnection [46] – significantly impair visual perception. Impairment is more pronounced for global rather than local contour inconsistencies [319, 11] (Figure 2.35). Across line drawings created by many people, occluding contours are depicted more uniformly than other contours [286], demonstrating that they provide especially vital visually diagnostic information. Vision is also sensitive to changes in contour curvature, T-junction geometry, and relative contour and junction placement [46, 318], suggesting that multiresolution contour geometry and topology strongly influence line drawing perception. However, vision readily accommodates global contour distortions [27], indicating that perception degrades with local material feature inconsistencies, but remains robust when coherent groups of material features change uniformly. This trait is likely relevant to hand-drawn frame perception, particularly in cases of squash-and-stretch (Figure 2.12).

Shape sensitivity is a pivotal element of visual cognition [253, 318], fusing lower-level contour information with higher-level abstract visual reasoning, prior knowledge, and symbolic comprehension [213, 277, 155]. Both adults and children easily glean volumetric shape understandings from line drawings of 3D objects [116, 92, 166, 156], comporting with findings that shape – rather than depth or object category – features underpin early visual processing [228]. Humans are also strongly attuned to relationships between contours, explaining medial axes’ salience to visual perception. Medial axis similarity is an effective predictor of shape similarity perception [8, 89, 188], also predicting contour grouping cognition during scene perception [331]. This comports with findings that contour inconsistencies cause perceptual impairment, as such inconsistencies alter shapes’ medial axes. Although perception is impaired when shape features are scrambled (Figure 2.35, middle), scenes remain coherent when viewed upside-down [177], suggesting that contour consistency perception may be isometry-invariant. Human vision is also heavily attuned to global shape features, integrating them with prior visual knowledge during scene perception. Scenes depicting objects in familiar spatial configurations far easier to classify [253, 176, 178, 229], suggesting that line drawings’ preservation of global shape features found in natural scenes may aid visual perception [206, 319, 228].

The human visual system is deeply dependent upon multiresolution integration of visual data.

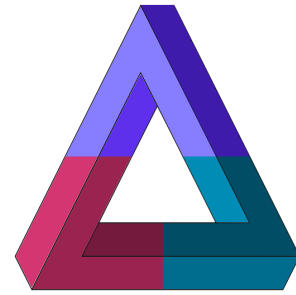


Figure 2.34: Depiction of an impossible object, demonstrating how impossibility perception arises from multiresolution feature integration. Local patches of material features (color-coded) depict physically possible object components, becoming impossible only when perceived collectively.

Visual data initially enters the brain in small, local feature patches [45], coalescing into globally coherent visual understandings via simultaneous local-to-global and global-to-local feature processing and synthesis [65, 45, 129, 310, 205]. Multiresolution integration enables comprehension of both local scene components and global composition of components within the scene, with each modulating perception of the other [253]. Additionally, prior knowledge integration during perception supports comprehension of structural and functional relationships between object parts [212, 164]. Perception thus bidirectionally passes visual data between low-level brain areas processing local, material features and high-level brain areas processing global, abstract features until multiresolution visual and relational consistency is achieved [45, 129]. This is well-illustrated by hand-drawn frame perception: a stroke viewed alone is meaningless, yet is readily perceived as, for instance, an arm’s occluding contour when viewed alongside all other strokes in a frame. Visual perception is impaired when scenes are scrambled, even when local features are maintained (Figure 2.35, middle), demonstrating how local features’ meanings degrade when their global spatial and relational consistency is violated [326, 177, 11]. Multiresolution integration is thought to lend enormous flexibility to human vision, potentially explaining why unnatural visual perspective in stylized art is fluently perceived (Figure 2.12). A new theory proposes that perspective perception is specific to the local material feature patch on which the eye is focused, recalibrating as the eye moves about the image [159]. The global image understanding gained is then abstract, and thus highly flexible under local perspective inconsistencies (Figure 2.34).

Cognitive representations of visual data evolve as they pass between lower- and higher-level brain areas [316], yielding heterogeneous visual understandings with which humans fluently extrapolate missing visual information. For instance, deleting arbitrary global contours from line drawings impedes scene perception [319], but this effect is mitigated when deletion does not alter image features’ medial axes [331]. Inversely, global shape remains perceptually coherent even when local contours are perturbed (Figure 2.35, bottom) [11]. Visual extrapolation also facilitates common perceptual tasks, enabling understanding of scene fragments via contour extension [251] and occlusion handling via contour grouping and shape completion [228, 46, 310, 73]. At the global level, visual extrapolation appears to leverage prior knowledge of natural scene composition, alongside logical scene reasoning [73, 215, 242]. Importantly, extrapolation of local, lower-level features and global, higher-level features appears to operate bidirectionally: both types of visual information modulate one another until optimal interpretations of local and global features are mutually reached [253].

Artists’ perception Because professional artists’ brains are so practiced at focused and precise visual perception, they possess many unique perceptual skills. Laypeople’s shape perception is fully invariant to planar translations and mildly invariant to planar reflection and rotation [252]. Professional artists significantly improve these invariances by practicing drawing [184, 317, 56]. Drawing practice also gives artists specialized multiresolution perceptual acuity [54, 53]; i.e., conscious restriction of visual attention to isolated multiscale parts of an image. Relatedly, artists make superior choices to laypeople regarding visual information included in line drawings, making their drawings especially easy to interpret [185, 138]. However, artists do not appear to have specialized access to visual information captured directly by the retina [256], implying that their specialized perceptual skills arise from mid- and high-level visual processing proficiency. This makes sense, as creating and perceiving line drawings always involves visual abstraction processing [289, 94, 111]. Research suggests that artists are uniquely adept at visual abstraction, and that this skill arises from their superior integration of visual data across different brain regions while drawing [55, 26, 112].

2.4.2 Deep Learning & Visual Perception

Deep neural networks are presently considered the gold standard for effectively any complex computer vision task. Despite their biologically-inspired architecture, a growing body of evidence suggests that DNNs' reported similarities to human vision do not hold up to rigorous scientific scrutiny. Critics argue that DNNs fail to emulate basic principles and characteristics of human visual cognition [37, 213, 212, 166], especially those most essential to hand-drawn frame perception.



Figure 2.35: Figures from [11]: Top: original shape; middle: globally scrambled shape, impairing human recognition but not DNN object classification; bottom: local contour perturbations, impairing DNN classification but not human recognition.

A principal schism between human vision and DNNs is the latter's heavy reliance on non-shape image features. Whereas human vision is enormously sensitive to shape, DNNs appear unable to exploit shape features even when explicitly trained to do so [37, 213]. For optimal shape comprehension performance, even basic aspects of shape cognition must be hard-coded into DNNs [211], implying that human-level shape comprehension is not an emergent property of DNNs. Conversely, and further distinguishing DNNs from visual cognition, human shape awareness is not an emergent property of learning to classify objects, which is precisely how DNNs supposedly learn shape features [212]. DNNs' preference for local texture features over multiresolution shape features is well-documented [140, 133, 38], posing major challenges for processing data-sparse hand-drawn frames [66]. Texture reliance potentially explains DNNs' sensitivity to pixel patch modifications, even those invisible to the human eye [211]. Moreover, metrics commonly used to demonstrate similarities between DNNs' and human brains' visual data representations can be manipulated via pixel modifications [95]. Evidence also suggests that texture fixation prevents DNNs from learning shape-based concepts, often precluding abstract visual reasoning [140].

As hand-drawn frame perception relies considerably on shape comprehension and abstract visual reasoning, these limitations do not bode well for our specific application. Unlike real-world photos and videos, hand-drawn frames' absence of texture information and extreme data sparsity hinders DNN-based frame processing (Figure 2.31) [284, 292], persisting even for colored frames, which also lack texture [66]. DNN performance also evidently degrades on deformed shapes [212, 133, 3], likely impeding their handling of hand-drawn characters' common bodily deformations (Figure 2.12). Moreover, training DNNs to accommodate hand-drawn character deformations is challenging, since Principles of Animation do not reflect physical laws of motion, thereby lacking data regularities which DNNs can easily exploit [162]. DNNs also demonstrate inability to recognize key visual features governing perception of volumetric form from 2D line drawings (Figure 2.36a-c) [166],

which are vital to establishing relationships between hand-drawn frames' material and depicted features. This lack of volumetric awareness prevents six state-of-the-art DNNs from distinguishing between line drawings of possible and impossible objects (Figure 2.34) [156], further diverging from human vision.

Although DNN architectures can be designed to integrate information across multiple resolutions (Section 2.3), DNNs are often insensitive to multiresolution visual features. Whereas human vision retains global shape awareness when contours are locally perturbed, DNNs appear to lack with global shape awareness, struggling to classify shapes under such perturbations (Figure 2.35, bottom) [11, 230]. While scrambling local features impairs human perception, DNNs classify scrambled and unscrambled

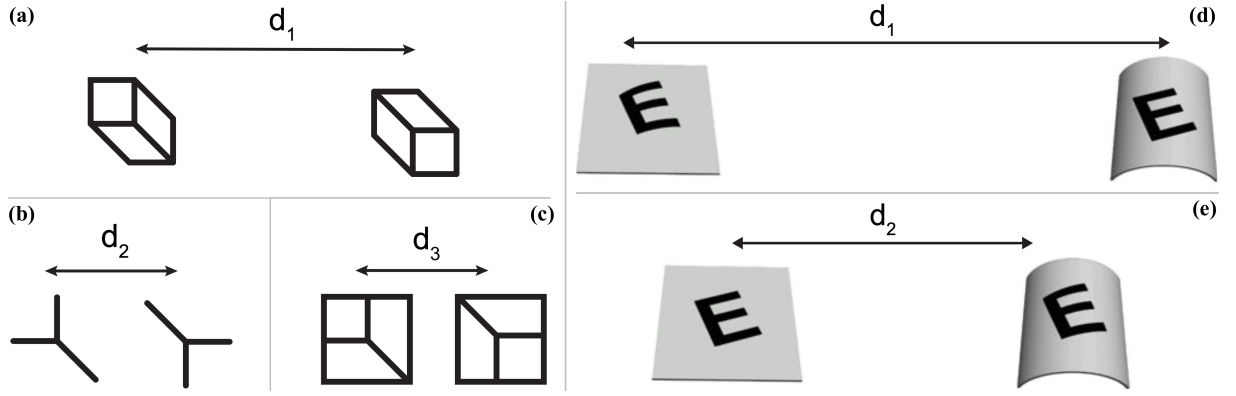


Figure 2.36: Stimuli from [166]. Left: All figures share material features, but due to 3D shape comprehension, humans perceive the figures in (a) as less similar than the respective figures in (b) or (c); DNNs treated figures (a), (b), and (c) as equally similar. Right: Humans perceive figures depicting consistent surface-pattern curvatures (e) as more similar than figures with inconsistent surface-pattern curvatures (d); DNNs do not appear to exhibit this property, indicating that they cannot detect surface-pattern curvature inconsistency.

images with similar ease (Figure 2.35, middle) [11]. Their lack of global awareness is so profound that convolutional neural networks have been approximated by bag-of-local-features models [38]. DNNs also appear insensitive to multiresolution relationships between visual features, particularly those involved in occlusion perception [166, 230], feature consistency detection [166, 156], and natural feature organization within scenes [178]. Strikingly, evidence suggests that DNNs lack sensitivity to relational features even when trained on relevant relationships, or when relational features are uniquely diagnostic to the given computational task [212, 166]. This is likely the cause of their inability to identify impossible objects [156], a task contingent on multiresolution relational processing (Figure 2.34). This also potentially explains DNNs’ difficulties recognizing familiar objects in novel contexts and poses, a task requiring decomposition of globally novel input into locally familiar components [37, 3]. Multiresolution relationship comprehension could potentially address DNNs’ vulnerability to adversarial attacks from nonsense images [237]; while resistance to attacks improves with training, humans do not require training to recognize nonsense images.

DNNs also struggle with abstract, compositional visual reasoning, an essential ingredient in humans’ fluent perception of stylized art. This limitation is thought to arise from DNNs’ difficulties with dynamically accessing and distributing salient information throughout their architecture [140]. Again, this behavior departs from visual cognition, which mobilizes and distributes information across numerous diverse brain areas during multiresolution integration. Representational entanglement is thought to be a key barrier to DNN-based compositional abstract reasoning, whereby salient visual features are not parsed, separately perceived, and abstractly related as they are in human vision. Consequently, salient features cannot be independently modulated or evaluated within DNNs, and inconsistencies between related, entangled features are not easily detected (Figure 2.36d-e). This may preclude DNNs from learning about salient features individually, prospectively explaining their poor visual generalization and abstraction, and therefore their tendency for overfitting [226, 227]. Even if DNNs were attuned to perceptually salient features like contour shape, part-wise relationships, and depicted volumetric form, representational entanglement precludes controllably modeling the complex interactions between these features underpinning visual cognition. While DNNs show improvements when explicitly trained for abstract symbolic reasoning [140, 14], substantial generalization arises when abstract relations are hard-coded into their architectures [15, 57], suggesting that human-level abstract reasoning is not an emergent property of DNNs. Abstraction also improves when DNN ‘neurons’ are endowed with specialized structural traits evocative of dendrites

in the human brain, potentially indicating that DNNs’ standard biologically-inspired designs remain insufficient for faithfully modeling human neuronal activity [58].

2.5 Topological Data Analysis

As we will discuss thoroughly in Section 2.6, hand-drawn animation technology research substantially benefits from sophisticated topological ideas. In fact, particularly successful frame processing strategies often intersect precisely with key elements of Topological Data Analysis (TDA). TDA is a rapidly evolving field with applications far beyond hand-drawn animation technology, as discussed in this section. With respect to our own research, TDA offers topological structures and data analytical strategies enabling dramatic elevation of successful topological approaches to hand-drawn animation technology, including stroke complex analysis, lattice and mesh topology comprehension, Laplacian diffusion, spectral analysis, and DNN integration into topological frame representations. TDA excels in providing sound conversion mechanisms between diverse data representation and analysis strategies, paving computable through-lines between stroke complexes, lattices and meshes, RAGs, point clouds, spectral analysis, Laplacians, and more. Because TDA systematically connects these data analytical settings, data representations can be freely chosen to optimally suit analysis needs, while retaining a more computable version of the chosen representation. Since TDA is rooted in pure mathematics, data characteristics are guaranteeably transferred from the chosen data representation to its more computable approximation. While TDA’s abstract topological foundations offer higher-order, dimensionality-reducing analysis capabilities largely absent from hand-drawn animation technology, it is distinct from classical topology in its incorporation of multiresolution analysis. Like visual perception, this captures critical structural information about how data features interact and combine at varying scales.

Homology Homology is a topological concept foundational to TDA. Computable with simple linear algebra yet connected to deep conceptual and mathematical ideas, homology captures fundamental features of a topological space. Moreover, it relates these features across varying dimensions, providing highly compact and expressive feature description. The features captured by homology are called *invariants* – meaning they are preserved under any deformation besides cutting and gluing – and typically include connected components (0-dimensional features), holes (1D features), cavities (2D features), and so on [102].

The homology of a topological space K encodes all features of dimension n as a mathematical group C_n , connecting them across descending dimensions with a chain of group homomorphisms δ

$$\dots \xrightarrow{\delta_{n+2}} C_{n+1} \xrightarrow{\delta_{n+1}} C_n \xrightarrow{\delta_n} C_{n-1} \xrightarrow{\delta_{n-1}} \dots$$

such that $\delta_n \delta_{n+1} = 0$. This equality guarantees that δ_n maps each n -dimensional feature in C_n to its $(n - 1)$ -dimensional boundary in C_{n-1} : edges map to their bounding vertices, surfaces map to their bounding edges, volumes map to their bounding surfaces, and so on. In fact, δ is the same boundary operator utilized in hand-drawn animation technology research to relate strokes and regions, encode depth relationships between strokes, and relate 2(.5)D and 3D frame features.

More precisely, $\text{im} \delta_{n+1}$ comprises all n -dimensional boundaries, and $\ker \delta_n$ comprises all “loops” of n -dimensional features whose boundaries cancel themselves out – for example, a line segment is bounded by its endpoints, but when formed into a circle, the line segment has no endpoints, and thus no boundary. The n^{th} homology group

$$H_n(K) = \ker \delta_n / \text{im} \delta_{n+1}$$

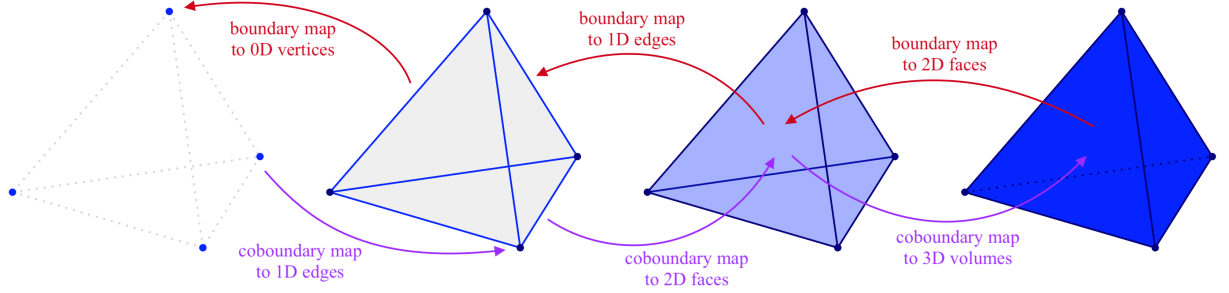


Figure 2.37: Illustration of the boundary and coboundary operators across topological spaces of descending and ascending dimension, respectively.

captures all n -dimensional topological holes in K , enabling the structure of a topological space to be traced across varying dimensions. Later, we will discuss the dual concept of *cohomology*, which is constructed similarly to homology but connects features across ascending, rather than descending, dimensions.

Cell complexes Many TDA techniques are designed for analyzing data represented as cell complexes, presenting another parallel with hand-drawn animation technology research. One such technique is persistent homology, which creates a collection of nested cell complexes from the input point cloud, capturing point proximity across local-to-global scales. Linear maps connect these complexes’ homology groups, enabling multiresolution tracking of multidimensional topological features [50, 102]. Persistent homology is highly popular in practice, with relevant applications including shape comparison [109, 121, 320], HPE and human action recognition [93, 187, 142], and many other computer vision applications [22, 246]. It also benefits from deep algebraic and category theoretic foundations [84, 20, 149], from which emerge many theoretical features desirable for data analysis: sensitivity to input data changes [77], high-quality computable metrics [43], and compatibility with numerous other topological descriptors useful in TDA [343]. Alternative persistence formulations specialize in point clouds with density variations [36, 299], dynamic point clouds [91, 335], and 2D-to-3D shape reconstruction [21, 314]. However, persistent homology often neglects critical geometric data, fails to record features’ spatial distribution, and is ill-suited for continuous data.

Laplacians on cell complexes are another active TDA subfield, dovetailing with the diffusion and spectral techniques increasingly popular in hand-drawn animation technology for their multiresolution analysis capabilities. This is no accident, as Laplacians and their TDA formulations are intimately connected with topological features. In particular, the combinatorial Laplacian $L_n = \delta_{n+1} \delta_{n+1}^T + \delta_n \delta_n^T$ generalizes the graph Laplacian to n -dimensional cell complexes, with $\dim(\ker L_n)$ giving the number of n -dimensional features [153, 150]. Recently, persistence and combinatorial Laplacians have been combined [220, 329] for protein analysis [221, 328] and viral evolution forecasting [63]. Spectral analysis of persistent Laplacians recovers both topological and geometric features of data [299], and can successfully characterize causes of DNN overfitting [175]. Sheaf Laplacians extend combinatorial Laplacians to sheaves – cell complexes with local data attached to vertices, edges, faces, etc. – with recent success in network analysis [355, 32, 13] alongside formidable spectral analysis capabilities [153]. Sheaves are important topological structures discussed at length in Chapter 4. They also happen to formalize hand-drawn animation technology research’s longstanding attempts to imbue bitmap frame stroke complexes with stroke geometry comprehension, without relying on stroke geometry data supplied within vector graphics settings.

Manifolds Although the relationship between Laplacians and the topology of differentiable manifolds has been known for decades [131], it has only recently become computable [88, 299]. This relationship,

formalized with differential geometry by de Rham-Hodge theory, is intimately connected to the discrete TDA concepts discussed above. In particular, geometric descriptors a manifold M can be decomposed into orthogonal components corresponding to elements of M 's cohomology groups $H^n(M)$, explicitly linking geometric and topological data [266, 334]. Moreover, such decomposition is possible only for descriptors that are trivial under the manifold's Hodge Laplacian, which is equivalent to the combinatorial Laplacian [153]. This enables analysis of continuous manifold data within the same discrete, highly computable frameworks used for persistent homology and combinatorial Laplacians, rendering TDA applicable to continuous data analysis tasks like medical imaging, fluid dynamics, and 3D modeling without requiring discretization of naturally continuous data. Applications include molecular analysis [267, 261], GPS analysis [276], manifold learning and topology evaluation [23, 137, 61], lattice and mesh analysis [360, 298, 259], and 3D mesh modeling [110, 134, 364].

Curves Curve analysis is an emerging realm of TDA, grounded in centuries of mathematical work on knot, braid, and link analysis, classifying interactions within and between 2D curves in 3D space. These fields have numerous applications in physics and biology [232], robotics [135, 219], and cryptography [283, 217], and a related concept called the winding number is foundational to computer graphics [90, 216]. Because curve interactions are challenging to compute when highly complex, TDA progress in this domain is very recent. Such advances are promising for hand-drawn animation technology, which often grapples with curve interaction analysis during stroke deformation, correspondence, and depth ordering [85]. A multiresolution formulation of a classical knot invariant was recently proposed, yielding a matrix representation of local-to-global curve interactions alongside prior knowledge specialized to the given use case [285]. The same invariant has been used for polymer [245] and protein [12, 244] analysis. Approaches adjacent to persistent homology are also used for materials analysis [254, 101] and shape analysis [290, 100].

Relational learning Abstract topology studies deformable geometric structures alongside relationships within and between these structures, rendering TDA especially suitable for explainable and controllable deep learning. Most notably, geometric deep learning has exploded in recent years, with equivariant neural networks and GNNs enjoying particular success [42, 48, 20, 307, 147, 41]. Equivariance leverages group theory to imbue neural networks with prior knowledge about data behavior and guarantee desirable data properties in analysis results, while graph learning models the topological structure of pairwise data relationships [246]. With TDA, multiway data relationships inaccessible to GNNs [19] can be modeled on diverse topological spaces including cell complexes [145, 31, 180, 16], sheaves [157, 152, 32], and hypergraphs [96, 119, 9]. Within these spaces, it becomes possible to capture multiresolution features of data relationships [146, 31], augmenting existing capabilities of frameworks like equivariant GNNs [157, 18, 163, 233] and yielding more accurate models of real-world phenomena [19]. Although this thesis does not directly concern machine learning, Chapter 4 explores ideas related to those described above in hopes of furthering TDA-based relational learning in future.

2.6 Discussion

In this chapter, we comprehensively assess the present state of hand-drawn animation technology research. This analysis – alongside knowledge gained from overviews of human pose estimation, visual perception, and topological data analysis – enables us to address Research Question 1 of this thesis (Section 1.2), which we do here and in Section 5.2. To understand present artistic and technical barriers to industry adoption of hand-drawn animation technology research, we identify the field's overarching successes and limitations by critically evaluating our survey findings and knowledge gained throughout this chapter. Our

interdisciplinary analysis illuminates viable strategies for extending hand-drawn animation technology’s successes and mitigating limitations to better accommodate industry needs.

Our review of hand-drawn animation technology research establishes a simple yet profound fact: hand-drawn frames are best computationally formulated as topological spaces. Since frames exclusively comprise stroke/region geometry and topology information, these material features are undeniably vital to visual perception of frames’ depicted features. Accordingly, most hand-drawn animation technology research focuses on relating material features within and between frames. Volumetric modeling methods exploit stroke relationships salient to occlusion perception [46, 73], as well as medial axes’ connections to volumetric shape perception [89, 188]. Likewise, most Deformation and Correspondence methods preserve depicted features by preserving material features, comporting with findings that material feature changes impact line drawing perception [318, 331]. Keane and Mullery describe how animators are also highly attuned to material features as they draw, with Mullery reporting that clean-up animators’ quasi-topological material feature understandings align with popular topological frame processing approaches. Yet both experts emphasize that animators’ material and depicted feature perceptions coexist and continually inform one another, a process increasingly reflected in topological HPE approaches but rarely emulated by hand-drawn animation technologies. With this dual feature understanding, animators effectively preserve characters’ depictions across articulate motion, despite their lack of ground-truth body representations.

To understand why hand-drawn animation technologies often struggle to capture essential relationships between material and depicted features, we turn again to their topological foundations. Most methods exploit very simple topological structures, such as stroke graphs, stroke complexes, and RAGs. Simply put, these structures are not sufficiently mathematically sophisticated or expressive to capture aspects of visual cognition essential for relating material and depicted features. For instance, while half-edge stroke graphs’ depth comprehension provides marginal awareness of material and depicted feature relationships, their lack of multiresolution data integration capabilities yields failure cases (Figure 2.25). These cases are potentially resolvable via global depth reasoning, similarly to humans’ impossible object perception (Figure 2.34). As discussed, depicted features emerge from interactions between material features and the human visual system; thus, without computationally modeling higher-level cognitive processes, technologies often cannot process depicted features alongside material features. While Volumetric Modeling methods advance considerably towards these aims, their topological approaches are often too naive to accommodate articulate motion [104, 98, 35], squash-and-stretch [25], or view-dependent geometry [25, 39].

Without the topological machinery needed to emulate perceptual processes generating material and depicted feature relationships, Deformation and Correspondence methods heavily rely on material feature isomorphism. Since material feature preservation ensures depicted feature preservation, isomorphism acts as a proxy for true depicted feature comprehension, whereby isomorphic features guaranteeably depict the same body part (Figure 2.14). Consequently, frames’ frequent material feature changes often require manual resolution via user interventions like bijective stroke matching, drawing unseen strokes, or redefining stroke depth relationships after each stroke topology change [169]. Mullery discusses how such interventions can cost valuable production time, drain animators’ energy, and draw their focus away from crucial artistic tasks, potentially disrupting studios’ production schedules and budgets. Bijectivity constraints also hinder bipartite stroke match approaches [51, 322, 195] and RAG methods [197, 198], despite regions’ strong relationships with depicted body structure. Although not all hand-drawn animators are interested in automatic frame generation, often preferring to maintain artistic agency over line quality – Keane calls the line “a seismograph of my soul” – a sizable market exists for automatic region colorization, exemplified by integration of Sýkora et al.’s work into TVPaint animation software [306, 304].

To access this and other markets, such as Keane’s desired sculptural drawing capabilities, hand-

drawn animation technology must forgo its reliance on material feature isomorphism and instead seek to establish genuine relationships between material and depicted features, which will support handling of frames’ frequent material feature changes. Such advancements will also invite further computer vision research on hand-drawn frames, which is a key theme of this thesis. By automatically comprehending and comparing hand-drawn frames’ depicted features, computer vision-based frame processing could automatically organize artistic assets by their depicted information, including facial expression, costume, depicted body motion, and camera angle. Such processing methods could significantly ameliorate hand-drawn studios’ immense data management challenges without inhibiting animators’ artistic agency, yielding many potential technologies desirable to hand-drawn studios.

Next, we explore how best to pursue this research objective. Given DNNs’ extraordinary success on myriad computer vision tasks, alongside their recent contributions to hand-drawn animation technology, they may appear well-poised to model perceptual processes governing material and depicted feature relationships. However, DNNs do not leverage key visual features or emulate cognitive processes vital to frame perception, and appear resistant to acquiring these capabilities even when specially engineered to do so. Since hand-drawn frames are so data-sparse, their material features are likely uninterpretable without 2D and 3D shape comprehension, topology comprehension, multiresolution data integration, and abstract visual reasoning; as discussed, these analytical capabilities are often inaccessible to DNNs. DNNs appear highly promising for hand-drawn animation technology when incorporated into primarily topological approaches [40, 225], but perform poorly when used independently (Figure 2.31). Fortunately, existing trends across hand-drawn animation technology research suggest alternative and potentially transformative pathways towards relating material and depicted features in future work.

Naive topological structures effectively capture local material feature interactions like stroke connectivity and T-junction depth awareness, even modeling body part-wise shape comprehension when accompanied by additional handling [344, 208]. However, they cannot capture high-level perceptual processes like multiresolution data integration, functional relationship comprehension, and depicted feature consistency evaluation. These are exactly the kinds of capabilities offered by emerging TDA techniques. Moreover, frame processing methods already achieve better results by elevating their topological sophistication, frequently (albeit likely unintentionally) coinciding with key TDA ideas. For instance, (co)homology theory’s boundary operator mathematically formalizes 2D region formation from loops of 1D boundary strokes. The boundary operator thereby imbues strokes with regions’ superior stability over articulate motion, handling material feature topology change better than other approaches [86, 87]. This both enhances artistic freedom and reveals profound connections between frame perception and stroke complex (co)homology. Half-edge stroke graphs also use boundary comprehension to encode region depth relationships [105, 265, 98, 169]. When accompanied by coarser-resolution structure awareness from medial axes, they support 3D character skeleton comprehension [208, 24]. Here, as in visual perception, heterogeneous material features combine to produce volumetric comprehension of depicted body topology. Moving beyond simple depth relationships, many methods attach local geometric data to stroke/region complexes’ vertices, edges, or faces, aiding stroke-wise transformation [330, 344, 346], lattice topology reconstruction [128], mesh deformation localization [325, 324, 60, 347], and frame feature correspondence [292, 198, 70, 210]. As discussed in Section 2.5, these are examples of sheaves, an increasingly important TDA approach to multiresolution data fusion and relational learning. Relatedly, graph diffusion and spectral analysis are gaining popularity for the fluency with which they relate local and global frame topology [336, 292, 198, 210], mirroring GCNs’ success in HPE. In future, these methods will likely benefit from TDA’s profound fluency within these domains

Alternatively yet relatedly, other successful methods modify topological frame representations enough

to relax keyframe similarity requirements while retaining their feature preservation guarantees, topological operators, and data representation capabilities. These techniques achieve superior balance between artistic expressivity, computability, and powerful mathematical functionality. Such approaches will be particularly critical as hand-drawn animation technology’s topological sophistication increases, likely improving frame processing expressivity beyond capabilities of even the most advanced TDA methods. Examples include recent lattice formulations vastly improving articulation and stroke naturalness [107, 108, 64], coarse stroke graph diffusion approximating isomorphism without stroke bijectivity/connectivity constraints [336], compatible deformation approximation [344, 346], and compatible triangulation reconstruction over extreme deformation and topology change [367]. Although these methods still exhibit certain limitations in practice, they demonstrate how careful and creative adaptations of mathematical ideas can more effectively leverage relationships between computational topology and hand-drawn animation. Best practices for these informal topological frame modeling techniques have yet to be established; to this end, researchers should collaborate with professional hand-drawn animators and study the perceptual processes involved in their work. As showcased in this thesis, collaboration between industry professionals and technical researchers can yield deep insights into how mathematical frame representations can be tailored to best meet hand-drawn studios’ needs.

Specific suggestions for future work along these paths of inquiry are given in Section 5.2; for now, we return our attention to the research presented in this thesis. As discussed, hand-drawn characters’ lack of ground-truth body representations renders them simultaneously visually inconsistent yet physically coherent (Figure 2.23), realistic yet symbolic, materially dynamic yet volumetrically stable (Figure 2.4), governed by yet constantly violating body motion constraints (Figure 2.12). As revealed by present limitations of Volumetric Modeling methods, these qualities are so mercurial and elusive that they may not even admit explicit formulations. Yet without accommodating them in computational frame representations, it is likely impossible to establish perceptually salient relationships between frames’ material and depicted features. We circumvent this paradox by leveraging our observation that depicted frame features are emergent properties of interactions between material frame features and human visual cognition. Our research seeks to model these interactions as a complex system, from which depicted frame features can implicitly emerge, rather than being explicitly encoded within frame representations.

Guided by our analysis of hand-drawn animation technology detailed in this chapter, we begin by considering which perceptual factors are salient to animators’ RFQ perception. Motivated by perception researchers’ growing calls for perceptual computing frameworks capturing empirically established traits of human vision [37, 166, 213], alongside hand-drawn animation technology’s material feature processing techniques, we aim to devise a computer vision system for extracting high-level, global depicted features from local material features salient to low- and mid-level visual cognition. Thereafter, we depart from prior hand-drawn animation technology research, endeavoring to model higher-level visual processes – including multiresolution data integration, part-wise relationship comprehension, visual consistency analysis, and abstract visual reasoning – from which material and depicted feature relationships emerge. Chapter 3 details perceptual studies conducted to ascertain relationships between material and depicted features influencing professional animators’ RFQ perception, guiding our design of perceptually-based TDA strategies tailored to industry needs. Chapter 4 synthesizes this knowledge into our TDA-based RFQ analysis framework which, like human vision, synthesizes local material frame features across multiple resolutions and visual abstraction levels, enabling depicted feature comprehension. Unlike prior volumetric modeling frameworks, we do not attempt to explicitly model relationships between material and depicted features. Instead, we mathematically guarantee stability – formalized as equivariance – between material and depicted features, capturing the feature relationships themselves within system behavior.

Chapter 3

Perceptual Experiments

Despite the wealth of knowledge about human visual perception of line drawings and professional artists’ specialized perceptual skills on generic visual tasks (Section 2.4), controlled perceptual experiments have yet to explore how these perceptual characteristics practically impact professional artists’ perception of their own art as they work. Such experiments are particularly essential for designing perceptually-based technology for high-efficiency, highly technologized collaborative artistic environments like hand-drawn animation studios. As discussed, hand-drawn animation technology researchers’ sometimes incomplete understandings of hand-drawn studios’ artistic workflows and technical needs can beget technologies unattractive to industry professionals.

Hand-drawn animation frames exhibit many properties atypical of line drawings studied in perception research, which usually comprise simple shape depictions [4, 27], rudimentary sketches [319, 289], or unstylized drawings derived from photographs [331, 318]. Moreover, the artistic workflows guiding hand-drawn frame creation require specialized perceptual processes far beyond those recruited for producing line drawings studied in prior research, most notably pose resolution and staying *on-model* (i.e., strictly retaining model sheet design specifications in every frame). Consequently, we cannot assume that prior perceptual studies’ findings faithfully translate to professional animators’ perception of hand-drawn frames’ unique artistic qualities within industry-based artistic workflows. Accordingly, controlled experimental evaluation of professional animators’ hand-drawn frame perception can provide indispensable insights into devising technologies desirable to hand-drawn studios.

This chapter details two perceptual experiments conducted on professional Cartoon Saloon animators, both of which are designed to faithfully emulate animators’ studio working conditions. Through this novel methodological framework, we investigate perceptual factors influencing animators’ judgments of Reference Frame Quality (RFQ), providing information vital for effective RFQ analysis framework design. Our first experiment (Section 3.1) explores whether animators’ RFQ perception is exclusively dependent on characters’ depicted pose, or if it is invariant up to local body part-wise isometry. We find evidence of the latter, demonstrating that animators’ RFQ perception is robust to local material feature changes preserving body part-wise pose resolution strategies, and concluding that jointly comparing frames’ material and depicted features is essential for modeling animators’ RFQ perception. Our second experiment (Section 3.2) investigates whether animators’ perceptions of hand-drawn characters’ depicted volumetric body poses increasingly diverge as frames are increasingly distorted by Principles of Animation, such as squash-and-stretch, bone bending/breaking, and stylized perspective (Figure 2.12). Instead, we find evidence that animators’ pose perceptions increasingly diverge as frame features grow less perceptually consistent, increasingly hindering perceptual synthesis of frames’ visual information into a globally cohesive

pose interpretation. This implies that our RFQ analysis framework must anticipate animators’ subjective pose interpretations when analyzing frames’ depicted features, and therefore must detect multiresolution frame feature inconsistencies from which heightened subjectivity arises.

To conduct both experiments, we first propose perceptually-based criteria for systematically evaluating material and depicted frame features. Motivated by Chapter 2’s discussions of hand-drawn frames’ material features (Section 2.1.2), manifold evidence of contour geometry and topology’s vital roles in visual perception (Section 2.4), the many successes of stroke complex and RAG frame processing methods (Section 2.2), we propose that material frame feature geometry and topology are salient to professional animators’ frame perception. Moreover, Glen Keane and Mark Mullery both assert that characters’ depicted body motions and animators’ own internal proprioception are at the forefront of animators’ minds during hand-drawn frame creation (Appendix B). Our experiments are therefore based upon three classes of proposed *perceptual cues* for hand-drawn animation frames:

1. *Geometric cues*, including body part-wise geometry, region foreshortening, and occluding contours;
2. *Topological cues*, including region connectivity and region presence/absence;
3. *Motion cues*, indicating the body weight distribution needed to physically realize a character’s depicted pose and/or body movement.

While our experimental findings suggest that these cue classes are salient to professional animators’ hand-drawn frame perception, our studies exclusively employ observational, rather than empirical, perceptual cue control and evaluation strategies. Our justification is twofold: first, existing empirical image processing techniques and metrics cannot guaranteeably emulate professional animators’ perception of hand-drawn frames. Therefore, observational cue assessment – assisted by Cartoon Saloon animation professionals – was deemed most prudent within the scope of this thesis, which does not concern perceptual quality evaluation of image processing metrics. Second, hand-drawn animation frame creation entails immeasurable intentional and subliminal expert artistic choices presently impossible to model computationally. Even one such choice can make the difference between director approval or a retake, and we endeavor to exclusively study animators’ perception of director-approved frames. Accordingly, we could not synthetically generate hand-drawn frames of our desired artistic quality in which perceptual cues could be empirically controlled. Controlling cues by editing frames would have heavily corrupted frames’ artistic integrity and almost certainly impacted animators’ frame perception beyond our experimental control. In future, we aim to pursue empirical perceptual cue control and evaluation techniques.

3.1 Perceptual Experiment 1: Local Frame Isometry

When first considering reference frame retrieval system design, it may seem sensible to simply perform body pose estimation on hand-drawn characters. Then, for a given character, the system could return frames depicting poses closest to a given input pose, which the user aims to depict in a hand-drawn frame. In this case, hand-drawn frame HPE methods (Section 2.3) might prospectively support RFQ analysis.

Although frames depicting similar body poses can be good drawing reference for one another, two key observations suggest that animators’ RFQ perception involves both depicted and material feature similarity between frames. First, frames depicting physically similar poses for the same body part – i.e., poses differing by small articulate motion – may be poor drawing references for each other when their local (i.e., body part-wise) material features differ non-isometrically (Figure 3.2, top, arm). Second, frames depicting physically distinct body part-wise poses may be good drawing reference for one another when their local material features differ isometrically; i.e., by 2D rotation/reflection (Figure 3.3, bottom, arm).

Human vision is somewhat isometry-invariant [252], and artists develop enhanced rotation invariance [184, 317, 56] and multiresolution perceptual acuity [54, 53] – i.e., conscious restriction of visual attention to isolated multiscale parts of an image – by practicing their trade. It is then reasonable to expect that frames equivalent up to local material feature isometry are good drawing references for each other with respect to the isometrically-transformed body part. However, the practical impact of these perceptual characteristics on professional artists’ understandings of their own artistic work has not been studied previously. This preliminary experiment tests our second observation: professional animators perceive frames depicting physically different body part-wise poses as good drawing reference when the body part’s depictions in each frame differ by local material feature isometry.

3.1.1 Variable & Data Selection

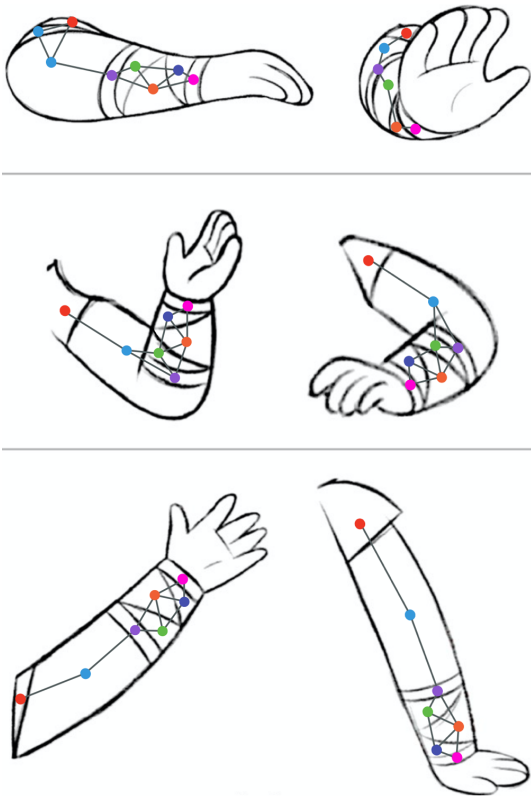


Figure 3.1: Experiment frame pairs demonstrating how each transformation variable affects geometric and topological cues, with RAGs superimposed for instructional purposes (not shown in experiment). Top = (high NIT, low Reflection, low Rotation). Middle = (low NIT, high Reflection, low Rotation). Bottom = (low NIT, low Reflection, high Rotation). RAG topology is preserved only in Middle and Bottom cases.

Only frames showing Mebh at eye-level (not from above or below) and facing front (not from behind or side profile) were used. Experiment frames were uncolored to match the majority of frames available at the time in production when animators need drawing references.

To test our hypothesis, we selected three region transformation variables: Rotation, Reflection, and Non-Isomorphic Transformation (NIT). Generally, some combination of these variables fully describes material feature differences between depictions of the same body part for a single character design across different poses and frames (Figure 3.1). Because local Rotation and Reflection preserve frames’ geometry and topology cues up to local isometry, the pose changes they convey do not alter a body part’s relative angle to the camera (drawn as foreshortening). Only NIT induces the non-isometric region transformations necessary for conveying foreshortening change. We therefore hypothesized that animators’ RFQ perception would inversely correlate to NIT more strongly than to Rotation and Reflection.

To optimize variable control, all hand-drawn frames used in our experiment depicted the same character: Mebh, from Cartoon Saloon animation studio’s feature film *Wolfwalkers*; frames were provided by Cartoon Saloon. Although Mebh appears in diverse poses across tens of thousands of frames, a frame set offering full variable control over her entire body did not exist. Consequently, participants were asked to consider only one of Mebh’s arms in each frame when answering experiment questions (Figure 3.3). Mebh’s arm was chosen because it is a complex object comprising multiple regions, appears in myriad articulate poses, and is stylistically consistent throughout the film. All experiment frames show Mebh in the same costume, with the same hairstyle, and without props. Frames were reasonably consistent in scale, line quality, and adherence to model sheet character design.

Most experiment frames were manually selected from the 245 frames of Mebh within a database of 649 *Wolfwalkers* frames, built by manually scrubbing through the film to find frames with high-quality depicted body pose (e.g., no head shots or blurred motion) and without prop, environment, or other character interactions. Using a custom-built compositing program, these frames were exported from TV-Paint animation files as uncolored line drawings depicting Mebh in isolation. After finalizing experiment design, other frames depicting arm poses needed for full variable control were found and exported as described above. For the reasons described earlier in this chapter, use of raw, unedited clean-up frames was essential to the validity of our experiment design; synthetic or manually edited frames would intractably confound our perceptual results. 64 frames were selected for the experiment and organized into pairs of reference frames.

Rotation, Reflection, and NIT change between each frame pair’s respective arms was evaluated observationally for the reasons given earlier in this chapter. Observational variable change measurement guidelines were as follows: Rotation: low = 0-10°; medium = 11-89°; high = 90-180°. Reflection: low = no reflection; medium = cross-body reflection (Figure 3.2, bottom); high = inside-outside reflection (Figure 3.1, middle). NIT: low = no cue change; medium = mild geometric cue change without topological cue change (Figure 3.3); high = strong geometric cue change with topological cue change (Figure 3.2, top). Professional animators who did not participate in the experiment confirmed that our evaluations were likely to comport with animators’ hand-drawn frame perception.

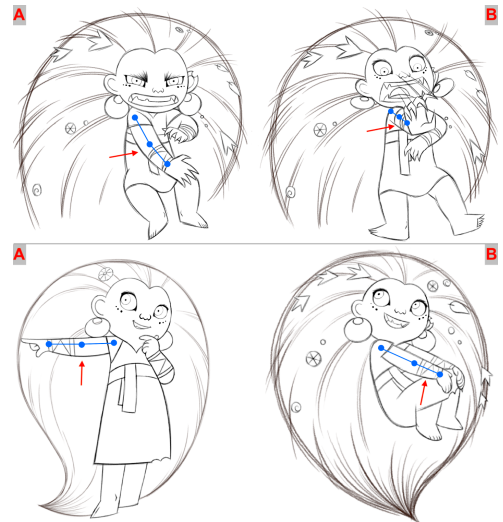


Figure 3.2: Experiment frame pairs with superimposed skeleton poses (in blue), demonstrating that physically similar poses can provide poor drawing reference (top; average RFQ rating = 2.09) and physically different poses can provide good drawing reference (bottom; average RFQ rating = 5.0).

3.1.2 Method

Rotation, Reflection, and NIT were each modulated across low, medium, and high levels (Figure 3.1), creating 27 cases overall. Two frame pairs were selected for each case, yielding 54 experiment questions measuring animators’ RFQ perception. For each question, both frames were resized and cropped to maximize their visual similarity. A red arrow showed participants which arm to consider in each frame when answering the question (Figure 3.3). The experiment questions were presented on Qualtrics software. Each RFQ question clearly described the participants’ task: *Given two frames A and B, rate the helpfulness of Frame B as drawing reference for Frame A*. Ratings were recorded on a 1-7 scale slider, with 1 being “Not at all helpful” and 7 being “Extremely helpful” (Figure 3.3). This rating was our measurement of animators’ RFQ perception between Frames A and B — specifically, of Frame B with respect to Frame A. Qualtrics was used to randomize the order in which questions appeared during the experiment.

Out of 30 professional hand-drawn animators employed by Cartoon Saloon at the time of the experiment, 10 volunteered to participate. Animators were recruited by internal email sent by Cartoon Saloon’s CTO, and were not paid to participate. Before starting the experiment, participants provided information on their professional history. Seven participants had 3-6 years of experience working for hand-drawn animation studios and the remaining three participants had 14-30+ years of experience; on average, participants had 9.6 years of industry experience. Seven participants had an additional 2-4 years of hand-drawn animation experience from attending animation college. Eight participants worked on

Wolfwalkers and had experience drawing Mebh. Since requiring participants to draw all 54 Frame A's during the experiment was impractical, especially as the experiment took place during working hours, participants were asked to imagine that they were drawing Frame A using Frame B as drawing reference.

If you were trying to draw Mebh's **RIGHT ARM** (see arrow) in Frame A, how helpful would Mebh's **LEFT ARM** (see arrow) in Frame B be as drawing reference?

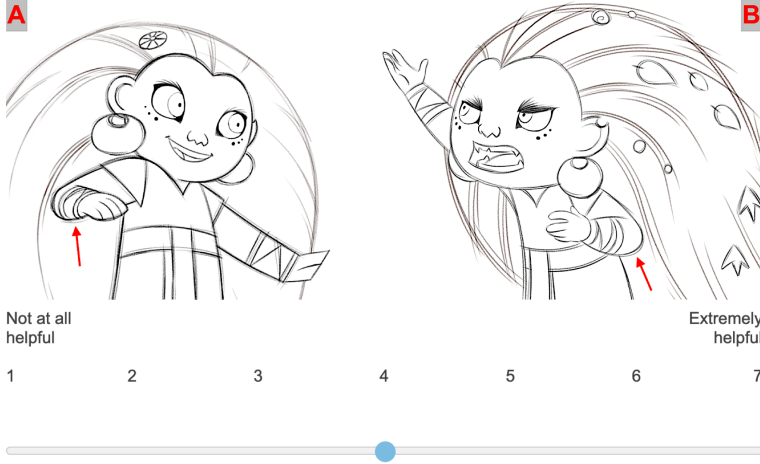


Figure 3.3: Example experiment question (average RFQ rating = 4.5).

Participants were asked to complete the experiment on the computer monitor or drawing tablet they used while animating, further emulating their working conditions (participants' reported screen widths = 15-32"). Because Mebh's hand pose was too variable across experiment frames to be empirically controlled, participants were asked to ignore differences in hand pose when answering experiment questions. Hands could not be edited out of frames, as this would impact frames' artistic quality, and thus our perceptual results as well. Participants completed two RFQ training questions to familiarize themselves with the question format, and were required to confirm their understanding of all experiment directions before proceeding to the RFQ perception experiment questions. After completing all RFQ questions, participants answered debriefing questions. They were asked to verbally describe which factors determined their answers to RFQ perception questions, and to report how much they referenced Mebh's model sheets during the experiment.

3.1.3 Results

We analyzed animators' RFQ scores via Three-Way ANOVA with repeated measures: NIT (low, medium, high), Reflection (low, medium, high), and Rotation (low, medium, high). We conducted Mauchly's test for sphericity, and where the assumption was violated, we applied the Greenhouse-Geisser correction (indicated with an asterisk). We found a main effect of NIT ($F^*(1.18, 10.64) = 74.40, p^* < 0.001, \eta_p^2 = 0.892$), where posthoc analysis showed that all three levels of NIT were rated significantly differently from each other: low (5.139), medium (4.556) and high (2.778) ($p < 0.004$ in all cases). We also found a main effect of Reflection ($F^*(1.27, 11.49) = 11.773, p < 0.001, \eta_p^2 = 0.567$), where posthoc showed that low (5.556) was rated the same as medium (4.239), while high (3.678) was significantly different to both ($p < 0.03$ in both cases). We also found a main effect of Rotation ($F^*(1.79, 16.08) = 14.652, p \approx 0, \eta_p^2 = 0.619$), where posthoc showed that low (4.344) was rated the same as medium (4.389), while high (3.739) was significantly different to both ($p < 0.02$ in both cases). The higher effect size of NIT compared to those

To model real-life animation studio conditions, participants were given Mebh's model sheets and asked to reference them during the experiment, as they would while animating. Participants were asked to imagine that they could manipulate Frame B in their drawing software as they would while animating (e.g., via rotation, reflection, etc.) to optimize its drawing reference quality for Frame A. Because this experiment aimed to measure animators' RFQ perception with respect to their professional workflows, it was important for their in-practice software usage to impact experiment results. Par-

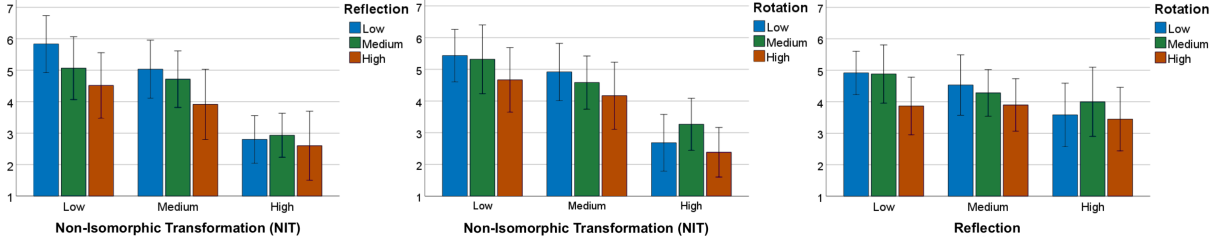


Figure 3.4: Significant Interactions between (left) NIT and Reflection, (center) NIT and Rotation, and (right) Reflection and Rotation. The Y-axis shows average RFQ ratings across participants on a 1-7 scale slider, with 1 being “Not at all helpful” and 7 being “Extremely helpful.”

of Reflection and Rotation supports our hypothesis that animators’ RFQ perception is impacted more strongly by pose changes depicted by NIT than by pose changes depicted by local Rotation or Reflection.

We also found a significant interaction between NIT and Reflection ($F(4, 36) = 4.127$, $p < 0.007$, $\eta_p^2 = 0.314$): pairwise comparisons showed us that, at low NIT, there were three significantly different ratings for Reflection levels low, medium and high ($p < 0.04$ in all cases). For medium NIT, low and medium Reflection were rated the same, while high Reflection was rated significantly worse than the other two ($p < 0.04$ in both cases). For high NIT, there was no effect of Reflection level, indicating that the more NIT present, the less important the differences in Reflection levels (Figure 3.4, left).

A significant interaction between NIT and Rotation was also found ($F^*(2.04, 18.38) = 3.999$, $p < 0.035$, $\eta_p^2 = 0.308$): at low NIT, low and medium Rotation were rated the same, while high Rotation was significantly worse than both ($p < 0.01$ in both cases). For medium NIT, low Rotation was rated significantly higher than either medium or high Rotation ($p < 0.02$ in all cases), while at high NIT, medium Rotation was rated significantly higher than low or high Rotation (Figure 3.4, center). This last result was most likely an artifact of particular selected frames.

Finally, a significant interaction occurred between Rotation and Reflection ($F(4, 36) = 4.088$, $p < 0.008$, $\eta_p^2 = 0.312$), where the pattern matched that of the NIT and Rotation interaction (Figure 3.4, right). No three-way interaction occurred.

Qualitative Results Participants reported that their RFQ ratings depended on foreshortening difference (6 participants), body pose difference (4), overall visual similarity (3), and how well the reference frame matched model sheet character design (2). One participant remarked that “even though an arm could look similar, a tilt in volume or rotation in space that is different might induce mistakes of interpretation of the design style.” As NIT corresponds to foreshortening changes, this further indicates NIT’s predominant effect on RFQ perception. 4 participants reported that referencing model sheets during the experiment helped them preserve character design — which naturally varies across animators and hand-drawn frames — and 3 reported that model sheets helped them draw basic poses. According to one participant, “if the foreshortening/exaggeration was too much for the *B* pose, I found it not very helpful [as drawing reference] and referred to the model sheets for clarity on the design on the arm.” Animators with ≤ 6 years of industry experience gave more detailed answers about model sheet helpfulness, possibly indicating that they are still actively learning skills for extrapolating novel poses from drawing references.

3.1.4 Discussion

Our preliminary experiment results support our hypothesis that animators’ RFQ perception is sensitive to geometric and topological cue change (NIT) and moderately stable under local isometry (Reflection

and Rotation). Therefore geometric and topological cue similarity is a key factor in animators’ RFQ perception, superseding body pose similarity in several cases (e.g., Figure 3.2). As such, our RFQ analysis framework must assess frames’ RAG topology/geometry similarity alongside their depicted pose similarity. However, future studies are needed to ascertain the exact relationship between pose and perceptual cue similarity in animators’ RFQ perception. Pairwise interactions indicate that local isometry’s influence on RFQ may diminish as NIT intensifies, but further testing is needed to confirm. We expect this to be true, since NIT alters visual information available in frames, and most participants reported that visual information preservation between frames positively influenced their RFQ ratings. More specifically, we interpret these reports to mean that frame pairs with arms differing by local material feature isometries optimally preserve director-approved pose resolution strategies, though this inference requires experimental validation in future. Therefore to model animators’ perceptual behavior, our RFQ analysis framework cannot simply measure joint angle differences between frames’ depicted poses; rather, it must recognize material feature equivalences up to body part-wise isometry.

Prior perceptual experiments on professional artists study their mental rotation abilities alongside general visual perception tasks like mentally combining objects [317], Gestalt completion [184], and multiresolution perceptual acuity [54, 53], finding that artists are experts in these tasks and their expertise improves with drawing practice over time [56]. As detailed inferences about professional hand-drawn animators’ RFQ perception cannot be made from these studies, our experiment was conducted to address this knowledge gap. We offer the first controlled evaluation of how professional artists’ visual perception expertise impacts their choices of reference frames as they work, allowing them to obtain high-quality pose resolution information from frames depicting physically dissimilar poses. Our preliminary findings build upon and cohere with prior research on artists’ perception, specifically regarding mental rotation and multiresolution perceptual acuity. We also endeavor to replicate hand-drawn animation studio conditions within our experiment design, which we consider vital for understanding artists’ perceptual processes in studio settings. This methodological approach can support future creation of effective perceptually-based artistic tools, potentially mitigating hand-drawn animation technology’s present barriers to industry uptake.

Limitations The small number of examples per variable case likely affected pairwise interactions. Additionally, the small number of participants working at the same animation studio, alongside testing of a single character and body part prevent, us from generalizing our findings to all hand-drawn animators, art styles, characters, and body parts. Because our experiment only tested local – rather than multiresolution – RFQ isometry invariance, we were unable to assess how isometry scale interacts with RFQ perception; for instance, frames differing by global pose reflection may be better references for one another than frames differing by many small, local limb rotations. However, prior perceptual studies provide strong reasons to expect our findings to generalize across characters, styles, and isometry scales. Professional artists are exceptionally skilled at localizing visual attention across multiscale image features [54, 53], increasing the likelihood that our results generalize across multiscale isometries. Moreover, artists’ expertise at mentally rotating, decomposing, and recomposing generic visual data [317, 56] imply that our findings will generalize across diverse characters and artistic styles.

Classifying variable cases by eye was appropriate for our preliminary study, but may have impacted our results, although every effort was made to prevent this. Moreover, our variable classifications may have relied on our perception of visual features specific to the character design and artistic style we evaluated, possibly rendering observational classification unsuitable for larger studies evaluating diverse characters and styles. Future work can use empirical shape comparison metrics for classification to address

these issues, but such metrics’ differences to human visual perception may create new limitations. An alternative, albeit time-consuming, approach is for animators to rate shape similarity between frames’ corresponding regions shown in isolation; these ratings can then serve to classify geometric cue change between frames. Topological cue change can likely be evaluated via empirical RAG comparison metrics – particularly those assessing shape similarity – although optimal choice of metric should be determined with another perceptual study. More broadly, for studies like our own to rigorously evaluate professional artists’ visual perception in practice, perceptually-based empirical metrics on artistically stylized and symbolic visual data must be thoroughly researched and tested.

Our findings offer compelling evidence that professional animators’ RFQ perception is, within their professional workflows, moderately invariant to local (i.e., body part-wise) isometry. Although this experiment exclusively studies local isometry invariance on a single body part, prior studies demonstrating professional artists’ multiresolution perceptual acuity [54, 53], alongside informal discussions with Cartoon Saloon animators on RFQ analysis framework design, lead us to pursue RFQ analysis up to multiresolution isometry invariance. In practice, this means that our RFQ framework’s frame comparison metrics should detect isometry invariance between frames at any scale, from one local bone to the entire global body (Figure 1.4). In our Figure 3.2 examples, our metrics will treat the top frames as similar everywhere but their left arms and the bottom frames as different everywhere but their left arms, comporting with their respective RFQ ratings. However, our results indicate that treating frames as fully multiresolution isometry-invariant will not accurately emulate animators’ RFQ perception. Since high Rotation and Reflection cases’ ratings were significantly distinct from those of low and medium cases, our RFQ framework should combine both isometry-invariant multiresolution material feature metrics with depicted pose metrics to optimally model animators’ RFQ perception. In practice, multiresolution isometry invariance will support body part-wise reference frame searches. This will enable animators to reference multiple frames depicting different parts of their desired pose when, as is often the case, a suitable full-body reference frame does not exist. Despite the extreme popularity of equivariant data analysis frameworks – especially EqNNs – prior research does not offer a framework in which heterogeneous isometries can be detected and compared across multiple scales within an articulate body model. The following chapter details our novel theory of heterogeneous multiresolution equivariance, which, unlike prior work, models animators’ RFQ perception subject to multiresolution isometry invariance.

3.2 Perceptual Experiment 2: 3D Body Pose Perception

Because hand-drawn characters lack ground-truth body representations, they undergo continual material feature deformations as they move. Such deformations can alter characters’ depicted body geometry and apparent functional body part-wise relationships across frames, and thus likely influence perception of hand-drawn characters’ depicted body poses. By studying how frame deformation impacts professional animators’ perception of characters’ depicted body poses, we gain insights into how best to design our RFQ analysis framework’s pose comprehension functionalities.

Deformations arise from either pose resolution or Principles of Animation. The former involves subtle, often imperceptible character design modifications made to better convey the given depicted pose, engendering view-dependent geometry in extreme cases (Figure 2.23). Such deformations can create inconsistencies between related perceptual cues, with each cue indicating a slightly different pose than the others. However, animators are so skilled at pose resolution that such inconsistencies often become noticeable under scrutiny, appearing natural to casual viewers. Crucially, inconsistency is a ‘global’ perceptual property: cues are ‘locally’ consistent when perceived independently, only becoming

inconsistent when perceived collectively via multiresolution visual integration (Figure 2.34). While prior studies evidence humans’ perceptual sensitivity to multiresolution cue consistency [166, 177, 27] and artists’ exceptional multiresolution perceptual acuity [54, 53], these traits’ practical impact on artists’ perception of their own work has not yet been investigated.

Deformations involving Principles of Animation – especially squash-and-stretch, stylized perspective, and bone bending/breaking – distort characters overtly because they are employed for artistic effect (Figure 2.12). Such distortions frequently generate visual ambiguities – for instance, a squashed limb could be perceived as foreshortened – which are heightened by frames’ data sparsity and artistic stylization [94, 158]. Since individuals differently interpret perceptual meanings of visually ambiguous stimuli [282, 83], it is possible that professional animators’ perceptions of characters’ depicted body poses increasingly diverge when frames are increasingly distorted. On the other hand, since perceptual expertise improves with artistic practice [55, 56], professional animators’ extensive experience deforming hand-drawn characters via Principles of Animation could help them uniformly interpret distorted frames’ depicted poses. Our Experiment 1 findings and a small study showing significant variation among professional 3D animators’ 3D pose perceptions of animated characters [24] make the former option more likely, though either possibility will influence how our RFQ framework analyzes characters’ depicted body poses.

This experiment tests the hypothesis that professional hand-drawn animators’ perceptions of hand-drawn characters’ depicted poses increasingly diverge as characters become increasingly distorted. Although visual ambiguities in artistic works is extensively studied [214, 94], controlled evaluation of how these ambiguities impact artists’ perception of artistic works is lacking. As discussed throughout this chapter, such insights are critical for understanding collaborative artistic environments like hand-drawn animation studios, particularly when designing technology for teams of artists. In our study, Cartoon Saloon animators pose a 3D Blender rig – designed to emulate wooden maquettes that animators use for posing reference (Figure 3.5) – to match character poses depicted in hand-drawn animation frames, also rating frames’ Distortion intensity and posing task Difficulty. We envision that animators will pose similar rigs when querying our proposed reference frame retrieval system. Thus, our RFQ analysis framework on which the system will be based must be able to understand animators’ perceptual relationships between input rig poses and hand-drawn frames. Our quantitative results fail to support our hypothesis, instead indicating that pose perception deviation increases as perceptual cue consistency decreases. This has deep topological implications for our RFQ analysis framework, highlighting the importance of detecting obstructions which prevent local perceptual cues from being collectively perceived as representing a single, globally consistent body pose.

3.2.1 Variable & Data Selection

This experiment’s primary objective was to compare professional animators’ ratings of hand-drawn frames’ Distortion intensity to variability of their depicted pose perceptions for each frame. In Section 2.1.2, we noted the subtle distinctions between animators’ mental conceptions of volumetric pose vs. material 3D body pose, with the former being far more flexible and abstract than the latter. However, because studying animators’ cognitive pose representations is beyond the scope of this thesis, and since our reference frame retrieval system will take 3D rig poses as input, this experiment uses 3D pose estimates as practical proxies for animators’ volumetric pose perceptions.

Like Experiment 1 (Section 3.1), this experiment was conducted at Cartoon Saloon during working hours, requiring us to accommodate active production schedules in our study’s design. Because our experimental setup included both frame rating and Blender posing tasks – the latter taking animators 10-25 minutes per frame, depending on pose complexity and their 3D modeling proficiency – only 12 frames were

used in the experiment. This was smallest number of frames needed to reasonably assess our hypothesis’ validity, ensuring that our experiment minimally interfered with participants’ professional duties. While a larger number of experiment frames would have been preferable, this would have required a longer experiment run, which Cartoon Saloon production schedules prevented at the time of the experiment. Since posing tasks were so time-consuming, only a few frames were tested per day.

To optimize variable control, all 12 of our experiment frames depicted the same character: Robyn, from the Cartoon Saloon feature film *Wolfwalkers*. All frames showed Robyn with the same costume and hairstyle, and without other characters, significant prop interaction, or self-occlusion obscuring portions of her body. Only frames showing Robyn at eye-level (not from above or below) and facing front (not from behind or side profile) were used. As *Wolfwalkers* employs multiple line styles, experiment frames used either solid or slightly sketchy line styles to maximize visual consistency. Frames were manually selected from the *Wolfwalkers* clean-up frame database used in Experiment 1, which contains 404 frames of Robyn and 649 frames total. As described previously, frames were collected through manual search and exported from TVPaint files via our custom compositing program. Experiment frames were manually edited to remove minor prop interactions without altering artistic quality – for instance, Robyn’s crossbow was removed when it protruded from behind her shoulder or arm, potentially impacting perception of her upper body pose.

To emulate reference frame availability during production, seven full-body and five partial-body frames were used in the experiment. Although full-body shots offer complete body pose information and are therefore preferable, partial-body shots are more common in practice. Selected partial-body frames showed Robyn from the hips up to prevent participants from extrapolating her leg poses, which would extend experiment runtime and confound pose perception results. Frames fell into three categories: Static Full-body (SF), Static Partial-body (SP), and Motion Full-body (MF); Static frames depicted Robyn standing still, while Motion frames depicted her walking, running, or jumping. Each category included frames that were expected to vary in perceived Distortion intensity; selections were guided by Cartoon Saloon personnel who were not experiment participants. While every effort was made to include cases of extreme Distortion, the found while manually searching the film involved purely geometric and not topological cue change. Therefore, our study primarily examines perception of purely geometric Distortion. As an aside, our inability to find Distortion cases involving topological cue change was somewhat surprising. Since perception is robust to deformation yet sensitive to part-wise relationship change [11, 27, 297], geometric cue change may be easier to interpret than topological cue change. Because animators typically aim to make frames as easy to visually interpret as possible, they may favor geometric Distortion to preserve characters’ perceptual clarity.



Figure 3.5: Posable wooden maquette used for drawing reference by professional animators.

3.2.2 Method

Participants Of the 16 professional animators employed by Cartoon Saloon at the time of the experiment, 12 agreed to participate. Since the experiment ran during working hours, participation was unpaid. Animators were recruited via internal email sent by Cartoon Saloon’s CTO, with full cooperation of all production coordinators and animation department supervisors. Before starting the experiment, participants provided information on their professional history. Four participants had 2-5 years of

animation industry experience, five participants had 7-16 years' experience, and the remaining three had 30+ years' experience. Participants had on average 18.5 years of industry experience, with most reporting an additional 1-3 years of hand-drawn animation experience during college. Seven participants worked on *Wolfwalkers* and had experience drawing Robyn. Five participants were Cleanup Animators, three were Rough Animators, and five were Posing Artists (one of whom is primarily a Storyboard Artist).

To minimize workday disruption and participant fatigue, the experiment took place over three consecutive days, with four frames tested per day. Before starting the experiment, participants answered training questions to familiarize themselves with the experiment's format. They were also required to correctly answer questions about the experiment's format and instructions, which were especially crucial for questions about frame Distortion. Because professional animators are so skilled at making physically awkward or impossible poses appear natural, distortion can be difficult to recognize. Participants were directed to carefully look for three types of frame distortion:

1. Visual Distortion: *How much is Robyn distorted away from her model?*
2. Physical Distortion: *How physically possible is Robyn's body pose?*
3. Animation techniques involving distortion, like squash-and-stretch, bone bending/breaking, and stylized perspective.

Each day comprised a Distortion rating section followed by a 3D posing section. Experiment questions and posing instructions were presented on Qualtrics software. To ensure an even mix of stimuli throughout the experiment, each day tested frames expected to depict low, medium, and high Distortion cases, as well as both full- and partial-body frames.

Posing tasks were performed on a generic 3D Blender rig modeling the human body, initially obtained from an online template and customized by the Cartoon Saloon software department. The rig was designed to emulate the posable wooden human body maquettes which animators routinely reference for posing

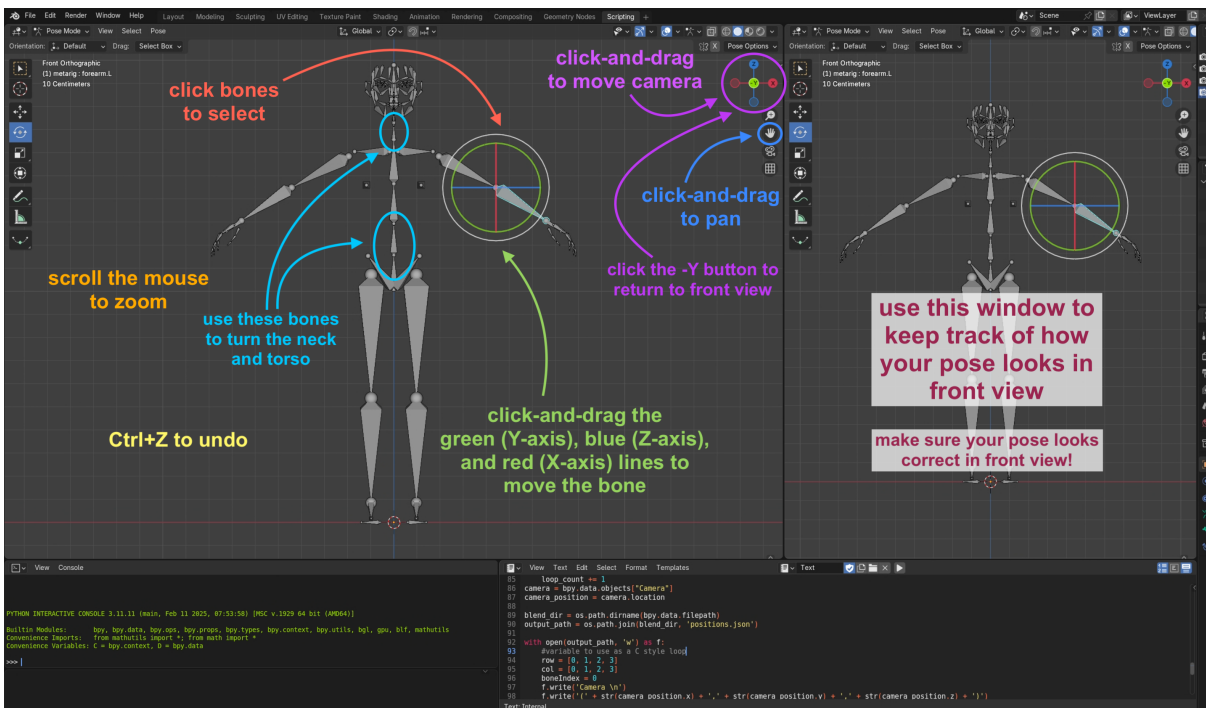


Figure 3.6: Blender instruction sheet given to participants. Instructions are superimposed over the Blender file containing the skeleton rig and viewport setup used during the experiment.

guidance while drawing frames (Figure 3.5). The rig lacked human body joint motion constraints, leaving participants free to create physically impossible poses. Although participants were given the option to complete each day’s Distortion rating section at their own desks, they were required to complete posing tasks on one of three software department computers with the correct version of Blender installed, ensuring technical setting uniformity for all posing tasks. All three computers were next to each other, enabling close supervision of the experiment and immediate technical and instructional support for participants as needed. Because most participants lacked 3D modeling experience, they were each given verbal, visual, and written instructions on how to pose the Blender rig before beginning the experiment (Figure 3.6).

Each day’s Distortion rating section comprised four questions, each involving a different experiment frame. Question order was randomized for each participant by Qualtrics. For each question, participants were shown the frame and asked: *In this frame, how intensely is Robyn distorted from her character model? (Look for Visual Distortion, Physical Distortion, and distorting animation techniques applied to her body pose.)* This question was answered on a 1-7 slider scale, with 1 labeled “No distortion” and 7 labeled “Intense distortion.” Participants were given Robyn’s model sheets for reference when answering Distortion rating questions.

Participants completed four posing tasks per day, each involving a frame from that day’s Distortion rating section. Again, the order in which frames’ posing tasks appeared was randomized for each participant by Qualtrics. During each posing task, animators were instructed to open a specific Blender file corresponding to the experiment frame specified on Qualtrics. Each file contained a copy of the Blender rig in a standard T-pose (Figure 3.6). Participants were shown a large image of the experiment frame, and asked to pose the Blender rig to match the pose depicted in the frame. Each participant completed posing tasks using two computer monitors, so the Blender file and the Qualtrics window showing the reference frame were visible simultaneously. Every Blender file contained two viewports: a working viewport where participants could change the camera view angle to better manipulate the rig, and a viewport fixed in front-view. Because all experiment frames depicted Robyn in front-view, participants were instructed to pose the rig such that its appearance in the front-view viewport optimally matched the given frame.

Although posing tasks were not time-constrained, participants were encouraged to complete each Blender rig pose within 20 minutes. Upon completing each rig pose, participants were instructed to save their work, which activated a Python script automatically exporting the rig’s bone positions to a JSON file. After returning to the Qualtrics interface, they were again shown the experiment frame and asked: *How difficult was posing the Blender model to match this frame?* This question was answered on a 1-7 slider scale, with 1 labeled “Extremely easy” and 7 labeled “Extremely difficult.”

3.2.3 Results

For every frame, each joint’s position was averaged across all 12 participants’ pose estimates, generating an average 3D pose estimate. Joint-wise 3D and 2D Deviations were then found by computing unweighted RMSD joint angle differences (in degrees) between each participant’s pose and the frame’s average pose; 2D Deviations were computed after projecting poses to the *XY*-plane. 3D and 2D joint-wise Deviations were separately averaged, yielding respective, whole-body 3D and 2D Deviation scores for every frame. 3D Deviations were analyzed because the posing tasks took place in 3D; 2D Deviations were analyzed because animators were instructed to optimally match the Blender rig’s pose to the given frame subject to the orthographic 2D front-view viewport. One participant’s Blender pose was missing for frame MF4, so the 11 remaining Deviations were averaged to fill the missing entry. As animators were not required to pose the rig’s head, neck, hands, and feet, these body parts’ Deviations were excluded from statistical analysis.

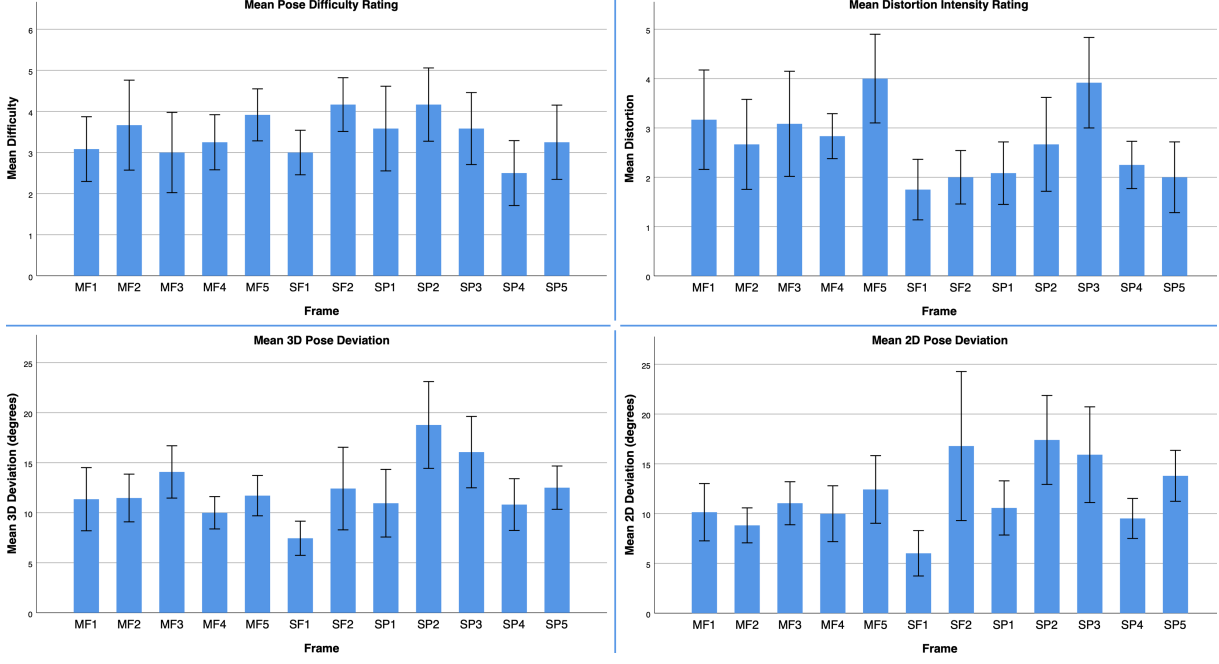


Figure 3.7: Average per-frame Difficulty ratings (top left), Distortion ratings (top right), 3D pose Deviation (bottom left), and 2D pose Deviation (bottom right). Error bars: 95% CI.

3D Deviation

To determine the effects of Distortion and Difficulty ratings on 3D Deviation, a linear mixed effects model was constructed with 3D joint angle Deviation as the dependent variable, Distortion and Difficulty ratings as fixed effects, random intercepts for each participant, and diagonal covariance structure. Although an significant effect of Distortion on 3D Deviation was not found ($F(1, 88.502) = 0.695$, $p = 0.407$, slope est. = 0.242), Difficulty significantly predicted 3D Deviation ($F(1, 103.465) = 8.847$, $p = 0.004$, slope est. = 0.901). These findings comport with Pearson correlations across all frames and users, with a stronger positive relationship between 3D Deviation and Difficulty ($r(144) = 0.249$, $p = 0.003$) than between 3D Deviation and Distortion ($r(144) = 0.176$, $p = 0.035$). While the Pearson correlations suggest that Distortion ratings have a marginal effect on 3D Deviation, the effect becomes insignificant when accounting for repeated measures in our mixed effects model. Contrastingly, the association between larger 3D joint Deviations and higher Difficulty ratings remains consistent in both our Pearson and mixed effects models, indicating that Difficulty's effect on 3D Deviation likely reveals a genuine relationship between these variables.

2D Deviation

The same analysis procedure conducted for 3D Deviations was also used for 2D Deviations. Pearson correlations for all frames and users between 2D joint Deviation and, respectively, Distortion ($r(144) = 0.14$, $p = 0.095$) and Difficulty ($r(144) = 0.234$, $p = 0.005$) suggested similar findings to our 3D case, but our mixed effects model revealed that neither Distortion ($F(1, 74.634) = 1.44$, $p = 0.234$) nor Difficulty ($F(1, 58.884) = 1.008$, $p = 0.32$) induced significant effects on 2D Deviation.

One-Way ANOVA

In order to determine if differences in ratings occurred across the frames, One-Way Repeated measures ANOVAs were conducted on participant ratings for Distortion and Difficulty, and on the calculated 3D

Deviation and 2D Deviation scores. Shapiro-Wilk tests found that normality was violated in every case, so non-parametric tests were conducted. Friedman tests revealed a significant effect of Frame on Distortion ($\chi^2 = 49.144$, $p < 0.001$), Difficulty ($\chi^2 = 21.113$, $p = 0.004$), 3D Deviation ($\chi^2 = 38.474$, $p < 0.001$), and 2D Deviation ($\chi^2 = 40.885$, $p < 0.001$). Post-hoc pairwise Wilcoxon Signed-Rank tests with Benjamini-Hochberg corrections found multiple significant differences among frame pairs (Table 3.1), with the highest pairwise Wilcoxon p -value for Distortion ($p = 0.01$), Difficulty ($p = 0.018$), 3D Deviation ($p = 0.008$), and 2D Deviation ($p = 0.005$) greater than each measure’s main effect p -value. Therefore the respective main effects of Frame on Distortion, Difficulty, 3D Deviation, and 2D Deviation are stronger than any pairwise contrast between frames.

Level	Distortion	Difficulty	3D Deviation	2D Deviation
High	MF5, SP3	MF2, MF5, SF2, SP1, SP2, SP3	MF3, SP2, SP3	MF5, SF2, SP2, SP3, SP5
Medium	MF1, MF2, MF3, MF4, SP2		MF1, MF2, MF4, MF5, SF2, SP1, SP4, SP5	MF1, MF2, MF3, MF4, SP1, SP4
Low	SF1, SF2, SP1, SP4, SP5	MF1, MF3, MF4, SF1, SP4, SP5	SF1	SF1

Table 3.1: Significant relationships between frames from pairwise Wilcoxon Signed-Rank tests. Groupings are relative to each variable and not indicative of rating/joint angle values, which are shown in Figure 3.7.

Qualitative results

The lack of relationship between Distortion and 3D Deviation is surprising, particularly given geometric and topological cues’ impact on animators’ frame perception, as found in Experiment 1 (Section 3.1.4). However, experiment frames’ high Distortion cases alter geometric but not topological cues (e.g., MF1 left leg, SP3 right arm), which Experiment 1 found to only moderately impact animators’ frame perception. It is therefore possible that professional animators recover pose information reasonably well in cases of purely geometric cue change. This effect may have also influenced 2D Deviation scores, potentially explaining the lack of statistical relationship between Distortion and 2D Deviation. Because Difficulty ratings reflected animators’ experiences with 3D rather than 2D posing tasks, the lack of statistical relationship between Difficulty and 2D Deviation is understandable.

To discern why Difficulty significantly predicted 3D Deviation in the linear mixed-effects model, we observationally evaluate frames with similar poses and varied Difficulty ratings. For each frame in the group rated highest Difficulty (Table 3.1), we identified similarly posed frames with lower Difficulty ratings, then compared and contrasted perceptual cues across higher and lower Difficulty frames (Figures 3.8 and 3.9). In the following analysis, we report body part-wise Deviations (in degrees) averaging constituent joint-wise 3D Deviations (e.g., arm Deviation averages elbow and shoulder Deviations). L/R denote left/right. While our analysis is limited to experiment frames and lacks formal experimental control of perceptual cues and depicted poses, we observe an overall trend of Difficulty and Deviation rising as perceptual cues become increasingly inconsistent.

For example, MF5 (high Difficulty) and MF1’s (low Difficulty) weight-bearing legs have identical foreshortening, indicating they are both posed vertically under Robyn’s body. Since this leg pose enables Robyn to land her jump safely in MF1, geometric and motion cues agree. Animators uniformly pose MF1’s leg accordingly (hipR = 1.61°). Yet MF5’s foreshortening and motion cues conflict: the leg should extend forward at the depicted point in a run cycle. Animators are thus divided between posing the leg

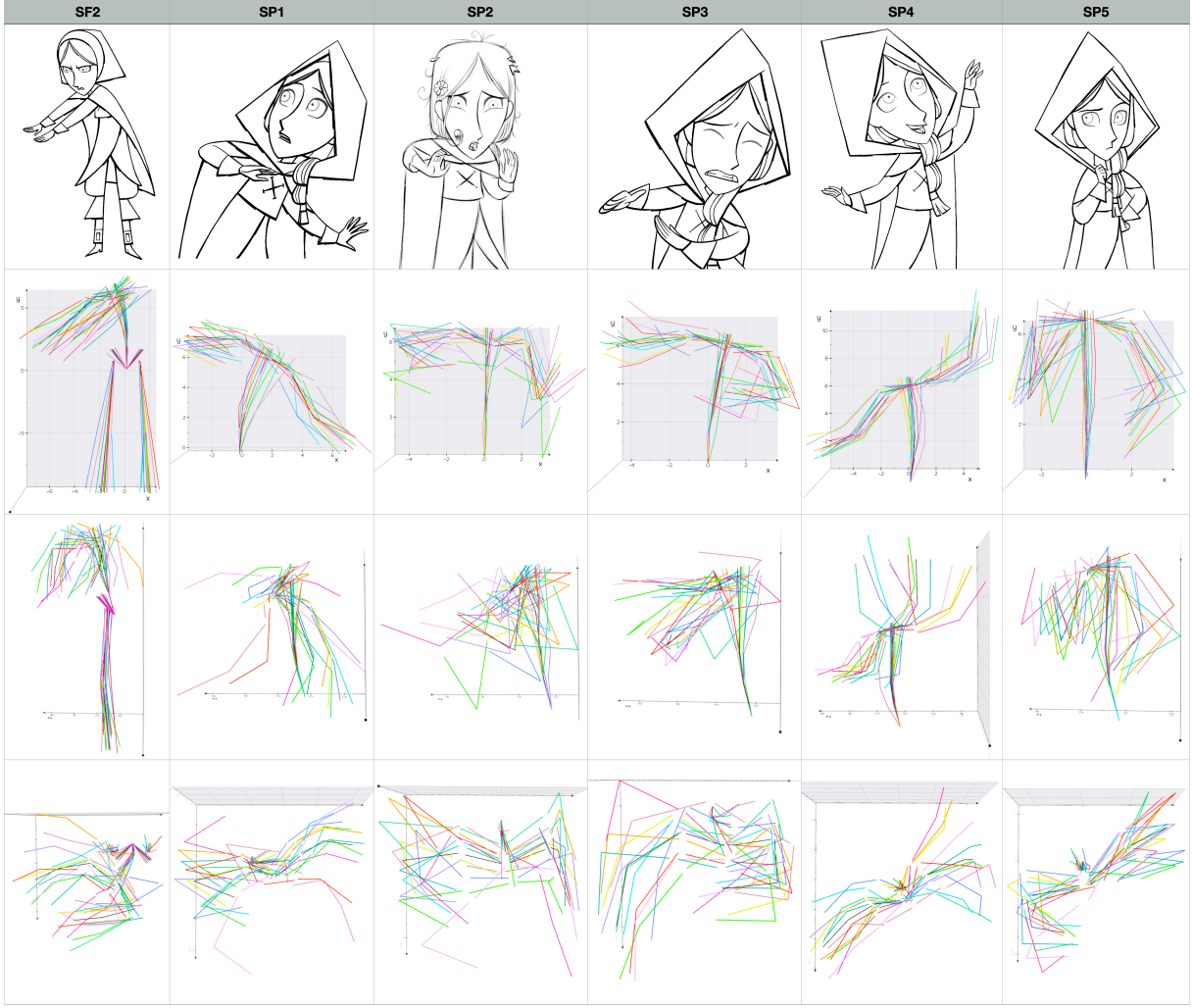


Figure 3.8: Experiment frames (top row) and participants' pose estimates for each frame, shown from front-view (2nd row from top), side-view (3rd row from top), and aerial-view (bottom row). Pose estimates are color-coded by participant, with average pose shown in black. Pose estimates' heads, necks, hands, and feet are not shown since participants were not required to pose these bones. Continued in Figure 3.9.

vertically or extending it forward ($\text{hipL} = 10.49^\circ$). Note that MF5's upper body cannot explain its high Difficulty, as low Difficulty frames MF3 and SF1 have similar upper bodies. Meanwhile, MF1 and MF2 (high Difficulty) depict identical motion, but MF2's weight-bearing leg foreshortening and visible boot sole imply forward leg extension. This conflicts with the vertical pose required for landing the jump, dividing animators between posing MF2's leg vertically to satisfy the frame's motion cues, or extending it forward to satisfy the frame's geometric and topological cues ($\text{hipR} = 8.73^\circ$).

While left arms in frames SP2 (high Difficulty) and SP5 (low Difficulty) share similar poses and hand/bodice connectivity, cue consistency again predicts Deviation and Difficulty. SP5's foreshortening and arm/torso topology cues agree, indicating the arm is maximally bent and flush with the torso ($\text{armL} = 8.44^\circ$). Aside from shoulder rotation, SP2 differs from SP5 primarily in its forearm foreshortening. Forearm geometry is physically incompatible with hand/bodice topology; perceived forearm pose thus varies extensively ($\text{armL} = 18.08^\circ$). SP2's right arm resembles MF4's (low Difficulty) left arm; for both of these arms, upper arm body part-wise geometry and foreshortening indicate mild forward extension. This agrees with MF4's forearm cues ($\text{armL} = 9.77^\circ$), whereas SP2's forearm foreshortening and topology

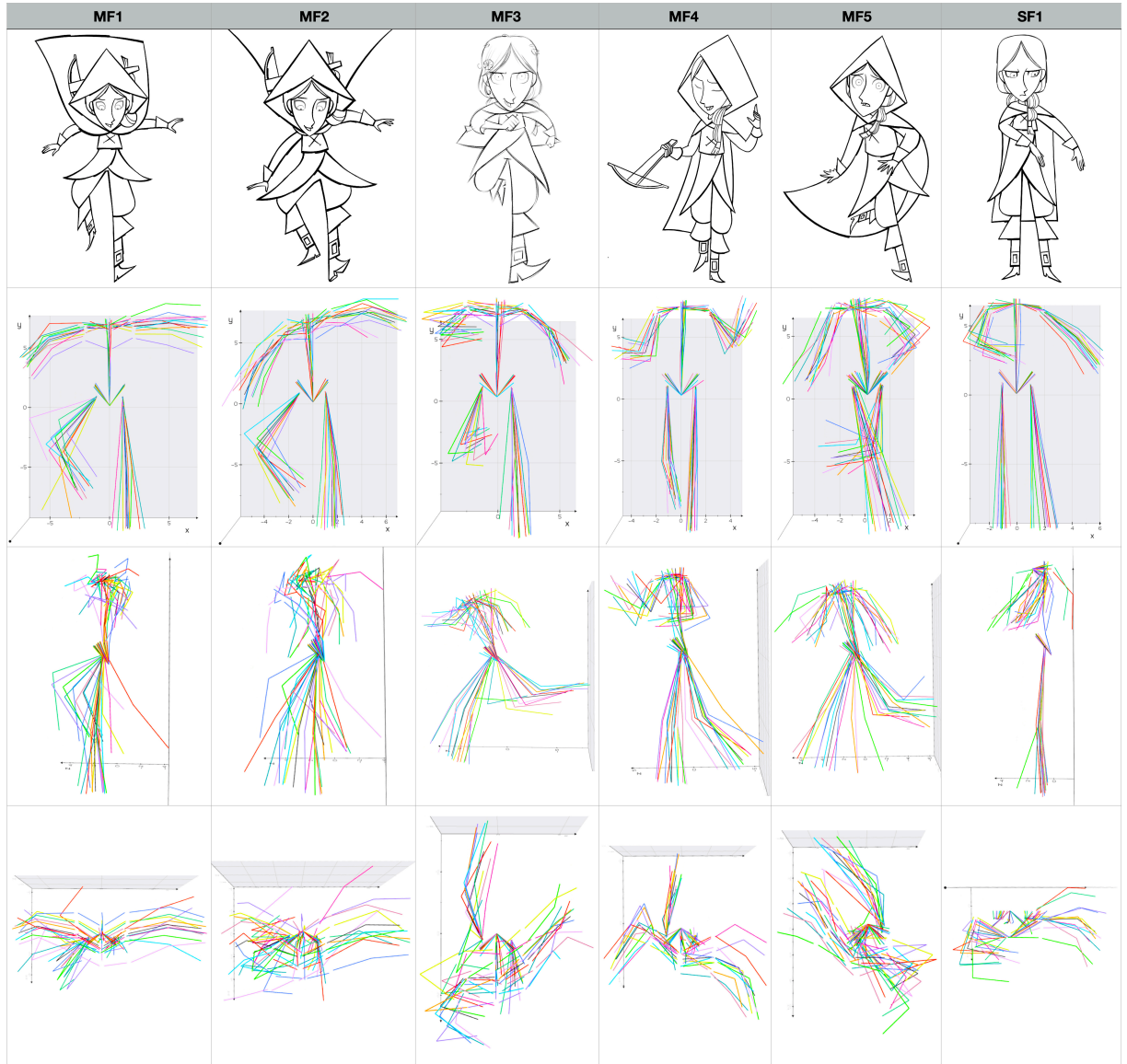


Figure 3.9: Experiment frames (top row) and participants’ pose estimates for each frame, shown from front-view (2nd row from top), side-view (3rd row from top), and aerial-view (bottom row). Pose estimates are color-coded by participant, with average pose shown in black. Pose estimates’ heads, necks, hands, and feet are not shown since participants were not required to pose these bones. Continued from Figure 3.8.

contradictorily imply backward arm extension ($\text{armR} = 20.17^\circ$).

SF2’s (high Difficulty) upper body geometry and topology uniquely arise from Robyn’s narrow-shouldered character design, and are thus not replicable on the wide-shouldered Blender rig (shoulders = 22.19°). SF1’s (low Difficulty) right arm is similarly posed but exhibits no character/rig cue conflicts, decreasing Deviation ($\text{shoulderR} = 10.7^\circ$). Interestingly, the same inconsistency likely produced an outlier 3D Deviation and Difficulty relationship in SP5 (low Difficulty): right forearm foreshortening and elbow crease convey backwards arm extension, but Robyn’s narrow shoulders make arm pose geometry impossible unless the arm is camera-parallel. Accordingly, animators were evenly divided between these poses ($\text{armR} = 14.32^\circ$). However, the rig’s shoulder width preserves SP5’s pose geometry when the arm is back-extended, likely preventing animators from noticing inconsistencies between these various geometric cues, and potentially explaining SP5’s low Difficulty. It is unlikely that rig/character incompatibility

considerably impacted our results beyond these cases; otherwise, frames with similar upper body poses would all have similar Deviations and Difficulty ratings. As we have illustrated, this is not at all the case.

3.2.4 Discussion

Our experiment results failed to validate our initial hypothesis of a positive relationship between animators' perception of frame Distortion and 3D Deviation across animators' pose estimates. However, we found a clear predictive relationship between 3D Deviation and animators' perception of posing task Difficulty. Observational analysis of experiment frames revealed that both Difficulty and 3D Deviation elevation appear related to increased perceptual cue inconsistency. Practically speaking, this implies that Difficulty ratings increased when animators struggled to integrate local perceptual cues into a globally consistent pose interpretation.

While future studies employing more systematic cue evaluation approaches are needed to confirm this relationship, our findings comport with extensive evidence that material and depicted feature inconsistencies impair line drawing perception [253, 318, 331, 46]. Prior work also demonstrates humans' perceptual sensitivity to multiresolution feature consistency [166, 177, 27] and part-wise relationships [253, 212, 164], further corroborating our theory of cue consistency's impact on animators' pose perception. For our theory to hold, animators must be able to (i) localize their visual attention to individual perceptual cues, (ii) perceptually interpret the local patch of depicted pose information conveyed by each individual perceptual cue, and (iii) identify inconsistencies between related cues' depicted pose information. We also expect that, like impossible object perception (Figure 2.34), cognitively integrating body part-wise pose perceptions subject to prior knowledge of human motion constraints supports physically impossible pose recognition. Both of these inferences have substantial implications for our RFQ analysis framework.

We also found evidence that geometric incompatibilities between our generic Blender rig and Robyn's character design elevated 3D Deviation and – when incompatibility was noticeable during the posing task – Difficulty. Our reference frame retrieval system will then benefit from character-specific rigs for specifying input poses. Since our rig emulated animators' real-life posing reference maquettes (Figure 3.5), these cases invite further study of animators' (in)ability to perceptually translate poses between generic articulate body references and character-specific body geometries as they work.

Experiment frame Distortion cases involved purely geometric cue change, which Experiment 1 found to only moderately impact animators' frame perception (Section 3.1.4). It is therefore possible that pose perception is less susceptible to Principles of Animation than we initially predicted, although extreme Distortion may elevate Deviation as hypothesized (e.g., MF1 legL = 16.67° , SP3 elbowR = 14.66°). To interpret these results, observe that many experiment frame Distortion cases were accompanied by other perceptual cues constraining feasible pose interpretations. If our theory of cue consistency's impact on pose perception is sound, then animators may interpret distorted body parts' poses more uniformly in such cases. This would agree with prior studies on perceptual mechanisms of visual ambiguity resolution [182] and possibly explaining the lack of statistical relationship between Distortion and 3D Deviation. Animators' extensive practice drawing distorted hand-drawn characters may also enhance the fluency of their distorted pose interpretations, comporting with findings that drawing practice enhances perceptual skill [55, 56]. This would explain comparable 3D Deviations between certain cases of non-distorted vs. mildly distorted frames with similar poses and cues. For instance, the average elbow Deviation for slightly curved arms in SF1, SP1, MF1, MF2, and MF3 is extremely low (7.68°), and nearly identical to that of straight arms in SF2 and MF5 (7.44°).

Limitations Despite the many benefits of conducting this study in an industry setting, real-world time and resource constraints created several limitations. Though our frame and participant sample sizes exceeded those of prior studies on professional animators [24, 86], it is possible that our statistical results are underpowered, and should be validated through larger-scale studies in future. To accommodate active production schedules and minimize participant fatigue, we only collected whole-body Distortion and Difficulty ratings, rather than individual ratings for each body part. While this strategy would have sufficed if our initial hypothesis held, it was not sensitive enough to statistically verify our qualitative findings. Since Deviation often varied substantially between different body parts in a single frame, future studies should employ rigorous body part-wise posing task difficulty and visual consistency metrics. It is also possible that the different types of frame distortion we tested – Visual Distortion, Physical Distortion, and distorting animation techniques – may differently impact animators’ pose perception, and should be separately evaluated in future work.

Alongside our small participant and frame sample sizes, our evaluation of only one hand-drawn character and artistic style prevents confident generalization of our findings to other characters, films, and studios. However, corroborating evidence from prior studies on visual inconsistency perception [253, 318, 156] increases the likelihood that our results will generalize successfully. Since dissimilarities between Robyn’s character design and our Blender rig appeared to impact both Deviation and Difficulty scores, similarities and differences between characters and rigs should be minimized and carefully controlled in experiments involving more characters, frames, and artistic styles. This will likely improve detection of patterns in animators’ pose perception which persist across multiple styles and human body types. Relatedly, our study exclusively concerns human characters, preventing generalization of our findings to non-human characters. Given animators’ reliance on their own internal proprioception while drawing, their perceptions of human vs. non-human character pose might substantially diverge. Lastly, as computational evaluation of character distortion and visual cues may not comport with animators’ frame perception, observational analysis was appropriate for our study. Future work should explore how these variables might be empirically controlled via perceptually-representative computational metrics.

This is the first perceptual study demonstrating that professional hand-drawn animators’ perception of hand-drawn characters’ 3D body poses can deviate significantly, and that the extent of Deviation varies significantly between frames. Then, a single frame could provide suitable drawing reference for multiple 3D body poses inputted into proposed reference frame retrieval system. Our RFQ analysis framework must therefore assess each frame’s space of perceptually feasible 3D poses, rather than simply obtaining a single pose estimate. Although our whole-body Distortion and Difficulty ratings precluded linear mixed effects analysis of body part-wise pose Deviations, observational analysis indicates that Deviation also varies across different body parts within each frame. Accordingly, perceptually feasible pose assessment should be body part-specific. This approach starkly contrasts with leading HPE methods, which output a single pose estimate for each input image and do not evaluate perceptual quality differences across various body parts (Section 2.3).

We also offer preliminary insight into specific factors engendering and increasing pose perception subjectivity for each frame, namely perceptual cue (in)consistency across multiple visual resolutions. As such, our RFQ analysis framework should not only anticipate potential variation across animators’ multiresolution pose perceptions, but should also detect obstructions preventing local perceptual cues from being collectively perceived as representing a globally consistent body pose. Obstruction detection must also discern (in)feasible pose relationships between body parts – thus enabling physically impossible pose detection – without breaking when impossible poses and extreme Distortion arise. While both

multiresolution perceptually feasible pose assessment and multiresolution obstruction detection may initially appear to be ill-defined problems, both concepts are topologically formalizable with cohomology (Section 2.5), as we discuss in the following chapter.

3.3 Conclusions

As discussed throughout Chapter 2, the relationships between frames’ material and depicted features lie at the heart of hand-drawn animation technology research. More specifically, these relationships underpin the challenge set forth in Research Question 2 (Section 1.2): to identify and empirically characterize the perceptual factors influencing professional animators’ judgments of RFQ. To this end, both perceptual experiments detailed in this chapter follow our novel methodology of modeling professional animators’ working environments within experiment design. Moreover, both systematically examine the same proposed set of geometric, topological, and motion perceptual cues for hand-drawn frames. However, each experiment studies distinct aspects of the relationships between frames’ material and depicted features regulating animators’ RFQ perception. Experiment 1 investigates conditions under which frames’ material feature similarity supersedes their depicted pose similarity in determining animators’ RFQ perception, finding that RFQ perception is moderately invariant to local material feature isometries. Experiment 2 explores how animators, even as experts in visual perception, vary in their subjective interpretations of depicted pose for the same material frame features, finding evidence of interpretations growing more diverse as perceptual cues grow less consistent. These findings point to extraordinarily sensitive and complex relationships between frames’ material and depicted features, which is unsurprising given hand-drawn characters’ lack of ground-truth body representations.

Reflecting on our findings yield further insights into animators’ RFQ perception. Our Experiment 2 findings indicate that pose resolution is likely an informationally destructive process, whereby the exact body pose envisioned by the animator who drew each frame cannot always be reliably recovered by other animators, despite their own perceptual pose resolution expertise. This further illuminates our Experiment 1 findings, as isometries uniquely preserve director-approved pose resolution strategies. Because these strategies intimately rely on relationships between frames’ material and depicted features, our findings suggest that animators cannot necessarily infer these relationships from frames with similar depicted poses and different perceptual cues to the frame they aim to draw. Therefore, animators are not innately fluent in relationships between material and depicted features, instead constructing them collaboratively by trial-and-error throughout production. Our RFQ analysis framework must accordingly prioritize pose resolution strategy preservation during reference frame retrieval, rendering purely HPE-based solutions – and DNNs generally, given their poor shape comprehension – insufficient. Yet performing dozens of material feature comparisons between hundreds of frames for each reference frame query would be grossly inefficient; moreover, depicted pose cannot be wholly discounted during RFQ analysis. Our framework must therefore model robust, perceptually-based relationships between frames’ material and depicted features such that pose comparisons efficiently and guaranteeably approximate material feature comparisons.

Experiments 1 and 2 both suggest that visual composability and decomposability factor strongly into animators’ RFQ perception. Experiment 1 illustrates animators’ ability to localize their comprehension of pose resolution strategies to a specific body part, regardless of global differences between frames’ depicted poses. More subtly, our Experiment 2 findings suggest that animators’ pose perception is sensitive to multiresolution symbolic relationships between local material perceptual cues. Frames’ perceptual cues are generally materially consistent (i.e., free from contour degradation and junction misplacement) and each symbolize coherent physical information about some local aspect of depicted pose. However, when

related local cues' symbolic meanings conflict with one another or with depicted body motion, animators struggle to integrate local cues into globally coherent pose interpretations. Together, these findings indicate that material frame features function similarly to a visually symbolic language for conveying volumetric character pose. Material features can be decomposed into smaller parts, with each part retaining its local symbolic information. Visual meaning arises from local-to-global material feature integration, but is degraded by inconsistent multiscale relationships between material features' symbolic meanings. Pose resolution thus involves translating animators' conceptual understandings of physical form into the film's visual symbolic language, which can be decomposed, restructured, and recomposed to engender different depicted pose perceptions, the coherence of which hinges upon consistency between local depicted pose information conveyed by local material components.

These cognitive processes – visual composability and decomposability, dual material and depicted feature similarity awareness, pose resolution strategy preservation, pose interpretation subjectivity, local-to-global perceptual cue integration, and multiresolution consistency evaluation – form the bedrock of our RFQ analysis approach, detailed in the following chapter.

Chapter 4

GENEO Sheaves

The perceptual experiment results of the previous chapter – alongside those discussed in Section 2.4 – are highly challenging to model computationally. Aside from 3D pose estimation for uncolored hand-drawn frames being a presently unsolved problem, an accurate model of professional animators’ RFQ perception must incorporate dual material and depicted feature similarity awareness, RFQ equivalence up to multiresolution isometry, perceptual cue consistency assessment, feasible pose and motion constraints, stylized deformation handling, and anticipation of animators’ various subjective pose perceptions for each frame. These perceptual features arise from core traits of human vision – including multiresolution integration, multiresolution contour geometry and topology awareness, part-wise relationship comprehension, and abstract visual reasoning – which are not adequately emulated by state-of-the-art DNN-based models of visual cognition. However, they are highly compatible with emerging subfields of Topological Data Analysis.

In this chapter, we step back from the specialized problem of RFQ analysis for hand-drawn frames and focus instead on the far more general problem of modeling and analyzing complex real-world systems. Rather than developing a computational framework exclusively suited to hand-drawn frame analysis, as is standard practice in hand-drawn animation technology research (Section 2.2), we devise a novel, general-purpose TDA framework, illustrating its utility through our application to RFQ perception modeling. This application successfully and explainably models all requisite perceptual features listed above, while our general-purpose framework invites myriad potential applications far beyond hand-drawn animation technology. These include but are not limited to motion capture analysis, macromolecular mechanics modeling, and 3D scene reconstruction. Pivotaly, we introduce new data modeling and evaluation features unattainable by previous methods, shifting our focus from studying hand-drawn frame data to studying the theory of data analysis itself.

In the latter context, a data analysis framework is a mechanism for converting observations of a real-world system into meaningful information about the system. It comprises an ordered collection of data transformations extracting and preserving information and data relationships relevant to the task at hand. The primary objects of study are thus data transformations, data relationships, and interactions between relationships and transformations. These are approached from two distinct perspectives, each with a robust topological formalization and significant practical success.

The first perspective examines frameworks whose transformations preserve relationships – formalized as group symmetries – within a homogeneous data set. Relationship preservation, or *equivariance*, guarantees that symmetries on input data induce corresponding symmetries on analysis results. Symmetries encode real-world properties of observed systems pertinent to analysis objectives – for example, convolu-

tional neural networks’ translation equivariance facilitates object recognition regardless of spatial location. Equivariant analysis frameworks, including CNNs, are formalized with Group Equivariant Non-expansive Operators (GENEOs). GENEOs define how best to select and exploit equivariant data relationships and transformations [20], achieving state-of-the-art performance in protein analysis [28], subgraph classification [2], and scene understanding [186] using less than 20 parameters. Although GENEOs can model local relationships between states of a singular subsystem, they cannot model relationships between states of multiple, independent subsystems of a given system, or, for instance, ensure that translating one subsystem and rotating another yields a globally valid system state. In many cases, this precludes GENEO analysis of complex real-world systems.

The second perspective concerns frameworks analyzing complex systems observed by multiple heterogeneous sensors, each recording a subsystem’s local state. To assess the system, sensor data must be fused into a globally valid system state respecting all relationships between subsystems. Heterogeneous data fusion is formalized with (cellular) sheaves: networks of sensor data alongside data transformations encoding relationships between subsystems [271]. Sheaves stitch together local information into a global picture of the system, enabling multiresolution consistency evaluation and optimization of related subsystems’ states; that is, sheaves can analyze systems at any level of granularity. They succeed in various domains including opinion dynamics [154], image feature detection [161], and geolocation [173]. Yet sheaves exclusively model relationships between subsystems, lacking GENEOs’ valuable insights into relationships within subsystems. Moreover, their most desirable fusion and analysis mechanisms require vector-valued data representations and linear data relationships, limiting their utility when vector representations are unsuitable and relationships between subsystems are nonlinear.

This chapter introduces GENEO sheaves, a novel data analysis framework synthesizing equivariant data analysis and heterogeneous data fusion. Combining sheaves’ and GENEOs’ mathematical and computational advantages, GENEO sheaves present a unified scheme for modeling, modulating, and evaluating heterogeneous multiresolution data relationships within and between subsystems of complex real-world systems. Whereas prior equivariant sheaf-like frameworks are limited to modeling systems with homogeneous data representations and only local or global homogeneous symmetries [288, 163, 18, 80], GENEO sheaves stitch together heterogeneous multiresolution subsystems and symmetries via equivariant maps encoding subsystem relationships (Figure 4.1). This setting undergirds our novel GENEO sheaf heterogeneous data fusion algorithm for function-valued – rather than vector-valued – data representations and non-linear data relationships, improving upon sheaf expressivity. Our algorithm prevents subsystems’ local behavior from violating global system state validity, emulating the tension between part-wise independence and system-wide coherence characteristic of real-world systems. This is formalized within our novel mathematical and data analytical theory of heterogeneous multiresolution equivariance, with which subsystems’ system-wide influence can be tracked and controlled. Most importantly, our theory ensures that GENEO sheaf data fusion is equivariant to heterogeneous multiresolution symmetries on (collections of) subsystems, guaranteeing

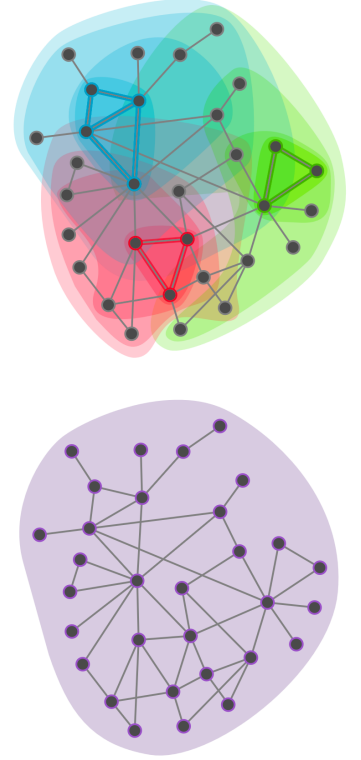


Figure 4.1: Conceptual diagram of heterogeneous multiresolution symmetries (top) acting over multiscale heterogeneous subsystems (red, green, blue) vs. homogeneous symmetries (bottom) exclusively acting either locally (purple-outlined vertices) or globally (purple overlay).

preservation of real-world system properties, behavior, and relationships in analysis results.

After overviewing sheaves and GENEOS (Section 4.1), we introduce GENEOS sheaves (Section 4.2) and our GENEOS sheaf heterogeneous data fusion algorithm (Section 4.3), improving upon computational and analytical capabilities of prior work. Next, we formalize our theory of heterogeneous multiresolution equivariance for GENEOS sheaves and prove our data fusion algorithm’s equivariance to heterogeneous multiresolution symmetries (Section 4.4). GENEOS sheaf pseudometrics and analytical tools are introduced for comparing and evaluating sensor observations of real-world systems, extending prior work on sheaf and GENEOS pseudometrics (Section 4.5). We then detail our GENEOS sheaf model of animators’ RFQ perception, proof-of-concept RFQ GENEOS sheaf implementation estimating hand-drawn characters’ 3D arm poses (Section 4.6), and proposed reference frame retrieval industry pipeline. Finally, we discuss the chapter’s contributions and opportunities for further development (Section 4.7). This research has been conducted in collaboration with Faraz Ahmad.

4.1 Sheaves & Group Equivariant Non-expansive Operators

4.1.1 Sheaves

For a given system, let Γ be a graph comprising vertex $v \in V(\Gamma)$ for each sensor observing a subsystem and edge $e \in E(\Gamma)$ connecting $v, u \in V(\Gamma)$ whenever the respective subsystems observed by v and u are related. Γ is therefore a topological space (specifically, a cellular complex) capturing the structure – but not nature – of relationships between subsystems observed by sensors $V(\Gamma)$. To specify the nature of subsystem relationships with respect to sensor observations, we construct a (*cellular*) *sheaf* over Γ .

Definition 4.1.1. A (*cellular*) *sheaf* $\mathcal{F}$ over graph Γ comprises:

1. Vector space $\mathcal{F}(v)$ for each vertex $v \in V(\Gamma)$, encoding each sensor’s local space of subsystem observations;
2. Vector space $\mathcal{F}(e)$ for each edge $e \in E(\Gamma)$, where observations of sensors $v, u \in V(\Gamma)$ incident to e can be meaningfully compared;
3. Linear map $\mathcal{F}_{v \leq e} : \mathcal{F}(v) \rightarrow \mathcal{F}(e)$ for each incident vertex-edge pair $v \leq e$, encoding the relationship from sensor v to sensor $u \in V(\Gamma)$ incident to e .

Vector spaces $\mathcal{F}(v)$ and $\mathcal{F}(e)$ are called *stalks*, and linear maps $\mathcal{F}_{v \leq e}$ are called *restriction maps*. The latter terminology conveys an important topological meaning, founded on the premise that related subsystems must share common information, and the information they share depends on their relationship. Topologically speaking, related subsystems overlap, and their intersection contains their shared information. Restriction maps transform subsystem observations $x_v \in \mathcal{F}(v)$ and $x_u \in \mathcal{F}(u)$ into data representations $\mathcal{F}_{v \leq e}(x_v)$ and $\mathcal{F}_{u \leq e}(x_u)$ comprising only the subsystems’ shared information, literally restricting related subsystems to their informational and topological intersection $\mathcal{F}(e)$.

Restriction maps homogenize heterogeneous sensor observations according to subsystem relationships, making it possible to meaningfully compare the local states of related subsystems. In practice, such comparisons quantify how close a collection of sensor observations is to representing a globally valid system state. If all sensors observe the system simultaneously, we have local observation $x_v \in \mathcal{F}(v)$ for each $v \in V(\Gamma)$. Aggregating these local observations forms an element x of the space of *0-cochains* $C^0(\Gamma; \mathcal{F}) = \bigoplus_{v \in V(\Gamma)} \mathcal{F}(v)$. The space of *1-cochains* $C^1(\Gamma; \mathcal{F}) = \bigoplus_{e \in E(\Gamma)} \mathcal{F}(e)$ is similarly defined. The *coboundary map*

$$\delta : C^0(\Gamma; \mathcal{F}) \rightarrow C^1(\Gamma; \mathcal{F}) \quad (4.1)$$

is defined $\delta(x)_e = \mathcal{F}_{v \leq e}(x_v) - \mathcal{F}_{u \leq e}(x_u)$ at each (arbitrarily oriented) edge $e = u \rightarrow v$, measuring consistency between the observed local states of related subsystems. Because δ locally quantifies consistency over each edge $e \in E(\Gamma)$, it becomes possible to geometrically identify where within the system inconsistencies arise.

Due to noise, errors, and other confounding factors, real-world simultaneous sensor observations are rarely globally consistent. Local observations must therefore be fused into a globally valid system state, called a *global section* of $\mathcal{F}$. Global sections arise when all states of related subsystems are entirely consistent with subsystem relationships; that is, when $\delta(x)_e = 0$ for every $e \in E(\Gamma)$. Different types of sheaves require different fusion techniques. Fusion on cellular sheaves is performed by generalizing the graph Laplacian to the sheaf-theoretic context.

Definition 4.1.2. For sheaf $\mathcal{F}$ over graph Γ , the (*cellular*) *sheaf Laplacian* $L_{\mathcal{F}} = \delta^T \delta : C^0(\Gamma; \mathcal{F}) \rightarrow C^0(\Gamma; \mathcal{F})$ at vertex v is defined

$$L_{\mathcal{F}}(x)_v = \sum_{v, u \leq e} \mathcal{F}_{v \leq e}^T (\mathcal{F}_{v \leq e}(x_v) - \mathcal{F}_{u \leq e}(x_u)) \quad (4.2)$$

Theorem 4.1 in [154] guarantees that $L_{\mathcal{F}}$ converges to a global section $x' \in \ker \delta$ that is the orthogonal projection of 0-cochain x – the system observation aggregating all initial sensor observations – onto $\ker \delta$. Recent work details $L_{\mathcal{F}}$ ’s myriad formulations, diffusion dynamics, and analytical capabilities [154, 13, 157], all of which hinge upon its linear algebraic formulation.

A key drawback of the sheaf Laplacian is its reliance on stalks of (vector) data spaces, rather than of (possibly finite and/or non-vector) data sets. A sheaf with data set stalks can be formalized as a *sheaf of pseudometric spaces* if every stalk admits a pseudometric [272]. Fusion on pseudometric sheaves is an optimization problem: for initial system observation x and pseudometric D_v on stalk $\mathcal{F}(v)$ for every $v \in V(\Gamma)$, find

$$\operatorname{argmin}_{x' \in \ker \delta} D_{\mathcal{F}}(x, x') = \operatorname{argmin}_{x' \in \ker \delta} \sup_{v \in V(\Gamma)} D_v(x_v, x'_v) \quad (4.3)$$

However, the space $\ker \delta$ of global sections must be known in advance. Global sections can either be manually specified or computed with linear algebra, although the latter requires a cellular sheaf [271].

While sheaves are highly computationally expressive, their real power lies in their topological theory. An algebraic feature called *cohomology* encodes fundamental obstructions to global data consistency and measures subsystems’ independence from the rest of the system [275, 154]. As discussed in Chapter 2, cohomology is dual to homology, and sheaves’ coboundary map is dual to the boundary operator relating topological spaces’ multi-dimensional features. Rather than evaluating and modulating sensor data, cohomology captures essential properties of the system itself, illustrated by successes of boundary operators and Laplacians in hand-drawn animation technology. Yet like the sheaf Laplacian, cohomology cannot be defined or computed without vector space stalks and linear restriction maps, creating a trade-off between accessing sheaves’ full suite of features vs. gaining expressivity beyond linear algebraic data. Although the benefits of reformulating these features for more expressive types of sheaves are recognized [275, 271], pseudometric sheaves have been the only effort towards this end.

Sheaves support diverse real-world applications, including signal processing [139, 273, 173], knowledge modeling [257, 132], and computer graphics. In sheaf-based scene analysis and image mosaicing, vertex stalks are populated with vectorized data extracted from local scene images. Restriction maps encode relationships between local images – including camera position, lighting, and object locations – needed to evaluate the scene [271]. Local images are then stitched together into a mosaic, while inconsistencies within the sheaf reveal scene condition differences among local images, indicating how the scene might change

over time. Sheaves also support 3D shape reconstruction, stitching together 2D shape data from multiple view angles of a 3D object. Shape data at each view angle can be represented by persistent homology, which “scans” the shape’s surface along each given view axis. Sheaf theory enforces quality guarantees on the 3D surface reconstructed from these scans, as well as parametrizing the space of reconstructed 3D shapes for accurate and efficient shape comparison [7]. In these applications and our own (Section 4.6.2), sheaves employ controllable and explainable spatial and functional reasoning to synthesize a collection of isolated visual data into a globally coherent understanding of a physical entity.

4.1.2 Group Equivariant Non-expansive Operators (GENEOs)

Sensor observations of a system are often expressible as bounded functions $\varphi : X \rightarrow \mathbb{R}$, where the structure of set $X \neq \emptyset$ is determined by the sensor itself. For instance, CT scans are functions from a 3D helix to the quantity of matter encountered by the X-ray beam at each point in space [123]. The set Φ of *admissible functions* comprises all possible observations φ that can be made by the given sensor. Observations $\varphi, \varphi' \in \Phi$ are compared using pseudometric

$$D_{\Phi}(\varphi, \varphi') = \sup_{x \in X} |\varphi(x) - \varphi'(x)| \quad (4.4)$$

Studying Φ entails studying data relationships – formalized as group symmetries – arising from the interaction between the sensor and observed system behavior. The group $\text{Aut}_{\Phi}(X)$ of Φ -*preserving automorphisms* contains all symmetries g on X such that $\varphi \circ g, \varphi \circ g^{-1} \in \Phi$ for every $g \in \text{Aut}_{\Phi}(X)$, $\varphi \in \Phi$. However, not all data relationships are useful in every data analysis context. To define a Group Equivariant Non-expansive Operator (GENEO) for a specific analysis task, we begin by selecting subgroup $G \leq \text{Aut}_{\Phi}(X)$ encoding relationships relevant to the analysis objectives at hand – e.g., CNNs use group G of planar translations. The tuple (Φ, G) is called a *perception pair*.

Next, we consider the information we wish to extract from Φ . Data analysis generally involves data transformations which jettison irrelevant information, leaving only information pertinent to the given analysis task. This is critical for meaningfully comparing sensor observations, since irrelevant information (e.g., noise) hinders accurate data evaluation. Let admissible function space Ψ contain the target data representations we aim to obtain from Φ . Data transformation $F : \Phi \rightarrow \Psi$ must be *non-expansive*, meaning that F decreases dissimilarity among data in Φ by virtue of discarding irrelevant information. F must also preserve data relationships defined by G so that they remain useful when evaluating data in Ψ . These stipulations yield the notion of a GENEO.

Definition 4.1.3. For perception pairs (Φ, G) and (Ψ, H) with group homomorphism $T : G \rightarrow H$, the function tuple $(F, T) : (\Phi, G) \rightarrow (\Psi, H)$ is a *GENEO* if for all $\varphi, \varphi' \in \Phi$ and all $g \in G$,

1. $F(\varphi \circ g) = F(\varphi) \circ T(g)$ (equivariance)
2. $D_{\Psi}(F(\varphi), F(\varphi')) \leq D_{\Phi}(\varphi, \varphi')$ (non-expansivity)

In many applications, equivariance group G defines not only data relationships, but also data equivalences with which observations can be classified. For example, drug design benefits from classifying molecules by physical and chemical properties invariant to their 3D spatial orientations, given by $G = SO(3)$. The *ground-truth pseudometric*

$$d_G(\varphi, \varphi') = \inf_{g \in G} D_{\Phi}(\varphi, \varphi' \circ g) \quad (4.5)$$

measures differences between observations $\varphi, \varphi' \in \Phi$ up to data equivalences defined by G [20].

Because the space of GENEOs is closed under composition and convex combination [263], sophisti-

cated data transformations can be constructed by combining simple GENEOS. Moreover, compactness of the space of GENEOS ensures that any data transformation can be approximated by a finite set of sampled GENEOS. Together, compactness and convexity provide strategies for choosing and hierarchically connecting GENEOS into operator networks, explainably modeling equivariant neural networks [20]. In practice, these strategies dramatically reduce the number of parameters needed for state-of-the-art performance on object recognition [20], protein pocket detection [28], and point cloud feature detection [186]. Although various GENEOS formulations exist, including graph GENEOS [2], set equivariance [120], and linear GENEOS [29], these cannot model symmetries acting on isolated subsystems, i.e., local parts of system data.

4.2 GENEOS Sheaves

Although sheaves are considered the canonical framework for heterogeneous sensor integration [271], they are ill-equipped to gauge interactions between sensors themselves and system behavior. This is the domain of GENEOS, which capture the geometric regularities arising from real-world phenomena and the mechanisms by which they are observed. Because these regularities so profoundly benefit data analysis [42], it makes sense to consider data transformations stitching together heterogeneous sensor observations according to subsystem relationships alongside data transformations encoding heterogeneous multiresolution symmetries on (collections of) observed subsystems.

Cellular sheaves lack a mechanism for defining, synthesizing, or propagating heterogeneous stalk-wise symmetries, comporting only with global homogeneous symmetries [80, 157]. Analogously, existing equivariant analysis frameworks cannot localize symmetries to subsystems, precluding learning and equivalence detection on subparts of data. To achieve heterogeneous multiresolution equivariance, local symmetries and sensor data must be jointly stitched together, ensuring that local symmetry modulation preserves global data consistency. Moreover, sheaves' data analysis capabilities – including multiresolution data evaluation, obstruction detection, and diffusion – should extend to symmetry analysis.

For a given system, let Γ be a graph comprising vertex $v \in V(\Gamma)$ for each sensor observing a subsystem and edge $e \in E(\Gamma)$ connecting $v, u \in V(\Gamma)$ whenever the respective subsystems observed by v and u are related. For each sensor, the set Φ_v of admissible observations comprises bounded functions $\varphi_v : X_v \rightarrow \mathbb{R}$, with the structure of set $X_v \neq \emptyset$ determined by the sensor. The group $\text{Aut}_{\Phi_v}(X_v)$ comprises Φ_v -preserving automorphisms $g : X_v \rightarrow X_v$ such that $\varphi_v \circ g, \varphi_v \circ g^{-1} \in \Phi_v$ for all $\varphi_v \in \Phi_v, g \in \text{Aut}_{\Phi_v}(X_v)$. We assume that $\text{Aut}_{\Phi_v}(X_v)$ is a finite group (or a finite approximation of an infinite group) and thus admits an orthogonal representation; we do not distinguish between groups and their orthogonal representations for notational and expository simplicity. We also assume that the sensor data is not arbitrary, and encodes real system observations taken concurrently. A GENEOS sheaf connects this heterogeneous collection of local symmetries and sensor data into a global system model.

Definition 4.2.1. For directed graph Γ , a *GENEOS sheaf* $\mathbf{F}$ on Γ comprises

1. Perception pair stalks $(\Phi_v, \text{Aut}_{\Phi_v}(X_v))$ for each $v \in V(\Gamma)$ and $(\Psi_e, \text{Aut}_{\Psi_e}(Y_e))$ for each $e \in E(\Gamma)$,
2. Equivariance group $G_{v,e} \leq \text{Aut}_{\Phi_v}(X_v)$ for each incident $v \triangleleft e$ with $v \in V(\Gamma), e \in E(\Gamma)$,
3. GENEOS $(F_{v \triangleleft e}, T_{v \triangleleft e}) : (\Phi_v, G_{v,e}) \rightarrow (\Psi_e, H_e)$ and $(F_{u \triangleleft e}, T_{u \triangleleft e}) : (\Phi_u, G_{u,e}) \rightarrow (\Psi_e, H_e)$ with $H_e \leq \text{Aut}_{\Psi_e}(Y_e)$ for edge $e \in E(\Gamma)$ incident to $v, u \in V(\Gamma)$.

To emphasize their connection with sheaf restriction maps, we call $(F_{v \triangleleft e}, T_{v \triangleleft e})$ and $(F_{u \triangleleft e}, T_{u \triangleleft e})$ *restriction GENEOS*, and groups $G_{v,e}, G_{u,e}$ *restriction equivariance groups*.

GENEOS sheaves unify sheaves' and GENEOS' topological and data analytical perspectives. Like

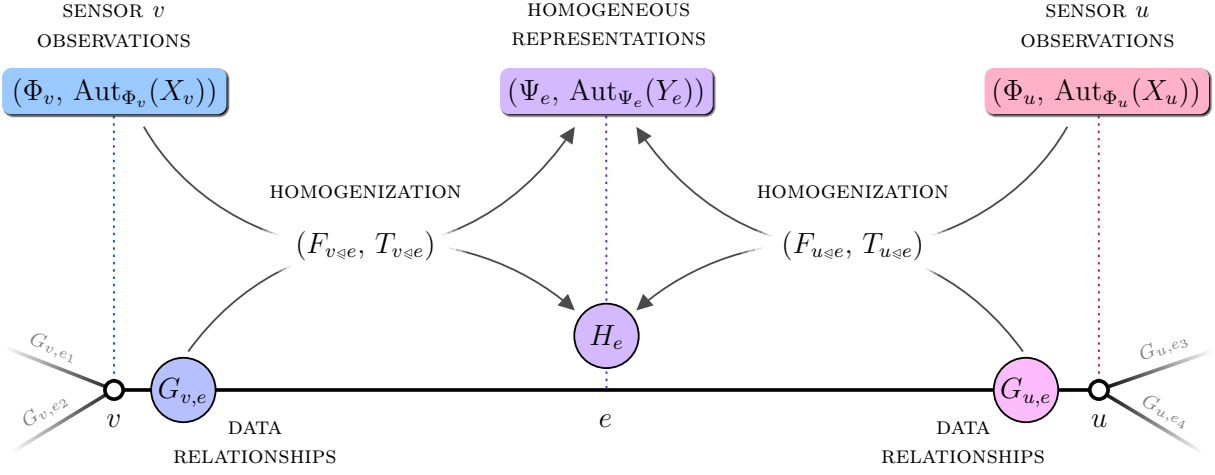


Figure 4.2: Illustration of a GENEO sheaf $\mathbf{F}$ on graph Γ (Definition 4.2.1) at adjacent $v, u \in V(\Gamma)$ sharing edge $e \in E(\Gamma)$. Mathematical notations are given alongside conceptual descriptions for expository clarity.

cellular sheaves, the topology of Γ captures the structure of relationships between subsystems observed by sensors $V(\Gamma)$. Groups $\text{Aut}_{\Phi_v}(X_v)$ encode local subsystem regularities engendered by interactions between each sensor $v \in V(\Gamma)$ and subsystem behavior recorded by observations Φ_v . Choice of unique $G_{v,e} \leq \text{Aut}_{\Phi_v}(X_v)$ for each incident $v \leq e$ defines regularities relevant to the local data analysis objectives over each edge $e \in E(\Gamma)$, encoding subsystem behavior common to both Φ_v and Ψ_e . Importantly, adjacent vertices v, u sharing edge e can have different edge-wise restriction equivariance groups $G_{v,e}, G_{u,e}$ so long as $T_{v \leq e} : G_{v,e} \rightarrow H_e \leftarrow G_{u,e} : T_{u \leq e}$. Restriction GENEOs stitch together related subsystems' heterogeneous observations Φ_v, Φ_u and regularities $G_{v,e}, G_{u,e}$, homogenizing the local topologies on X_v and X_u respectively induced by ground-truth pseudometrics $d_{G_{v,e}}$ and $d_{G_{u,e}}$ (Equation 4.5).

From a data analysis standpoint, restriction GENEOs convert heterogeneous sensor observations $\varphi_v \in \Phi_v, \varphi_u \in \Phi_u$ into homogeneous data representations $F_{v \leq e}(\varphi_v), F_{u \leq e}(\varphi_u) \in \Psi_e$ which can be meaningfully compared. Meaningful comparison of heterogeneous data is possible when (i) observations φ_v, φ_u contain shared information and (ii) there exist data transformations $F_{v \leq e}, F_{u \leq e}$ extracting and homogenizing this shared information. For instance, 2D and 3D shapes can be meaningfully compared when 3D shapes' occluding contours are projected to 2D outlines. Because GENEO sheaves are not confined to vector data and linear restriction maps, modes of data representation and homogenization are far more expressive than those of cellular sheaves.

Following from both sheaf and GENEO theories, we study the system modeled by GENEO sheaf $\mathbf{F}$ by studying the sets of raw and homogenized sensor observations. These are respectively the sets of *admissible 0-assignments* $\Phi_{\mathbf{F}}$ and *admissible 1-assignments* $\Psi_{\mathbf{F}}$, defined

$$\Phi_{\mathbf{F}} = \prod_{v \in V(\Gamma)} \Phi_v \quad \Psi_{\mathbf{F}} = \prod_{e \in E(\Gamma)} \Psi_e \quad (4.6)$$

Like a sheaf 0-cochain, admissible 0-assignment $\varphi \in \Phi_{\mathbf{F}}$ aggregates choice of observation $\varphi_v \in \Phi_v$ for every $v \in V(\Gamma)$.

To understand how well φ portrays a globally valid picture of the observed system, we must evaluate consistency between related subsystems' observed states $\varphi_v \in \Phi_v$ and $\varphi_u \in \Phi_u$ according to subsystem

relationships $(F_{v \preceq e}, T_{v \preceq e})$ and $(F_{u \preceq e}, T_{u \preceq e})$. That is, we measure how close φ is to being a global section. Adapting GENEOS' admissible function pseudometric (Equation 4.4), we say that φ is a *global section of $\mathbf{F}$* if

$$D_{\Psi_e}(F_{v \preceq e}(\varphi_v), F_{u \preceq e}(\varphi_u)) = \sup_{y \in Y_e} |F_{v \preceq e}(\varphi_v)(y) - F_{u \preceq e}(\varphi_u)(y)| = 0 \quad (4.7)$$

for every edge $e \in E(\Gamma)$. An *A-local section* on subgraph $A \subseteq \Gamma$ is an admissible 0-assignment $\varphi \in \Phi_{\mathbf{F}}$ for which $D_{\Psi_e}(F_{v \preceq e}(\varphi_v), F_{u \preceq e}(\varphi_u)) = 0$ for every $e \in E(A)$.

As discussed, raw sensor observations are rarely globally consistent, requiring data fusion to optimize consistency. Like the relationship between the cellular sheaf coboundary map and sheaf Laplacian, our GENEOS sheaf consistency measurement should comport with a data fusion strategy. By exploiting restriction GENEOS equivariance, we can convert consistency-optimizing transformations on Ψ_e into respective consistency-optimizing transformations on Φ_v and Φ_u . The *$H_{\mathbf{F}}$ -coboundary map* $\delta_{H_{\mathbf{F}}} : \Phi_{\mathbf{F}} \rightarrow H_{\mathbf{F}} = \bigoplus_{e \in E(\Gamma)} H_e$ is defined $\delta_{H_{\mathbf{F}}}(\varphi) = h$, where h is a choice of transformation $h_e \in H_e$ for each $e \in E(\Gamma)$ such that

$$D_{\Psi_e}(F_{u \preceq e}(\varphi_u), F_{v \preceq e}(\varphi_v) \circ h_e) = \inf_{h \in H_e} D_{\Psi_e}(F_{u \preceq e}(\varphi_u), F_{v \preceq e}(\varphi_v) \circ h) = d_{H_e}(F_{u \preceq e}(\varphi_u), F_{v \preceq e}(\varphi_v)) \quad (4.8)$$

at each (arbitrarily oriented) edge $e = u \rightarrow v$. The set of *global $H_{\mathbf{F}}$ -sections* of $\mathbf{F}$ is given by

$$\ker \delta_{H_{\mathbf{F}}} = \{\varphi \in \Phi_{\mathbf{F}} \mid \delta_{H_{\mathbf{F}}}(\varphi)_e = \text{id}_{H_e} \text{ for all } e \in E(\Gamma)\} \quad (4.9)$$

By restriction GENEOS equivariance, $\varphi_v \circ T_{v \preceq e}^{-1}(h_e)$ optimizes data consistency at v over edge e while maintaining the quality, structure, and integrity of sensor observation φ_v for all $h \in H_e$, $v \in V(\Gamma)$, $e \in E(\Gamma)$. In the next section, we will see how transformations $T_{v \preceq e}^{-1}(h_e)$ underpin our data fusion algorithm.

Modeling complex relationships Because convex combinations of GENEOS are GENEOS [20], combining simple data transformations lets us model complex local relationships. For instance, multiple GENEOS encoding physical and chemical molecular properties can be combined for protein pocket detection [28], and multiple GENEOS detecting simple 3D scene geometries can be combined for object detection within 3D scenes [186]. To extend this functionality to GENEOS sheaves, consider the space of all possible restriction GENEOS from $(\Phi_v, G_{v,e})$ to (Ψ_e, H_e) for incident $v \preceq e$ with $v \in V(\Gamma)$, $e \in E(\Gamma)$. From this space, select $(F_1, T_{v \preceq e}), \dots, (F_n, T_{v \preceq e})$ all sharing homomorphism $T_{v \preceq e} : G_{v,e} \rightarrow H_e$. Let $\gamma_1, \dots, \gamma_n \in \mathbb{R}$ with $\sum_{1 \leq i \leq n} |\gamma_i| = 1$. Define restriction GENEOS $(F_{v \preceq e}, T_{v \preceq e}) : (\Phi_v, G_{v,e}) \rightarrow (\Psi_e, H_e)$ such that

$$F_{v \preceq e}(\varphi_v) = \sum_{1 \leq i \leq n} \gamma_i F_i(\varphi_v)$$

This is a straightforward technique for constructing sophisticated restriction GENEOS from combinations of simple, easily interpretable restriction GENEOS.

4.3 GENEOS Sheaf Data Fusion

Heterogeneous data fusion constructs a globally consistent picture of a system from interrelated heterogeneous sensor observations. As discussed, raw observations are virtually never globally consistent. Fusion thus aims to match the collection of raw observations to the globally consistent system state they best resemble. In practice, researchers often work with real-world sensor data alongside a computational system model. The model comprises model sensors, whose observations can and should be globally consistent.

Our heterogeneous data fusion algorithm plugs raw sensor observations into a GENEIO sheaf system model, then optimizes consistency throughout the system model and between corresponding raw and model observations. Consistency optimization is performed via the GENEIO sheaf Laplacian, the first adaptation of the cellular sheaf Laplacian to non-cellular sheaves.

4.3.1 The GENEIO Sheaf Laplacian

Adapting the cellular sheaf Laplacian's myriad desirable features to non-cellular sheaves is a valuable endeavor. However, progress has been stymied because the cellular sheaf Laplacian is inherently linear algebraic, generalizing tenuously to other mathematical settings. Although the cellular sheaf Laplacian has a nonlinear formulation [154], its nonlinear component is applied after vector data is homogenized by linear restriction maps. Rather, a truly non-cellular sheaf Laplacian should homogenize non-vector data via nonlinear restriction maps. GENEIO sheaves can support such a construction by leveraging restriction GENEIO equivariance.

The GENEIO sheaf Laplacian must convert consistency-optimizing transformations on homogeneous data representations $F_{v \leq e}(\varphi_v)$, $F_{u \leq e}(\varphi_u)$ into respective consistency-optimizing transformations on heterogeneous sensor observations $\varphi_v \in \Phi_v$, $\varphi_u \in \Phi_u$. It should also be $\Phi_{\mathbf{F}}$ -preserving; that is, it must output (approximately) valid sensor observations $\varphi'_v \in \Phi_v$ for all $v \in V(\Gamma)$, thereby preventing data distortion and degradation during Laplacian diffusion.

To motivate the GENEIO sheaf Laplacian, consider the general bundle (sheaf) Laplacian [13] written $L_{\mathcal{B}}(\mathbf{x})_v = \sum_{u, v \leq e} \mathbf{x}_v - O_{vu} \mathbf{x}_u$ for consistency-optimizing orthogonal transformation O_{vu} . In other words, $L_{\mathcal{B}}(\mathbf{x})_v$ averages the (orthogonal Euclidean) isometries minimizing edge-wise disagreement for each incident $v \leq e$. Recall that the $H_{\mathbf{F}}$ -coboundary map outputs edge-wise consistency-optimizing transformation $h_e \in H_e$ on Ψ_e for every $e \in E(\Gamma)$, from which we can define vertex-wise consistency-optimizing transformation $T_{v \leq e}^{-1}(h_e)$ on Φ_v for every $v \in V(\Gamma)$. Because $T_{v \leq e}^{-1}(h_e) \in \text{Aut}_{\Phi_v}(X_v)$, it is an isometry on X_v [20]. Moreover, restriction GENEIO equivariance guarantees that

$$D_{\Psi_e}(F_{u \leq e}(\varphi_u), F_{v \leq e}(\varphi_v) \circ h_e) = D_{\Psi_e}(F_{u \leq e}(\varphi_u), F_{v \leq e}(\varphi_v \circ T_{v \leq e}^{-1}(h_e)))$$

for $h_e = \delta_{H_{\mathbf{F}}}(\varphi)_e$ at every $v \in V(\Gamma)$ and $e \in E(\Gamma)$.

Definition 4.3.1. Let $\varphi \in \Phi_{\mathbf{F}}$. For each $v \in V(\Gamma)$, let $g_{v,e} = T_{v \leq e}^{-1}(\delta_{H_{\mathbf{F}}}(\varphi)_e) \in G_{v,e}$ and

$$a_{v,e} = \underset{a \in \text{Aut}_{\Phi_v}(X_v) \setminus G_{v,e}}{\text{argmin}} \quad D_{\Psi_e}(F_{u \leq e}(\varphi_u), F_{v \leq e}((\varphi_v \circ g_{v,e}) \circ a)). \quad (4.10)$$

The *GENEIO sheaf Laplacian* at each vertex $v \in V(\Gamma)$ is defined

$$L_{\mathbf{F}}(\varphi)_v = \sum_{v \leq e} g_{v,e} \circ a_{v,e} \quad (4.11)$$

At each vertex $v \in V(\Gamma)$, the GENEIO sheaf Laplacian aggregates Φ_v -preserving isometries on sensor observation $\varphi_v \in \Phi_v$, with each isometry optimizing consistency between v and one of its neighbors u . Put differently, $L_{\mathbf{F}}(\varphi)_v$ measures the aggregate disagreement between φ_v and observations φ_u for all adjacent $u \in V(\Gamma)$. Like cellular sheaf diffusion, our implementation of $L_{\mathbf{F}}$ diffusion (Section 4.6.2) transforms $\varphi(t)_v$ by a small increment of $L_{\mathbf{F}}(\varphi(t))_v$ at each diffusion step t . Since these increments are scaled-down averages of (orthogonal representations of) isometries in Φ_v , $L_{\mathbf{F}}$ diffusion is approximately $\Phi_{\mathbf{F}}$ -preserving.

Like the cellular sheaf Laplacian, $L_{\mathbf{F}}$'s diffusion dynamics are designed to steer input heterogeneous

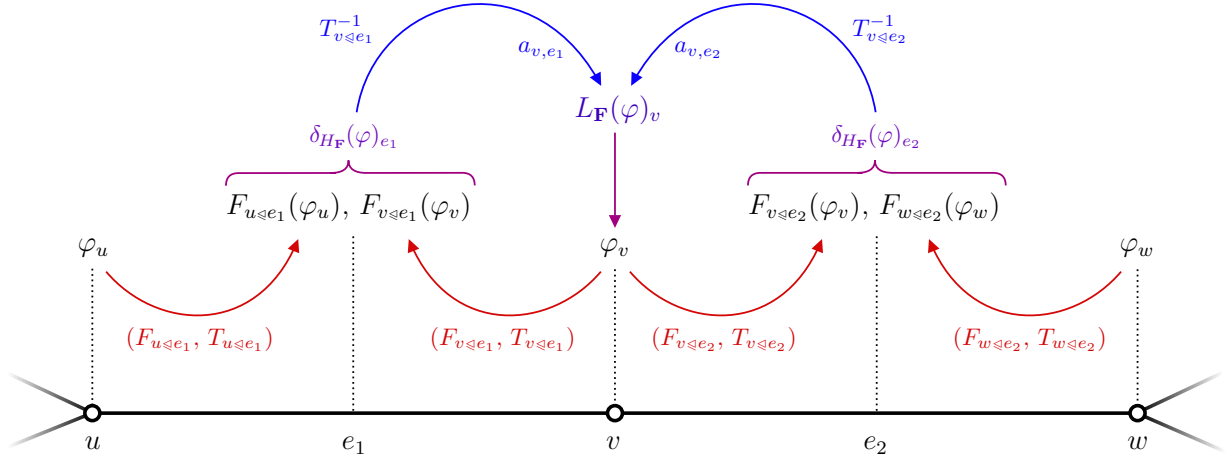


Figure 4.3: Illustration of the GENEO sheaf Laplacian (Definition 4.3.1) at vertex $v \in V(\Gamma)$, showing the cycle of data exchange and integration across vertex and edge stalks at each diffusion step.

sensor observations towards a globally valid system state. In particular, we conjecture that $L_{\mathbf{F}}$ converges to global $H_{\mathbf{F}}$ -sections.

Conjecture 4.3.2. *For initial $\varphi(0) \in \Phi_{\mathbf{F}}$, solutions $\varphi(t)$ to the GENEO sheaf Laplacian diffusion dynamics equation $\frac{d\varphi}{dt} = -\alpha L_{\mathbf{F}}(\varphi)$ for $\alpha > 0$ converge as $t \rightarrow \infty$ towards a global $H_{\mathbf{F}}$ -section $\varphi' \in \ker \delta_{H_{\mathbf{F}}}$.*

Although a formal proof of Conjecture 4.3.2 is left to future work, we have strong reasons to expect it is true. Crucially, our computational implementation of $L_{\mathbf{F}}$ in our data fusion algorithm (Section 4.6.2) exhibits the Conjecture’s predicted convergence behavior. Moreover, the GENEO sheaf Laplacian is constructed analogously to the cellular sheaf Laplacian, with both operators minimizing neighborhood disagreement at each diffusion step. We can therefore reasonably expect that as $t \rightarrow \infty$, $L_{\mathbf{F}}$ pushes φ_v towards $\varphi_v \circ L_{\mathbf{F}}(\varphi)_v = \varphi_v \circ \frac{1}{\deg(v)} \sum_{v \leq e} (g_{v,e} \circ a_{v,e})$; that is, towards φ_v transformed by the average of consistency-optimizing automorphisms $g_{v,e} \circ a_{v,e}$ over each incident edge $e \ni v$. If this is the case, Equation 4.8 ensures that $\delta_{H_{\mathbf{F}}}(\varphi(t))_e$ converges to id_{H_e} as edge-wise consistency is optimized. Our mathematical reasoning and computational validation justifies our assumption that Conjecture 4.3.2 holds.

The relationship between $L_{\mathbf{F}}$ and global $H_{\mathbf{F}}$ -sections is described by the following lemma.

Lemma 4.3.3. $\ker L_{\mathbf{F}} \subseteq \ker \delta_{H_{\mathbf{F}}}$.

Proof. For $\varphi \in \ker L_{\mathbf{F}}$, we have $L_{\mathbf{F}}(\varphi)_v = \sum_{v \leq e} g_{v,e} \circ a_{v,e} = \text{id}_{\text{Aut}_{\Phi_v}(X_v)}$ for all $v \in V(\Gamma)$. $a_{v,e} \neq g_{v,e}^{-1}$ because $a_{v,e} \notin G_{v,e}$, implying that $a_{v,e} = \text{id}_{\text{Aut}_{\Phi_v}(X_v)} = g_{v,e}$. Since $T_{v \leq e}(g_{v,e}) = \text{id}_{H_e}$, we have $\varphi \in \ker \delta_{H_{\mathbf{F}}}$. $\square$

4.3.2 Harmonic Extension for Data Fusion

Presently, the only data fusion algorithm for non-cellular sheaves matches real-world sensor observations to the nearest global section of the sheaf system model [271]. This requires the space of global sections to be known in advance, which is not always feasible in practice. Unless global sections can be manually specified, they must be computed via linear algebra from a cellular sheaf approximating the original sheaf, potentially degrading the original sensor data and system model. We propose an alternative approach to

data fusion via the GENEO sheaf Laplacian and harmonic extension.

Definition 4.3.4. For GENEO sheaf $\mathbf{F}$ on graph Γ and subgraph $A \subseteq \Gamma$, let $\varphi_A \in \Phi_A$ be a choice of $\varphi_u \in \Phi_u$ for all $u \in V(A)$. A $H_{\mathbf{F}}$ -harmonic extension of φ_A to the rest of Γ is an admissible 0-assignment $\varphi \in \Phi_{\mathbf{F}}$ such that $\varphi|_A = \varphi_A$ and $L_{\mathbf{F}}(\varphi)_v = \text{id}_{\text{Aut}_{\Phi_v}(X_v)}$ for all $v \in V(\Gamma \setminus A)$.

Let subgraph A be our collection of real-world sensors and φ_A be our raw sensor observations. Due to noise, errors, and the impossibility of extrapolating a complex whole from one of its parts, we cannot correctly interpret real observation φ_u for $u \in V(A)$ simply by finding its closest model observation $\varphi_v \in \Phi_u$ for adjacent $v \in V(\Gamma \setminus A)$. Rather, correct interpretation of a single input sensor observation requires consideration of all input sensor observations, as each sensor's meaning depends on the meanings of all other sensors observing the system. Philosophically speaking, observations of real-world phenomena are immutable, but interpretation of their meaning depends upon the given informational context – which, in our case, is supplied by data diffused across model system $\Gamma \setminus A$. Our algorithm takes the $H_{\mathbf{F}}$ -harmonic extension of raw observations φ_A to the rest of Γ , yielding a globally consistent model system state matching input observations.

The following theorem assumes that Conjecture 4.3.2 holds, and has been computationally validated by our data fusion algorithm implementation (Section 4.6.2).

Theorem 4.3.5. For $u \in V(A)$, $v \in V(\Gamma \setminus A)$, and $\alpha > 0$, the A -restricted dynamics $L_{\mathbf{F}}^A$ defined

$$\left. \frac{d\varphi}{dt} \right|_u = \text{id}_{\text{Aut}_{\Phi_u}(X_u)} \quad \left. \frac{d\varphi}{dt} \right|_v = -\alpha L_{\mathbf{F}}(\varphi)_v \quad (4.12)$$

on initial $\varphi(0) \in \Phi_{\mathbf{F}}$ converge toward a $H_{\mathbf{F}}$ -harmonic extension of φ_A .

Proof. For adjacent $u, v \in V(\Gamma)$, let $u \in V(A)$, $v \in V(\Gamma \setminus A)$. Assuming that Conjecture 4.3.2 holds, $\lim_{t \rightarrow \infty} \varphi(t)_v = \varphi(0)_v \circ (g_{v,e} \circ a_{v,e})$ such that

$$g_{v,e} \circ a_{v,e} = \underset{g \in G_{v,e}, a \in \text{Aut}_{\Phi_v}(X_v) \setminus G_{v,e}}{\text{argmin}} D_{\Psi_e}(F_{u \leq e}(\varphi_u), F_{v \leq e}((\varphi_v \circ g) \circ a)).$$

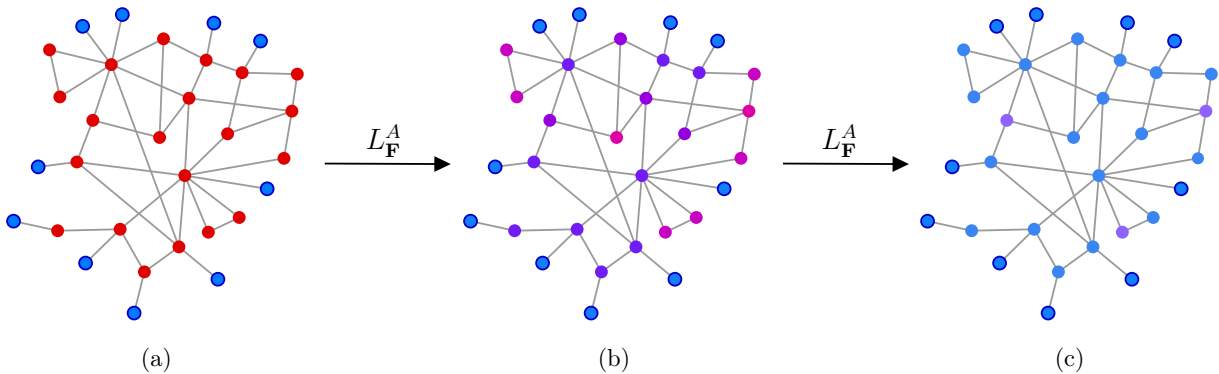


Figure 4.4: GENEO sheaf data fusion. (a) Initial admissible 0-assignment comprising input raw, real-world sensor observations (blue nodes) and model sensor states (red nodes). (b) $\mathbf{F}$ during data fusion, with model sensors (pink and purple nodes) evolving to match fixed real-world sensor observations φ_A over subgraph A (blue nodes). (c) $\mathbf{F}$ after data fusion, showing the $H_{\mathbf{F}}$ -harmonic extension of real-world sensor observations (nodes outlined dark blue) to the model system. Obstructions to global consistency are recorded in the model system (lavender nodes).

Then the A -restricted dynamics on initial $\varphi(0) \in \Phi_{\mathbf{F}}$ converge as $t \rightarrow \infty$ toward a $H_{\mathbf{F}}$ -harmonic extension of φ_A . $\square$

Unlike prior work, our fusion algorithm does not require advance knowledge or computation of our GENEIO sheaf system model's global sections. This in turn circumvents the need to cellularly approximate our GENEIO sheaf, fully preserving our sensor data and system model during fusion. Moreover, our algorithm optimizes system-wide consistency without requiring the output model system state to be a global section, creating new analytical opportunities. Whereas prior work identifies obstructions to system-wide consistency directly from raw sensor data, harmonic extension propagates obstructions to the model system (Figure 4.4c) where their causes and consequences are easily assessed. Our algorithm also accommodates flexible relationships between sensor observations and the system model, which may present as certain kinds of obstructions to global consistency. This benefits applications involving variable part-wise relationships, such as hand-drawn characters' bodily deformations (Section 4.6).

4.4 Heterogeneous Multiresolution Equivariance

Equivariant data analysis frameworks guarantee that symmetries on input data induce corresponding symmetries on analysis results [42]. Despite their growing prominence, existing frameworks exclusively admit global symmetries on data – for instance, rotating or reflecting a photograph. Yet many real-world systems admit heterogeneous multiresolution symmetries – for instance, rotating a human arm or exchanging left and right sides of a body pose (Figure 1.4). In applications such as our own (Section 4.6), analyzing, relating, and controlling heterogeneous multiresolution symmetries is essential for data analysis. GENEIO sheaves provide the setting for such functionality, yielding the novel theory of heterogeneous multiresolution equivariance detailed in this section. After defining our theory's underlying mathematical machinery, we introduce two types of heterogeneous multiresolution symmetries. Most importantly, we prove our data fusion algorithm's equivariance to both types of heterogeneous multiresolution symmetries, guaranteeing that symmetries on raw heterogeneous sensor data induce corresponding symmetries on our algorithm's output model system state.

4.4.1 GENEIO Sheaf Morphisms

Before introducing symmetries on GENEIO sheaves, we take a brief category theoretic interlude. Heterogeneous multiresolution symmetries demand a rigorous mathematical formulation to underpin our proofs of their behavior during data fusion. They are best formulated as *morphisms*, or structure-preserving maps, between GENEIO sheaves.

Definition 4.4.1. For GENEIO sheaves $\mathbf{F}, \mathbf{F}'$ on graph Γ , a *GENEIO sheaf morphism* $(\mu, M) : \mathbf{F} \rightarrow \mathbf{F}'$ comprises functions $(\mu_v, M_v) : (\Phi_v, \text{Aut}_{\Phi_v}(X_v)) \rightarrow (\Phi'_v, \text{Aut}_{\Phi'_v}(X'_v))$ for all $v \in V(\Gamma)$ and functions $(\mu_e, M_e) : (\Psi_e, H_e) \rightarrow (\Psi'_e, H'_e)$ for all $e \in E(\Gamma)$ such that (μ, M) commutes with restriction GENEIOs. That is, $\mu_e(F_{v \preceq e}(\varphi_v)) = F'_{v \preceq e}(\mu_v(\varphi_v))$ and $M_e(T_{v \preceq e}(g)) = T'_{v \preceq e}(M_v(g))$ for all $\varphi_v \in \Phi_v$, all $g \in G_{v,e}$, and each incident $v \preceq e$ with $v \in V(\Gamma)$, $e \in E(\Gamma)$.

For all $\varphi_v, \varphi'_v \in \Phi_v$, $\psi_e, \psi'_e \in \Psi_e$, $v \in V(\Gamma)$, and $e \in E(\Gamma)$, we say that (μ, M) is:

- equivariant if $\mu_e(F_{v \preceq e}(\varphi_v \circ g)) = F'_{v \preceq e}(\mu_v(\varphi_v)) \circ M_e(T_{v \preceq e}(g))$.
- non-expansive if $D_{\Phi'_v}(\mu_v(\varphi_v), \mu_v(\varphi'_v)) \leq D_{\Phi_v}(\varphi_v, \varphi'_v)$ and $D_{\Psi'_e}(\mu_e(\psi_e), \mu_e(\psi'_e)) \leq D_{\Psi_e}(\psi_e, \psi'_e)$.

Remark. If (μ, M) is an equivariant non-expansive GENEIO sheaf morphism, then (μ_v, M_v) and (μ_e, M_e)

are GENEOS for all $v \in V(\Gamma)$, $e \in E(\Gamma)$.

GENEO sheaf morphisms are closed under composition. For GENEOS $\mathbf{F}$, $\mathbf{F}'$, and $\mathbf{F}''$ on Γ and GENEOS sheaf morphisms $(\mu, M) : \mathbf{F} \rightarrow \mathbf{F}'$ and $(\mu', M') : \mathbf{F}' \rightarrow \mathbf{F}''$, define map $(\mu'', M'') = (\mu' \circ \mu, M' \circ M) : \mathbf{F} \rightarrow \mathbf{F}''$. We can easily confirm that (μ'', M'') is a GENEOS sheaf morphism:

$$\begin{aligned}\mu'_e(\mu_e(F_{v \preceq e}(\varphi_v))) &= \mu'_e(F'_{v \preceq e}(\mu_v(\varphi_v))) = F''_{v \preceq e}(\mu'_v(\mu_v(\varphi_v))) \\ M'_e(M_e(T_{v \preceq e}(g))) &= M'_e(T'_{v \preceq e}(M_v(g))) = T''_{v \preceq e}(M'_v(M_v(g)))\end{aligned}$$

By Proposition 3.1.1 in [263], composing GENEOS $(F, T) : (\Phi, G) \rightarrow (\Psi, H)$ and $(F', T') : (\Psi, H) \rightarrow (\Omega, K)$ yields a GENEOS. Therefore under these conditions, equivariant non-expansive GENEOS sheaf morphisms are closed under composition as well.

4.4.2 Stalk Symmetries

Our first type of heterogeneous multiresolution symmetry acts locally on (collections of) subsystem observations, uniformly transforming vertex stalk data over subgraphs of Γ . These so-called stalk symmetries arise when modulating part, but not all, of a system – for instance, when repositioning a satellite within a satellite network, rotating a molecule within a macromolecule, or rotating a limb within a body pose (Figure 4.13, left column). Stalk symmetries provide fine-tuned control over system behavior.

Definition 4.4.2. For subgraph $A \subseteq \Gamma$, s_A is a *stalk symmetry* over A if $s_A \in \bigcap_{v \preceq e} G_{v,e}$ whenever $v \in V(A)$ and $e \in E(A)$.

Despite heterogeneity of data representations $\varphi_v \in \Phi_v$ for $v \in V(A)$, existence of nontrivial stalk symmetries over A implies some degree of functional homogeneity among data within A . In fact, group $S_A = \bigcap_{v \preceq e \in A} G_{v,e}$ of stalk symmetries over A describes the extent to which the subsystems in A (as observed by sensors $V(A)$) function as a unit. If S_A is nontrivial, then the smaller (high-resolution) subsystems observed by individual sensors $v \in V(A)$ can be combined into a larger (lower-resolution) subsystem over A . Choice of restriction equivariance groups $G_{v,e}$ determines existence and behavior of these multiresolution subsystems. Stalk symmetry groups thus provide a local-to-global picture of how heterogeneous subsystems arise within the system modeled by GENEOS sheaf $\mathbf{F}$. Because local subsystems within A are free to act collectively (via S_A) or individually (via $\text{Aut}_{\Phi_v}(X_v)$), multiresolution subsystems are not hard-coded into $\mathbf{F}$ and can emerge and dissipate as needed via application of stalk symmetries.

Theorem 4.4.3. *Stalk symmetries induce GENEOS sheaf morphisms.*

Proof. For GENEOS sheaf $\mathbf{F}$ on Γ , let s_A be a stalk symmetry on $\mathbf{F}$ over subgraph $A \subseteq \Gamma$. Define map $(\mu, M) : \mathbf{F} \rightarrow \mathbf{F}$ comprising functions $(\mu_v, M_v) : (\Phi_v, \text{Aut}_{\Phi_v}(X_v)) \rightarrow (\Phi_v, \text{Aut}_{\Phi_v}(X_v))$ for all $v \in V(\Gamma)$ and functions $(\mu_e, M_e) : (\Psi_e, H_e) \rightarrow (\Psi_e, H_e)$ for all $e \in E(\Gamma)$ such that

1. $\mu_u(\varphi_u) = \varphi_u \circ s_A$ for all $u \in V(A)$ and $\mu_v(\varphi_v) = \varphi_v$ for all $v \in V(\Gamma \setminus A)$,
2. $\mu_e(F_{u \preceq e}(\varphi_u)) = F_{u \preceq e}(\varphi_u) \circ T_{u \preceq e}(s_A)$ for all $u \in V(A)$ and $\mu_e(F_{v \preceq e}(\varphi_v)) = F_{v \preceq e}(\varphi_v)$ for all $v \in V(\Gamma \setminus A)$,
3. $D_{\Phi_v}(\mu_v(\varphi_v), \mu_v(\varphi'_v) \circ M_v(g)) = D_{\Phi_v}(\varphi_v, \varphi'_v \circ g)$ and $D_{\Psi_e}(\mu_e(\psi_e), \mu_e(\psi'_e) \circ M_e(h)) = D_{\Psi_e}(\psi_e, \psi'_e \circ h)$ for all $g \in \text{Aut}_{\Phi_v}(X_v)$ and $h \in H_e$, respectively, and all $v \in V(\Gamma)$, $e \in E(\Gamma)$.

$\mu_e(F_{v \preceq e}(\varphi_v)) = F_{v \preceq e}(\varphi_v) \circ T_{v \preceq e}(s_A) = F_{v \preceq e}(\varphi_v \circ s_A) = F_{v \preceq e}(\mu_v(\varphi_v))$, so μ commutes with restriction

GENEOs.

$$\begin{aligned}
D_{\Psi_e}(\mu_e(F_{v \preceq e}(\varphi_v)), \mu_e(F_{v \preceq e}(\varphi'_v)) \circ M_e(T_{v \preceq e}(g))) &= D_{\Psi_e}(F_{v \preceq e}(\varphi_v), F_{v \preceq e}(\varphi'_v) \circ T_{v \preceq e}(g)) \\
&= D_{\Psi_e}(F_{v \preceq e}(\varphi_v), F_{v \preceq e}(\varphi'_v \circ g)) \\
&= D_{\Psi_e}(F_{v \preceq e}(\mu_v(\varphi_v)), F_{v \preceq e}(\mu_v(\varphi'_v) \circ M_v(g))) \\
&= D_{\Psi_e}(\mu_e(F_{v \preceq e}(\varphi_v)), \mu_e(F_{v \preceq e}(\varphi'_v)) \circ T_{v \preceq e}(M_v(g))),
\end{aligned}$$

so $M_e(T_{v \preceq e}(g)) = T_{v \preceq e}(M_v(g))$. The third equality follows from Equation 6 in [20]. Therefore (μ, M) is a GENEO sheaf morphism. $\square$

The following result implies that stalk symmetries on $\mathbf{F}$ preserve the stalk symmetries of $\mathbf{F}$.

Lemma 4.4.4. *Let $(\mu, M) : \mathbf{F} \rightarrow \mathbf{F}'$ be a GENEO sheaf morphism. If s_A is a stalk symmetry of $\mathbf{F}$ over $A \subseteq \Gamma$, then $M_v(s_A)$ is a stalk symmetry of $\mathbf{F}'$ over A .*

Proof. By definition of GENEO sheaf morphism, $M_v(s_A) \in \text{im } T_{v \preceq e}'^{-1} = G'_{v,e}$ for all incident $v \preceq e$ with $v \in V(A)$, $e \in E(A)$. Then $M_v(s_A) \in \bigcap_{v \preceq e} G'_{v,e}$ for all $v \in V(A)$, so $M_v(s_A)$ is a stalk symmetry of $\mathbf{F}'$ over A . $\square$

By formalizing stalk symmetries as GENEO sheaf morphisms, we can study their interactions with the GENEO sheaf Laplacian. First, we show that s_A preserves data consistency over A because it acts uniformly over A . This makes sense in practice: if observations φ_v, φ_u on adjacent $v, u \in V(A)$ undergo the same symmetry transformation, then the local relationship between $F_{v \preceq e}(\varphi_v), F_{u \preceq e}(\varphi_u)$ over edge $e \supseteq v, u$ should remain unchanged. Formally speaking, s_A commutes with $L_{\mathbf{F}}(\varphi)_A$.

Lemma 4.4.5. *Let $(\mu, M) : \mathbf{F} \rightarrow \mathbf{F}$ be the GENEO sheaf morphism induced by stalk symmetry s_A on $\mathbf{F}$ over $A \subseteq \Gamma$. Then $L_{\mathbf{F}}(\mu(\varphi))_A = M(L_{\mathbf{F}}(\varphi)_A)$.*

Proof. Let $v, u \preceq e$ for $v, u \in V(A)$, $e \in E(A)$, and $g_{v,e} = T_{v \preceq e}^{-1}(\delta_{H_{\mathbf{F}}}(\varphi)_e) \in G_{v,e}$. By Equation (6) in [20], Equation 4.8, and Theorem 4.4.3,

$$\begin{aligned}
d_{H_e}(F_{u \preceq e}(\mu(\varphi)_u), F_{v \preceq e}(\mu(\varphi)_v)) &= d_{H_e}(F_{u \preceq e}(\varphi_u \circ s_A), F_{v \preceq e}(\varphi_v \circ s_A)) \\
&= d_{H_e}(F_{u \preceq e}(\varphi_u), F_{v \preceq e}(\varphi_v)) \\
&= D_{\Psi_e}(F_{u \preceq e}(\varphi_u), F_{v \preceq e}(\varphi_v) \circ T_{v \preceq e}(g_{v,e})) \\
&= D_{\Psi_e}(\mu_e(F_{u \preceq e}(\varphi_u)), \mu_e(F_{v \preceq e}(\varphi_v)) \circ M_e(T_{v \preceq e}(g_{v,e}))) \\
&= D_{\Psi_e}(F_{u \preceq e}(\mu_u(\varphi_u)), F_{v \preceq e}(\mu_v(\varphi_v) \circ M_v(g_{v,e}))).
\end{aligned}$$

Then $L_{\mathbf{F}}(\mu(\varphi))_v = \sum_{v \preceq e} M_v(g_{v,e}) \circ a'_{v,e}$ such that

$$a'_{v,e} = \underset{a' \in \text{Aut}_{\Phi_v}(X_v) \setminus G_{v,e}}{\text{argmin}} D_{\Psi_e}(F_{u \preceq e}(\mu_u(\varphi_u)), F_{v \preceq e}((\mu_v(\varphi_v) \circ M_v(g_{v,e})) \circ a')).$$

Let $a_{v,e} \in G_{v,e}$ such that $L_{\mathbf{F}}(\varphi)_v = \sum_{v \preceq e} g_{v,e} \circ a_{v,e}$.

$$\begin{aligned}
D_{\Psi_e}(F_{u \preceq e}(\mu_u(\varphi_u)), F_{v \preceq e}((\mu_v(\varphi_v) \circ M_v(g_{v,e})) \circ M_v(a_{v,e}))) \\
&= D_{\Psi_e}(F_{u \preceq e}(\varphi_u), F_{v \preceq e}((\varphi_v \circ g_{v,e}) \circ a_{v,e})) \\
&= \inf_{a \in \text{Aut}_{\Phi_v}(X_v) \setminus G_{v,e}} D_{\Psi_e}(F_{u \preceq e}(\varphi_u), F_{v \preceq e}((\varphi_v \circ g_{v,e}) \circ a)) \\
&= \inf_{a \in \text{Aut}_{\Phi_v}(X_v) \setminus G_{v,e}} D_{\Psi_e}(F_{u \preceq e}(\mu_u(\varphi_u)), F_{v \preceq e}((\mu_v(\varphi_v) \circ M_v(g_{v,e})) \circ M_v(a))),
\end{aligned}$$

so $a'_{v,e} = M_v(a_{v,e})$. Then $L_{\mathbf{F}}(\mu(\varphi))_v = \sum_{v \preceq e} M_v(g_{v,e} \circ a_{v,e}) = M_v(L_{\mathbf{F}}(\varphi)_v)$ for all $v \in V(A)$. $\square$

Next, we examine how $L_{\mathbf{F}}$ propagates stalk symmetries s_A across A 's boundary. For adjacent $v, u \in V(\Gamma)$ with $u \in V(A)$ and $v \in V(\Gamma \setminus A)$, there is no guarantee that $s_A \in \text{Aut}_{\Phi_v}(X_v)$. However, s_A is related by restriction GENEIO homomorphisms $T_{v \preceq e}, T_{u \preceq e}$ to corresponding symmetry $T_{v \preceq e}^{-1}(T_{u \preceq e}(s_A)) \in \text{Aut}_{\Phi_v}(X_v)$ whenever $s_A \in G_{u,e}$. Therefore $L_{\mathbf{F}}$ propagates heterogeneous versions of subsystem stalk symmetries to the rest of the system.

Theorem 4.4.6. *Let $(\mu, M) : \mathbf{F} \rightarrow \mathbf{F}$ be the GENEIO sheaf morphism induced by stalk symmetry s_A on $\mathbf{F}$ over $A \subseteq \Gamma$. For $v \in V(\Gamma \setminus A)$ with adjacent vertices $u \in V(A)$, $w \in V(\Gamma \setminus A)$, suppose $s_A \in G_{u,e}$ for each adjacent $u \in V(A)$. If $L_{\mathbf{F}}(\varphi)_v = \sum_{v \preceq e} g_{v,e} \circ a_{v,e}$ then*

$$L_{\mathbf{F}}(\mu(\varphi))_v = \sum_{u, v \preceq e; u \in V(A)} g_{v,e} \circ a_{v,e} \circ T_{v \preceq e}^{-1}(T_{u \preceq e}(s_A)) + \sum_{w, v \preceq e'; w \in V(\Gamma \setminus A)} g_{v,e'} \circ a_{v,e'} \quad (4.13)$$

Proof. By Equation (6) in [20],

$$\begin{aligned} d_{H_e}(F_{u \preceq e}(\mu_u(\varphi_u)), F_{v \preceq e}(\mu_v(\varphi_v))) &= d_{H_e}(F_{u \preceq e}(\varphi_u), F_{v \preceq e}(\varphi_v)) \\ &\geq D_{\Psi_e}(F_{u \preceq e}(\varphi_u), F_{v \preceq e}(\varphi_v \circ g_{v,e} \circ a_{v,e})) \\ &= D_{\Psi_e}(F_{u \preceq e}(\varphi_u) \circ T_{u \preceq e}(s_A), F_{v \preceq e}(\varphi_v \circ g_{v,e} \circ a_{v,e}) \circ T_{u \preceq e}(s_A)) \\ &= D_{\Psi_e}(F_{u \preceq e}(\mu_u(\varphi_u)), F_{v \preceq e}(\varphi_v \circ g_{v,e} \circ a_{v,e} \circ T_{v \preceq e}^{-1}(T_{u \preceq e}(s_A)))). \end{aligned}$$

Therefore $g_{v,e} \circ a_{v,e} \circ T_{v \preceq e}^{-1}(T_{u \preceq e}(s_A))$ is the term in $L_{\mathbf{F}}(\mu(\varphi))_v$ optimizing consistency over edge $e \triangleright u, v$. Because $\delta_{H_{\mathbf{F}}}(\mu(\varphi))_{e'} = \delta_{H_{\mathbf{F}}}(\varphi)_{e'}$ for all $e' \triangleright w, v$ such that $w \in V(\Gamma \setminus A)$, Equation 4.13 holds. $\square$

Lemma 4.4.7. *Let $(\mu, M) : \mathbf{F} \rightarrow \mathbf{F}$ be the GENEIO sheaf morphism induced by stalk symmetry s_A on $\mathbf{F}$ over $A \subseteq \Gamma$. For (arbitrarily oriented) edge $e = u \rightarrow v$ with $u \in V(A)$ and $v \in V(\Gamma \setminus A)$, we have $\delta_{H_{\mathbf{F}}}(\mu(\varphi))_e = T_{v \preceq e}(g_{v,e}) \circ T_{u \preceq e}(s_A)$.*

Proof. Let $\delta_{H_{\mathbf{F}}}(\mu(\varphi))_e = h_e \in H_e$; then by Equation 4.8,

$$\begin{aligned} d_{H_e}(F_{u \preceq e}(\varphi_u), F_{v \preceq e}(\varphi_v)) &= d_{H_e}(F_{u \preceq e}(\mu_u(\varphi_u)), F_{v \preceq e}(\mu_v(\varphi_v))) \\ &= D_{\Psi_e}(F_{u \preceq e}(\mu_u(\varphi_u)), F_{v \preceq e}(\mu_v(\varphi_v)) \circ h_e) \\ &= D_{\Psi_e}(F_{u \preceq e}(\varphi_u \circ s_A), F_{v \preceq e}(\varphi_v) \circ h_e) \\ &= D_{\Psi_e}(F_{u \preceq e}(\varphi_u) \circ T_{u \preceq e}(s_A), F_{v \preceq e}(\varphi_v) \circ h_e) \\ &= D_{\Psi_e}(F_{u \preceq e}(\varphi_u), F_{v \preceq e}(\varphi_v) \circ h_e \circ T_{u \preceq e}(s_A)^{-1}). \end{aligned}$$

Then $T_{v \preceq e}(g_{v,e}) = h_e \circ T_{u \preceq e}(s_A)^{-1} \Rightarrow T_{v \preceq e}(g_{v,e}) \circ T_{u \preceq e}(s_A) = h_e = \delta_{H_{\mathbf{F}}}(\mu(\varphi))_e$. $\square$

During GENEIO sheaf data fusion, symmetries on real-world sensor observations should induce corresponding symmetries on the model system state. This heterogeneous multiresolution equivariance guarantees that real-world system behavior is reliably reflected in the system model, ensuring faithful computational approximation of real-world phenomena. This is especially critical when corresponding real and model sensor observations are heterogeneous: for instance, when performing 3D reconstruction (model system state) from 2D data (raw sensor observations). The following result guarantees that GENEIO sheaf data fusion converts stalk symmetries on raw sensor observations into corresponding stalk symmetries within the system model (Figure 4.5).

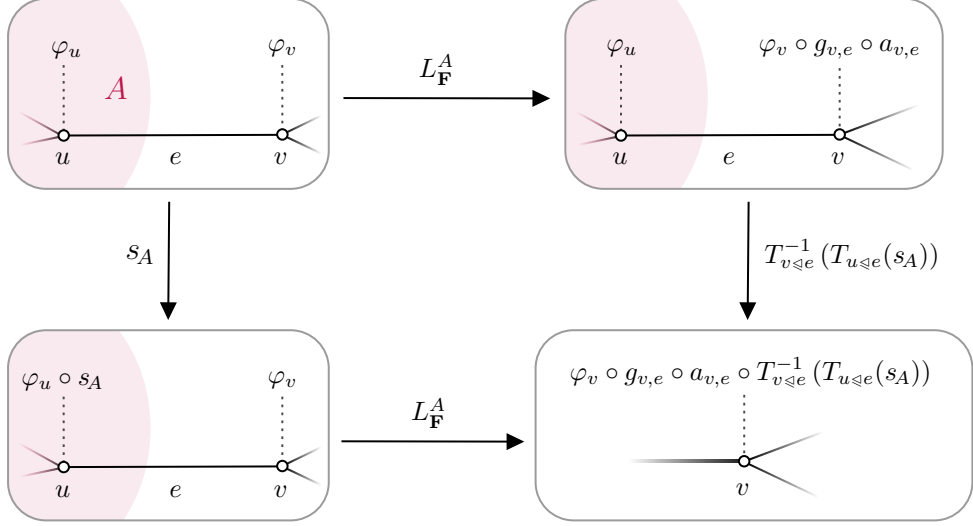


Figure 4.5: Illustration of Theorem 4.4.8, stating that GENEО sheaf data fusion is stalk symmetry-equivariant. Note that $\varphi'_v = \varphi_v \circ g_{v,e} \circ a_{v,e}$ in the theorem statement.

Theorem 4.4.8 (Data fusion is stalk symmetry-equivariant). *For GENEО sheaf $\mathbf{F}$ on graph Γ and subgraph $A \subseteq \Gamma$, let $(\mu, M) : \mathbf{F} \rightarrow \mathbf{F}$ be the GENEО sheaf morphism induced by stalk symmetry s_A on $\mathbf{F}$ over $A \subseteq \Gamma$. If $\lim_{t \rightarrow \infty} \varphi(t) = \varphi'$ for $\varphi, \varphi' \in \Phi_{\mathbf{F}}$ under A -restricted dynamics $L_{\mathbf{F}}^A$, then*

$$\lim_{t \rightarrow \infty} \mu(\varphi)(t)_v = \varphi'_v \circ T_{v \leq e}^{-1}(T_{u \leq e}(s_A)) \quad (4.14)$$

under $L_{\mathbf{F}}^A$ for every $v \in V(\Gamma \setminus A)$ adjacent to $u \in V(A)$ with $s_A \in G_{u,e}$ for $e \geq v, u$.

Proof. Let $g_{v,e} \in G_{v,e}$ and $a_{v,e} \in \text{Aut}_{\Phi_v}(X_v) \setminus G_{v,e}$ such that $\varphi'_v = \varphi_v \circ g_{v,e} \circ a_{v,e}$. By Theorem 4.4.6 and assuming Conjecture 4.3.2 holds, the A -restricted dynamics on $\mu(\varphi)$ converge at v to $\varphi_v \circ (g_{v,e} \circ a_{v,e}) \circ T_{v \leq e}^{-1}(T_{u \leq e}(s_A)) = \varphi'_v \circ T_{v \leq e}^{-1}(T_{u \leq e}(s_A))$. $\square$

4.4.3 Stalk Permutations

Our second type of heterogeneous multiresolution symmetry exchanges subsystems which are structurally and functionally equivalent up to some stalk symmetry – for instance, left and right arm poses are interchangeable up to reflection (Figure 4.13, right column). These so-called stalk permutations exchange isomorphic vertex stalk data defined over isomorphic subgraphs of Γ . Stalk permutations are vital for modeling systems with interchangeable subsystems – such as self-assembling robots and macromolecular structures – and analyzing important relationships between system states.

Definition 4.4.9. For GENEО sheaf $\mathbf{F}$ on graph Γ , let $p \in \text{Aut}(\Gamma)$ be a graph automorphism exchanging isomorphic subgraphs $A, A^* \subseteq \Gamma$. Let s_A be a stalk symmetry of $\mathbf{F}$ over A inducing equivariant GENEО sheaf morphism $(\mu, M) : \mathbf{F} \rightarrow \mathbf{F}$ such that $\mu_v : \Phi_v \rightarrow \Phi_{v^*}$, $\mu_e : \Psi_e \rightarrow \Psi_{e^*}$, $M_v : G_{v,e} \rightarrow G_{v^*,e^*}$, and $M_e : H_e \rightarrow H_{e^*}$ for all $v \in V(A)$, $e \in E(A)$. A *stalk permutation* $\pi_A : \mathbf{F} \rightarrow \mathbf{F}$ over A, A^* is defined

$$\begin{aligned} p \circ (\mu, M) &= \pi_A : (\Phi_v, \text{Aut}_{\Phi_v}(X_v)) \mapsto (\Phi_{v^*}, \text{Aut}_{\Phi_{v^*}}(X_{v^*})) \\ & : (\Psi_e, H_e) \mapsto (\Psi_{e^*}, H_{e^*}) \end{aligned} \quad (4.15)$$

such that for $g_{v^*} = p(M_v(g_v)) \in \text{Aut}_{\Phi_{v^*}}(X_{v^*})$ and $h_{e^*} = p(M_e(h_e)) \in H_{e^*}$,

1. $D_{\Phi_v}(\varphi_v, \varphi'_v \circ g_v) = D_{\Phi_{v^*}}(\pi_A(\varphi_v), \pi_A(\varphi'_v) \circ g_{v^*}) \quad \forall \varphi_v, \varphi'_v \in \Phi_v, \forall g_v \in \text{Aut}_{\Phi_v}(X_v)$
2. $D_{\Psi_e}(\psi_e, \psi'_e \circ h_e) = D_{\Psi_{e^*}}(\pi_A(\psi_e), \pi_A(\psi'_e) \circ h_{e^*}) \quad \forall \psi_e, \psi'_e \in \Psi_e, \forall h_e \in H_e$

These conditions ensure that stalk-wise data and automorphisms are both faithfully exchanged by stalk permutations. Continuing our earlier example, π_A reflects then exchanges left and right arm poses, while M exchanges clockwise and counterclockwise rotations between left and right arms' vertex stalks. In practice, this means that rotating the left arm clockwise before exchanging left and right arm poses is equivalent to exchanging left and right arm poses before rotating the right arm counterclockwise. In other words, stalk symmetries commute with stalk permutations.

Lemma 4.4.10 (Stalk symmetries commute with stalk permutations). *For GENEIO sheaf $\mathbf{F}$ on graph Γ and stalk permutation π_A of $\mathbf{F}$ over subgraphs $A, A^* \subseteq \Gamma$, let s_B be a stalk symmetry of $\mathbf{F}$ over $B \subseteq A$ such that $s_{B^*} = p(M_v(s_B)) \in \text{Aut}_{\Phi_{v^*}}(X_{v^*})$ for all $v \in V(B)$. Then $\pi_A(\varphi \circ s_B) = \pi_A(\varphi) \circ s_{B^*}$.*

Proof. For all $v \in V(A)$, we have $\pi_A(\varphi_v \circ s_B) = p(\mu_v(\varphi_v \circ s_B)) = p(\mu_v(\varphi_v) \circ M_v(s_B)) = \pi_A(\varphi_v) \circ s_{B^*}$ by Definition 4.4.1. The same argument holds for all $e \in E(A)$, so s_B commutes with π_A . $\square$

Like stalk symmetries, stalk permutations act uniformly over the subsystems they exchange. Accordingly, data consistency is preserved under stalk permutation.

Lemma 4.4.11. *For stalk permutation π_A of $\mathbf{F}$ over subgraphs $A, A^* \subseteq \Gamma$, we have $L_{\mathbf{F}}(\pi_A(\varphi))_{A^*} = \pi_A(L_{\mathbf{F}}(\varphi)_A)$.*

Proof. $L_{\mathbf{F}}(\pi_A(\varphi))_{A^*} = L_{\mathbf{F}}(p(\mu_A(\varphi)))_{A^*} = p(L_{\mathbf{F}}(\mu_A(\varphi))_A) = p(M_A(L_{\mathbf{F}}(\varphi)_A)) = \pi_A(L_{\mathbf{F}}(\varphi)_A)$. The second equality follows from commutativity of graph automorphisms with sheaf Laplacians [150]. The third equality follows from Lemma 4.4.5. $\square$

For stalk permutation π_A over subgraphs $A, A^* \subseteq \Gamma$, suppose we have isomorphic subsystems of raw sensor observations on $B \subseteq A, B^* \subseteq A^*$, with corresponding model sensor observations on $A \setminus B, A^* \setminus B^*$. The following result states that permuting B, B^* before data fusion is equivalent to permuting A, A^* after data fusion (Figure 4.6). This ensures that GENEIO sheaf data fusion converts stalk permutations observed within the system into corresponding stalk permutations recorded in the system model.

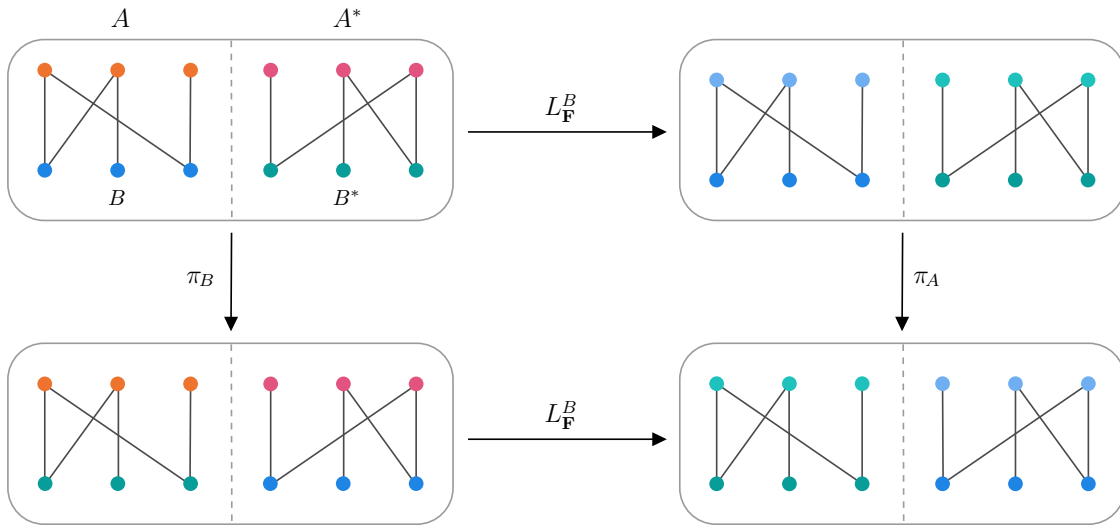


Figure 4.6: Illustration of Theorem 4.4.12, stating that GENEIO sheaf data fusion is stalk permutation-equivariant.

Theorem 4.4.12 (Data fusion is stalk permutation-equivariant). *For GENEIO sheaf $\mathbf{F}$ on graph Γ , let π_A be a stalk permutation over subgraphs $A, A^* \subseteq \Gamma$, and let $B \subseteq A$. If $\lim_{t \rightarrow \infty} \varphi(t) = \varphi'$ for $\varphi, \varphi' \in \Phi_{\mathbf{F}}$ under B -restricted dynamics $L_{\mathbf{F}}^B$, then for stalk permutation $\pi_B = \pi_A|_B$,*

$$\lim_{t \rightarrow \infty} \pi_B(\varphi)(t)_v = \pi_A(\varphi')_v \quad (4.16)$$

under $L_{\mathbf{F}}^B$ for all $v \in V(A \setminus B)$ adjacent to $u \in V(B)$.

Proof. Let $v \in V(A \setminus B)$ with adjacent $u \in V(B)$. Because $L_{\mathbf{F}}^B$ leaves φ_u fixed for all $u \in V(B)$,

$$\begin{aligned} D_{\Psi_e}(F_{u \leq e}(\varphi'_u), F_{v \leq e}(\varphi'_v)) &= D_{\Psi_e}(F_{u \leq e}(\varphi_u), F_{v \leq e}(\varphi'_v)) \\ &= D_{\Psi_{e^*}}(\pi_B(F_{u \leq e}(\varphi_u)), \pi_A(F_{v \leq e}(\varphi'_v))) \\ &= D_{\Psi_{e^*}}(F_{u^* \leq e^*}(\pi_B(\varphi_u)), F_{v^* \leq e^*}(\pi_A(\varphi'_v))) \end{aligned}$$

Since $\Phi_v = \Phi_{v^*}$ and π_A preserves stalk pseudometrics, $F_{v^* \leq e^*}(\pi_A(\varphi'_v))$ minimizes $L_{\mathbf{F}}^B(\pi_B(\varphi))_e$. $\square$

4.5 Pseudometrics on GENEIO Sheaves

Both sheaves and GENEIOs admit pseudometrics which facilitate data analysis in various ways. GENEIO pseudometrics D_{Φ} and d_G respectively ensure data stability during analysis and enable ground-truth data comparison up to equivariance group G [20]. Sheaf pseudometrics compare 0-assignments to assess system state similarity, facilitate heterogeneous data fusion, and identify obstructions to global data consistency [271, 272]. The last of these is especially important because non-cellular sheaves lack cohomology, which summarizes multiresolution obstructions to patching local data into globally consistent states. Multiresolution obstruction analysis is essential for understanding systems, since certain obstructions only emerge at certain resolutions. For example, the Penrose triangle appears locally consistent at each corner, revealing its physical impossibility only when perceived holistically (Figure 2.34). In this section, we synthesize and extend existing work on sheaf and GENEIO pseudometrics, yielding a suite of GENEIO sheaf tools for data comparison, obstruction analysis, and tracking subsystem influence.

A *sheaf of pseudometric spaces* on graph Γ is a sheaf comprising set K_v with pseudometric D_v for each $v \in V(\Gamma)$, set K_e with pseudometric D_e for each $e \in E(\Gamma)$, and continuous restriction maps $f_{v \leq e} : K_v \rightarrow K_e$ for every incident $v \leq e$ in Γ [271]. GENEIO sheaves comport with this definition: for every $v \in V(\Gamma)$, $e \in E(\Gamma)$, the sets Φ_v and Ψ_e are respectively equipped with pseudometrics D_{Φ_v} and D_{Ψ_e} , and restriction GENEIOs are continuous by construction. To compare system states given by admissible 0-assignments $\varphi, \varphi' \in \Phi_{\mathbf{F}}$, we use *admissible 0-assignment pseudometric*

$$D_{\Phi_{\mathbf{F}}}(\varphi, \varphi') = \sup_{v \in V(\Gamma)} D_{\Phi_v}(\varphi_v, \varphi'_v) = \sup_{v \in V(\Gamma)} \sup_{x \in X_v} |\varphi_v(x) - \varphi'_v(x)| \quad (4.17)$$

synthesizing sheaf pseudometric $D_{\mathcal{F}}$ (Equation 4.3) with admissible function pseudometric D_{Φ} (Equation 4.4) extended to product pseudometric space $\Phi_{\mathbf{F}} = \prod_{v \in V(\Gamma)} \Phi_v$. We can localize comparison to subgraphs $A \subseteq \Gamma$ using pseudometric

$$D_{\Phi_A}(\varphi, \varphi') = \sup_{v \in V(A)} D_{\Phi_v}(\varphi_v, \varphi'_v) = \sup_{v \in V(A)} \sup_{x \in X_v} |\varphi_v(x) - \varphi'_v(x)| \quad (4.18)$$

Symmetry-aware GENEIO pseudometrics also fluently extend to GENEIO sheaves. For GENEIO sheaf $\mathbf{F}$ on Γ and subgraph $A \subseteq \Gamma$, let groups S_A and Π_{A, A^*} respectively comprise all stalk symmetries

over A and all stalk permutations over A, A^* . The respective groups of all stalk symmetries $S_{\mathbf{F}}$ on $\mathbf{F}$ and all stalk permutations $\Pi_{\mathbf{F}}$ on $\mathbf{F}$ are defined

$$S_{\mathbf{F}} = \prod_{A \subseteq \Gamma} S_A \quad \Pi_{\mathbf{F}} = \prod_{A, A^* \subseteq \Gamma} \Pi_{A, A^*} \quad (4.19)$$

For admissible function space Φ , GENE theory compares Φ -preserving automorphisms using pseudometric [263]

$$D_{\text{Aut}}(g, g') = \sup_{\varphi \in \Phi} D_{\Phi}(\varphi \circ g, \varphi \circ g'), \quad \forall g, g' \in \text{Aut}_{\Phi}(X)$$

Accordingly, for heterogeneous multiresolution symmetries $g, g' \in S_{\mathbf{F}} \times \Pi_{\mathbf{F}}$, define *symmetry pseudometric*

$$D_{S_{\mathbf{F}} \times \Pi_{\mathbf{F}}}(g, g') = \sup_{\varphi \in \Phi_{\mathbf{F}}} D_{\Phi_{\mathbf{F}}}(\varphi \circ g, \varphi \circ g') \quad (4.20)$$

To measure how well subsystem states φ_A, φ'_A locally align up to stalk symmetries $s_A \in S_A$, we use pseudometric

$$d_{S_A}(\varphi_A, \varphi'_A) = \inf_{s_A \in S_A} D_{\Phi_A}(\varphi_A, \varphi'_A \circ s_A) \quad (4.21)$$

Definition 4.5.1. For GENE sheaf $\mathbf{F}$ on graph Γ , let $\text{Aut}_{\Phi_{\mathbf{F}}}(X_{\mathbf{F}}) = \prod_{v \in V(\Gamma)} \text{Aut}_{\Phi_v}(X_v)$ comprise all $\Phi_{\mathbf{F}}$ -preserving automorphisms of $\mathbf{F}$. For the given analysis objectives, choose global equivariance group $G_{\mathbf{F}} \leq S_{\mathbf{F}} \times \Pi_{\mathbf{F}} \leq \text{Aut}_{\Phi_{\mathbf{F}}}(X_{\mathbf{F}})$ of stalk symmetries and permutations up to which admissible 0-assignments are considered equivalent. The *ground-truth pseudometric* measuring global data similarity between every $\varphi, \varphi' \in \Phi_{\mathbf{F}}$ up to equivalences $g \in G_{\mathbf{F}}$ is defined

$$d_{G_{\mathbf{F}}}(\varphi, \varphi') = \inf_{g \in G_{\mathbf{F}}} D_{\Phi_{\mathbf{F}}}(\varphi, \varphi' \circ g) = \inf_{g \in G_{\mathbf{F}}} \sup_{v \in V(\Gamma)} D_{\Phi_v}(\varphi_v, \varphi'_v \circ g|_v) \quad (4.22)$$

A core advantage of equivariant analysis frameworks – and GENEOS specifically – is data classification up to symmetries comporting with the given analysis objectives. We have seen how GENE sheaves support heterogeneous multiresolution symmetry modulation, whereas other equivariant analysis frameworks support only homogeneous global symmetry modulation. The ground-truth pseudometric extends heterogeneous multiresolution equivariance to data comparison, assessing how well admissible 0-assignments can be aligned via heterogeneous multiresolution symmetries $G_{\mathbf{F}}$ selected to match the analysis task at hand. In Section 4.6, we describe how the ground-truth pseudometric supports GENE sheaf analysis tailored to specialized applications. Note that while the ground-truth pseudometric is expensive to compute, it can be robustly and efficiently approximated via persistent homology [20].

4.5.1 Obstruction Analysis

In place of cohomology, sheaves of pseudometric spaces admit a global data consistency measure which we adapt for GENE sheaves. For admissible 0-assignments $\varphi \in \Phi_{\mathbf{F}}$ of GENE sheaf $\mathbf{F}$, the *consistency radius* [272] is defined

$$c_{\mathbf{F}}(\varphi) = \sup_{e \in E(\Gamma)} D_{\Psi_e}(F_{u \leq e}(\varphi_u), F_{v \leq e}(\varphi_v)) = \sup_{e \in E(\Gamma)} \sup_{y \in Y_e} |F_{u \leq e}(\varphi_u)(y) - F_{v \leq e}(\varphi_v)(y)| \quad (4.23)$$

Whereas the $H_{\mathbf{F}}$ -coboundary map (Equation 4.8) evaluates edge-wide data alignment up to symmetry groups H_e , the consistency radius precisely measures differences between homogenized data representations $F_{v \leq e}(\varphi_v)$ and $F_{u \leq e}(\varphi_u)$ at each edge $e \in E(\Gamma)$. The set $\ker c_{\mathbf{F}}$ comprises $\mathbf{F}$'s global sections, and is a

subset of $\mathbf{F}$'s global $H_{\mathbf{F}}$ -sections (Equation 4.9). The difference between $\mathbf{F}$'s global $H_{\mathbf{F}}$ -sections and global sections is the difference between globally optimal data alignment and global data consistency. GENEIO sheaf data fusion optimally aligns input data, after which $c_{\mathbf{F}}$ quantifies unresolved obstructions to global consistency within the model system (Figure 4.4c).

Bounding obstructions By Proposition 23 of [271], the consistency radius for $\varphi \in \Phi_{\mathbf{F}}$ has upper bound

$$\frac{c_{\mathbf{F}}(\varphi)}{2} \leq D_{\Phi_{\mathbf{F}}}(\varphi, \varphi^*)$$

for every global section $\varphi^* \in \ker c_{\mathbf{F}}$. For admissible 0-assignments $\varphi, \varphi' \in \Phi_{\mathbf{F}}$, define *consistency radius pseudometric*

$$D_{c_{\mathbf{F}}}(\varphi, \varphi') = \sup_{e \in E(\Gamma)} |c_{\mathbf{F}}(\varphi)_e - c_{\mathbf{F}}(\varphi')_e| \quad (4.24)$$

measuring maximum edge-wise difference in data consistency between system states φ and φ' .

Lemma 4.5.2. *For GENEIO sheaf $\mathbf{F}$ on Γ and $\varphi, \varphi' \in \Phi_{\mathbf{F}}$, let $(v, u) = \operatorname{argmax}_{e \in E(\Gamma)} D_{c_{\mathbf{F}}}(\varphi, \varphi')_e$. Then*

$$D_{c_{\mathbf{F}}}(\varphi, \varphi') \leq D_{\Phi_u}(\varphi_u, \varphi'_u) + D_{\Phi_v}(\varphi_v, \varphi'_v) \quad (4.25)$$

Proof.

$$\begin{aligned} D_{c_{\mathbf{F}}}(\varphi, \varphi') &= \sup_{e \in E(\Gamma)} |c_{\mathbf{F}}(\varphi)_e - c_{\mathbf{F}}(\varphi')_e| \\ &= |D_{\Psi_e}(F_{u \leq e}(\varphi_u), F_{v \leq e}(\varphi_v)) - D_{\Psi_e}(F_{u \leq e}(\varphi'_u), F_{v \leq e}(\varphi'_v))| \\ &\leq D_{\Psi_e}(F_{u \leq e}(\varphi_u), F_{u \leq e}(\varphi'_u)) + D_{\Psi_e}(F_{v \leq e}(\varphi_v), F_{v \leq e}(\varphi'_v)) \\ &\leq D_{\Phi_u}(\varphi_u, \varphi'_u) + D_{\Phi_v}(\varphi_v, \varphi'_v) \end{aligned}$$

The first inequality is derived from the reverse triangle inequality, and the second inequality follows from restriction GENEIO non-expansivity. $\square$

Suppose we want to measure how subsystem modulation – in the form of stalk symmetry $s \in S_{\mathbf{F}}$ – impacts difference in data consistency between system states $\varphi, \varphi' \in \Phi_{\mathbf{F}}$. The following result bounds data consistency difference under stalk symmetry with respect to restriction GENEIO homomorphisms.

Lemma 4.5.3. *For GENEIO sheaf $\mathbf{F}$ on Γ , $s \in S_{\mathbf{F}}$, and $\varphi, \varphi' \in \Phi_{\mathbf{F}}$, let $\operatorname{argmax}_{e \in E(\Gamma)} D_{c_{\mathbf{F}}}(\varphi, \varphi' \circ s)_e = (v, u)$. Then*

$$D_{c_{\mathbf{F}}}(\varphi, \varphi' \circ s) \leq D_{\Psi_e}(F_{u \leq e}(\varphi_u), F_{u \leq e}(\varphi'_u) \circ T_{u \leq e}(s|_u)) + D_{\Psi_e}(F_{v \leq e}(\varphi_v), F_{v \leq e}(\varphi'_v) \circ T_{v \leq e}(s|_v)) \quad (4.26)$$

Proof.

$$\begin{aligned} D_{c_{\mathbf{F}}}(\varphi, \varphi' \circ s) &= \sup_{e \in E(\Gamma)} |c_{\mathbf{F}}(\varphi)_e - c_{\mathbf{F}}(\varphi' \circ s)_e| \\ &= |D_{\Psi_e}(F_{u \leq e}(\varphi_u), F_{v \leq e}(\varphi_v)) - D_{\Psi_e}(F_{u \leq e}(\varphi'_u \circ s|_u), F_{v \leq e}(\varphi'_v \circ s|_v))| \\ &\leq D_{\Psi_e}(F_{u \leq e}(\varphi_u), F_{u \leq e}(\varphi'_u \circ s|_u)) + D_{\Psi_e}(F_{v \leq e}(\varphi_v), F_{v \leq e}(\varphi'_v \circ s|_v)) \\ &= D_{\Psi_e}(F_{u \leq e}(\varphi_u), F_{u \leq e}(\varphi'_u) \circ T_{u \leq e}(s|_u)) + D_{\Psi_e}(F_{v \leq e}(\varphi_v), F_{v \leq e}(\varphi'_v) \circ T_{v \leq e}(s|_v)) \end{aligned}$$

The inequality is derived from the reverse triangle inequality, and the final equality follows from restriction GENEIO equivariance. $\square$

To minimize data consistency difference between $\varphi, \varphi' \in \Phi_{\mathbf{F}}$ up to stalk symmetries and permutations relevant to analysis objectives, choose suitable $G_{\mathbf{F}} \leq S_{\mathbf{F}} \times \Pi_{\mathbf{F}}$ and define pseudometric

$$d_{c_{\mathbf{F}}, G_{\mathbf{F}}}(\varphi, \varphi') = \inf_{g \in G_{\mathbf{F}}} D_{c_{\mathbf{F}}}(\varphi, \varphi' \circ g) = \inf_{g \in G_{\mathbf{F}}} \sup_{e \in E(\Gamma)} |c_{\mathbf{F}}(\varphi)_e - c_{\mathbf{F}}(\varphi' \circ g)_e| \quad (4.27)$$

To maximize global data consistency for $\varphi \in \Phi_{\mathbf{F}}$ with respect to $G_{\mathbf{F}}$, define pseudometric

$$c_{\mathbf{F}, G_{\mathbf{F}}}(\varphi) = \inf_{g \in G_{\mathbf{F}}} c_{\mathbf{F}}(\varphi \circ g) \quad (4.28)$$

With these metrics, it becomes possible to measure how real-world multiresolution (sub)system behavior defined by $G_{\mathbf{F}}$ interacts with obstructions to system-wide data consistency.

Multiresolution obstruction analysis Following data fusion on input real-world sensor observations, we want to perform multiresolution obstruction analysis – that is, we want to know where in the model system inconsistencies arise, from which we can infer why they arise. The consistency radius numerically quantifies global system state inconsistency, but cannot detect multiresolution obstructions. To this end, we present a natural extension of Robinson’s work on multiresolution obstruction detection for sheaves of pseudometric spaces [272] to the GENEIO sheaf context.

For GENEIO sheaf $\mathbf{F}$ on Γ with subgraph $A \subseteq \Gamma$, the *A-local consistency radius* of $\varphi \in \Phi_{\mathbf{F}}$ is defined [272]

$$c_{\mathbf{F}}(\varphi)_A = \sup_{u, v \leq e \in E(A)} D_{\Psi_e}(F_{v \leq e}(\varphi_v), F_{u \leq e}(\varphi_u))$$

Lemma 3 of [272] states that if $\varphi \in \Phi_{\mathbf{F}}$ and $B \subseteq A \subseteq \Gamma$, then

$$c_{\mathbf{F}}(\varphi)_B \leq c_{\mathbf{F}}(\varphi)_A \leq c_{\mathbf{F}}(\varphi)$$

Local consistency radii distinguish between widespread inconsistency and localized inconsistency. For instance, if $\varphi, \varphi' \in \Phi_{\mathbf{F}}$ with $c_{\mathbf{F}}(\varphi) = c_{\mathbf{F}}(\varphi')$, then it is possible that φ is inconsistent over many edges while φ' is consistent over all edges except one. For $\epsilon > 0$, we call $A \subseteq \Gamma$ an *ϵ -consistent subgraph* of Γ with respect to $(\mathbf{F}, \varphi)$ if $c_{\mathbf{F}}(\varphi)_A < \epsilon$. Then, φ and φ' have distinct ϵ -consistent subgraphs for some $\epsilon > 0$, distinguishing between φ ’s widespread inconsistency and φ' ’s localized inconsistency.

By Lemma 5 of [272], there exists for every $\epsilon > 0$ a unique (maximally coarse) ϵ -consistent subgraph $A_{\mathbf{F}, \varphi}(\epsilon)$ such that every other ϵ -consistent subgraph of Γ with respect to $(\mathbf{F}, \varphi)$ is a topological refinement of $A_{\mathbf{F}, \varphi}(\epsilon)$. It follows that for $0 < \epsilon_1 \leq \epsilon_2$,

$$B_{\mathbf{F}, \varphi}(\epsilon_1) \subseteq A_{\mathbf{F}, \varphi}(\epsilon_2)$$

That is, lower-consistency subgraphs contain higher-consistency subgraphs. The relationship between multiresolution data consistency and Γ ’s geometric structure yields the *consistency filtration* for $(\mathbf{F}, \varphi)$, comprising nested ϵ -consistent subgraphs for increasing ϵ -consistency thresholds $0 < \epsilon_1 < \dots < \epsilon_{n-1} < \epsilon_n = c_{\mathbf{F}}(\varphi)$ [272]. By tracking the geometry of data consistency across local-to-global scales, the consistency filtration enables multiresolution obstruction detection on GENEIO sheaves. In particular, by identifying substantially different consecutive thresholds ϵ_i and ϵ_{i+1} , we can localize a likely non-trivial obstruction (i.e., an obstruction arising from system behavior rather than noise) to the edges $e \in E(\Gamma)$ with $e \in E(B_{\mathbf{F}, \varphi}(\epsilon_i))$ and $e \notin E(A_{\mathbf{F}, \varphi}(\epsilon_{i+1}))$.

4.5.2 Tracking Subsystem Influence

We have demonstrated GENEIO sheaves' ability to apply, diffuse, and compare data up to heterogeneous multiresolution symmetries on subsystems. However, we have yet to analyze heterogeneous multiresolution symmetries' system-wide influence. In real-world systems, each subsystem's behavior can influence states of other subsystems, even when subsystems and their behaviors are heterogeneous. We evaluate these system dynamics by tracking obstructions to subsystems' ability to influence the rest of the system. This last contribution to our theory of heterogeneous multiresolution equivariance facilitates analysis of subsystems' relationships and interdependent behavior at various levels of granularity.

Definition 4.5.4. For GENEIO sheaf $\mathbf{F}$ on graph Γ , let S_A be the group of stalk symmetries over subgraph $A \subseteq \Gamma$. For $S_A^i \subseteq S_A$, let $A_i \subseteq \Gamma$ be the largest subgraph of Γ with $A \subseteq A_i$ such that for every $v \in V(A)$, $u \in V(A_i)$ connected by path $e_0, \dots, e_{k-1}, e_k \in E(A_i)$, there exists group homomorphism $T_{v \rightarrow u} : G_{v, e_0} \rightarrow G_{u, e_k}$ defined

$$T_{v \rightarrow u} = T_k \circ T_{k-1} \circ \dots \circ T_0, \quad T_j = T_{v_{j+1} \preceq e_j}^{-1} \circ T_{v_j \preceq e_j} \quad \text{for all } 0 \leq j \leq k \quad (4.29)$$

such that $s_A \notin \ker T_{v \rightarrow u}$ for all nontrivial $s_A \in S_A^i$. We call A_i the S_A^i -reachable subgraph of Γ for $\mathbf{F}$.

Stalk symmetries' reachable subgraphs geometrically define their maximal sphere of influence over the system, and homomorphisms $T_{v \rightarrow u}$ describe stalk symmetries' heterogeneous manifestations on different parts of A_i . In other words, A_i bounds how far local symmetries S_A^i on subsystem A can heterogeneously propagate across – and thereby impact – the rest of the system. In practice, stalk symmetries may not thoroughly propagate across their reachable subgraphs; propagation extent and intensity depends upon choice of restriction GENEIOs and $L_{\mathbf{F}}$ diffusion dynamics. Theorems 4.4.6 and 4.4.8 show how different $L_{\mathbf{F}}$ dynamics yield different propagation intensities, with $H_{\mathbf{F}}$ -harmonic extension being the most extreme case. Additionally, stalk symmetries may not propagate if they fail to improve edge-wise data consistency with respect to restriction GENEIO relationships between subsystems.

Just as ϵ -consistent subgraphs underpin analysis of multiresolution obstructions to data consistency, S_A^i -reachable subgraphs underpin analysis of multiresolution obstructions to subsystem influence. Further, for nested sets of local subsystem behaviors – encoded as sets of stalk symmetries – we can construct a local-to-global filtration of these behaviors' spheres of influence. This filtration geometrically tracks obstructions to system-wide propagation of heterogeneous multiresolution stalk symmetries.

Definition 4.5.5. For GENEIO sheaf $\mathbf{F}$ on Γ and stalk symmetry group S_A over subgraph $A \subseteq \Gamma$, the collection

$$\{\text{id}_{S_A}\} \subseteq S_A^n \subseteq S_A^{n-1} \subseteq \dots \subseteq S_A^1 \subseteq S_A \quad (4.30)$$

induces *stalk symmetry filtration* $A \subseteq A_1 \subseteq \dots \subseteq A_{n-1} \subseteq A_n \subseteq \Gamma$ for $\mathbf{F}$.

It follows from Definition 4.5.4 that stalk symmetry filtrations are indeed filtrations, since $A_i \subseteq A_{i+1}$ whenever $S_A^{i+1} \subseteq S_A^i$. Whereas consistency filtrations are specific to $(\mathbf{F}, \varphi)$, stalk symmetry filtrations are only specific to $\mathbf{F}$. However, the extent to which S_A^i propagates throughout A_i under $L_{\mathbf{F}}$ may vary for different $\varphi \in \Phi_{\mathbf{F}}$.

Stalk symmetry filtrations geometrically describe how different types of local subsystem behavior heterogeneously manifest throughout the rest of the system, bolstering our understanding of heterogeneous multiresolution equivariance. Multiple filtrations can be defined for a single subsystem, enabling precise comparison of various behaviors' spheres of influence. By comparing reachable subgraphs A_i and A_{i+1} , we identify edges $e \in E(\Gamma)$ with $e \in E(A_i)$ and $e \notin E(A_{i+1})$ where propagation of stalk symmetries $S_A^i \setminus S_A^{i+1}$

breaks down. As $L_{\mathbf{F}}$ diffusion is animated by restriction GENEIO equivariance, stalk symmetry filtrations also offer some indication of system behavior during heterogeneous data fusion. Accordingly, we can assess how different subsystem behaviors impact global data alignment during GENEIO sheaf data fusion.

4.6 GENEIO Sheaves for Reference Frame Quality Analysis

Now that GENEIO sheaves, heterogeneous multiresolution equivariance, and their data analytical features have been mathematically formalized, we illustrate their conceptual depth and computational utility through an application to RFQ analysis on hand-drawn frames. First, we detail a GENEIO sheaf model of animators’ RFQ perception, showcasing GENEIO sheaves’ deep capacity for perceptual computing and ability to model Chapter 3’s perceptual experiment findings. We then discuss the design and results of our proof-of-concept RFQ GENEIO sheaf implementation, which estimates 3D arm poses depicted in hand-drawn animation frames. Lastly, we outline how our proposed RFQ GENEIO sheaf-based reference frame retrieval system will integrate into hand-drawn animation studios’ film production pipelines.

4.6.1 GENEIO Sheaf Model of RFQ Perception

Presently, state-of-the-art computational models of human visual perception are overwhelmingly DNN-based. As discussed in Section 2.4, these models struggle to exhibit core traits of visual cognition, including multiresolution data integration, shape awareness, part-wise relationship comprehension, and abstract visual reasoning. DNNs’ lack of explainability preclude decisive insights into why these failures occur and how they can be provably resolved [156, 11]. Recent research reveals profound gaps between DNN vision models and subtleties of visual system neurology [58], implying that modeling brain activity at a granular, neuronal level may be a tenuous perceptual computing paradigm. Given TDA’s many parallels with visual perception discussed throughout this thesis, it is unsurprising that TDA is rapidly advancing explainable relational learning [246] and can provably imbue DNNs with perceptually-based functionalities [146, 42, 231, 278]. Yet despite TDA’s considerable success in neuroscience research [52], TDA-based perceptual computing is still in its infancy [313, 194]. In response to perception researchers’ growing calls for perceptual computing frameworks exhibiting scientifically established traits of human vision [37, 166, 213], we propose GENEIO sheaves as a novel TDA-based model of human visual perception, which we illustrate subject to hand-drawn frame perception. Rather than modeling individual neurons, GENEIO sheaves can be programmed to emulate heterogeneous functionalities of and working relationships between visual brain areas. Within this fully explainable framework, essential traits of human vision inaccessible to DNNs can be provably installed.

In our illustrative example, GENEIO sheaf $\mathbf{F}$ on graph Γ models the abstract system of the human body as understood within the visual brain, with system states comprising physically feasible body poses. $\mathbf{F}$ is thus designed to approximate the perceptual processes involved in interpreting hand-drawn characters’ depicted volumetric body poses across diverse artistic styles, alongside professional animators’ RFQ perceptual judgments. Although $\mathbf{F}$ ’s perceptual formulation is RFQ-specific for expository clarity, the analytical principles described generalize to many other computer vision scenarios.

$\mathbf{F}$ ’s collection of raw subsystem sensor observations encode local patches of the retinal image, i.e., interactions between photons and retinal neurons when viewing a hand-drawn frame. Patch locality comports with human visual biology, since retinal neurons cannot integrate visual data across resolutions larger than a couple degrees of visual field [45] (Figure 4.8). Instead, local retinal image patches are globally integrated into a cohesive perceptual interpretation via simultaneous bidirectional data exchange between lower-level and higher-level brain areas (Figure 4.7) [45, 316, 65]. During perception, local lower-level

visual data initially derived from the retinal image – including contour geometry, topology, and relative placement – undergoes representational transformations as it is fed-forward to higher-level brain areas [316]. There, it is integrated at coarser visual field resolutions. Concurrently, higher-level visual processing – including abstract visual reasoning, structural and functional part-wise relationship assessment, and prior knowledge injection – is performed on these coarser-resolution higher-level representations. Resulting perceptual insights are fed-back to lower-level brain areas, where local lower-level data is recalibrated subject to the received higher-level signals, before it is again fed-forward to higher-level areas. This continues until both feed-forward and feed-back signals cease to be sent, meaning that a globally consistent perceptual interpretation has been reached.

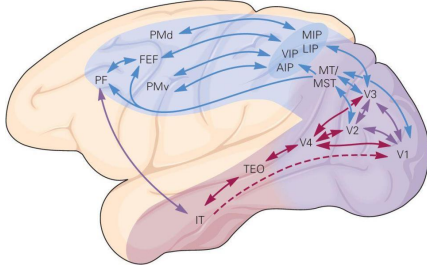


Figure 4.7: Bidirectional visual data exchange between lower-level (purple) and higher-level (blue and pink) brain areas. Figure from [358].

This bidirectional visual data integration process is congruent to heterogeneous data fusion on GENEIO sheaves. Suppose $\mathbf{F}$'s restriction GENEIOs model various low- and high-level visual processing operations applicable to diverse artistic styles, such as contour extraction, region adjacency detection, foreshortening assessment, medial axis computation, 3D shape comprehension, and prior knowledge injection of joint motion constraints. Let $\mathbf{F}$'s model sensors each encode a model subsystem comprising one isolated 3D skeleton bone. During harmonic extension on $\mathbf{F}$, raw sensor observations (local retinal image patches from a hand-drawn frame) are transformed by $\mathbf{F}$'s heterogeneous restriction GENEIOs (representational transformations) as they map to $\mathbf{F}$'s edge stalks (lower- and higher-level brain areas), with Γ 's topology directing data passage between brain areas. Since both lower- and higher-level brain areas are active extremely early in visual processing [65], $\mathbf{F}$'s simultaneous data passage to edges modeling lower- and higher-level areas is neurologically justified. Edges modeling higher-level areas evaluate abstract functional and relational consistency (pose feasibility) between neighboring 3D skeleton bones' respective local states (perceived volumetric positions and orientations). Edges modeling lower-level areas recalibrate each bone's local state to optimally align its occluding contours with material frame features from the corresponding retinal image patch. At each bone vertex stalk, heterogeneous signals from both higher- and lower-level brain areas are progressively integrated until a globally consistent perception of the hand-drawn character's depicted body pose emerges and signals cease to be sent; that is, until $L_{\mathbf{F}} = \text{id}_{H_{\mathbf{F}}}$. A proof-of-concept implementation of this model is detailed in the following subsection (see Figure 4.10).

Multiresolution data integration Synthesizing a collection of local data (admissible 0-assignment; retinal image) into a globally valid understanding of the collection's holistic meaning (global section; perceptual interpretation) is foundational to both GENEIO sheaves and human vision. In both contexts, sophisticated global meaning can be extracted from simple collections of local data because the analysis framework itself – rather than the input data – is highly sophisticated and multifaceted. Moreover, the analysis framework profoundly influences how each piece of local data is interpreted within the global visual setting. In this way, sophisticated perceptual interpretations are obtained from simple material features, a core trait of human line drawing perception [65, 326, 177]. Since multiresolution data integration enforces structural and functional part-wise relationships across multiple scales, both GENEIO sheaves and human vision can detect multiresolution visual inconsistencies missed by DNNs (Figure 2.35) [11].

GENEIO sheaves support multiscale analysis of part-wise visual and functional relationships (restriction GENEIOs) and subsystem behavior (described by the space of A -local sections and by stalk

symmetries/permutations over $A \subseteq \Gamma$). Likewise, humans – especially artists – can localize visual reasoning to multiple scales while retaining global understanding of visual data [54, 53]. This supports extrapolation of how local subsystem changes globally impact the the abstract system (human body) represented by the given visual data (hand-drawn frame) [164]. On $\mathbf{F}$, this comprehension is modeled by stalk symmetry filtrations. Moreover, differences between $L_{\mathbf{F}}$ diffusion dynamics for various $\varphi \in \Phi_{\mathbf{F}}$ model human visual reasoning of how specific subsystem changes differently impact certain global system states: for instance, bending a leg unbalances the body (produces non-trivial $L_{\mathbf{F}}$ dynamics) in some poses but not others.

Context-aware interpretation is a key visual cognition trait arising from multiresolution data integration [94, 182], and is faithfully modeled by harmonic extension on $\mathbf{F}$. Perceptual interpretation of each retinal image patch depends upon multiscale visual data relationships across the entire visual field: a circle, for instance, has numerous interpretations when viewed in isolation, but represents a human head when viewed in the context of restroom signage. Likewise, for subgraphs $B \subset A \subseteq \Gamma$, $\mathbf{F}$'s space of $H_{\mathbf{F}}$ -harmonic extensions of φ_A is a subset of the space of $H_{\mathbf{F}}$ -harmonic extensions of φ_B , mirroring how perceptual interpretations are generally further constrained by larger quantities of visual data. Humans can also meaningfully interpret visual data depicting part, but not all, of a known whole, facilitating pose perception for both partial-body and full-body hand-drawn frames (Figures 3.8 and 3.9). Whereas partial-body frames frustrate DNN HPE methods (Figure 2.33), our proof-of-concept GENEEO sheaf implementation demonstrates that, like human vision, $\mathbf{F}$ meaningfully interprets partial-body pose at local visual scales.

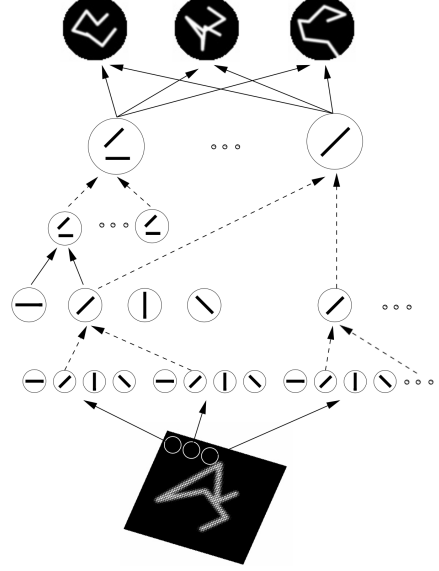


Figure 4.8: Multiresolution integration and material feature composability over increasingly coarse resolutions of visual field. Figure from [268].

Visual ambiguity Human viewers can adopt unique subjective interpretations of visual data while remaining aware of its multiple feasible perceptual interpretations [182, 83]. Analogously, A -restricted dynamics converge towards a unique $H_{\mathbf{F}}$ -harmonic extension of φ_A over Γ , while φ_A 's space of $H_{\mathbf{F}}$ -harmonic extensions captures all globally valid system states preserving φ_A . Spaces of $H_{\mathbf{F}}$ -harmonic extensions thus encode feasible subjective interpretations of visual data, modeling animators' varied pose estimates for each frame in Experiment 2 (Section 3.2). Since a frame is unlikely to engender a finite set of perceptual interpretations, the admissible 0-assignment pseudometric $D_{\Phi_{\mathbf{F}}}$ (Equation 4.17) can serve to construct spaces of approximate – rather than exact – $H_{\mathbf{F}}$ -harmonic extensions up to some similarity threshold.

Because A -restricted dynamics optimize, but do not require, global data consistency, pose estimates computed on $\mathbf{F}$ optimally approximate body motion constraints without necessarily satisfying them explicitly. This naturally accommodates visual ambiguities created by hand-drawn characters' frequent stylistic deformations, which confound pose estimators assuming rigid body geometry [25]. In our model, Principles of Animation like squash-and-stretch, bone bending/breaking, and stylized perspective (Figure 2.12) simply elevate inconsistency between adjacent skeleton bones in $\mathbf{F}$'s pose estimates. Analogously, human vision registers these impossible bodily deformations while maintaining a globally coherent understanding of character pose. Pose estimates' admissible 1-assignments thus quantifiably encode input frames' body part-wise deformations, and consistency filtrations over $\mathbf{F}$'s model system subgraph geometrically track deformation intensity across local-to-global scales. This means that similarly deformed

frames should have similar consistency filtrations, facilitating reference frame retrieval subject to specified body part-wise applications of Principles of Animation.

GENEO sheaves inherit sheaves’ ability to extrapolate missing information from available data by leveraging part-wise relationships between subsystems [173, 143]. Feasible states of an unobserved subsystem are constrained by the observed states of its related subsystems, alongside the nature of these subsystem relationships. Comparably, human vision resolves ambiguities arising from object occlusion through inferences made from nearby geometric data [228, 46] and part-wise relationships between visual features [73, 215]. These strategies also support other types of visual ambiguity resolution [182], potentially including animators’ possible ability to constrain their perceptions of distorted body part pose using nearby perceptual cues, as theorized in our Experiment 2 Discussion. When estimating hand-drawn character pose, $\mathbf{F}$ leverages perceptual cues and body motion constraints salient to the aforementioned kinds of visual ambiguity resolution. In practice, this should facilitate reasonable handling of body part-wise occlusion over articulate motion and perceptually-based pose estimation on distorted body parts.

Perceptual cue inconsistency Our Experiment 2 findings indicate that in order to anticipate animators’ varied subjective pose perceptions for a given frame, $\mathbf{F}$ must detect perceptual cue inconsistencies. Although $L_{\mathbf{F}}^A$ always converges to a pose estimate optimally satisfying all perceptual cues in any given frame, frames with cue inconsistencies may have empty spaces of $H_{\mathbf{F}}$ -harmonic extensions [275]. To reconstruct an inconsistent frame’s space of perceptually feasible poses, a consistency filtration can be computed after data fusion over the subgraph of input material feature patches and their adjacent 3D bones. High-inconsistency areas show where perceptual cues disagree with the pose estimate, localizing cue inconsistencies to specific body parts. The space of feasible pose perceptions is found by weighting $D_{\Phi_{\mathbf{F}}}$ to locally relax similarity thresholds over inconsistent body parts, then finding poses sufficiently $D_{\Phi_{\mathbf{F}}}$ -similar to the frame’s pose estimate. This models body part-wise differences in the subjectivity of animators’ pose perceptions for a single frame, as observed in Experiment 2. Existing pose estimates for frames with similar inconsistency cases can also be used to infer possible pose perceptions for a given input frame, especially when other frames have additional perceptual cues better constraining the inconsistent body part’s feasible poses.

Multiresolution isometry invariance Retinal image data undergoes myriad representational transformations during passage between and analysis within lower- and higher-level visual brain areas. Because symmetries on visual data generally induce corresponding symmetries on perceptual interpretations of its representational meaning [258], representational transformations must preserve visual symmetries; that is, human vision exhibits equivariance. Analogously, restriction GENEO equivariance underpins GENEO sheaf data fusion equivariance (Theorems 4.4.8 and 4.4.12), modeling how symmetries’ perceptual meanings heterogeneously transform alongside data representations as information passes through the visual system. Heterogeneous equivariance is a core component of visual abstraction, enabling animators to convey a camera-parallel limb rotation in $SO(3)$ by isometrically rotating the limb’s representative regions in $SO(2)$ across a sequence of frames. Moreover, our Experiment 1 (Section 3.1) findings alongside artists’ multiresolution perceptual acuity [54, 53] imply that animators’ RFQ perception is moderately equivariant to heterogeneous multiresolution isometries on hand-drawn frames, modeled by stalk symmetries (body part-wise rotation/reflection) and stalk permutations (cross-body reflections) on $\mathbf{F}$ (Figure 4.13).

The GENEO sheaf ground-truth pseudometric $d_{G_{\mathbf{F}}}$ (Equation 4.22) fluently captures animators’ RFQ perception equivariance and can be weighted to favor certain symmetries $g \in G_{\mathbf{F}}$, tailoring reference frame retrieval to animators’ perceptual preferences and technical working environment. For instance, frames differing from the desired pose by small limb rotation or full-body reflection (which animators can reflect on their computers, improving RFQ) can be prioritized, while those differing by large local

isometries are disregarded, comporting with Experiment 1’s results. However, Experiment 1 reveals that animators’ RFQ perception can strongly depend on frames’ material rather than depicted feature similarity (Figure 3.2). Comparing region geometry and topology across hundreds of frames for each reference frame query would be extraordinarily inefficient. Because GENEО sheaf data fusion equivariance ensures that $\mathbb{R}^2$ frame isometries are faithfully converted into corresponding $\mathbb{R}^3$ pose isometries during pose estimation on $\mathbf{F}$, efficient pose estimate comparisons provably approximate cumbersome material feature comparisons during reference frame retrieval. Additionally, restricting $d_{G_{\mathbf{F}}}$ to particular subgraphs will support body part-specific reference frame retrieval, allowing animators to accrue high-quality body part-wise reference frames when, as is often the case, a suitable full-body reference frame does not exist.

Equivariance can also support study of relationships between frames’ material and depicted features. As Experiment 1 (Section 3.1) suggests that animators’ RFQ perception is best modeled by flexibly combining depicted and material feature metrics, $d_{G_{\mathbf{F}}}$ can be dynamically weighted to prioritize pose similarity, region shape similarity, or some combination of the two. Various spaces of related frames generated by different weightings of $d_{G_{\mathbf{F}}}$ can then be compared to systematically study relationships between frames’ material and depicted features with respect to animators’ RFQ perception. To enhance this analysis, future perceptual studies on animators can be conducted to identify weightings especially salient to RFQ perception. This can also be pursued for other GENEО sheaf pseudometrics and other frame perception tasks.

Because human vision involves part-wise relationship comprehension, prior knowledge injection, and abstract visual reasoning, multiresolution symmetries are understood both representationally and functionally. For instance, humans can understand region rotations as representing limb rotations, as well as determine whether a certain volumetric limb rotation is physically feasible within a frame’s depicted pose. While GENEО sheaves cannot yet fully emulate the depths of human visual reasoning, the physical feasibility of applying symmetry $g \in G_{\mathbf{F}}$ to pose φ can be determined by comparing consistency filtrations for φ and $\varphi \circ g$. When g is a stalk symmetry, analyzing $\mathbf{F}$ ’s g -reachable subgraph for $\varphi \circ g$ reveals body parts whose local poses may change under $L_{\mathbf{F}}$ while restoring global pose feasibility. Given sheaves’ success in knowledge and logic modeling [257, 132], extending these capabilities to higher-level abstract reasoning presents promising paths for future work.

4.6.2 RFQ GENEО Sheaves for Hand-drawn Character Pose Estimation

To demonstrate GENEО sheaves’ utility, computationally corroborate our theoretical results, and illustrate our GENEО sheaf perceptual model’s viability, we present a proof-of-concept GENEО sheaf implementation estimating 3D arm poses depicted in Cartoon Saloon hand-drawn frames. Despite our implementation’s simplicity, we showcase key features of our conceptual and analytical framework, including GENEО sheaf data fusion, stalk symmetry equivariance, and integrated lower- and higher-level visual processing. We focus on the arm because it is a structurally simple yet highly articulate body part, allowing us to test a wide variety of complex poses. Because we test the arm in isolation and model only one type of lower-level visual processing (occluding contour extraction) and higher-level visual processing (elbow pose feasibility), respectively, we lack the majority of visual information we intend to leverage in our forthcoming full-body implementation. Consequently, we test synthetic arm regions rather than arm regions from real hand-drawn frames, as our current implementation lacks the analytical features needed to accommodate the latter. However, our test cases were inspired by real *Wolfwalkers* frames, which appear alongside our results (Figure 4.11).

Implementation In our RFQ GENE sheaf (Figure 4.10), real-world sensor observations comprise respective 2D pixel images u and l of synthetic upper and lower arm regions, while model sensor observations comprise respective 3D cylinder meshes U and L representing upper and lower arm bones. Pose estimation involves heterogeneous data fusion on 2D regions and 3D bones; that is, we take the $H_{\mathbf{F}}$ -harmonic extension of input arm regions to the rest of the RFQ GENE sheaf.

We construct GENE sheaf $\mathbf{F}$ on (arbitrarily directed) graph $\Gamma = (V(\Gamma), E(\Gamma))$ with $V(\Gamma) = \{u, l, U, L\}$ and $E(\Gamma) = \{i, j, k\}$. Given a fixed camera position $c \in \mathbb{R}^3$ along the Z -axis, we evaluate visual similarity between each 2D region and its corresponding 3D bone on edges $i = u \rightarrow U$ and $j = l \rightarrow L$, respectively. On edge $k = L \rightarrow U$, we evaluate 3D arm pose feasibility with respect to elbow joint motion constraints. $\mathbf{F}$ has vertex perception pair stalks:

- $(\Phi_v, \text{Aut}_{\Phi_v}(X_v))$ at 2D region vertices $v \in \{l, u\}$, with $\Phi_v = \{\varphi_v : X_v \rightarrow [0, 1]\}$ and $X_v = \{(x, y) \in \mathbb{R}^2 : |x + y| + |x - y| \leq s\}$, and Φ_v comprising 2D binary pixel images of width and length s ,
- $(\Phi_V, \text{Aut}_{\Phi_V}(X_V))$ at 3D bone vertices $V \in \{L, U\}$, with $\Phi_V = \{\varphi_V : X_V \rightarrow [0, 1]\}$ for $X_V = \mathbb{R}^3$ such that for 3D mesh m_V ,

$$\varphi_V(x, y, z) = \begin{cases} 1 & \text{if } (x, y, z) \text{ lies on } m_V\text{'s surface} \\ 0 & \text{otherwise} \end{cases}$$

and edge perception pair stalks:

- $(\Psi_e, \text{Aut}_{\Psi_e}(Y_e))$ at edges $e \in \{i, j\}$ evaluating bone/region visual similarity, with $\Psi_e = \{\psi_e : Y_e \rightarrow [0, 1]\}$ for $Y_e = \{(x, y) \in \mathbb{R}^2 : |x + y| + |x - y| \leq s\}$, and Ψ_e comprising 2D binary pixel images of width and length s ,
- $(\Psi_k, \text{Aut}_{\Psi_k}(Y_k))$ at edge k evaluating elbow joint constraints, with $\Psi_k = \{\psi_k : Y_k \rightarrow [0, 1]\}$ for $Y_k = \mathbb{R}^3$ such that for 3D mesh m_k ,

$$\psi_k(x, y, z) = \begin{cases} 1 & \text{if } (x, y, z) \text{ lies on } m_k\text{'s surface} \\ 0 & \text{otherwise} \end{cases}$$

Meshes m_L and m_U are tapered cylinders representing lower and upper arm bones, and mesh m_k is a straight line segment in $\mathbb{R}^3$.

$\mathbf{F}$ has the following restriction GENEos:

- $(F_{u \leq i}, T_{u \leq i}) : (\Phi_u, SO(2)) \rightarrow (\Psi_i, SO(2))$, such that $F_{u \leq i}$ extracts 2D keypoints along the sides of upper arm region φ_u and $T_{u \leq i}$ is trivial.
- $(F_{U \leq i}, T_{U \leq i}) : (\Phi_U, SO(P_{XY})) \rightarrow (\Psi_i, SO(2))$, such that $F_{U \leq i}$ extracts the occluding contour (Equation 4.31) of upper arm bone φ_U for camera position c , projects the contour to $\mathbb{R}^2$, and returns 2D side keypoints. $T_{U \leq i}$ is the natural homomorphism from rotations in the XY -plane $P_{XY} \subset \mathbb{R}^3$ to $SO(2)$.
- $(F_{l \leq j}, T_{l \leq j}) : (\Phi_l, SO(2)) \rightarrow (\Psi_j, SO(2))$, such that $F_{l \leq j}$ extracts 2D keypoints along the top and bottom edges of lower arm region φ_l , and $T_{l \leq j}$ is trivial.

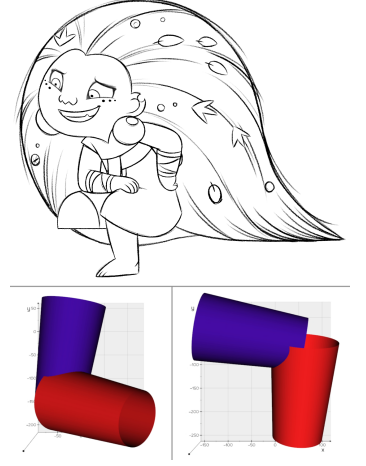


Figure 4.9: RFQ GENE sheaf left and right arm pose estimates (bottom) for a hand-drawn frame (top).

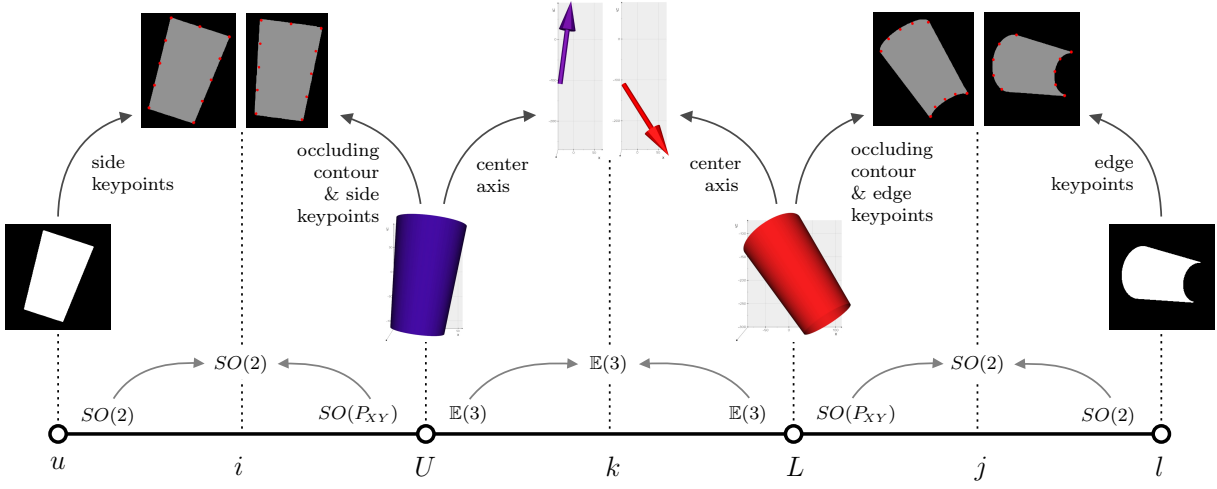


Figure 4.10: Diagram of our RFQ GENE sheaf implementation, addressing the problem of 3D (arm) pose estimation on 2D hand-drawn animation frames.

- $(F_{L \leq j}, T_{L \leq j}) : (\Phi_L, SO(P_{XY})) \rightarrow (\Psi_j, SO(2))$, such that $F_{L \leq j}$ extracts the occluding contour (Equation 4.31) of lower arm bone φ_L for camera position c , projects the contour to $\mathbb{R}^2$, and returns 2D top and bottom edge keypoints. $T_{L \leq j}$ is the natural homomorphism from rotations in the XY -plane $P_{XY} \subset \mathbb{R}^3$ to $SO(2)$.
- $(F_{L \leq k}, T_{L \leq k}) : (\Phi_L, \mathbb{E}(3)) \rightarrow (\Psi_k, \mathbb{E}(3))$ and $(F_{U \leq k}, T_{U \leq k}) : (\Phi_U, \mathbb{E}(3)) \rightarrow (\Psi_k, \mathbb{E}(3))$, such that both $F_{L \leq k}$ and $F_{U \leq k}$ extract the directed center axis line segment for their respective input bones $\varphi_L \in \Phi_L$ and $\varphi_U \in \Phi_U$. Both $T_{L \leq k}$ and $T_{U \leq k}$ are trivial.

For point p on a given 3D surface, its surface normal n , and camera position $c \in \mathbb{R}^3$, the occluding contour is computed as

$$\mathcal{O}(p) = \begin{cases} 1 & \text{if } n \cdot (c - p) = 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.31)$$

Before proceeding, we informally confirm that $\mathbf{F}$'s restriction GENEOS are in fact GENEOS. Proving equivariance is straightforward, since all restriction GENEOS are either trivial or natural. To show non-expansivity, we observe that each restriction GENEOS decreases dissimilarity between any pair of inputs. Specifically, different 2D shapes can have equivalent sets of edge and side keypoints, different 3D surfaces can have equivalent occluding contours, and cylinders of different circumferences can have equivalent center axis line segments.

We implemented $\mathbf{F}$ in Python with Vedo [106] for 3D mesh processing, and tested our method on 32 pairs of synthetic upper and lower arm regions. Regions were generated by projecting upper and lower 3D bone meshes positioned to match arm poses from real hand-drawn animation frames from *Wolfwalkers* (Figure 4.9, top), and depicted a wide variety of arm poses. All frames depicted the same character, and side/edge keypoint restriction GENEOS were defined subject to the character's design.

3D pose estimation on each input region pair was performed via GENEOS sheaf data fusion: for subgraph $A \subseteq \Gamma$ of raw sensor observations with $V(A) = \{l, u\}$, $E(A) = \emptyset$, we take the $H_{\mathbf{F}}$ -harmonic extension of φ_A to the rest of Γ via A -restricted dynamics $L_{\mathbf{F}}^A$ of the GENEOS sheaf Laplacian $L_{\mathbf{F}}$. Once $L_{\mathbf{F}}^A$ converges, the model system state φ_B for model system subgraph $B = \Gamma \setminus A$ gives our output arm pose estimate $\varphi_U \cup \varphi_L$ for $\{U, L\} = V(B)$; that is, the union of respective upper and lower arm bone subsystem states. Unlike a typical 3D character rig, bone meshes m_U and m_L are not hierarchically connected, nor

do they reside in a shared 3D space. Rather, m_U and m_L are localized to their respective vertex stalks and housed within respective 3D spaces X_U and X_L , interacting only when consistency between their center axis line segments is evaluated subject to elbow motion constraints in Ψ_k . Elbow constraints are thus both angular and positional.

As discussed in Section 2.3, estimating hand-drawn character pose by fitting preexisting 3D body geometry to each input frame [25, 39] can yield degenerate results (Figure 2.33). This is because, as seen in Experiment 2, hand-drawn characters’ body geometry is extremely fluid, undergoing ceaseless minor alterations during pose resolution alongside frequent overt distortion from Principles of Animation. Human vision effortlessly handles these irregularities precisely because it does not employ geometry-fitting [212]. Instead, vision leverages multiresolution data integration and part-wise relational reasoning to obtain globally cohesive geometric comprehension from geometrically incongruous visual data, as illustrated by the Kanizsa triangle illusion (Figure 4.12). By allowing local character geometry to self-assemble into a globally cohesive pose estimate according to both low-level shape and high-level relational cues, our RFQ GENE0 sheaf faithfully models key facets of human vision and is well-equipped to handle volatile and inconsistent character geometry.

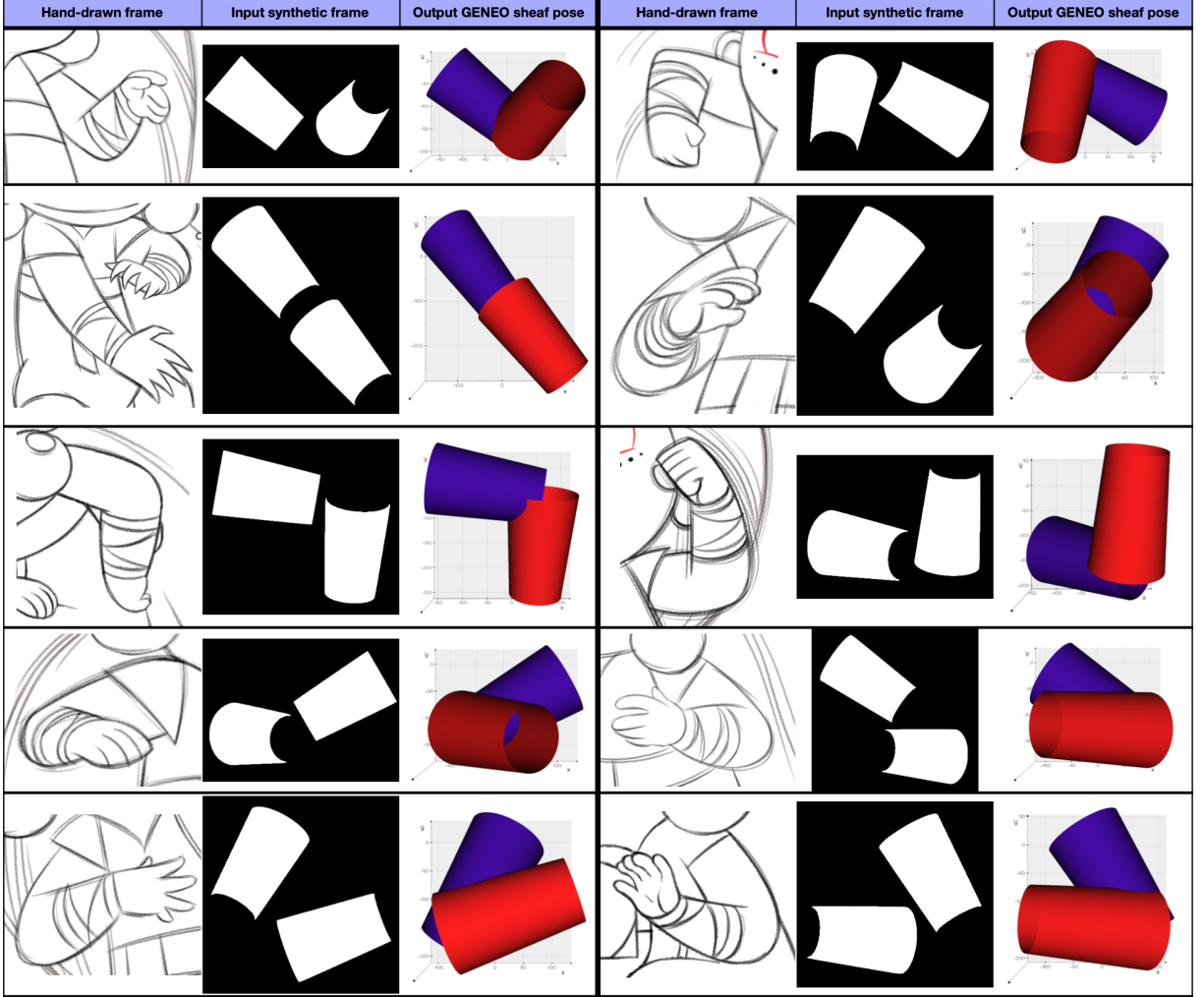


Figure 4.11: Example 3D arm pose estimates from our RFQ GENE0 sheaf implementation, shown alongside input pairs of synthetic upper and lower arm regions, as well as the *Wolfwalkers* frames on which they are based. Complete results are available in Appendix A.

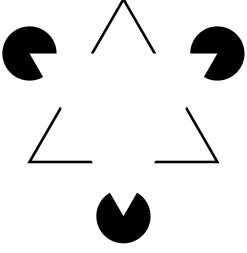


Figure 4.12: Kanizsa triangle illusion.

Complete results of our implementation tests are available in Appendix A. While formal evaluation of our results was not possible due to lack of suitable perceptual metrics and 3D pose baselines for hand-drawn frames, we obtained satisfactory arm pose estimates in every case tested. Our results validate GENEIO sheaves as a suitable framework for 3D pose estimation of hand-drawn characters, and confirm that the basic premises of our conceptual GENEIO sheaf model of RFQ perception (Section 4.6.1) behave as expected. They also demonstrate the real-world efficacy of our GENEIO sheaf data fusion algorithm (Conjecture 4.3.2 and Theorem 4.3.5), which is the first sheaf Laplacian-based fusion algorithm capable of integrating heterogeneous non-vector data representations across nonlinear

subsystem relationships. Our work improves upon the expressivity of prior sheaf Laplacian formulations [154, 32], but is less efficient than these formulations due to the higher complexity of the data that we process. However, our algorithm’s demonstrated efficacy suggests many desirable paths for future work, such as spectral analysis on GENEIO sheaves, which could build upon hand-drawn animation technology research’s growing interest in spectral methods (Section 2.6).

Although our implemented GENEIO sheaf was not large enough to thoroughly test all facets of heterogeneous multiresolution equivariance, our results suggest that data fusion is equivariant to stalk symmetries (Theorem 4.4.8) and stalk permutations (Theorem 4.4.12), computationally modeling our Experiment 1 findings (Section 3.1). In particular, the frame pair shown in Figure 4.13’s right column was tested in Experiment 1, achieving an average RFQ rating of 6.4 out of 7, the 2nd highest rating of all 54 frame pairs tested. An Experiment 1 frame pair comparable to Figure 4.13’s left column (Figure 1.4, top left) was also tested, achieving the 3rd highest average RFQ rating of 6.36. Simply put, our results indicate that input 2D regions differing by a 2D isometry yield output 3D poses differing by a corresponding 3D isometry.

Computational evidence of GENEIO sheaf data fusion equivariance implies several key points. First, our RFQ GENEIO sheaf’s ability to model our Experiment 1 findings substantiate GENEIO sheaves as viable perceptual computing framework. Second and most crucially, our RFQ GENEIO sheaf’s faithful conversion of local $\mathbb{R}^2$ isometries on frames into corresponding local $\mathbb{R}^3$ isometries on pose estimates ensures that pose estimate comparison up to ground-truth pseudometric d_{G_F} can approximate material frame feature comparison up to d_{G_F} . Then, as discussed in the previous subsection, our reference frame retrieval system can use pose estimates as computationally efficient proxies for assessing the geometric and topological perceptual cues found to be salient to animators’ frame perception in Experiments 1 and

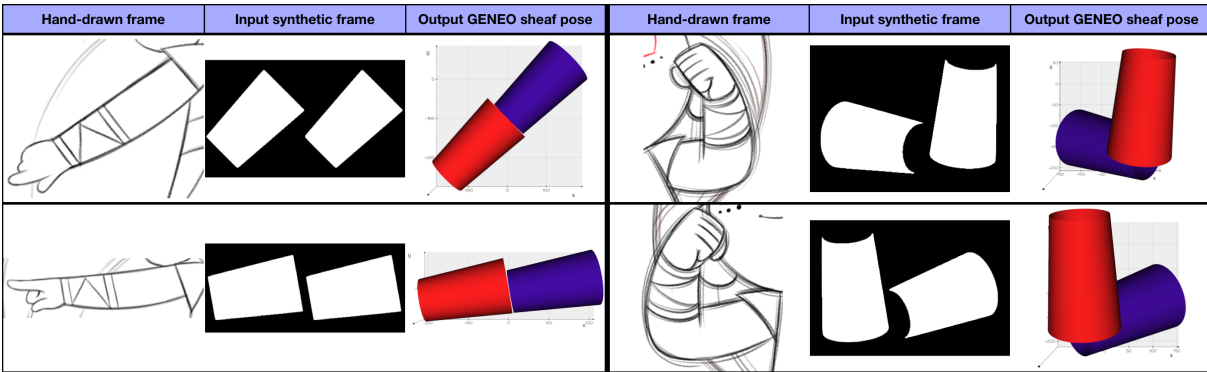


Figure 4.13: Results of our RFQ GENEIO sheaf implementation indicating that stalk symmetries (left column) and stalk permutations (right column) computationally behave as anticipated by Theorems 4.4.8 and 4.4.12.

2. Heterogeneous multiresolution equivariance thus establishes the stability between frames’ material and depicted features described in Chapter 2’s Discussion. From this stability, material and depicted feature relationships can emerge naturally as described in the previous subsection, rather than being hard-coded into our analysis framework. Moreover, by comparing pose estimates up to multiresolution isometry, our reference frame retrieval system should be able to return frames preserving director-approved pose resolution strategies relevant to the given input query.

Our results successfully demonstrate body part-wise pose estimation, implying that RFQ GENE0 sheaves will function as truly local-to-global pose estimators. Our framework emulates key traits of human vision – like visual composability and decomposability, part-wise structural and functional relationship comprehension, and multiresolution perceptual acuity – which appeared to play pivotal roles in animators’ frame perception during Experiments 1 and 2. Unlike DNN-based pose estimators, many of which lack these traits and thus struggle on partial-body frames, these and other perceptual factors influencing animators’ RFQ perception are explainably and controllably programmed into $\mathbf{F}$ ’s mathematical construction. In future, we expect $\mathbf{F}$ ’s multiresolution visual acuity to fluently handle both partial-body and full-body frames and, eventually, to support body part-wise reference frame retrieval queries.

Limitations Our RFQ GENE0 sheaf’s most significant limitation is that body structure and character design must be encoded into $\mathbf{F}$ prior to analysis, making our method less dynamic than other hand-drawn character pose estimators. Consequently, RFQ GENE0 sheaves must be manually defined for each model sheet, although we expect this process to become partially automatable in future. Note that DNN-based HPE methods for hand-drawn characters also tend to require substantial manual set-up, as they may need training on hundreds or thousands of manually annotated frames to achieve high-quality performance. Our current implementation fails when the arm is highly foreshortened, although this issue is not systemic. It is caused by keypoint overlaps, and can be addressed by improving our restriction GENE0 formulations. Unlike other HPE methods, $\mathbf{F}$ is not a combined body part recognition and pose estimation framework, meaning that body part identification will need to be computed separately in order to define φ_A . Although our implementation of $L_{\mathbf{F}}$ exhibits the convergence behavior described by Conjecture 4.3.2, a full proof of this Conjecture will be formulated in future work to guarantee that this behavior persists in all circumstances. Convergence of $L_{\mathbf{F}}^A$ is reasonably efficient in our implementation (and can be made more so), but not as fast as DNN-based pose estimators. In practice, however, this should not be a problem since animators can only draw so many frames in a day, and pose estimation can run overnight on studio servers. Despite these drawbacks in comparison to DNN-based pose estimators, we note that both sheaves and GENE0s are highly compatible with DNNs [150, 20, 32]. Therefore in future, we will likely be able to integrate learning-based approaches into our RFQ GENE0 sheaf while retaining its existing advantages over purely DNN-based HPE approaches for hand-drawn frames.

4.6.3 Proposed Industry Pipeline

We conclude by overviewing how our proposed RFQ GENE0 sheaf-based reference frame retrieval system will integrate into hand-drawn animation studios’ film production pipelines. Although building our proof-of-concept implementation into a complete reference frame retrieval system will take considerable time and effort, a plan for its eventual deployment in an industry setting has already been carefully developed through long-running collaboration with Cartoon Saloon animators and technical personnel. As discussed throughout this thesis, a key shortcoming of many existing hand-drawn animation technologies is their incompatibilities with professional animators’ artistic and technical needs. Technologies also might conflict with hand-drawn studios’ established software ecosystems, data management protocols,

and technological best-practices painstakingly developed over years of trial-and-error. It is therefore essential for hand-drawn animation technology researchers to work with animators and studios directly when conceiving their research goals to ensure their output is tailored to industry needs.

Our pipeline design makes reasonable assumptions about the softwares and file structures of other hand-drawn studios. However, each studio’s technology usage is unique, and in-house software developers frequently customize general-purpose technologies to the studio’s preferences. Our proposed pipeline is based on Cartoon Saloon’s technology usage, but easily generalizes to other studios’ technical set-ups, particularly with assistance from in-house software developers. Technical specifications, such as animation file layer names, can be enacted by studios to comport with our proposed pipeline; it is usual for studios to undertake low-effort technical protocols in order to use specific software.

Our proposed hand-drawn studio reference frame retrieval system pipeline is as follows:

1. Animation supervisor or film director approves an animation scene, comprising uncolored hand-drawn animation frames. Approval is typically recorded in studios’ production tracking softwares, such as Shotgrid.
2. Approval triggers a customized computer program linked to the production tracking software, which obtains the approved scene’s animation file from within the studio’s file system.
3. Frames are extracted from the animation file via the proprietary compositing program used for data selection in Experiments 1 and 2 (Sections 3.1 and 3.2), assuming the animation file’s layers are labeled by depicted character name. If a frame depicts multiple characters, the compositing program separates them into different frames according to the animation file’s layer names. If multiple characters are drawn on the same layer, a face detector such as [291, 165] can be used to identify characters’ faces, and diffusion-based RAG comprehension can serve to identify each character’s body.
4. For each frame, a combination of HRNet 2D pose estimation trained on Human-Art [174], sheaf-based shape recognition, and existing proprietary frame processing programs identifies body parts for each character and extracts their local depictions.
5. Each frame’s set of local body part depictions are inputted into an RFQ GENEEO sheaf specific to the given character design, which is determined from the production tracking software’s metadata for the given scene. 3D pose estimation is performed analogously to our proof-of-concept implementation.
6. Frames and their pose estimates are stored in a database on studio servers, which can be queried by our reference frame retrieval system. Because GENEEO sheaves require no training, our system will be usable early in production.
7. To retrieve a reference frame, users pose a 3D rig like the one used in Experiment 2 (Figure 3.6) to match their desired pose, then optionally customize their query according to desired body part(s), multiresolution isometry preferences, body part-wise distortion preferences, and more.
8. The reference frame retrieval system queries the frame/pose database via GENEEO sheaf pseudometrics relevant to the given query.
9. Upon retrieving a set of frames matching search conditions, users can select particular frames to see additional results visually similar to the selected frame.

Each character design’s associated RFQ GENEEO sheaf will be designed through collaboration between in-house software developers and character designers before production begins. Each RFQ GENEEO sheaf will primarily employ restriction GENEEOs from a predefined library of simple perceptual cue evaluation functions, including occluding contour extractors, body part self-contact detectors, and

joint motion constraints. These will be applied to particular body parts subject to the given character design, with custom restriction GENEOS defined as needed. We aim to eventually devise automated RFQ GENEOS sheaf construction directly from model sheets (Figure 1.1) and accompanying body part identification annotations on model sheets supplied by character designers.

Although a fully automated RFQ GENEOS sheaf construction and reference frame retrieval system pipeline will involve considerable future work, our proposed pipeline requires minimal interaction from professional animators and other artistic personnel, avoiding interference with studios’ established artistic and technical workflows. Cartoon Saloon character designers confirm that the aforementioned annotations are easily supplied, unlike the hundreds of skeleton-annotated training frames requisite for most DNN-based pose estimators; such substantial user interventions may conflict with production schedules and personnel budgets. Even if we are unable to fully automate our pipeline, the manual interventions will be performed by hand-drawn studios’ in-house software developers, rather than artistic personnel. Since software developers are already responsible for in-house software customization, this division of labor aligns well with existing professional responsibilities within hand-drawn studios.

4.7 Discussion

GENEOS sheaves offer the first comprehensive theoretical and computational unification of heterogeneous data fusion and equivariant data analysis, two influential yet largely separate data analytical perspectives. Both perspectives primarily reside within the increasingly prominent field of Geometric Deep Learning (GDL), which exploits topological and algebraic structures to model properties and behavior of real-world data [42]. Interestingly, GDL shares parallels with hand-drawn animation technology research: both fields model real-world systems as mathematical structures, which capture many essential system features and offer deep insights into system behavior, but can also limit modeling and analytical flexibility. However, TDA is increasingly enhancing already successful mathematical approaches to data analysis by deepening and interconnecting their underlying topological ideas, often vastly improving analytical capabilities and mitigating practical limitations. Sheaves and GENEOS are both prime examples of this phenomenon, with the former substantially enhancing graph neural networks’ (GNNs) expressivity [150] and the latter enabling explainable EqNN construction for any selection of symmetry groups [20].

GENEOS sheaves continue this path of inquiry, offering a novel mathematical setting in which systems too complex for either sheaves or GENEOS can be explainably modeled and evaluated. GENEOS sheaf equivariance underpins the first Laplacian-based data fusion algorithm for non-cellular sheaves, integrating heterogeneous function-valued subsystem data via nonlinear maps encoding subsystem relationships. Our novel theory of heterogeneous multiresolution equivariance stitches together subsystems’ local data, relationships, and behavior into a global system model. This enables fine-tuned subsystem modulation, local-to-global subsystem influence tracking, and guaranteed preservation of heterogeneous multiresolution symmetries during GENEOS sheaf data fusion. Established sheaf and GENEOS pseudometrics meld into customizable, symmetry-aware comparison and consistency evaluation mechanisms for GENEOS sheaf data, facilitating a wide variety of insights into system behavior. GENEOS sheaves’ mathematical depth fluently translates into conceptual depth, showcased by our model of RFQ perception. These features present considerable opportunities for further theoretical development, including GENEOS sheaf Laplacian dynamics [154], restriction GENEOS construction algorithms [79], and maps between GENEOS sheaves on different graphs [2, 151].

GENEOS sheaves invite myriad potential applications, following from and expanding upon sheaves’ and GENEOS’ many respective use cases. GENEOS sheaves suit any application involving well-understood

systems that can be modeled computationally, especially those with heterogeneous subsystem data and behavior. As our own application illustrates, GENEО sheaves leverage prior knowledge of the system to extrapolate its state from sparse or soft – as well as noisy or uncertain – sensor observations. Our use case directly extends to human pose estimation from photos and videos, motion capture analysis, action recognition, and 2D-to-3D reconstruction. Sheaf-based scene analysis from image, text, and numerical data [271] may benefit from GENEО sheaves’ superior visual data handling, prospectively aiding 3D scene reconstruction and generation, medical imaging analysis, and object recognition. Sheaves [288], GENEОs [28], and related frameworks [281] all succeed in molecular analysis, laying groundwork for GENEО sheaf macromolecular modeling and drug design. This dovetails with GENEО sheaves’ suitability for decomposable systems of interchangeable subparts, indicating avenues into protein folding, nanotechnology, and self-assembling robotics.

While many of these applications follow naturally from our own application to RFQ analysis, broader paths of inquiry exist. GENEО sheaves fall within the purview of Topological Signal Processing, a subfield of TDA employing topological structures and techniques to evaluate sensor data [270]. GENEО sheaf approaches in this vein may enhance GENEО-based subgraph classification [2] and sheaf-based uncertainty quantification [143], satellite synchronization [273], object tracking [287], and knowledge modeling [132, 257]. Yet the most promising research directions may stem from adapting sheaf and GENEО neural network formulations [20, 152, 13] to GENEО sheaves. As discussed, the key limitations of our GENEО sheaf computational model are its mathematical complexity and user-specified construction, both of which restrict scalability. Such limitations are common in geometric and topological data analysis frameworks, but are effectively addressed by integrating deep learning functionality. The resulting GDL and Topological Deep Learning (TDL) models can reconstruct sophisticated mathematical machinery from sparse, user-defined mathematical constraints, vastly improving scalability and off-the-shelf applicability [32, 15]. GDL and TDL models also exhibit critical advantages over standard DNNs, including controllability, explainability, generalizability, higher-order relational reasoning, decreased training needs, and significant parameter reduction [42, 246, 28, 20].

The growing success of GDL and TDL sheds light on GENEО sheaves’ analytical and computational potential. Like prior syntheses of TDA and DNNs, we can reasonably expect GENEО sheaf neural networks (GSNNs) to ameliorate GENEО sheaves’ present scalability limitations. While EqNNs are restricted to global modulation and analysis of symmetries common to all subsystems, heterogeneous multiresolution equivariance represents, modulates, propagates, and tracks diverse subsystem symmetries across multiple system resolutions. Alongside sheaf neural networks’ established representational, relational, and dynamical advantages over GNNs, heterogeneous multiresolution equivariance may enable entirely novel modes of machine learning. Like other GDL and TDL models, GSNN learning will be highly explainable and controllable, potentially addressing increasingly prominent transparency and efficiency challenges within real-world deep learning sectors. Massive training needs, data center construction regulation, energy consumption concerns, hardware limitations, and raw materials availability will further galvanize interest in explainable and controllable DNNs in coming years, especially as DNNs’ analytical limitations become increasingly clear (Section 2.4.2). In particular, GSNNs adapted from RFQ GENEО sheaves may guaranteeably imbue learning-based computer vision with human-like visual processing, including shape awareness, part-wise relationship comprehension, multiresolution data integration, and abstract visual reasoning.

4.8 Conclusions

This chapter synthesizes insights into hand-drawn animation technology presented throughout Chapters 2 and 3 into a novel data analytical framework. It endeavors to answer Research Question 3 of this thesis (Section 1.2) – that is, how perceptual factors influencing animators’ RFQ perception can be computationally modeled to systematically compare hand-drawn frames. To contextualize this chapter’s contributions, recall that Chapter 2 established how hand-drawn animation technology research often focuses on processing and manipulating frames’ material features without robustly relating them to depicted pose. Meanwhile, Chapter 3’s experimental findings reveal that relationships between frames’ material and depicted features influence animators’ RFQ perception and perception of hand-drawn characters’ volumetric poses. However, these relationships can be so subtle and complex that even expert animators learn them gradually throughout production, rather than innately and instantaneously comprehending them. Moreover, pose resolution strategy preservation, dual material and depicted feature similarity awareness, visual composability and decomposability, pose interpretation subjectivity, and multiresolution perceptual cue integration, and cue consistency evaluation all appear relevant to animators’ RFQ perception. Then, by computationally modeling these perceptual factors, it becomes possible to capture perceptually salient aspects of relationships between frames’ material and depicted features.

Chapter 2’s state-of-the-art review finds that especially successful trends in hand-drawn animation technology comport precisely with this strategy. Sophisticated topological frame processing techniques yield superior usability and visual results, which we attribute to their primitive emulation of some of the perceptual factors described above, from which the crucial relationships between material and depicted features emerge. This chapter substantially advances this strategy, presenting the first application of TDA to hand-drawn animation technology. By anchoring our work in perceptual factors directly linked to relationships between material and depicted features, we are able to select advanced topological approaches precisely expressing the perceptual qualities – and thus, the feature relationships – we aim to model. This strategy proves successful, yielding a deep, unified mathematical and computational framework modeling perceptual factors salient to animators’ RFQ perception. Our proof-of-concept implementation demonstrates this success by leveraging these modeled perceptual factors to accurately infer depicted pose from material frame features. This indicates that our topology-based perceptual modeling approach captures relationships between frames’ material and depicted features, as intended. Furthermore, RFQ GENEOS sheaves’ perceptually-based data analytical metrics and tools enable not only modeling, but intentional exploitation and modulation of these perceptual factors, supporting perceptually-based RFQ analysis and, in future, reference frame retrieval.

Our novel theory of heterogeneous multiresolution equivariance and novel heterogeneous data fusion algorithm form the bridge between frames’ material and depicted features within our RFQ GENEOS sheaf framework. While restriction GENEOS extract local physical pose information symbolized by each material perceptual cue, harmonic extension fuses this local information together into a globally cohesive understanding of depicted pose. This process was essential to our Experiment 2 (Section 3.2) findings, underpinning both pose perception subjectivity and perceptual cue (in)consistency’s impact on pose perception. For each frame’s material features, our data fusion algorithm is designed to find a pose interpretation optimally satisfying both material features’ symbolic meanings and humans’ innate understandings of volumetric form, functional relationships between body parts, and spatial reasoning informing occlusion and self-contact perception. Concurrently, heterogeneous multiresolution equivariance ensures that relationships between frames’ material features – specifically, those Experiment 1 (Section 3.1) found salient to RFQ perception – faithfully translate into corresponding relationships between frames’ pose estimates, enabling pose similarity to efficiently approximate material feature similarity

during RFQ analysis. Heterogeneous multiresolution equivariance thus establishes a stable relationship between material and depicted features; in practice, this stability underpins our perception of hand-drawn characters as having ground-truth physical representations, despite no such representations existing. As discussed throughout Section 4.6, unifying heterogeneous data fusion and heterogeneous multiresolution equivariance supports deep investigation of further connections between material and depicted features within our RFQ GENEO sheaf framework.

Chapter 5

Conclusions

This thesis approaches a single problem – the stagnation of technological advancements within the hand-drawn animation industry – from three distinct yet interconnected perspectives: methodological, psychological, and mathematical. The first investigates the present state of hand-drawn animation technology research, seeking explanations for hand-drawn studios’ frequent disinterest in the field’s technological outputs. The second studies perceptual factors influencing professional hand-drawn animators’ artistic and technical workflows, aiming to improve hand-drawn animation technology researchers’ understandings of animators’ practical needs. The third constructs a novel data analytical framework, building upon successful hand-drawn animation technology research approaches – specifically, those strategically deploying advanced topological ideas – to computationally modeling our perceptual findings, tailoring our framework to industry needs. The latter two perspectives concern a particular research objective: understanding and systematically modeling animators’ perception of reference frame quality, supporting our ultimate aim of building a reference frame retrieval system for hand-drawn studios. This chapter overviews our contributions, findings, and directions for future work.

5.1 Contributions

As detailed throughout our review of the field (Section 2.2), hand-drawn animation technology research has long understood hand-drawn frames as topological spaces. Yet researchers primarily utilize basic topological constructs, which offer limited conceptual and computational expressivity when processing hand-drawn frames. While TDA is dramatically improving the computability of mathematically sophisticated and conceptually expressive topological ideas, these advancements have yet to permeate hand-drawn animation technology research. Moreover, despite topology’s immense popularity and success in hand-drawn frame processing, prior work has not thoroughly addressed why connections between topology and hand-drawn animation exist, what can be learned from them, and how they can be strategically cultivated to allay challenges plaguing the field. This thesis takes initial steps towards addressing these knowledge gaps, investigating and interconnecting artistic, mathematical, perceptual, and computational dimensions of relationships between hand-drawn animation and topology.

By surveying hand-drawn animation technology through the novel, dual lens of its topological foundations alongside artistic and practical insights from industry experts, we find that present research principally focuses on hand-drawn frames’ material stroke/region features without systematically relating them to characters’ depicted volumetric pose features. Throughout Chapter 2, we discuss the process whereby characters’ depicted body topology/geometry dictates their articulate motion over hand-drawn

frames, inducing material feature changes. In turn, material changes are perceived as depicting volumetric articulate body motion. Without automatic strategies for relating frames’ depicted and material features, frame processing methods cannot anticipate perceptually coherent material feature changes, nor interpret volumetric meanings of complex material feature relationships. Instead, many methods require frequent and meticulous user interventions which may negatively impact animators’ artistry, studio workflows, production schedules, and budgets. For instance, Mark Mullery describes how performing manual stroke correspondence and depth ordering could deplete animators’ energy for artistic tasks, which may deter industry uptake of softwares involving such interventions (Appendix B). Yet methods employing creative or advanced topological techniques often more effectively relate frames’ material and depicted features, outperforming naive topological methods. Such techniques emulate aspects of human vision likely vital to hand-drawn frame perception.

Therefore, to develop topological frame processing solutions tailored to industry needs, perceptual mechanisms underpinning professional animators’ artistic processes must be well-understood. Though line drawing perception and professional artists’ perception have been rigorously studied, we contribute the first controlled experiments investigating perceptual factors practically impacting professional animators’ visual understandings of their own artistic work. We focus specifically on animators’ RFQ perception, with our findings guiding design of our RFQ analysis framework. From our joint topological and industry-based understanding of hand-drawn frames, we propose a novel set of perceptual cues – geometric, topological, and motion – for systematic frame classification and comparison within our experiments. Our experimental findings substantiate these cues’ perceptual saliency, dovetailing with our state-of-the-art review findings. Experiment 1 (Section 3.1) finds that animators’ RFQ perception is moderately invariant to local, body part-wise isometry – likely generalizing to multiresolution isometry invariance on collections of body parts – indicating that material feature preservation positively influences animators’ RFQ perception. Experiment 2 (Section 3.2) finds evidence that animators’ perceptions of frames’ depicted poses grow more subjective as perceptual cue consistency decreases.

Together, our findings illuminate several insights into animators’ hand-drawn frame perception. First, relationships between frames’ material and depicted features are challenging even for animators, as frame perception experts, to discern. This explains why isometry – which uniquely preserves director-approved pose resolution strategies – predicts RFQ more reliably than pose similarity. Animators thus collaboratively craft and learn relationships between material and depicted frame features throughout production, rather innately understanding such relationships by virtue of their expertise. Yet these relationships’ perceptual clarity depends upon the consistency of pose information conveyed by related cues, with inconsistencies degrading animators’ pose perception uniformity. Understanding animators’ frame perception thus involves understanding how depicted volumetric pose interpretations form via multiresolution integration of visual information conveyed by local material features.

Our insights into hand-drawn animation technology research and professional animators’ frame perception culminate in our novel topological model of RFQ perception, designed to underpin a future reference frame retrieval system for hand-drawn studios. We contribute RFQ GENEIO sheaves (Section 4.6), a new TDA frame processing framework advancing hand-drawn animation technology’s burgeoning trends towards TDA-based perceptual modeling. Unlike DNNs, RFQ GENEIO sheaves explainably relate frames’ material and depicted features through known tenets of visual perception, leveraging our proposed perceptual cues to estimate hand-drawn characters’ depicted 3D body poses without any prior training. RFQ GENEIO sheaves’ data analysis and comparison tools computationally model our perceptual experiment findings, including multiresolution isometry-invariant frame comparison, material and depicted feature relationship awareness, perceptual cue (in)consistency detection, functional part-wise relationship

comprehension, and pose perception subjectivity. In practice, these capabilities will provide numerous advantages over DNN-based frame processing, including usability early in production, pose estimation for both partial-body and full-body frames, pose resolution strategy retrieval up to multiresolution isometry, and fluent accommodation of Principles of Animation. This will support development of our RFQ GENEО sheaf-based reference frame retrieval system tailored to hand-drawn studios’ artistic and technical needs.

Theoretical contributions Rather than uniquely applying to RFQ analysis, GENEО sheaves contribute a novel, general-purpose TDA framework for complex system analysis across interlinked mathematical and computational dimensions (Chapter 4). GENEО sheaves unify the increasingly prominent yet largely separate TDA subfields of heterogeneous data fusion and equivariant data analysis. The former synthesizes independent yet interrelated heterogeneous subsystem data representations into globally valid system states, while the latter models regularities, equivalences, and relationships between various system states as group symmetries preserved across data analysis transformations. Prior to GENEО sheaves, Laplacian-based heterogeneous data fusion was mathematically formalized only for vector-valued subsystem data and linear subsystem relationships, while equivariant analysis frameworks primarily modeled homogeneous global symmetries acting uniformly across a system. Improving upon previous frameworks’ expressivity, GENEО sheaves support Laplacian-based heterogeneous data fusion over function-valued subsystem data and nonlinear subsystem relationships. They also model heterogeneous multiresolution symmetries acting independently across heterogeneous (collections of) subsystems, while ensuring that independent subsystem behavior preserves global system state validity.

These features underpin our contribution of heterogeneous multiresolution equivariance, a novel mathematical and data analytical theory. Our theory enables heterogeneous subsystems’ local behavior and system-wide influence to be modeled, modulated, compared, and tracked via diverse, explainable, and controllable GENEО sheaf metrics and analytical tools. GENEО sheaves thus offer advanced system modeling capabilities within a unified mathematical and computational framework. Most essentially, our theory guarantees that GENEО sheaf data fusion is equivariant to heterogeneous multiresolution symmetries on (collections of) subsystems, ensuring real-world system behavior, characteristics, and complex internal dynamics faithfully translate into analysis results. Given sheaves’ and GENEОs success in myriad domains including sensor integration, protein analysis, and knowledge modeling, we expect GENEО sheaves and heterogeneous multiresolution equivariance to have many exciting applications across diverse research fields.

Methodological contributions This thesis introduces multiple new methodological approaches to understanding hand-drawn animation technology. In our state-of-the-art review, we contribute a novel analytical lens through which to view the subject, focusing jointly on the field’s topological foundations and insights from industry experts. By identifying and evaluating methods’ underlying mathematical ideas – rather than simply their algorithmic contributions and practical functionalities – we directly connect topological frame representation and analysis techniques to their artistic advantages and limitations. To contextualize these insights within hand-drawn animators’ and studios’ artistic and technical workflows, we conduct interviews with hand-drawn animation industry experts Glen Keane and Mark Mullery. These interviews (Appendix B) explore practical, philosophical, artistic, and technological barriers to industry adoption of hand-drawn animation technologies, alongside desirable traits of such technologies. By synthesizing topological, artistic, and industry perspectives on hand-drawn animation technology, we diagnose impediments to industry uptake, gain novel insights into successful research trends, and provide actionable, sound advice for future work. As many of the field’s challenges stem from disconnects between technology researchers and industry experts, we advise researchers to work closely with hand-drawn

animation industry professionals when developing prospective technologies.

In our perceptual experiments, we contribute a novel experimental design methodology optimally emulating animators’ working studio environment. This methodology supports our controlled evaluations of perceptual factors specific to practical tasks that animators encounter during hand-drawn animated film production. Our approach deviates from prior work, which studies professional artists’ perceptual abilities on abstract visual tasks within generic experimental settings. While valuable in their own right, such studies do not necessarily illuminate perceptual factors influencing animators’ visual understandings of their own artistic work subject to their professional duties. Modeling working environments within perceptual experiment design is essential for understanding how best to develop technologies for collaborative artistic workplaces like hand-drawn animation studios. For instance, Experiment 1 instructs animators to imagine that they can improve frames’ RFQ via image manipulations in their drawing software. This ensures that our RFQ GENE0 sheaf metrics comport with animators’ industry-based artistic workflows, rather than with their generic perceptual abilities. Likewise, Experiment 2 asks animators to record their pose perceptions on digital versions of the posable wooden maquettes they commonly reference when drawing frames. By recording pose perceptions on generic body structures already familiar to professional animators, we gain insight into animators’ (in)ability to perceptually translate poses between posable maquettes and character-specific body geometries as they work.

Material contributions Our primary material contribution is our proof-of-concept RFQ GENE0 sheaf implementation, substantiating a number of key claims within this thesis. Our implementation’s efficacy validates our overarching research approach of topologically-based perceptual modeling (Section 4.6.2). We demonstrate RFQ GENE0 sheaves’ ability to explainably model critical facets of human vision, supporting our conceptual RFQ GENE0 sheaf model of hand-drawn frame perception (Section 4.6.1) and departing significantly from DNN-based models of visual cognition. Our results indicate that RFQ GENE0 sheaves emulate animators’ local isometry-invariant RFQ perception as found in Experiment 1 (Section 3.1), increasing the likelihood that RFQ GENE0 sheaves will provide a suitable foundation for our reference frame retrieval system. More generally, our implementation computationally verifies our novel Laplacian-based GENE0 sheaf heterogeneous data fusion algorithm and its associated theorems. It also provides evidence supporting the validity of our key heterogeneous multiresolution equivariance theorems (Figure 4.13). By computationally validating core elements of GENE0 sheaves’ mathematical theory, we gain confidence in their data analytical efficacy within and far beyond hand-drawn animation technology. Our code is owned by Cartoon Saloon, and can be made available upon request.

We also contribute a proprietary database of 649 raw clean-up animation frames from Cartoon Saloon’s feature film *Wolfwalkers*, from which all perceptual experiment frame data presented in this thesis was sourced. The database was generated via custom compositing software, which automatically obtained, prepped, and edited frames, rendering them suitable for use in perceptual and computer vision experiments. All frames within the database were manually selected to meet experimental needs, exclusively depicting high-quality full-body or partial-body front-view poses (i.e., no close-up shots), exactly one character per frame, and no character-prop interactions. This database has been made available to other Trinity College Dublin students for academic research purposes, with permission from Cartoon Saloon.

5.2 Future Work & Limitations

Hand-drawn animation technology To complete our review of hand-drawn animation technology, we overview the many exciting avenues for future work, informed by guidance from industry experts

and our topological perspective on the field. First and foremost, we recommend that researchers and hand-drawn studios pursue technological advancements in collaboration with one another. This will greatly benefit researchers’ understandings of industry needs and ultimately greatly improve industry adoption of research outputs. Researchers should also recognize that, as keepers of a highly traditional art form rooted in human craftsmanship, hand-drawn studios are often both excited about and weary of artistic technologies, with the latter sentiment growing in the dawning era of generative AI. Direct collaboration and communication between researchers and studios will likely cultivate mutual understanding between the two parties, making industry professionals more receptive to researchers’ work. Increased in-house technological research and development can also tremendously benefit hand-drawn studios, following from successful examples set by the 3D animation industry.

Moreover, we recommend that researchers seeking industry adoption of their technological outputs shift their focus away from automating artistic tasks, and instead pursue computer vision research on hand-drawn frames. According to both Keane and Mullery, artistic task automation is often antithetical to hand-drawn studios’ philosophical, artistic, and business objectives. Meanwhile, computer vision research has enormous potential to alleviate studios’ data management challenges and extend animators’ artistic capabilities in novel and desirable ways. Many hand-drawn studios purchase specialized softwares and employ computer programmers for production data management, indicating that they may be eager to adopt computer vision-based data management technologies. Mullery emphasizes the need to technologically support animators as artists rather than machinists, suggesting potential benefits of technologies ‘seeing’ frames as animators do. Keane speaks at length about his long-held wish for high-quality sculptural drawing technologies, which also will likely benefit from advanced computer vision. This dovetails with our assessment of the field’s most successful research trends, all of which involve employing sophisticated or creative topological techniques – which happen to emulate key aspects of visual cognition – to relate frames’ material and depicted features.

We expect that the most effective advancements along this path of inquiry will arise from incorporating TDA into hand-drawn animation technology. As discussed, TDA can extend many existing hand-drawn animation techniques to improve multiresolution data integration, part-wise relationship comprehension, abstract visual reasoning, and more. Specifically, the many stroke/region deformation and correspondence methods attaching local geometric data to stroke complexes and RAGs are formalized by sheaf theory [292, 198, 344, 346]. Sheaf theory, combinatorial Laplacians on higher-dimensional cell complexes, coarser topological features like medial axes, and other TDA-based relational learning strategies can imbue frame processing methods with the visual reasoning and multiresolution data evaluation capabilities they need to circumvent present failure cases (Figure 2.25) by relating material and depicted features [208, 24]. TDA also advances data diffusion and spectral analysis on both discrete and continuous topological spaces, potentially elevating sophistication and flexibility of these techniques in frame processing [336, 292, 198, 210]. de Rham-Hodge theory could extend boundary operator-based stroke/region relationship modeling [86, 87] to mesh-based depicted feature extraction methods [40], improving their topological awareness. Both de Rham-Hodge theory and combinatorial Laplacians could advance topologically flexible mesh processing [367] and lattice topology reconstruction under extreme deformation [128]. Higher-order feature relationships could also be modeled by extending present uses of boundary operators to higher dimensions, potentially relating RAG subgraphs in multiple frames to the same volumetric body part. Although TDA-based curve processing is still embryonic, its potential to improve non-isomorphic stroke correspondence handling should be investigated in future. DNN-based frame processing approaches also prove effective when combined with primarily topological techniques [225, 40, 39], encouraging further development of these strategies in concert with TDA’s many DNN-compatible approaches.

The primary limitation of our exploration of hand-drawn animation technology is our exclusive focus on methods processing uncolored line drawings of articulate animated characters, excluding highly relevant frame processing subfields including frame segmentation [239, 33, 350], frame vectorization [117, 62, 249], frame auto-coloring [305, 122, 250], and sketch clean-up [240, 247, 342]. Given the success of our novel methodological approach combining topological evaluation with industry-based insights, we recommend that broader realms of hand-drawn animation technology be evaluated via similar means in future.

Hand-drawn frame perception Both of our perceptual experiments were limited by their small numbers of experiment frames, small numbers of participants from the same animation studio, and use of a single character and artistic style. Expanding our existing experiment design to include more frames, participants, characters, and art styles will address most of these limitations, and will be the primary goal of future work. Moreover, frames’ perceptual cues were classified by eye in both experiments, rather than by empirical metrics. Such metrics can be used in future studies, although their incongruities to human vision may beget further limitations. A more reliable alternative involves animators manually rating shape similarity between frames’ corresponding regions displayed in isolation (Experiment 1) and perceptual cue consistency for each body part depicted in each frame (Experiment 2). RFQ ratings and 3D pose estimates, respectively, can then be empirically evaluated subject to these ratings. However, it will be highly challenging to collect animators’ perceptual cue consistency ratings within a controlled experimental setting.

Experiment 1 only tested local isometry across arm poses, so future studies involving isometries on varied (combinations of) body parts are needed to establish whether animators’ RFQ perception is fully multiresolution isometry-invariant. While it was not possible for our participants to draw target frames during the experiment, future work might explore this option. Our experiment was designed for body parts with left-right symmetry, and could be reformulated for non-symmetric character designs in future. Participants reported that reference frames’ visual clarity, character design correctness, and pose resolution strategy preservation strongly influenced their RFQ ratings, so these factors should be studied in future work. Further experiments can also compare animators’ and laypeople’s RFQ perception or provide more reference frames per question to supplement available visual information. For our own research objectives, it will be important to compare how RFQ ratings are influenced by depicted pose vs. material feature changes. Such insights can guide construction of specialized GENE sheaf frame comparison pseudometrics in our future reference frame retrieval system.

Because Experiment 2 ran during studio working hours and animators’ participation time windows were limited, we could not collect body part-wise Distortion and Difficulty ratings. In future, we may benefit from rerunning this experiment to collect these ratings, enabling formal evaluation of our perceptual cue consistency theory. To ensure that high-quality pose estimates are obtained for all experiment frames, future studies should be conducted over several weeks, with only 1-3 frames tested per day. Written responses should also be collected from animators in future, reporting specific factors determining their perceived posing task difficulty for each frame. Another viable line of inquiry can examine whether professional animators have particular facility in interpreting frame Distortion compared to laypeople. This would be especially interesting to investigate with child participants, as children are often the primary audience for hand-drawn animated films, and children’s perception of line drawings is an active research area [92, 203, 200]. Although our findings agree with prior work on line drawing perception, hand-drawn frames’ articulation, stylization, distortion, and integration of both visual symbolism and realism sets them apart from typical line drawings. Accordingly, future hand-drawn frame perception studies on perceptual cue consistency should be carefully designed, potentially requiring alternate, industry-specific

methodologies to those used in prior line drawing perception experiments.

GENEO sheaves GENEIO sheaves and heterogeneous multiresolution equivariance offer exciting, diverse, and far-reaching avenues for future work. Mathematically, GENEIO sheaves’ greatest limitation is their lack of a true cohomology theory, without which it is impossible to algebraically classify and compute core properties of the given data. Although prior work on sheaves of pseudometric spaces [271, 272] underpins GENEIO sheaves’ obstruction detection capabilities in the absence of cohomology, this line of inquiry is underdeveloped and presents exciting opportunities for mathematical and computational maturation. Specifically, existing work on the theory of linear GENEIOs for vector-valued data [29] may facilitate links between GENEIO sheaves and cellular sheaf cohomology theory. Cohomology theory is also intimately connected with the cellular sheaf Laplacian [154, 150], which, unlike the GENEIO sheaf Laplacian, has a matrix presentation. This endows cellular sheaves with a profound and computable spectral theory, which offers substantial data analytical advantages [153]. Future work should therefore investigate whether any of this analytical machinery can be transferred to GENEIO sheaf Laplacians. A complete proof of Conjecture 4.3.2 is a critical element of future work, and may reveal whether the GENEIO sheaf Laplacian or another Laplacian-based GENEIO sheaf data fusion operator exhibits mathematical properties related to sheaf cohomology. Representation theory, in which group symmetries are encoded as matrices [42], also presents a promising infrastructure for fluently translating cellular sheaves’ cohomology and spectral theories to GENEIO sheaves.

Cellular sheaf Laplacians [154, 355] and, more generally, combinatorial Laplacians [220, 329] have myriad formulations for specialized data analysis objectives, all of which should be explored subject to the GENEIO sheaf Laplacian. Beyond harmonic extension, sheaf Laplacian diffusion dynamics have many expressive and flexible incarnations controlling modes of system-wide data passage and subsystem influence [154], which can be adapted for GENEIO sheaf Laplacians in future. To enable higher-order subsystem relationship modeling, sheaf Laplacians, combinatorial Laplacians, and persistent Laplacians have all been generalized beyond graphs and applied to higher-dimensional cell complexes [173, 246], paving the way for similar generalizations of GENEIO sheaf Laplacians. Subsystem modeling expressivity and control might also be bolstered by combining GENEIO sheaves with other adaptations of GENEIO theory, including set equivariance [20] and partial equivariance [120]. A particularly exciting research direction is further theoretical development of GENEIO sheaf morphisms between different GENEIO sheaves, which has the potential to unite sheaves’ and GENEIOs’ respective category theoretic formulations [151, 20] and will likely dovetail with emerging research on GENEIOs between non-isomorphic graphs [2]. If successful, this line of inquiry will enable RFQ GENEIO sheaf pose estimates to map between characters with different ground-truth region complexes, possibly also opening new insights into relationships between hand-drawn frames’ material and depicted features.

In future, heterogeneous multiresolution equivariance will be explored from group theoretic and representation theoretic perspectives, prospectively facilitating development of mathematical infrastructure, analytical tools, and conceptual insights. In particular, we aim to further study and expand upon stalk symmetry filtrations and other mechanisms of subsystem influence tracking, perhaps in concert with particular types of GENEIO sheaf Laplacian diffusion dynamics. These investigations may reveal group theoretic system analysis opportunities beyond stalk symmetries and permutations, potentially bolstering GENEIO sheaves’ system modeling capabilities. Given heterogeneous multiresolution equivariance’s potent links to GENEIO sheaf morphisms, we anticipate that rich and illuminating connections may arise between category theoretic, group theoretic, and representation theoretic GENEIO sheaf formulations.

Although this thesis does not explore learning-based GENEIO sheaf formulations, both sheaves

[13, 32, 17, 96, 152] and GENEOS [20, 28, 186, 30] are highly compatible with DNNs. Consequently, combining GENEOS sheaves with DNNs is a highly exciting prospect for future work, especially given the increasing interest in TDA-based relational learning [246]. Given the myriad applications of sheaves’ and GENEOS’ respective learning-based formulations (Section 4.7), we expect that GENEOS sheaf-based DNNs will be similarly applicable to many diverse domains. DNNs may also improve GENEOS sheaves’ computational efficiency, which is another limitation of our proof-of-concept implementation. Given GENEOS-based DNNs’ state-of-the-art performance with minimal training data [28, 186], GENEOS sheaf-based DNNs may enable high-accuracy hand-drawn character pose estimation with minimal training frames. Alongside yielding DNN-based frame analysis usable early in animation production, this prospect extends frame processing methods’ recent success at integrating DNNs into primarily topological techniques [225, 40, 39]. Like future manifestations of RFQ GENEOS sheaf-based DNNs, these approaches leverage DNNs’ significant computational advantages while ensuring that core aspects of human vision are topologically emulated during hand-drawn frame analysis.

Reference frame retrieval Although this thesis takes crucial steps towards developing a reference frame retrieval system for hand-drawn studios, our work comprises only a proof-of-concept RFQ GENEOS sheaf implementation. Our present implementation is designed to process synthetic arm regions, and will require significant future development to estimate the entire pose depicted in a raw clean-up frame. Though our results are highly promising and computationally validate our key GENEOS sheaf theorems, our implementation was too small to test our entire suite of GENEOS sheaf theoretical results and analytical tools. Consequently, we were unable to verify every aspect of our RFQ GENEOS sheaf frame perception model. Verification will be completed in future after further development of our RFQ GENEOS sheaf implementation. Moreover, the perceptual quality of our implementation results was not empirically tested. In future, we aim to evaluate result quality through another perceptual study on Cartoon Saloon animators. While our implementation fails when the arm is highly foreshortened, this is not a systemic issue; it is caused by our current restriction GENEOS formulations. Improved restriction GENEOS will be devised in future. Though we tested sufficiently many heterogeneous data representations and restriction GENEOS to verify our framework’s efficacy, several other restriction GENEOS will need to be developed in future to support complete pose estimation for each frame. These will likely include medial axis-based region shape analysis, superficial region adjacency comprehension (e.g., for body part self-contacts and occlusions), and relative region size analysis for depth comprehension.

Although we believe that RFQ GENEOS sheaf construction can eventually be partially automated, our present implementation requires manual specification of body part-wise restriction GENEOS for each model sheet. In future, we aim to pursue automated construction by automatically assigning restriction GENEOS from a basic function library – including those described above – to model sheets using model sheet skeleton annotations, RAG comprehension, and GENEOS sheaf-based DNN computer vision. To fully automate RFQ GENEOS sheaf pose estimation, we must develop a body part identification framework in future. Given the success of Human-Art-trained HRNet pose estimation [174] on uncolored clean-up frames, we aim to combine this method with region shape recognition and sheaf-based RAG analysis to automatically identify each frame’s body parts. Alternatively, unifying GENEOS sheaves and DNNs may support an even more elegant solution, which can be trained on Human-Art to ensure robust performance. To complete our reference frame retrieval system, we will need to develop a user interface, frame database storage, and other software elements. These will be pursued in collaboration with Cartoon Saloon over the coming years. As we develop our system, we will continue to solicit input from Cartoon Saloon animators and technical personnel. We will also test our prototypes in real Cartoon Saloon production environments throughout the software development process.

5.3 Conclusions

This thesis endeavors to deeply investigate and interconnect disparate facets of hand-drawn animation technology, presenting novel perspectives on a complex field intersecting technological advancement and a resolutely human art form. Tension between technology and artistry is palpable throughout the research field and within hand-drawn studios, albeit manifesting in distinct ways. Hand-drawn animation technology research often pursues algorithmic solutions for artistic tasks, sometimes improving efficiency, but other times requiring technical interventions which animators may be ill-equipped or unwilling to undertake. Meanwhile, animators are weary of technologies potentially degrading their artistic agency – thereby corrupting the humanity at the core of their work’s appeal – yet eager for technologies enhancing their artistic capabilities or alleviating the burdens of their non-artistic tasks. This thesis provides research foundational to automating the non-artistic task of reference frame retrieval, approaching the problem from artistic and theoretical – rather than algorithmic – perspectives.

Our research seeks to understand the artistic process of hand-drawn animation itself as a formal complex system, positioning animators and the perceptual mechanisms undergirding their artistry at the heart of our work. While this perspective is largely absent from prior hand-drawn animation technology research, our work suggests its potential to benefit and advance the field. Frame processing methods’ common topological foundations provide our empirical framework for assessing where, how, and why researchers’ and animators’ understandings of hand-drawn frames diverge. Ultimately, we find that naive topological approaches cannot emulate perceptual factors governing relationships between frames’ material and depicted features. Without computational knowledge of these relationships, methods often cannot fully interpret material frame features according to their perceptual meanings, and thus may struggle to anticipate animators’ perceptions of depicted frame features.

Our research focuses on these relationships, studying the unified system of human vision’s interactions with hand-drawn frames’ material features. We posit that material and depicted feature relationships implicitly arise from heterogeneous data dynamics and relationships within this system. This notion is motivated by humans’ perception of hand-drawn characters as coherent physical entities, despite their lack of ground-truth body representations. Thus, volumetric pose interpretation is an emergent property of material frame features’ interactions with the human visual system. Moreover, since hand-drawn characters’ bodies have no ground-truth volumetric representations, relationships between frames’ material and depicted features cannot be established directly. Instead, this thesis pursues a system model of frame perception where volumetric pose arises as an emergent, rather than explicit, feature of hand-drawn frames. Despite our research’s limitations, our system-oriented approach to hand-drawn animation technology yields highly promising results and many compelling prospects for future work. Our novel frame analysis framework, RFQ GENEEO sheaves, faithfully, expressively, and controllably models every aspect of frame perception we sought to capture, including those discovered through our own perceptual experiments.

The success of our approach has exciting implications for the future of hand-drawn animation technology research, particularly research seeking industry adoption. Building off the connections between hand-drawn animation technology, topology, and visual perception detailed in this thesis, we make the following recommendations for future work. First, researchers should collaborate directly with professional animators and industry experts to identify prospective technologies desirable to hand-drawn studios. Strategies for integrating prospective technologies into existing studio pipelines and animators’ artistic workflows should also be established in advance. Researchers should then carefully analyze the problem at hand, identifying material and depicted frame features and feature relationships pertinent their desired software functionality. Next, researchers should identify perceptual factors and cognitive mechanisms

influencing perception of these features and relationships. This process may benefit from further perceptual studies on professional animators following our novel methodology modeling animators’ studio workflows and artistic tasks relevant to the given prospective technology. Finally, researchers should identify TDA frameworks faithfully modeling the perceptual factors and processes undergirding the frame feature relationships relevant to their desired software functionality. While some frame processing methods have employed similar approaches [169, 105], our work illustrates the benefits of advancing the mathematical and perceptual rigor of this research strategy.

As demonstrated in our research, this approach can enable analysis of depicted frame features as emergent properties of the resulting computational framework. This circumvents the need to represent depicted features explicitly, potentially improving functionality and consequently encouraging industry adoption. While this thesis takes a high-level and intensive approach to this procedure, it can also be applied to many simple frame processing tasks popular throughout the field. As discussed in Sections 2.6 and 5.2, many TDA solutions are ideally poised to extend existing methods’ capabilities and ameliorate their limitations without profound mathematical or perceptual groundwork.

To contextualize this thesis within the current landscape of hand-drawn animation technology, we compare our research approach to recent, rapid advancements in AI-based animation technology. RFQ GENEIO sheaves are obstinately a data analysis technology – rather than a data generation technology – which is precisely why Cartoon Saloon invested in this research. According to Cartoon Saloon technical personnel, data analysis technologies automating production data organization, tracking, and access tasks – like reference frame retrieval – are desirable because they perform menial, non-artistic duties, thereby maximizing studios’ available time and energy for artistic work. Contrastingly, Generative AI’s (Gen-AI) approach to hand-drawn animation technology minimizes time spent on artistic work by automating artistic tasks. This contradicts hand-drawn studios’ core ethos, whereby hand-drawn animation’s hand-crafted quality is both the means and the end of the filmmaking process. Many hand-drawn animators feel that delegating artistic work to a machine fundamentally defeats the purpose of making hand-drawn films. To Mullery, Gen-AI’s failure to support human artists extends beyond research contexts, and is beginning to compromise the value of legacy artistic technology companies like Adobe and ToonBoom. “We still paint our backgrounds physically, most of the time, because Photoshop is not yet good enough to produce the effects that paint, markers, and pencils have,” he says. By pivoting towards Gen-AI, these companies are neglecting functionalities most desirable for hand-drawn studios, while investing in functionalities antithetical to hand-drawn animation’s guiding principles.

Not only does Gen-AI largely fail to benefit hand-drawn studios, it may actively harm hand-drawn animation over time. As discussed throughout this thesis, novice animators often hone their skills on the low-level artistic tasks which Gen-AI primarily seeks to automate, including in-betweening, Ink & Paint, and shading. Hand-drawn studios may soon experience financial pressure to automate these tasks using Gen-AI, limiting their ability to cultivate robust internal talent pools. Moreover, Gen-AI’s trendiness could soon lead film funders to neglect hand-drawn studios in favor of Gen-AI hand-drawn-style productions. Cartoon Saloon Technical Directors Helga Bjarnadóttir, Fergal Brennan, and Anne Bolger express serious concern that, even if these changes are only temporary, they could rapidly erode the industry’s talent pool, triggering closure of studios and animation schools. Mullery also worries that Gen-AI will exponentially exacerbate challenges originally born from hand-drawn animation’s transition from drawing on paper to drawing on computers. In particular, digitizing animation made it easier to store and modify numerous versions of each shot and design asset. Often, this leads to choice paralysis for directors and supervisors, slowing and massively over-complicating the filmmaking process. Gen-AI will likely intensify this phenomenon, and may even create new issues in this vein.

The potential consequences of integrating Gen-AI into the hand-drawn animation industry beg a critical and seldom-asked question: what is the value of technological innovation that fails to address – or even actively contradicts – human needs? Computers are ultimately tools made by and for humans. They are worthy of study as self-contained entities, but this does not mean that every resulting innovation is necessarily fit for use across all realms of human life. Of course, transformational technical innovation will always subsume the version of human civilization prior to its invention. But this does not absolve researchers from the responsibility of interrogating how the technology they create best fits into the world around them, focusing specifically on the intended human users’ needs. Our work showcases how such interrogation can create exciting opportunities for technological innovation, even within an artistic, human-driven, and technologically-averse industry. Once AI for hand-drawn animation is considered through a user-focused lens, other cutting-edge technological opportunities become clear. Mullery, Bjarnadóttir, Brennan, and Bolger are strongly interested in AI-based character and asset design continuity tracking across scenes, asset position consistency checking across multiple view angles of a location, 3D modeling for previsualization and complex asset prototyping, automated retrieval of director-approved asset designs, and frame-by-frame visual error tracking as production data passes between artistic departments. AI-driven conversion of institutional production knowledge into a queryable internal search engine presents a particularly exciting prospect; studios have long attempted manual efforts towards this end without complete success. With careful, user-focused technical strategy and design, computer scientists will find many exciting opportunities for innovation within the hand-drawn animation industry.

To conclude, we reflect on an important point raised by our research about the future of TDA. While we initially endeavored to devise a system model of hand-drawn frame perception, we ultimately created a general-purpose data analytical framework addressing key limitations of heterogeneous data fusion and equivariant data analysis. GENE sheaves comprise a deep mathematical theory, effective computational framework, rich conceptual and philosophical motivation, and myriad potential applications across diverse data analytical domains. Although much future work is needed to thoroughly explore and establish the depth and breadth of these varied facets, it is notable that studying an extraordinarily niche computational problem – RFQ analysis for hand-drawn animation frames – yielded a promising general-purpose framework for complex system modeling. This may be a coincidence, or due to the intricacy of the system we initially aimed to model alongside our high-level mathematical approach. However, an alternative possibility exists: rigorous and novel TDA approaches to niche, complex applications may provide a highly effective starting point for exciting mathematical developments in future work. Much of TDA research devises computational frameworks for general topological constructs before pursuing specific applications. This thesis suggests that it may be equally effective to deeply study niche applications, identify relevant topological ideas, and develop novel TDA frameworks expressing them effectively. TDA’s success demonstrates the depth and diversity of potential connections between abstract mathematics and real-world systems, far surpassing those typically studied in applied mathematics and related disciplines. This thesis supports the possibility that these connections run both ways: not only can mathematics model deep truths of real-world phenomena, but creative, complex, highly-specialized real-world problems may in turn galvanize advancements in abstract mathematics.

Bibliography

- [1] M. Agrios. A very elementary introduction to sheaves. 2022. URL <https://arxiv.org/abs/2202.01379>.
- [2] F. Ahmad, M. Ferri, and P. Frosini. Generalized permutants and graph geneos. *Machine Learning and Knowledge Extraction*, 5(4):1905–1920, 2023. ISSN 2504-4990. doi:[10.3390/make5040092](https://doi.org/10.3390/make5040092).
- [3] M. Alcorn, Q. Li, Z. Gong, C. Wang, L. Mai, W.-S. Ku, and A. Nguyen. Strike (with) a pose: Neural networks are easily fooled by strange poses of familiar objects. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4840–4849, 06 2019. doi:[10.1109/CVPR.2019.00498](https://doi.org/10.1109/CVPR.2019.00498).
- [4] C. F. Altmann, H. H. Bülthoff, and Z. Kourtzi. Perceptual organization of local elements into global shapes in the human visual cortex. *Current Biology*, 13(4):342–349, 2003. ISSN 0960-9822. doi:[https://doi.org/10.1016/S0960-9822\(03\)00052-6](https://doi.org/10.1016/S0960-9822(03)00052-6).
- [5] L. Alzubaidi, J. Bai, A. Al-Sabaawi, J. Santamaría, A. Albahri, B. Al-dabbagh, M. Fadhel, M. Manoufali, J. Zhang, A. Al-Timemy, Y. Duan, A. Abdullah, L. Farhan, Y. Lu, A. Gupta, F. Albu, A. Abbosh, and Y. Gu. A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *Journal of Big Data*, 10, 04 2023. doi:[10.1186/s40537-023-00727-2](https://doi.org/10.1186/s40537-023-00727-2).
- [6] A. Andre and S. Saito. Single-view sketch based modeling. In *Proceedings of the Eighth Eurographics Symposium on Sketch-Based Interfaces and Modeling*, page 133–140, 2011. ISBN 9781450309066. doi:[10.1145/2021164.2021189](https://doi.org/10.1145/2021164.2021189).
- [7] S. Arya, J. Curry, and S. Mukherjee. A sheaf-theoretic construction of shape space. *Found. Comput. Math.*, 25(3):813–863, May 2024. ISSN 1615-3375. doi:[10.1007/s10208-024-09650-1](https://doi.org/10.1007/s10208-024-09650-1).
- [8] V. Ayzenberg, F. S. Kamps, D. D. Dilks, and S. F. Lourenco. Skeletal representations of shape in the human visual cortex. *Neuropsychologia*, 164:108092, 2022. ISSN 0028-3932. doi:[10.1016/j.neuropsychologia.2021.108092](https://doi.org/10.1016/j.neuropsychologia.2021.108092).
- [9] J. Bai, B. Gong, Y. Zhao, F. Lei, C. Yan, and Y. Gao. Multi-scale representation learning on hypergraph for 3d shape retrieval and recognition. *IEEE Transactions on Image Processing*, 30: 5327–5338, 2021. doi:[10.1109/TIP.2021.3082765](https://doi.org/10.1109/TIP.2021.3082765).
- [10] Y. Bai, D. M. Kaufman, C. K. Liu, and J. Popović. Artist-directed dynamics for 2d animation. *ACM Trans. Graph.*, 35(4), July 2016. ISSN 0730-0301. doi:[10.1145/2897824.2925884](https://doi.org/10.1145/2897824.2925884).
- [11] N. Baker, H. Lu, G. Erlikhman, and P. J. Kellman. Deep convolutional networks do not classify based on global object shape. *PLoS Comput. Biol.*, 14(12):e1006613, Dec. 2018. doi:[10.1371/journal.pcbi.1006613](https://doi.org/10.1371/journal.pcbi.1006613).
- [12] Q. Baldwin and E. Panagiotou. The local topological free energy of proteins. *Journal of Theoretical Biology*, 529:110854, 2021. ISSN 0022-5193. doi:[10.1016/j.jtbi.2021.110854](https://doi.org/10.1016/j.jtbi.2021.110854).

- [13] J. Bamberger, F. Barbero, X. Dong, and M. Bronstein. Bundle neural networks for message diffusion on graphs. 2024. URL <https://arxiv.org/abs/2405.15540>.
- [14] D. G. T. Barrett, F. Hill, A. Santoro, A. S. Morcos, and T. P. Lillicrap. Measuring abstract reasoning in neural networks. *CoRR*, 2018. URL <http://arxiv.org/abs/1807.04225>.
- [15] P. Battaglia, R. Pascanu, M. Lai, D. J. Rezende, and K. kavukcuoglu. Interaction networks for learning about objects, relations and physics. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, page 4509–4517, 2016. ISBN 9781510838819. doi:[10.5555/3157382.3157601](https://doi.org/10.5555/3157382.3157601).
- [16] C. Battiloro, L. Testa, L. Giusti, S. Sardellitti, P. D. Lorenzo, and S. Barbarossa. Generalized simplicial attention neural networks. *IEEE Transactions on Signal and Information Processing over Networks*, 10:833–850, 2024. doi:[10.1109/TSIPN.2024.3485473](https://doi.org/10.1109/TSIPN.2024.3485473).
- [17] C. Battiloro, Z. Wang, H. Riess, P. Di Lorenzo, and A. Ribeiro. Tangent bundle convolutional learning: From manifolds to cellular sheaves and back. *IEEE Transactions on Signal Processing*, 72: 1892–1909, 2024. doi:[10.1109/TSP.2024.3379862](https://doi.org/10.1109/TSP.2024.3379862).
- [18] C. Battiloro, E. Karaismailoğlu, M. Tec, G. Dasoulas, M. Audirac, and F. Dominici. E(n) equivariant topological neural networks. 2025. URL <https://arxiv.org/abs/2405.15429>.
- [19] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri. Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, 874:1–92, 2020. ISSN 0370-1573. doi:[10.1016/j.physrep.2020.05.004](https://doi.org/10.1016/j.physrep.2020.05.004).
- [20] M. Bergomi, P. Frosini, D. Giorgi, and N. Quercioli. Towards a topological–geometrical theory of group equivariant non-expansive operators for data analysis and machine learning. *Nature Machine Intelligence*, 09 2019. doi:[10.1038/s42256-019-0087-3](https://doi.org/10.1038/s42256-019-0087-3).
- [21] N. Bermingham, V. Robins, and K. Turner. Planar symmetry detection and quantification using the extended persistent homology transform. 07 2023. doi:[10.48550/arXiv.2307.08281](https://doi.org/10.48550/arXiv.2307.08281).
- [22] A. Bernstein, E. Burnaev, M. Sharaev, O. Kachan, and E. Kondrateva. Topological data analysis in computer vision. In *Twelfth International Conference on Machine Vision (ICMV 2019)*, volume 11433, pages 673–679, 2020. doi:[10.1117/12.2562501](https://doi.org/10.1117/12.2562501).
- [23] T. Berry and T. Sauer. Consistent manifold representation for topological data analysis. *Foundations of Data Science*, 1(1):1–38, 2019. doi:[10.3934/fods.2019001](https://doi.org/10.3934/fods.2019001).
- [24] M. Bessmeltsev, W. Chang, N. Vining, A. Sheffer, and K. Singh. Modeling character canvases from cartoon drawings. *ACM Trans. Graph.*, 34(5), Nov. 2015. ISSN 0730-0301. doi:[10.1145/2801134](https://doi.org/10.1145/2801134).
- [25] M. Bessmeltsev, N. Vining, and A. Sheffer. Gesture3d: Posing 3d characters via gesture drawings. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH Asia 2016)*, 35(6), 2016. doi:[10.1145/2980179.2980240](https://doi.org/10.1145/2980179.2980240).
- [26] J. Bhattacharya and H. Petsche. Drawing on mind’s canvas: Differences in cortical integration patterns between artists and non-artists. *Human brain mapping*, 26:1–14, 09 2005. doi:[10.1002/hbm.20104](https://doi.org/10.1002/hbm.20104).
- [27] I. Biederman. Recognition-by-components: a theory of human image understanding. *Psychol. Rev.*, 94(2):115–147, Apr. 1987. doi:[10.1037/0033-295X.94.2.115](https://doi.org/10.1037/0033-295X.94.2.115).
- [28] G. Bocchi, P. Frosini, A. Micheletti, A. Pedretti, C. Gratterer, F. Lunghini, A. Beccari, and C. Talarico. Geneonet: A new machine learning paradigm based on group equivariant non-expansive operators. an application to protein pocket detection. 01 2022. doi:[10.48550/arXiv.2202.00451](https://doi.org/10.48550/arXiv.2202.00451).
- [29] G. Bocchi, S. Botteghi, M. Brasini, P. Frosini, and N. Quercioli. On the finite representation of

- linear group equivariant operators via permutant measures. *Annals of Mathematics and Artificial Intelligence*, 91(4):465–487, Aug. 2023. doi:[10.1007/s10472-022-09830-1](https://doi.org/10.1007/s10472-022-09830-1).
- [30] G. Bocchi, M. Ferri, and P. Frosini. A novel approach to graph distinction through GENEOS and permutants. *Scientific Reports*, 15(1):6259, Feb. 2025. doi:[10.1038/s41598-025-90152-7](https://doi.org/10.1038/s41598-025-90152-7).
 - [31] C. Bodnar. Topological deep learning: Graphs, complexes, sheaves. 2022. doi:[10.17863/CAM.97212](https://doi.org/10.17863/CAM.97212).
 - [32] C. Bodnar, F. Di Giovanni, B. P. Chamberlain, P. Liò, and M. Bronstein. Neural sheaf diffusion: a topological perspective on heterophily and oversmoothing in gnns. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, 2022. ISBN 9781713871088. doi:[10.5555/3600270.3601616](https://doi.org/10.5555/3600270.3601616).
 - [33] A. Bonnici and K. P. Camilleri. Exploiting artistic cues to obtain line labels for free-hand sketches. In *Proceedings of the International Symposium on Sketch-Based Interfaces and Modeling*, SBIM ’12, page 77–86, Goslar, DEU, 2012. Eurographics Association. ISBN 9783905674422. doi:[10.5555/2331067.2331079](https://doi.org/10.5555/2331067.2331079).
 - [34] A. Bonnici, A. Akman, G. Calleja, K. P. Camilleri, P. Fehling, A. Ferreira, F. Hermuth, J. H. Israel, T. Landwehr, J. Liu, and et al. Sketch-based interaction and modeling: where do we stand? *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 33(4):370–388, 2019. doi:[10.1017/S0890060419000349](https://doi.org/10.1017/S0890060419000349).
 - [35] P. Borosán, M. Jin, D. DeCarlo, Y. Gingold, and A. Nealen. Rigmesh: automatic rigging for part-based shape modeling and deformation. *ACM Trans. Graph.*, 31(6), Nov. 2012. ISSN 0730-0301. doi:[10.1145/2366145.2366217](https://doi.org/10.1145/2366145.2366217).
 - [36] M. Botnan and M. Lesnick. An introduction to multiparameter persistence. 03 2022. doi:[10.48550/arXiv.2203.14289](https://doi.org/10.48550/arXiv.2203.14289).
 - [37] J. S. Bowers, G. Malhotra, M. Dujmović, M. Llera Montero, C. Tsvetkov, V. Biscione, G. Puebla, F. Adolphi, J. E. Hummel, R. F. Heaton, and et al. Deep problems with neural network models of human vision. *Behavioral and Brain Sciences*, 46:e385, 2023. doi:[10.1017/S0140525X22002813](https://doi.org/10.1017/S0140525X22002813).
 - [38] W. Brendel and M. Bethge. Approximating cnns with bag-of-local-features models works surprisingly well on imagenet. In *Seventh International Conference on Learning Representations (ICLR 2019)*, 2019. doi:[10.48550/arXiv.1904.00760](https://doi.org/10.48550/arXiv.1904.00760).
 - [39] K. Brodt and M. Bessmeltsev. Sketch2pose: estimating a 3d character pose from a bitmap sketch. *ACM Trans. Graph.*, 41(4), July 2022. ISSN 0730-0301. doi:[10.1145/3528223.3530106](https://doi.org/10.1145/3528223.3530106).
 - [40] K. Brodt and M. Bessmeltsev. Skeleton-driven inbetweening of bitmap character drawings. *ACM Trans. Graph.*, 43(6), Nov. 2024. ISSN 0730-0301. doi:[10.1145/3687955](https://doi.org/10.1145/3687955).
 - [41] M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst. Geometric deep learning: Going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017. doi:[10.1109/MSP.2017.2693418](https://doi.org/10.1109/MSP.2017.2693418).
 - [42] M. Bronstein, J. Bruna, T. Cohen, and P. Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. 2021. URL <https://arxiv.org/abs/2104.13478>.
 - [43] P. Bubenik and A. Elchesen. Universality of persistence diagrams and the bottleneck and wasserstein distances. *Computational Geometry*, 105-106:101882, 2022. ISSN 0925-7721. doi:[10.1016/j.comgeo.2022.101882](https://doi.org/10.1016/j.comgeo.2022.101882).
 - [44] P. Buchanan, R. Mukundan, and M. Doggett. Automatic single-view character model reconstruction. In *Proceedings of the International Symposium on Sketch-Based Interfaces and Modeling*, SBIM ’13, page 5–14, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450322058.

doi:[10.1145/2487381.2487385](https://doi.org/10.1145/2487381.2487385).

- [45] J. Bullier. Integrated model of visual processing. *Brain Research Reviews*, 36(2):96–107, 2001. ISSN 0165-0173. doi:[10.1016/S0165-0173\(01\)00085-6](https://doi.org/10.1016/S0165-0173(01)00085-6).
- [46] B. N. Bushnell, P. J. Harding, Y. Kosai, and A. Pasupathy. Partial occlusion modulates contour-based shape encoding in primate area v4. *Journal of Neuroscience*, 31(11):4012–4024, 2011. ISSN 0270-6474. doi:[10.1523/JNEUROSCI.4766-10.2011](https://doi.org/10.1523/JNEUROSCI.4766-10.2011).
- [47] Y. Cai, Z. Wang, Z. Luo, B. Yin, A. Du, H. Wang, X. Zhang, X. Zhou, E. Zhou, and J. Sun. Learning delicate local representations for multi-person pose estimation. In *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III*, page 455–472, 2020. ISBN 978-3-030-58579-2. doi:[10.1007/978-3-030-58580-8_27](https://doi.org/10.1007/978-3-030-58580-8_27).
- [48] W. Cao, Z. Yan, Z. He, and Z. He. A comprehensive survey on geometric deep learning. *IEEE Access*, 8:35929–35949, 2020. doi:[10.1109/ACCESS.2020.2975067](https://doi.org/10.1109/ACCESS.2020.2975067).
- [49] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 43(01):172–186, Jan. 2021. ISSN 1939-3539. doi:[10.1109/TPAMI.2019.2929257](https://doi.org/10.1109/TPAMI.2019.2929257).
- [50] G. Carlsson. Topology and data. *Bulletin of The American Mathematical Society*, 46:255–308, 04 2009. doi:[10.1090/S0273-0979-09-01249-X](https://doi.org/10.1090/S0273-0979-09-01249-X).
- [51] L. Carvalho, R. Marroquim, and E. Vital Brazil. Dilight: Digital light table – inbetweening for 2d animations using guidelines. *Computers & Graphics*, 65:31–44, 2017. ISSN 0097-8493. doi:[10.1016/j.cag.2017.04.001](https://doi.org/10.1016/j.cag.2017.04.001).
- [52] E. G. Z. Centeno, G. Moreni, C. Vriend, L. Douw, and F. A. N. Santos. A hands-on tutorial on network and topological neuroscience. *Brain Structure and Function*, 227(3):741–762, Apr. 2022.
- [53] R. Chamberlain and J. Wagemans. Visual arts training is linked to flexible attention to local and global levels of visual stimuli. *Acta Psychologica*, 161:185–197, 2015. ISSN 0001-6918. doi:[10.1016/j.actpsy.2015.08.012](https://doi.org/10.1016/j.actpsy.2015.08.012).
- [54] R. Chamberlain, I. C. McManus, H. Riley, Q. Rankin, and N. Brunswick. Local processing enhancements associated with superior observational drawing are due to enhanced perceptual functioning, not weak central coherence. *The Quarterly Journal of Experimental Psychology*, 66(7):1448–1466, 2013. doi:[10.1080/17470218.2012.750678](https://doi.org/10.1080/17470218.2012.750678).
- [55] R. Chamberlain, J. Drake, A. Kozbelt, R. Hickman, J. Siev, and J. Wagemans. Artists as experts in visual cognition: An update. *Psychology of Aesthetics, Creativity, and the Arts*, 13:58–73, 03 2018. doi:[10.1037/aca0000156](https://doi.org/10.1037/aca0000156).
- [56] R. Chamberlain, A. Kozbelt, J. Drake, and J. Wagemans. Learning to see by learning to draw: A longitudinal analysis of the relationship between representational drawing training and visuospatial skill. *Psychology of Aesthetics Creativity and the Arts*, 04 2019. doi:[10.1037/aca0000243](https://doi.org/10.1037/aca0000243).
- [57] M. B. Chang. A compositional Object-Based approach to learning physical dynamics. In *5th International Conference on Learning Representations (ICLR 2017)*. Toulon, France, ICLR, 2017.
- [58] S. Chavlis and P. Poirazi. Dendrites endow artificial neural networks with accurate, robust and parameter-efficient learning. *Nature Communications*, 16, 01 2025. doi:[10.1038/s41467-025-56297-9](https://doi.org/10.1038/s41467-025-56297-9).
- [59] H. Chen, R. Feng, S. Wu, H. Xu, F. Zhou, and Z. Liu. 2d human pose estimation: A survey. 2022. URL <https://arxiv.org/abs/2204.07370>.
- [60] J. Chen, L. Zhang, X. Li, B. Zhang, and Z. Ye. Locally controlled as-rigid-as-possible deformation for 2d characters. *Computer Animation and Virtual Worlds*, 28:e1750, 03 2017. doi:[10.1002/cav.1750](https://doi.org/10.1002/cav.1750).

- [61] J. Chen, R. Zhao, Y. Tong, and G.-W. Wei. Evolutionary de rham-hodge method. *Discrete & Continuous Dynamical Systems - B*, 22, 01 2017. doi:[10.3934/dcdsb.2020257](https://doi.org/10.3934/dcdsb.2020257).
- [62] J. Chen, M. Du, X. Qin, and Y. Miao. An improved topology extraction approach for vectorization of sketchy line drawings. *The Visual Computer*, 34(12):1633–1644, Dec. 2018. doi:[10.1007/s00371-018-1549-z](https://doi.org/10.1007/s00371-018-1549-z).
- [63] J. Chen, Y. Qiu, R. Wang, and G.-W. Wei. Persistent laplacian projected omicron ba.4 and ba.5 to become new dominating variants. *Computers in Biology and Medicine*, 151:106262, 2022. ISSN 0010-4825. doi:[10.1016/j.compbiomed.2022.106262](https://doi.org/10.1016/j.compbiomed.2022.106262).
- [64] J. Chen, X. Zhu, M. Even, J. Basset, P. Bénard, and P. Barla. Efficient interpolation of rough line drawings. *Computer Graphics Forum*, 42(7), 2023. doi:[10.1111/cgf.14946](https://doi.org/10.1111/cgf.14946).
- [65] M. Chen, Y. Yan, X. Gong, C. D. Gilbert, H. Liang, and W. Li. Incremental integration of global contours through interplay between visual cortical areas. *Neuron*, 82(3):682–694, 2014. ISSN 0896-6273. doi:[10.1016/j.neuron.2014.03.023](https://doi.org/10.1016/j.neuron.2014.03.023).
- [66] S. Chen and M. Zwicker. Improving the perceptual quality of 2d animation interpolation. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVII*, page 271–287, Berlin, Heidelberg, 2022. Springer-Verlag. ISBN 978-3-031-19789-5. doi:[10.1007/978-3-031-19790-1_17](https://doi.org/10.1007/978-3-031-19790-1_17).
- [67] S. Chen and M. Zwicker. Transfer learning for pose estimation of illustrated characters. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1638–1647, 2022. doi:[10.1109/WACV51458.2022.00170](https://doi.org/10.1109/WACV51458.2022.00170).
- [68] S. Chen and M. Zwicker. Match-free inbetweening assistant (miba): A practical animation tool without user stroke correspondence. In *Computer Vision – ACCV 2024: 17th Asian Conference on Computer Vision, Hanoi, Vietnam, December 8–12, 2024, Proceedings, Part VI*, page 89–103, 2024. ISBN 978-981-96-0959-8. doi:[10.1007/978-981-96-0960-4_6](https://doi.org/10.1007/978-981-96-0960-4_6).
- [69] S. Chen, K. Zhang, Y. Shi, H. Wang, Y. Zhu, G. Song, S. An, J. Kristjansson, X. Yang, and M. Zwicker. Panic-3d: Stylized single-view 3d reconstruction from portraits of anime characters. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21068–21077, 06 2023. doi:[10.1109/CVPR52729.2023.02018](https://doi.org/10.1109/CVPR52729.2023.02018).
- [70] S.-Y. Chen, J.-Q. Zhang, L. Gao, Y. He, S. Xia, M. Shi, and F.-L. Zhang. Active colorization for cartoon line drawings. *IEEE Transactions on Visualization and Computer Graphics*, 28(2): 1198–1208, 2022. doi:[10.1109/TVCG.2020.3009949](https://doi.org/10.1109/TVCG.2020.3009949).
- [71] Z. Chen, J. Dai, J. Bai, and J. Pan. Dgformer: Dynamic graph transformer for 3d human pose estimation. *Pattern Recognition*, 152:110446, 2024. ISSN 0031-3203. doi:[10.1016/j.patcog.2024.110446](https://doi.org/10.1016/j.patcog.2024.110446).
- [72] Y. Cheng, B. Yang, B. Wang, Y. Wending, and R. Tan. Occlusion-aware networks for 3d human pose estimation in video. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 723–732, 2019. doi:[10.1109/ICCV.2019.00081](https://doi.org/10.1109/ICCV.2019.00081).
- [73] H. Choi, A. Pasupathy, and E. Shea-Brown. Predictive coding in area v4: Dynamic shape discrimination under partial occlusion. *Neural Computation*, 30(5):1209–1257, 05 2018. ISSN 0899-7667. doi:[10.1162/neco_a_01072](https://doi.org/10.1162/neco_a_01072).
- [74] X. Chu, W. Ouyang, H. Li, and X. Wang. Structured feature learning for pose estimation. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4715–4723, 2016. doi:[10.1109/CVPR.2016.510](https://doi.org/10.1109/CVPR.2016.510).
- [75] X. Chu, W. Yang, W. Ouyang, C. Ma, A. L. Yuille, and X. Wang. Multi-context attention for

- human pose estimation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5669–5678. IEEE Computer Society, July 2017. doi:[10.1109/CVPR.2017.601](https://doi.org/10.1109/CVPR.2017.601).
- [76] H. Ci, C. Wang, X. Ma, and Y. Wang. Optimizing network structure for 3d human pose estimation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2262–2271, 2019. doi:[10.1109/ICCV.2019.00235](https://doi.org/10.1109/ICCV.2019.00235).
- [77] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37:263–271, 06 2005. doi:[10.1007/s00454-006-1276-5](https://doi.org/10.1007/s00454-006-1276-5).
- [78] F. Cole, A. Golovinskiy, A. Limpaecher, H. Stoddart Barros, A. Finkelstein, T. Funkhouser, and S. Rusinkiewicz. *Where Do People Draw Lines?* Association for Computing Machinery, New York, NY, USA, 1 edition, 2023. ISBN 9798400708978. doi:[10.1145/3596711.3596756](https://doi.org/10.1145/3596711.3596756).
- [79] F. Conti, P. Frosini, and N. Quercioli. On the construction of group equivariant non-expansive operators via permutants and symmetric functions. *Frontiers in Artificial Intelligence*, 5, 2022. ISSN 2624-8212. doi:[10.3389/frai.2022.786091](https://doi.org/10.3389/frai.2022.786091).
- [80] Z. Cooperband, M. Lopez, and B. Schulze. Equivariant cosheaves and finite group representations in graphic statics. 2024. URL <https://arxiv.org/abs/2401.09392>.
- [81] F. Cordier, H. Seo, J. Park, and J.-y. Noh. Sketching of mirror-symmetric shapes. *IEEE Trans. Vis. Comput. Graph.*, 17:1650–1662, 11 2011. doi:[10.1109/TVCG.2010.258](https://doi.org/10.1109/TVCG.2010.258).
- [82] J. Coutinho, B. A. D. Marques, and J. P. Gois. Puppeteering 2.5d models. In *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 1–8, 2016. doi:[10.1109/SIBGRAPI.2016.010](https://doi.org/10.1109/SIBGRAPI.2016.010).
- [83] A. Cretenoud, L. Grzeczowski, M. Kunchulia, and M. Herzog. Individual differences in the perception of visual illusions are stable across eyes, time, and measurement methods. *Journal of Vision*, 21:26, 05 2021. doi:[10.1167/jov.21.5.26](https://doi.org/10.1167/jov.21.5.26).
- [84] J. Curry. Topological data analysis and cosheaves. *Japan Journal of Industrial and Applied Mathematics*, 32, 11 2014. doi:[10.1007/s13160-015-0173-9](https://doi.org/10.1007/s13160-015-0173-9).
- [85] J. Curry, R. Gelnett, and M. C. B. Zaremsky. Configuration spaces of circles in the plane. 2024. URL <https://arxiv.org/abs/2411.04800>.
- [86] B. Dalstein, R. Ronfard, and M. van de Panne. Vector graphics complexes. *ACM Trans. Graph.*, 33(4), July 2014. ISSN 0730-0301. doi:[10.1145/2601097.2601169](https://doi.org/10.1145/2601097.2601169).
- [87] B. Dalstein, R. Ronfard, and M. van de Panne. Vector graphics animation with time-varying topology. *ACM Trans. Graph.*, 34(4), July 2015. ISSN 0730-0301. doi:[10.1145/2766913](https://doi.org/10.1145/2766913).
- [88] M. Desbrun, E. Kanso, and Y. Tong. Discrete differential forms for computational modeling. In *ACM SIGGRAPH 2006 Courses*, SIGGRAPH ’06, page 39–54, 2006. ISBN 1595933646. doi:[10.1145/1185657.1185665](https://doi.org/10.1145/1185657.1185665).
- [89] N. Destler, M. Singh, and J. Feldman. Skeleton-based shape similarity. *Psychological Review*, 130:1653–1671, 03 2023. doi:[10.1037/rev0000412](https://doi.org/10.1037/rev0000412).
- [90] S. L. Devados and J. O’Rourke. *Discrete and Computational Geometry*. Princeton University Press, Princeton, NJ, 2011.
- [91] T. Dey, M. Lipiński, M. Mrozek, and R. Slechta. Tracking dynamical features via continuation and persistence. 03 2022. doi:[10.48550/arXiv.2203.05727](https://doi.org/10.48550/arXiv.2203.05727).
- [92] M. R. Dillon and E. S. Spelke. Core geometry in perspective. *Developmental Science*, 18(6):894–908, 2015. doi:[10.1111/desc.12266](https://doi.org/10.1111/desc.12266).

- [93] A. Dirafzoon, N. Lokare, and E. Lobaton. Action classification from motion capture data using topological data analysis. In *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1260–1264, 2016. doi:[10.1109/GlobalSIP.2016.7906043](https://doi.org/10.1109/GlobalSIP.2016.7906043).
- [94] N. A. Dodgson. Abstract depiction of human and animal figures: examples from two centuries of art and craft. In *Proceedings of the Joint Symposium on Computational Aesthetics and Sketch-Based Interfaces and Modeling and Non-Photorealistic Animation and Rendering*, Expressive '18, 2018. ISBN 9781450358927. doi:[10.1145/3229147.3229152](https://doi.org/10.1145/3229147.3229152).
- [95] M. Dujmović, J. S. Bowers, F. Adolphi, and G. Malhotra. The pitfalls of measuring representational similarity using representational similarity analysis. *bioRxiv*, 2022. doi:[10.1101/2022.04.05.487135](https://doi.org/10.1101/2022.04.05.487135).
- [96] I. Duta, G. Cassarà, F. Silvestri, and P. Liò. Sheaf hypergraph networks. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2023. Curran Associates Inc. doi:[10.5555/3666122.3666651](https://doi.org/10.5555/3666122.3666651).
- [97] M. Dvorožňák, W. Li, V. G. Kim, and D. Šýkora. Toonsynth: example-based synthesis of hand-colored cartoon animations. *ACM Trans. Graph.*, 37(4), July 2018. ISSN 0730-0301. doi:[10.1145/3197517.3201326](https://doi.org/10.1145/3197517.3201326).
- [98] M. Dvorožňák, D. Šýkora, C. Curtis, B. Curless, O. Sorkine-Hornung, and D. Salesin. Monster mash: a single-view approach to casual 3d modeling and animation. *ACM Trans. Graph.*, 39(6), Nov. 2020. ISSN 0730-0301. doi:[10.1145/3414685.3417805](https://doi.org/10.1145/3414685.3417805).
- [99] M. Dvorožňák, S. S. Nejad, O. Jamriška, A. Jacobson, L. Kavan, and D. Šýkora. Seamless reconstruction of part-based high-relief models from hand-drawn images. In T. Aydın and D. Šýkora, editors, *Expressive: Computational Aesthetics, Sketch-Based Interfaces and Modeling, Non-Photorealistic Animation and Rendering*. ACM, 2018. ISBN 978-1-4503-5892-7. doi:[10.1145/3229147.3229153](https://doi.org/10.1145/3229147.3229153).
- [100] P. Dłotko. Ball mapper: a shape summary for topological data analysis. 2019. URL <https://arxiv.org/abs/1901.07410>.
- [101] P. Dłotko, D. Gurnari, and R. Sazdanovic. Mapper-type algorithms for complex data and relations. *Journal of Computational and Graphical Statistics*, 33(4):1383–1396, 2024. doi:[10.1080/10618600.2024.2343321](https://doi.org/10.1080/10618600.2024.2343321).
- [102] H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. 01 2010. ISBN 978-0-8218-4925-5. doi:[10.1007/978-3-540-33259-6_7](https://doi.org/10.1007/978-3-540-33259-6_7).
- [103] A. Elhagry, M. Saeed, and M. Araia. Lighter stacked hourglass human pose estimation. 2021. URL <https://arxiv.org/abs/2107.13643>.
- [104] E. Entem, L. Barthe, M.-P. Cani, F. Cordier, and M. van de Panne. Modeling 3d animals from a side-view sketch. *Computers & Graphics*, 46:221–230, 2015. ISSN 0097-8493. doi:[10.1016/j.cag.2014.09.037](https://doi.org/10.1016/j.cag.2014.09.037). Shape Modeling International 2014.
- [105] E. Entem, A. D. Parakkat, L. Barthe, R. Muthuganapathy, and M.-P. Cani. Automatic structuring of organic shapes from a single drawing. *Computers & Graphics*, 81:125–139, 2019. ISSN 0097-8493. doi:[10.1016/j.cag.2019.04.006](https://doi.org/10.1016/j.cag.2019.04.006).
- [106] M. M. et al. vedo, a python module for scientific analysis and visualization of 3d objects and point clouds. 2019. doi:[10.5281/zenodo.15630979](https://doi.org/10.5281/zenodo.15630979). URL <https://zenodo.org/records/15630979>.
- [107] M. Even, P. Bénard, and P. Barla. Non-linear rough 2d animation using transient embeddings. *Computer Graphics Forum*, 42(2):411–425, 2023. doi:[10.1111/cgf.14771](https://doi.org/10.1111/cgf.14771).
- [108] M. Even, P. Bénard, and P. Barla. Inbetweening with occlusions for non-linear rough 2d animation. Technical Report RR-9559, Inria ; Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, Nov.

2024. URL <https://inria.hal.science/hal-04797216>.

- [109] B. D. Fabio and C. Landi. A mayer-vietoris formula for persistent homology with an application to shape recognition in the presence of occlusions. *Foundations of Computational Mathematics*, 11(5): 499–527, 2011. doi:[10.1007/s10208-011-9100-x](https://doi.org/10.1007/s10208-011-9100-x).
- [110] A. Fabri and S. Pion. Cgal: the computational geometry algorithms library. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '09, page 538–539, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605586496. doi:[10.1145/1653771.1653865](https://doi.org/10.1145/1653771.1653865).
- [111] J. E. Fan, J. D. Wammes, J. B. Gunn, D. L. Yamins, K. A. Norman, and N. B. Turk-Browne. Relating visual production and recognition of objects in human visual cortex. *Journal of Neuroscience*, 40(8): 1710–1721, 2020. ISSN 0270-6474. doi:[10.1523/JNEUROSCI.1843-19.2019](https://doi.org/10.1523/JNEUROSCI.1843-19.2019).
- [112] J. E. Fan, W. A. Bainbridge, R. Chamberlain, and J. D. Wammes. Drawing as a versatile cognitive tool. *Nat. Rev. Psychol.*, 2(9):556–568, Sept. 2023. doi:[10.1038/s44159-023-00212-w](https://doi.org/10.1038/s44159-023-00212-w).
- [113] X. Fan, K. Zheng, Y. Lin, and S. Wang. Combining local appearance and holistic view: Dual-source deep neural networks for human pose estimation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1347–1355. IEEE Computer Society, June 2015. doi:[10.1109/CVPR.2015.7298740](https://doi.org/10.1109/CVPR.2015.7298740).
- [114] X. Fan, A. H. Bermano, V. G. Kim, J. Popović, and S. Rusinkiewicz. Tooncap: a layered deformable model for capturing poses from cartoon characters. In *Proceedings of the Joint Symposium on Computational Aesthetics and Sketch-Based Interfaces and Modeling and Non-Photorealistic Animation and Rendering*, Expressive '18, 2018. ISBN 9781450358927. doi:[10.1145/3229147.3229149](https://doi.org/10.1145/3229147.3229149).
- [115] H.-S. Fang, J. Li, H. Tang, C. Xu, H. Zhu, Y. Xiu, Y.-L. Li, and C. Lu. Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(6):7157–7173, June 2023. ISSN 0162-8828. doi:[10.1109/TPAMI.2022.3222784](https://doi.org/10.1109/TPAMI.2022.3222784).
- [116] M. Farshchi, A. Kiba, and T. Sawada. Seeing our 3D world while only viewing contour-drawings. *PLoS One*, 16(1):e0242581, Jan. 2021. doi:[10.1371/journal.pone.0242581](https://doi.org/10.1371/journal.pone.0242581).
- [117] J.-D. Favreau, F. Lafarge, and A. Bousseau. Fidelity vs. simplicity: a global approach to line drawing vectorization. *ACM Trans. Graph.*, 35(4), July 2016. ISSN 0730-0301. doi:[10.1145/2897824.2925946](https://doi.org/10.1145/2897824.2925946).
- [118] X. Feng and M. Shi. 2d cartoon character deformation by sketch skeleton. In *2010 Third International Joint Conference on Computational Science and Optimization*, volume 2, pages 447–451, 2010. doi:[10.1109/CSO.2010.32](https://doi.org/10.1109/CSO.2010.32).
- [119] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao. Hypergraph neural networks. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*, 2019. ISBN 978-1-57735-809-1. doi:[10.1609/aaai.v33i01.33013558](https://doi.org/10.1609/aaai.v33i01.33013558).
- [120] L. Ferrari, P. Frosini, N. Quercioli, and T. Francesca. A topological model for partial equivariance in deep learning and data analysis. *Frontiers in Artificial Intelligence*, 6, 2023. doi:[10.3389/frai.2023.1272619](https://doi.org/10.3389/frai.2023.1272619).
- [121] M. Ferri, F. Patrizio, and C. Landi. Stable shape comparison by persistent homology. In *Atti del Seminario Matematico e Fisico dell'Università di Modena e Reggio Emilia*, 2011. URL https://www.dm.unibo.it/~frosini/pdf/FerriFrosiniLandiFG60_15_dic_2011.pdf.
- [122] S. Fourey, D. Tschumperlé, and D. Revoy. A fast and efficient semi-guided algorithm for flat coloring line-arts. In *Proceedings of the Conference on Vision, Modeling, and Visualization*, EG VMV '18,

page 1–9, Goslar, DEU, 2018. Eurographics Association. doi:[10.2312/vmv.20181247](https://doi.org/10.2312/vmv.20181247).

- [123] P. Frosini. Towards an observer-oriented theory of shape comparison: position paper. In *Proceedings of the Eurographics 2016 Workshop on 3D Object Retrieval*, 3DOR '16, page 5–8, Goslar, DEU, 2016. Eurographics Association. ISBN 9783038680048. doi:[10.5555/3056462.3056465](https://doi.org/10.5555/3056462.3056465).
- [124] L. Fu, J. Zhang, and K. Huang. Orgm: Occlusion relational graphical model for human pose estimation. *IEEE Transactions on Image Processing*, 26(2):927–941, 2017. doi:[10.1109/TIP.2016.2639441](https://doi.org/10.1109/TIP.2016.2639441).
- [125] M. Fujita and S. Saito. Hand-drawn animation with self-shaped canvas. In *ACM SIGGRAPH 2017 Posters*, SIGGRAPH '17, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450350150. doi:[10.1145/3102163.3102212](https://doi.org/10.1145/3102163.3102212).
- [126] T. Fukusato and A. Maejima. View-dependent formulation of 2.5d cartoon models. 2021. URL <https://arxiv.org/abs/2103.15472>.
- [127] T. Fukusato and S. Morishima. Active comicing for freehand drawing animation. In Y. Dobashi and H. Ochiai, editors, *Mathematical Progress in Expressive Image Synthesis III*, pages 45–56, Singapore, 2016. Springer Singapore. ISBN 978-981-10-1076-7. doi:[10.1007/978-981-10-1076-7_6](https://doi.org/10.1007/978-981-10-1076-7_6).
- [128] T. Fukusato, A. Maejima, and T. Igarashi. Locality-preserving free-form deformation. 2024. URL <https://arxiv.org/abs/2405.07450>.
- [129] M. Furtak, L. Mudrik, and M. Bola. The forest, the trees, or both? hierarchy and interactions between gist and object processing during perception of real-world scenes. *Cognition*, 221:104983, 2022. ISSN 0010-0277. doi:[10.1016/j.cognition.2021.104983](https://doi.org/10.1016/j.cognition.2021.104983).
- [130] C. Furusawa, T. Fukusato, N. Okada, T. Hirai, and S. Morishima. Quasi 3d rotation for hand-drawn characters. In *ACM SIGGRAPH 2014 Posters*, SIGGRAPH '14, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450329583. doi:[10.1145/2614217.2633397](https://doi.org/10.1145/2614217.2633397).
- [131] J. Gallier and J. Quaintance. *Homology, cohomology, and sheaf cohomology for algebraic topology, algebraic geometry, and differential geometry*. World Scientific, New Jersey, 2022. ISBN 9789811245022.
- [132] T. Gebhart, J. Hansen, and P. Schrater. Knowledge sheaves: A sheaf-theoretic framework for knowledge graph embedding. 2023. URL <https://arxiv.org/abs/2110.03789>.
- [133] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. Wichmann, and W. Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. 11 2018. doi:[10.48550/arXiv.1811.12231](https://doi.org/10.48550/arXiv.1811.12231).
- [134] C. Geuzaine and J.-F. Remacle. Gmsh: A 3-d finite element mesh generator with built-in pre- and post-processing facilities. *International Journal for Numerical Methods in Engineering*, 79: 1309–1331, 09 2009. doi:[10.1002/nme.2579](https://doi.org/10.1002/nme.2579).
- [135] R. Ghrist. *CONFIGURATION SPACES, BRAIDS, AND ROBOTICS*, pages 263–304. 2010. doi:[10.1142/9789814291415_0004](https://doi.org/10.1142/9789814291415_0004).
- [136] H. Gohara, S. Sugimoto, and S. Morishima. Data driven in-betweening for hand drawn rotating face. In *ACM SIGGRAPH 2010 Posters*, SIGGRAPH '10, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781450303934. doi:[10.1145/1836845.1836853](https://doi.org/10.1145/1836845.1836853).
- [137] A. Gomez and J. Franco. Hodge laplacians and hodge diffusion maps. 04 2025. doi:[10.48550/arXiv.2504.07910](https://doi.org/10.48550/arXiv.2504.07910).
- [138] D. J. Graham and M. Meng. Artistic representations: clues to efficient coding in human vision. *Vis. Neurosci.*, 28(4):371–379, July 2011. doi:[0.1017/S0952523811000162](https://doi.org/10.1017/S0952523811000162).

- [139] R. Green, R. Cardona, J. Cleveland, J. Ozbolt, A. Hylton, R. Short, and M. Robinson. Dude where’s my stars: A novel topologically justified approach to star tracking. In *2021 IEEE Aerospace Conference (50100)*, pages 1–15, 2021. doi:[10.1109/AERO50100.2021.9438366](https://doi.org/10.1109/AERO50100.2021.9438366).
- [140] K. Greff, S. van Steenkiste, and J. Schmidhuber. On the binding problem in artificial neural networks. *CoRR*, 2020. URL <https://arxiv.org/abs/2012.05208>.
- [141] Y. Gryaditskaya, F. Hähnlein, C. Liu, A. Sheffer, and A. Bousseau. Lifting freehand concept sketches into 3d. *ACM Trans. Graph.*, 39(6), Nov. 2020. ISSN 0730-0301. doi:[10.1145/3414685.3417851](https://doi.org/10.1145/3414685.3417851).
- [142] W. Guo, K. Manohar, S. L. Brunton, and A. G. Banerjee. Sparse-tda: Sparse realization of topological data analysis for multi-way classification. *IEEE Transactions on Knowledge & Data Engineering*, 30(07):1403–1408, July 2018. ISSN 1558-2191. doi:[10.1109/TKDE.2018.2790386](https://doi.org/10.1109/TKDE.2018.2790386).
- [143] S. Gupta, Y. Zhang, X. Hu, P. Prasanna, and C. Chen. Topology-aware uncertainty for image segmentation. 2023. URL <https://arxiv.org/abs/2306.05671>.
- [144] F. Hahn, F. Mutzel, S. Coros, B. Thomaszewski, M. Nitti, M. Gross, and R. W. Sumner. Sketch abstractions for character posing. In *Proceedings of the 14th ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, page 185–191, 2015. doi:[10.1145/2786784.2786785](https://doi.org/10.1145/2786784.2786785).
- [145] M. Hajij, K. Istvan, and G. Zamzmi. Cell complex neural networks. 10 2020. doi:[10.48550/arXiv.2010.00743](https://doi.org/10.48550/arXiv.2010.00743).
- [146] M. Hajij, G. Zamzmi, T. Papamarkou, N. Miolane, A. Guzmán-Sáenz, K. N. Ramamurthy, T. Birdal, T. K. Dey, S. Mukherjee, S. N. Samaga, N. Livesay, R. Walters, P. Rosen, and M. T. Schaub. Topological deep learning: Going beyond graph data. 2023. URL <https://arxiv.org/abs/2206.00606>.
- [147] J. Han, J. Cen, L. Wu, Z. Li, X. Kong, R. Jiao, Z. Yu, T. Xu, F. Wu, Z. Wang, H. Xu, Z. Wei, D. Zhao, Y. Liu, Y. Rong, and W. Huang. A survey of geometric graph neural networks: data structures, models and applications. *Frontiers of Computer Science*, 19(11):1911375, May 2025. doi:[10.1007/s11704-025-41426-w](https://doi.org/10.1007/s11704-025-41426-w).
- [148] Q. Han, S. Zhang, and P. Wang. Dual chain dynamic hypergraph convolution network for 3D human pose estimation. *Scientific Reports*, 15(1):35211, Oct. 2025. doi:[10.1038/s41598-025-22261-2](https://doi.org/10.1038/s41598-025-22261-2).
- [149] H. Hang and W. Mio. Correspondence modules and persistence sheaves: A unifying perspective on one-parameter persistent homology. 2021. URL <https://arxiv.org/abs/2006.08557>.
- [150] J. Hansen. *Laplacians of Cellular Sheaves: Theory and Applications*. PhD thesis, University of Pennsylvania, 2020.
- [151] J. Hansen. A gentle introduction to sheaves on graphs. 2021. URL <https://jakobhansen.org/publications/gentleintroduction.pdf>.
- [152] J. Hansen and T. Gebhart. Sheaf neural networks. *CoRR*, abs/2012.06333, 2020. URL <https://arxiv.org/abs/2012.06333>.
- [153] J. Hansen and R. Ghrist. Toward a spectral theory of cellular sheaves. *Journal of Applied and Computational Topology*, 3, 12 2019. doi:[10.1007/s41468-019-00038-7](https://doi.org/10.1007/s41468-019-00038-7).
- [154] J. Hansen and R. Ghrist. Opinion dynamics on discourse sheaves. *SIAM Journal on Applied Mathematics*, 81(5):2033–2060, 2021. doi:[10.1137/20M1341088](https://doi.org/10.1137/20M1341088).
- [155] R. Hawkins, M. Sano, N. Goodman, and J. Fan. Visual resemblance and interaction history jointly constrain pictorial meaning. *Nature Communications*, 14, 04 2023. doi:[10.1038/s41467-023-37737-w](https://doi.org/10.1038/s41467-023-37737-w).
- [156] D. Heinke, P. Wachman, W. van Zoest, and E. C. Leek. A failure to learn object shape geometry:

- Implications for convolutional neural networks as plausible models of biological vision. *Vision Research*, 189:81–92, 2021. ISSN 0042-6989. doi:[10.1016/j.visres.2021.09.004](https://doi.org/10.1016/j.visres.2021.09.004).
- [157] F. Hernandez Caralt, G. Bernárdez Gil, I. Duta, P. Liò, and E. Alarcón Cot. Joint diffusion processes as an inductive bias in sheaf neural networks. In *Proceedings of the Geometry-grounded Representation Learning and Generative Modeling Workshop (GRaM)*, volume 251 of *Proceedings of Machine Learning Research*, pages 249–263. PMLR, 29 Jul 2024. doi:[10.48550/arXiv.2407.20597](https://doi.org/10.48550/arXiv.2407.20597).
 - [158] A. Hertzmann. Why do line drawings work? a realism hypothesis. *Perception*, 49(4):439–451, Mar. 2020. ISSN 1468-4233. doi:[10.1177/0301006620908207](https://doi.org/10.1177/0301006620908207).
 - [159] A. Hertzmann. Toward a theory of perspective perception in pictures. *Journal of Vision*, 24(4): 23–23, 04 2024. ISSN 1534-7362. doi:[10.1167/jov.24.4.23](https://doi.org/10.1167/jov.24.4.23).
 - [160] D. D. Hoffman. *Visual Intelligence: How We Create What We See*. Norton, 2000.
 - [161] C.-S. Hu and Y.-M. Chung. A sheaf and topology approach to detecting local merging relations in digital images. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 4391–4400, 2021. doi:[10.1109/CVPRW53098.2021.00496](https://doi.org/10.1109/CVPRW53098.2021.00496).
 - [162] L. Huang, S. Mukhopadhyay, M. Ehrlich, and A. Shrivastava. Computational in-betweening for line drawings in animation. 2023. URL https://lilhuang.github.io/docs/SIGGRAPH_Asia_2022_poster_abstract_FINAL.pdf. unpublished.
 - [163] W. Huang, J. Han, Y. Rong, T. Xu, F. Sun, and J. Huang. Equivariant graph mechanics networks with constraints. 2022. URL <https://arxiv.org/abs/2203.06442>.
 - [164] H. Huey, X. Lu, C. M. Walker, and J. E. Fan. Visual explanations prioritize functional properties at the expense of visual fidelity. *Cognition*, 236:105414, 2023. ISSN 0010-0277. doi:[10.1016/j.cognition.2023.105414](https://doi.org/10.1016/j.cognition.2023.105414).
 - [165] hysts. Anime face detector. 2021.
 - [166] G. Jacob, R. T. Pramod, H. Katti, and S. P. Arun. Qualitative similarities and differences in visual object representations between brains and deep networks. *Nature Communications*, 12(1):1872, Mar. 2021. doi:[10.1038/s41467-021-22078-3](https://doi.org/10.1038/s41467-021-22078-3).
 - [167] E. Jain, Y. Sheikh, M. Mahler, and J. Hodgins. Three-dimensional proxies for hand-drawn characters. *ACM Trans. Graph.*, 31(1), Feb. 2012. ISSN 0730-0301. doi:[10.1145/2077341.2077349](https://doi.org/10.1145/2077341.2077349).
 - [168] L. Jia, J. Yu, and J. Liao. Procedural Modeling of Water Caustics and Foamy Water for Cartoon Animation. In *Pacific Conference on Computer Graphics and Applications*, pages 1–4, Los Alamitos, CA, USA, Sept. 2010. IEEE Computer Society. doi:[10.1109/PacificGraphics.2010.8](https://doi.org/10.1109/PacificGraphics.2010.8).
 - [169] J. Jiang, H. S. Seah, and H. Z. Liew. Stroke-based drawing and inbetweening with boundary strokes. *Computer Graphics Forum*, 41(1):257–269, 2022. doi:[10.1111/cgf.14433](https://doi.org/10.1111/cgf.14433).
 - [170] B. Jin and W. Geng. Correspondence specification learned from master frames for automatic inbetweening. *Multimedia Tools and Applications*, 74(13):4873–4889, July 2015. doi:[10.1007/s11042-013-1847-4](https://doi.org/10.1007/s11042-013-1847-4).
 - [171] M. Johnson. *Ink & Paint: The Women of Walt Disney’s Animation*. Disney Publishing Worldwide, New York, 2017.
 - [172] B. Jones, J. Popovic, J. McCann, W. Li, and A. Bargteil. Dynamic sprites. In *Proceedings of Motion on Games*, MIG ’13, page 39–46, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450325462. doi:[10.1145/2522628.2522631](https://doi.org/10.1145/2522628.2522631).
 - [173] C. A. Joslyn, L. Charles, C. DePerno, N. Gould, K. Nowak, B. Praggastis, E. Purvine, M. Robinson,

- J. Strules, and P. Whitney. A sheaf theoretical approach to uncertainty quantification of heterogeneous geolocation information. *Sensors*, 20(12), 2020. ISSN 1424-8220. doi:[10.3390/s20123418](https://doi.org/10.3390/s20123418).
- [174] X. Ju, A. Zeng, W. Jianan, X. Qiang, and Z. Lei. Human-art: A versatile human-centric dataset bridging natural and artificial scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. doi:[10.1109/CVPR52729.2023.00067](https://doi.org/10.1109/CVPR52729.2023.00067).
- [175] Z. Kadkhodaie, F. Guth, E. P. Simoncelli, and S. Mallat. Generalization in diffusion models arises from geometry-adaptive harmonic representations. In *International Conference on Learning Representations*, volume 12, 2024. doi:[url=https://arxiv.org/abs/2310.02557](https://arxiv.org/abs/2310.02557).
- [176] D. Kaiser and M. V. Peelen. Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *NeuroImage*, 169:334–341, 2018. ISSN 1053-8119. doi:[10.1016/j.neuroimage.2017.12.065](https://doi.org/10.1016/j.neuroimage.2017.12.065).
- [177] D. Kaiser, G. Häberle, and R. M. Cichy. Real-world structure facilitates the rapid emergence of scene category information in visual brain signals. *Journal of Neurophysiology*, 124(1):145–151, 2020. doi:[10.1152/jn.00164.2020](https://doi.org/10.1152/jn.00164.2020).
- [178] D. Kaiser, G. Inciuraite, and R. M. Cichy. Rapid contextualization of fragmented scene information in the human visual system. *NeuroImage*, 219:117045, 2020. ISSN 1053-8119. doi:[10.1016/j.neuroimage.2020.117045](https://doi.org/10.1016/j.neuroimage.2020.117045).
- [179] R. H. Kazi, T. Grossman, N. Umetani, and G. Fitzmaurice. Motion amplifiers: Sketching dynamic illustrations using the principles of 2d animation. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI ’16, page 4599–4609, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450333627. doi:[10.1145/2858036.2858386](https://doi.org/10.1145/2858036.2858386).
- [180] A. Keros, V. Nanda, and K. Subr. Dist2cycle: A simplicial neural network for homology localization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36:7133–7142, 06 2022. doi:[10.1609/aaai.v36i7.20673](https://doi.org/10.1609/aaai.v36i7.20673).
- [181] M. Kitamura, Y. Kanamori, and R. Tsuruno. 2.5d modeling from illustrations of different views. *International Journal of Asia Digital Art and Design*, 18(4):74–79, 2014. doi:[10.20668/adada.18.4_74](https://doi.org/10.20668/adada.18.4_74).
- [182] C. Klink, R. Wezel, and R. Ee. United we sense, divided we fail: Context-driven perception of ambiguous visual stimuli. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367:932–41, 04 2012. doi:[10.1098/rstb.2011.0358](https://doi.org/10.1098/rstb.2011.0358).
- [183] J. Koenderink and A. van Doorn. The structure of visual spaces. *Journal of Mathematical Imaging and Vision*, 31:171–187, 2008. doi:[10.1007/s10851-008-0076-3](https://doi.org/10.1007/s10851-008-0076-3).
- [184] A. Kozbelt. Artists as experts in visual cognition. *Visual Cognition - VIS COGN*, 8:705–723, 12 2001. doi:[10.1080/13506280042000090](https://doi.org/10.1080/13506280042000090).
- [185] A. Kozbelt, A. Seidel, A. Elbassiouny, Y. Mark, and D. Owen. Visual selection contributes to artists’ advantages in realistic drawing. *Psychology of Aesthetics, Creativity, and the Arts*, 4:93–102, 05 2010. doi:[10.1037/a0017657](https://doi.org/10.1037/a0017657).
- [186] D. Lavado, C. Soares, and A. Micheletti. Scene-net v2: Interpretable multiclass 3d scene understanding with geometric priors. In *Proceedings of the Geometry-grounded Representation Learning and Generative Modeling Workshop (GRaM)*, volume 251 of *Proceedings of Machine Learning Research*, pages 222–232. PMLR, 29 Jul 2024. URL <https://proceedings.mlr.press/v251/lavado24a.html>.
- [187] J. L. Leon, R. Alonso, E. G. Reyes, and R. G. Diaz. Topological features for monitoring human activities at distance. In *Activity Monitoring by Multiple Distributed Sensing*, pages 40–51, 2014.
- [188] F. F. Leymarie and P. Aparajeya. Medialness and the perception of visual art. *Art & Perception*, 5

- (2):169 – 232, 2017. doi:[10.1163/22134913-00002064](https://doi.org/10.1163/22134913-00002064).
- [189] C. Li, H. Pan, Y. Liu, X. Tong, A. Sheffer, and W. Wang. Bendsketch: modeling freeform surfaces through 2d sketching. *ACM Trans. Graph.*, 36(4), July 2017. ISSN 0730-0301. doi:[10.1145/3072959.3073632](https://doi.org/10.1145/3072959.3073632).
 - [190] Y. Li, S. Yang, P. Liu, S. Zhang, Y. Wang, Z. Wang, W. Yang, and S.-T. Xia. Simcc: A simple coordinate classification perspective for human pose estimation. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VI*, page 89–106, Berlin, Heidelberg, 2022. Springer-Verlag. ISBN 978-3-031-20067-0. doi:[10.1007/978-3-031-20068-7_6](https://doi.org/10.1007/978-3-031-20068-7_6).
 - [191] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *Computer Vision – ECCV 2014*, pages 740–755. Springer International Publishing, 2014. doi:[10.1007/978-3-319-10602-1_48](https://doi.org/10.1007/978-3-319-10602-1_48).
 - [192] C. Liu and M. Bessmeltsev. State-of-the-art report in sketch processing. *Computer Graphics Forum*, 44(2), 2025. doi:[10.1111/cgf.70079](https://doi.org/10.1111/cgf.70079).
 - [193] C. Liu, P. Bénard, A. Hertzmann, and S. Hoshyari. Contesse: Accurate occluding contours for subdivision surfaces. *ACM Trans. Graph.*, 42(1), Jan. 2023. ISSN 0730-0301. doi:[10.1145/3544778](https://doi.org/10.1145/3544778).
 - [194] C. Liu, X. Ma, H. Zhang, S. Xie, and D. Yu. Dynamic neuropsychological approach for multi-quality image assessment using grey-topological data analysis. *IEEE Access*, 12:139609–139619, 2024. doi:[10.1109/ACCESS.2024.3446281](https://doi.org/10.1109/ACCESS.2024.3446281).
 - [195] D. Liu, Q. Chen, J. Yu, H. Gu, D. Tao, and H. S. Seah. Stroke correspondence construction using manifold learning. *Computer Graphics Forum*, 30(8):2194–2207, 2011. doi:[10.1111/j.1467-8659.2011.01969.x](https://doi.org/10.1111/j.1467-8659.2011.01969.x).
 - [196] K. Liu, R. Ding, Z. Zou, L. Wang, and W. Tang. A comprehensive study of weight sharing in graph networks for 3d human pose estimation. In *Computer Vision – ECCV 2020*, pages 318–334. Springer International Publishing, 2020. ISBN 978-3-030-58607-2. doi:[10.1007/978-3-030-58607-2_19](https://doi.org/10.1007/978-3-030-58607-2_19).
 - [197] S. Liu, X. Wang, Z. Wu, and H. S. Seah. Shape correspondence based on kendall shape space and rag for 2d animation. *The Visual Computer*, 36:2457–2469, 2020. doi:[10.1007/s00371-020-01958-7](https://doi.org/10.1007/s00371-020-01958-7).
 - [198] S. Liu, X. Wang, X. Liu, Z. Wu, and H. S. Seah. Shape correspondence for cel animation based on a shape association graph and spectral matching. *Computational Visual Media*, 9:633–656, 2023. doi:[10.1007/s41095-022-0298-0](https://doi.org/10.1007/s41095-022-0298-0).
 - [199] X. Liu, X. Mao, X. Yang, L. Zhang, and T.-T. Wong. Stereoscopizing cel animations. *ACM Trans. Graph.*, 32(6), Nov. 2013. ISSN 0730-0301. doi:[10.1145/2508363.2508396](https://doi.org/10.1145/2508363.2508396).
 - [200] Y. Liu and W. Wang. Body-structure-based cartoon character modeling from multi-view hand-drawings. In *2016 5th International Conference on Computer Science and Network Technology (ICCSNT)*, pages 89–93, 2016. doi:[10.1109/ICCSNT.2016.8070125](https://doi.org/10.1109/ICCSNT.2016.8070125).
 - [201] Y. Liu, M. Yu, Q. fu, W. Chen, and L. Xie. Cognitive mechanism related to line drawings and its applications in intelligent process of visual media: a survey. *Frontiers of Computer Science*, 10: 216–232, 06 2016. doi:[10.1007/s11704-015-4450-1](https://doi.org/10.1007/s11704-015-4450-1).
 - [202] Z. Liu, L. Zhou, H. Leung, and H. P. H. Shum. High-quality compatible triangulations and their application in interactive animation. *Computers & Graphics*, 76:60–72, 2018. ISSN 0097-8493. doi:[10.1016/j.cag.2018.07.002](https://doi.org/10.1016/j.cag.2018.07.002).
 - [203] B. Long, J. Fan, H. Huey, Z. Chai, and M. Frank. Parallel developmental changes in children’s production and recognition of line drawings of visual concepts. *Nature Communications*, 15(1191),

02 2024. doi:[10.1038/s41467-023-44529-9](https://doi.org/10.1038/s41467-023-44529-9).

- [204] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. Smpl: a skinned multi-person linear model. *ACM Trans. Graph.*, 34(6), Oct. 2015. ISSN 0730-0301. doi:[10.1145/2816795.2818013](https://doi.org/10.1145/2816795.2818013).
- [205] M. X. Lowe, J. P. Gallivan, S. Ferber, and J. S. Cant. Feature diagnosticity and task context shape activity in human scene-selective cortex. *NeuroImage*, 125:681–692, 2016. ISSN 1053-8119. doi:[10.1016/j.neuroimage.2015.10.089](https://doi.org/10.1016/j.neuroimage.2015.10.089).
- [206] M. X. Lowe, J. Rajsic, S. Ferber, and D. B. Walther. Discriminating scene categories from brain activity within 100 milliseconds. *Cortex*, 106:275–287, 2018. ISSN 0010-9452. doi:[10.1016/j.cortex.2018.06.006](https://doi.org/10.1016/j.cortex.2018.06.006).
- [207] Z. Luo, S. Cai, J. Dong, R. Ming, L. Qiu, X. Zhan, and X. Han. Rabbit: Parametric modeling of 3d biped cartoon characters with a topological-consistent dataset. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12825–12835, 2023. doi:[10.1109/CVPR52729.2023.01233](https://doi.org/10.1109/CVPR52729.2023.01233).
- [208] J. Ma, J. Wang, J. Li, and D. Zhang. Real-time skeletonization for sketch-based modeling. *Computers & Graphics*, 102:56–66, 2022. ISSN 0097-8493. doi:[10.1016/j.cag.2021.11.005](https://doi.org/10.1016/j.cag.2021.11.005).
- [209] M. D. B. Machuca, J. H. Israel, D. F. Keefe, and W. Stuerzlinger. Toward more comprehensive evaluations of 3d immersive sketching, drawing, and painting. *IEEE Transactions on Visualization and Computer Graphics*, 30(8):4648–4664, 2024. doi:[10.1109/TVCG.2023.3276291](https://doi.org/10.1109/TVCG.2023.3276291).
- [210] A. Maejima, H. Kubo, T. Funatomi, T. Yotsukura, S. Nakamura, and Y. Mukaigawa. Graph matching based anime colorization with multiple references. In *ACM SIGGRAPH 2019 Posters*, SIGGRAPH ’19, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450363143. doi:[10.1145/3306214.3338560](https://doi.org/10.1145/3306214.3338560).
- [211] G. Malhotra, B. D. Evans, and J. S. Bowers. Hiding a plane with a pixel: examining shape-bias in cnns and the benefit of building in biological constraints. *Vision Research*, 174:57–68, 2020. ISSN 0042-6989. doi:[10.1016/j.visres.2020.04.013](https://doi.org/10.1016/j.visres.2020.04.013).
- [212] G. Malhotra, M. Dujmović, J. Hummel, and J. S. Bowers. Human shape representations are not an emergent property of learning to classify objects. *bioRxiv*, 2022. doi:[10.1101/2021.12.14.472546](https://doi.org/10.1101/2021.12.14.472546).
- [213] G. Malhotra, M. Dujmović, and J. S. Bowers. Feature blindness: A challenge for understanding and modelling visual object recognition. *PLOS Computational Biology*, 18(5):1–27, 05 2022. doi:[10.1371/journal.pcbi.1009572](https://doi.org/10.1371/journal.pcbi.1009572).
- [214] P. Mamassian. Ambiguities and conventions in the perception of visual art. *Vision Research*, 48(20):2143–2153, 2008. ISSN 0042-6989. doi:[10.1016/j.visres.2008.06.010](https://doi.org/10.1016/j.visres.2008.06.010). Vision Research Reviews.
- [215] L. O. A. Marini. *Perception and representation in the visual brain and in visual art*. PhD thesis, University College London, 2004.
- [216] C. Martens and M. Bessmeltsev. One-shot method for computing generalized winding numbers. *Computer Graphics Forum*, 44(5):e70194, 2025. doi:[10.1111/cgf.70194](https://doi.org/10.1111/cgf.70194).
- [217] A. MARZUOLI and G. PALUMBO. Post quantum cryptography from mutant prime knots. *International Journal of Geometric Methods in Modern Physics*, 08(07):1571–1581, 2011. doi:[10.1142/S0219887811005798](https://doi.org/10.1142/S0219887811005798).
- [218] T. Matsukawa, H. Ogura, S. Yamashita, K. Shibasaki, and M. Ikehara. Pose estimation of artwork characters with series and parallel dilated convolution and style channel attention. In *2025 IEEE International Conference on Image Processing (ICIP)*, pages 857–862, 2025. doi:[10.1109/ICIP55913.2025.11084500](https://doi.org/10.1109/ICIP55913.2025.11084500).

- [219] C. Mavrogiannis, J. A. DeCastro, and S. S. Srinivasa. Abstracting road traffic via topological braids: Applications to traffic flow analysis and distributed control. *The International Journal of Robotics Research*, 43(9):1299–1321, 2024. doi:[10.1177/02783649231188740](https://doi.org/10.1177/02783649231188740).
- [220] F. Mémoli, Z. Wan, and Y. Wang. Persistent laplacians: Properties, algorithms and implications. *SIAM Journal on Mathematics of Data Science*, 4(2):858–884, 2022. doi:[10.1137/21M1435471](https://doi.org/10.1137/21M1435471).
- [221] Z. Meng and K. Xia. Persistent spectral-based machine learning (perspect ml) for protein-ligand binding affinity prediction. *Science Advances*, 7(19):eabc5329, 2021. doi:[10.1126/sciadv.abc5329](https://doi.org/10.1126/sciadv.abc5329).
- [222] Y. Miao, F. Hu, X. Zhang, J. Chen, and R. Pajarola. Symmsketch: Creating symmetric 3d free-form shapes from 2d sketches. *Comp. Visual Media*, 1:3–16, 08 2015. doi:[10.1007/s41095-015-0002-8](https://doi.org/10.1007/s41095-015-0002-8).
- [223] R. Miyauchi, T. Fukusato, H. Xie, and K. Miyata. Stroke correspondence by labeling closed areas. In *2021 Nicograph International (NicoInt)*, pages 34–41, 2021. doi:[10.1109/NICOINT52941.2021.00014](https://doi.org/10.1109/NICOINT52941.2021.00014).
- [224] R. Miyauchi, Y. Peng, T. Fukusato, and H. Xie. Skeleton2stroke: Interactive stroke correspondence editing with pose features. In *SIGGRAPH Asia 2021 Technical Communications*, SA ’21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450390736. doi:[10.1145/3478512.3488612](https://doi.org/10.1145/3478512.3488612).
- [225] H. Mo, C. Gao, and R. Wang. Joint stroke tracing and correspondence for 2d animation. *ACM Trans. Graph.*, 43(3), Apr. 2024. ISSN 0730-0301. doi:[10.1145/3649890](https://doi.org/10.1145/3649890).
- [226] M. Montero, C. Ludwig, R. Costa, G. Malhotra, and J. Bowers. The role of disentanglement in generalisation. In *International Conference on Learning Representations (ICLR)*, 05 2021.
- [227] M. Montero, J. Bowers, R. Costa, C. Ludwig, and G. Malhotra. Lost in latent space: Disentangled models and the challenge of combinatorial generalisation. 04 2022. doi:[10.48550/arXiv.2204.02283](https://doi.org/10.48550/arXiv.2204.02283).
- [228] A. T. Morgan, L. S. Petro, and L. Muckli. Scene representations conveyed by cortical feedback to early visual cortex can be described by line drawings. *Journal of Neuroscience*, 39(47):9410–9423, 2019. ISSN 0270-6474. doi:[10.1523/JNEUROSCI.0852-19.2019](https://doi.org/10.1523/JNEUROSCI.0852-19.2019).
- [229] Y. Morgenstern, K. R. Storrs, F. Schmidt, F. Hartmann, H. Tiedemann, J. Wagemans, and R. W. Fleming. High-level aftereffects reveal the role of statistical features in visual shape encoding. *Current Biology*, 34(5):1098–1106.e5, 2024. ISSN 0960-9822. doi:[10.1016/j.cub.2023.12.039](https://doi.org/10.1016/j.cub.2023.12.039).
- [230] H. N. H. Mortaja. In search of a new common ground: Applying the signature-testing approach to object recognition mechanisms in biological and artificial intelligence. Master’s thesis, University of Auckland, 2024.
- [231] C. A. Mueller and A. Birk. Conceptualization of object compositions using persistent homology. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, page 1095–1102. IEEE Press, 2018. doi:[10.1109/IROS.2018.8594516](https://doi.org/10.1109/IROS.2018.8594516).
- [232] K. Murasugi. *Knot Theory and Its Applications*. Birkhäuser, Boston, MA, 1996. doi:[10.1007/978-0-8176-4719-3](https://doi.org/10.1007/978-0-8176-4719-3).
- [233] A. Musaelian, S. Batzner, A. Johansson, L. Sun, C. J. Owen, M. Kornbluth, and B. Kozinsky. Learning local equivariant representations for large-scale atomistic dynamics. *Nature Communications*, 14(1): 579, Feb. 2023. doi:[10.1038/s41467-023-36329-y](https://doi.org/10.1038/s41467-023-36329-y).
- [234] M. Myronova, W. Neveu, and M. Bessmeltsev. Differential operators on sketches via alpha contours. *ACM Trans. Graph.*, 42(4), July 2023. ISSN 0730-0301. doi:[10.1145/3592420](https://doi.org/10.1145/3592420).
- [235] R. Nascimento, F. Queiroz, A. Rocha, T. Ing Ren, V. Mello, and A. Peixoto. Colorization and illumination of 2d animations based on a region-tree representation. In *2011 24th SIBGRAPI Conference on Graphics, Patterns and Images*, pages 9–16, 2011. doi:[10.1109/SIBGRAPI.2011.6](https://doi.org/10.1109/SIBGRAPI.2011.6).

- [236] A. Newell, K. Yang, and J. Deng. Stacked hourglass networks for human pose estimation. In *Computer Vision – ECCV 2016*, pages 483–499, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46484-8. doi:[10.1007/978-3-319-46484-8_29](https://doi.org/10.1007/978-3-319-46484-8_29).
- [237] A. Nguyen, J. Yosinski, and J. Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 427–436, 06 2015. doi:[10.1109/CVPR.2015.7298640](https://doi.org/10.1109/CVPR.2015.7298640).
- [238] G. Noris, D. Šýkora, S. Coros, B. Whited, M. Simmons, A. Hornung, M. Gross, and R. W. Sumner. Temporal noise control for sketchy animation. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*, NPAR ’11, page 93–98, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450309073. doi:[10.1145/2024676.2024691](https://doi.org/10.1145/2024676.2024691).
- [239] G. Noris, D. Šýkora, A. Shamir, S. Coros, B. Whited, M. Simmons, A. Hornung, M. Gross, and R. Sumner. Smart scribbles for sketch segmentation. *Computer Graphics Forum*, 31(8):2516–2527, 2012. doi:[10.1111/j.1467-8659.2012.03224.x](https://doi.org/10.1111/j.1467-8659.2012.03224.x).
- [240] G. Noris, A. Hornung, R. W. Sumner, M. Simmons, and M. Gross. Topology-driven vectorization of clean line drawings. *ACM Trans. Graph.*, 32(1), Feb. 2013. ISSN 0730-0301. doi:[10.1145/2421636.2421640](https://doi.org/10.1145/2421636.2421640).
- [241] L. Olsen, F. Samavati, and J. Jorge. Naturasketch: Modeling from images and natural sketches. *IEEE Computer Graphics and Applications*, 31(6):24–34, 2011. doi:[10.1109/MCG.2011.84](https://doi.org/10.1109/MCG.2011.84).
- [242] J. Ortiz-Tudela, J. Bergmann, M. Bennett, I. Ehrlich, L. Muckli, and Y. L. Shing. Concurrent contextual and time-distant mnemonic information co-exist as feedback in the human visual cortex. *NeuroImage*, 265:119778, 2023. ISSN 1053-8119. doi:[10.1016/j.neuroimage.2022.119778](https://doi.org/10.1016/j.neuroimage.2022.119778).
- [243] Y. Ostrovsky, P. Cavanagh, and P. Sinha. Perceiving illumination inconsistencies in scenes. *Perception*, 34(11):1301–1314, 2005. doi:[10.1068/p5418](https://doi.org/10.1068/p5418).
- [244] E. Panagiotou and K. Plaxco. *Topology and Geometry of Biopolymers*, chapter A topological study of protein folding kinetics, pages 223–234. Springer, 01 2020. ISBN 9781470448400. doi:[10.1090/conm/746/15010](https://doi.org/10.1090/conm/746/15010).
- [245] E. Panagiotou, K. C. Millett, and P. J. Atzberger. Topological methods for polymeric materials: Characterizing the relationship between polymer entanglement and viscoelasticity. *Polymers*, 11(3), 2019. ISSN 2073-4360. doi:[10.3390/polym11030437](https://doi.org/10.3390/polym11030437).
- [246] T. Papamarkou, T. Birdal, M. Bronstein, G. Carlsson, J. Curry, Y. Gao, M. Hajij, R. Kwitt, P. Liò, P. Di Lorenzo, V. Maroulas, N. Miolane, F. Nasrin, K. N. Ramamurthy, B. Rieck, S. Scardapane, M. T. Schaub, P. Veličković, B. Wang, Y. Wang, G.-W. Wei, and G. Zamzmi. Position: topological deep learning is the new frontier for relational learning. In *Proceedings of the 41st International Conference on Machine Learning, ICML’24*. JMLR.org, 2024. doi:[10.5555/3692070.3693671](https://doi.org/10.5555/3692070.3693671).
- [247] A. D. Parakkat, U. Bondi Pundarikaksha, and R. Muthuganapathy. A delaunay triangulation based approach for cleaning rough sketches. *Computers & Graphics*, 74:171–181, 2018. ISSN 0097-8493. doi:[10.1016/j.cag.2018.05.011](https://doi.org/10.1016/j.cag.2018.05.011).
- [248] A. D. Parakkat, S. Joshi, H. Gowtham, and R. Muthuganapathy. An interactive assistant for sketching and shading. *Visual Computing for Industry, Biomedicine, and Art*, 3(15), 06 2020. doi:[10.1186/s42492-020-00049-7](https://doi.org/10.1186/s42492-020-00049-7).
- [249] A. D. Parakkat, M.-P. Cani, and K. Singh. Color by numbers: Interactive structuring and vectorization of sketch imagery. In *Proceedings of the 2021 CHI Conference on Human Factors in*

- Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. doi:[10.1145/3411764.3445215](https://doi.org/10.1145/3411764.3445215).
- [250] A. D. Parakkat, P. Memari, and M.-P. Cani. Delaunay painting: Perceptual image colouring from raster contours with gaps. *Computer Graphics Forum*, 41(6):166–181, 2022. doi:[10.1111/cgf.14517](https://doi.org/10.1111/cgf.14517).
 - [251] S. Park, H. Intraub, D.-J. Yi, D. Widders, and M. M. Chun. Beyond the edges of a view: Boundary extension in human scene-selective visual cortex. *Neuron*, 54(2):335–342, 2007. ISSN 0896-6273. doi:[10.1016/j.neuron.2007.04.006](https://doi.org/10.1016/j.neuron.2007.04.006).
 - [252] A. Pasupathy, Y. El-Shamayleh, and D. V. Popovkina. Visual shape and object perception. 01 2018. doi:[10.1093/acrefore/9780190264086.013.75](https://doi.org/10.1093/acrefore/9780190264086.013.75).
 - [253] M. V. Peelen, E. Berlot, and F. P. de Lange. Predictive processing of scenes and objects. *Nat. Rev. Psychol.*, 3:13–26, Jan. 2024. doi:[10.1038/s44159-023-00254-0](https://doi.org/10.1038/s44159-023-00254-0).
 - [254] Y. Pei, J. Zhang, Y. Yuan, D. Yang, and F. T. Au. Ultrasonic guided wave topological data analysis for corrosion characterization of steel strand based on morse theory. *Mechanical Systems and Signal Processing*, 237:113119, 2025. ISSN 0888-3270. doi:[10.1016/j.ymssp.2025.113119](https://doi.org/10.1016/j.ymssp.2025.113119).
 - [255] J. Peng, Y. Zhou, and P. Y. Mok. Ehfusion: an efficient heterogeneous fusion model for group-based 3d human pose estimation: Ehfusion: an efficient heterogeneous fusion model for group-based 3d human pose estimation. *Vis. Comput.*, 41(8):5323–5345, Nov. 2024. ISSN 0178-2789. doi:[10.1007/s00371-024-03724-5](https://doi.org/10.1007/s00371-024-03724-5).
 - [256] F. Perdreau and P. Cavanagh. Is artists’ perception more veridical? *Frontiers in Neuroscience*, Volume 7 - 2013, 2013. ISSN 1662-453X. doi:[10.3389/fnins.2013.00006](https://doi.org/10.3389/fnins.2013.00006).
 - [257] S. Phillips. Sheaving — a universal construction for semantic compositionality. *Phil. Trans. R. Soc. B*, 375(1791), 2020. doi:[10.1098/rstb.2019.0303](https://doi.org/10.1098/rstb.2019.0303).
 - [258] Z. Pizlo and J. A. de Barros. The concept of symmetry and the theory of perception. *Frontiers in Computational Neuroscience*, 15, 2021. doi:[10.3389/fncom.2021.681162](https://doi.org/10.3389/fncom.2021.681162).
 - [259] K. Poelke and K. Polthier. Boundary-aware hodge decompositions for piecewise constant vector fields. *Computer-Aided Design*, 78:126–136, 2016. ISSN 0010-4485. doi:[10.1016/j.cad.2016.05.004](https://doi.org/10.1016/j.cad.2016.05.004).
 - [260] L. Qiu, X. Zhang, Y. Li, G. Li, X. Wu, Z. Xiong, X. Han, and S. Cui. Peeking into occluded joints: A novel framework for crowd pose estimation. In *Computer Vision – ECCV 2020*, pages 488–504, Cham, 2020. Springer International Publishing. ISBN 978-3-030-58529-7. doi:[10.1007/978-3-030-58529-7_29](https://doi.org/10.1007/978-3-030-58529-7_29).
 - [261] Y. Qiu and G.-W. Wei. Mathematics-assisted directed evolution and protein engineering. 2023. URL <https://arxiv.org/abs/2306.04658>.
 - [262] Z. Qiu, K. Qiu, J. Fu, and D. Fu. Dgcnn: Dynamic graph convolutional network for efficient multi-person pose estimation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07):11924–11931, Apr. 2020. doi:[10.1609/aaai.v34i07.6867](https://doi.org/10.1609/aaai.v34i07.6867).
 - [263] N. Quercioli. On the topological theory of group equivariant non-expansive operators. 2021. URL <https://amsdottorato.unibo.it/id/eprint/9770/1/Tesi-Dottorato-Quercioli.pdf>.
 - [264] G. Rai and O. Sharma. Sketch animation: State-of-the-art report. 2025. URL <https://arxiv.org/abs/2510.10218>.
 - [265] S. Ramos, D. F. Trevisan, H. C. Batagelo, M. Costa Sousa, and J. P. Gois. Contour-aware 3d reconstruction of side-view sketches. *Computers & Graphics*, 77:97–107, 2018. ISSN 0097-8493. doi:[10.1016/j.cag.2018.09.009](https://doi.org/10.1016/j.cag.2018.09.009).
 - [266] F. Razafindrazaka, K. Poelke, K. Polthier, and L. Goubergrits. A consistent discrete 3d hodge-type

- decomposition: implementation and practical evaluation. 11 2019. doi:[10.48550/arXiv.1911.12173](https://doi.org/10.48550/arXiv.1911.12173).
- [267] M. Reuter, F.-E. Wolter, and N. Peinecke. Laplace-beltrami spectra as ‘shape-dna’ of surfaces and solids. *Computer-Aided Design*, 38(4):342–366, 2006. ISSN 0010-4485. doi:[10.1016/j.cad.2005.10.011](https://doi.org/10.1016/j.cad.2005.10.011). Symposium on Solid and Physical Modeling 2005.
 - [268] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11):1019–1025, Nov. 1999. doi:[10.1038/14819](https://doi.org/10.1038/14819).
 - [269] A. Rivers, T. Igarashi, and F. Durand. 2.5d cartoon models. *ACM Trans. Graph.*, 29(4), July 2010. ISSN 0730-0301. doi:[10.1145/1778765.1778796](https://doi.org/10.1145/1778765.1778796).
 - [270] M. Robinson. *Topological Signal Processing*. Mathematical Engineering. Springer Berlin, Heidelberg, 1 edition, January 2014. doi:[10.1007/978-3-642-36104-3](https://doi.org/10.1007/978-3-642-36104-3).
 - [271] M. Robinson. Sheaves are the canonical data structure for sensor integration. *Information Fusion*, 36:208–224, 2017. ISSN 1566-2535. doi:[10.1016/j.inffus.2016.12.002](https://doi.org/10.1016/j.inffus.2016.12.002).
 - [272] M. Robinson. Assignments to sheaves of pseudometric spaces. *Compositionality*, 2:2, 06 2020. doi:[10.32408/compositionality-2-2](https://doi.org/10.32408/compositionality-2-2).
 - [273] M. Robinson and C. T. Capraro. Super-resolving star clusters with sheaves. *EURASIP Journal on Advances in Signal Processing*, 2022(1):26, Mar. 2022. doi:[10.1186/s13634-022-00861-8](https://doi.org/10.1186/s13634-022-00861-8).
 - [274] E. Rosales, J. Rodriguez, and A. SHEFFER. Surfacebrush: from virtual reality drawings to manifold surfaces. *ACM Trans. Graph.*, 38(4), July 2019. ISSN 0730-0301. doi:[10.1145/3306346.3322970](https://doi.org/10.1145/3306346.3322970).
 - [275] D. Rosiak. *Sheaf Theory through Examples*. The MIT Press, 10 2022. ISBN 9780262370424. doi:[10.7551/mitpress/12581.001.0001](https://doi.org/10.7551/mitpress/12581.001.0001).
 - [276] R. M. Rustamov. Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *Proceedings of the Fifth Eurographics Symposium on Geometry Processing*, SGP ’07, page 225–233, Goslar, DEU, 2007. Eurographics Association. ISBN 9783905673463. doi:[10.5555/1281991.1282022](https://doi.org/10.5555/1281991.1282022).
 - [277] M. Sablé-Meyer, L. Benjamin, C. P. Watkins, C. He, M. Pajot, T. Morfioise, F. Al Roumi, and S. Dehaene. A geometric shape regularity effect in the human brain. May 2025. doi:[10.7554/elife.106464.1](https://doi.org/10.7554/elife.106464.1).
 - [278] E. U. Samani and A. G. Banerjee. Persistent homology meets object unity: Object recognition in clutter. *IEEE Transactions on Robotics*, 40:886–902, 2024. doi:[10.1109/TRO.2023.3343994](https://doi.org/10.1109/TRO.2023.3343994).
 - [279] K. Sato, Y. Matsui, T. Yamasaki, and K. Aizawa. Reference-based manga colorization by graph correspondence using quadratic programming. In *SIGGRAPH Asia 2014 Technical Briefs*, SA ’14, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450328951. doi:[10.1145/2669024.2669037](https://doi.org/10.1145/2669024.2669037).
 - [280] B. Sayim and P. Cavanagh. What line drawings reveal about the visual brain. *Frontiers in Human Neuroscience*, volume 5 - 2011, 2011. ISSN 1662-5161. doi:[10.3389/fnhum.2011.00118](https://doi.org/10.3389/fnhum.2011.00118).
 - [281] A. Schneuing, C. Harris, Y. Du, K. Didi, A. Jamasb, I. Igashov, W. Du, C. Gomes, T. L. Blundell, P. Lio, M. Welling, M. Bronstein, and B. Correia. Structure-based drug design with equivariant diffusion models. *Nature Computational Science*, 4(12):899–909, Dec. 2024. doi:[10.1038/s43588-024-00737-x](https://doi.org/10.1038/s43588-024-00737-x).
 - [282] L. Scocchia, M. Valsecchi, and J. Triesch. Top-down influences on ambiguous perception: the role of stable and transient states of the observer. *Frontiers in Human Neuroscience*, Volume 8 - 2014, 2014. doi:[10.3389/fnhum.2014.00979](https://doi.org/10.3389/fnhum.2014.00979).
 - [283] S. Sconza and A. Wildi. A knot-based key exchange protocol. 2024. URL <https://eprint.iacr.org/2024/471>.

- [284] J. Shen, K. Hu, W. Bao, C. W. Chen, and Z. Wang. Bridging the gap: Sketch-aware interpolation network for high-quality animation sketch inbetweening. In *Proceedings of the 32nd ACM International Conference on Multimedia*, MM '24, page 10287–10295, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400706868. doi:[10.1145/3664647.3681146](https://doi.org/10.1145/3664647.3681146).
- [285] L. Shen, H. Feng, F. Li, F. Lei, J. Wu, and G.-W. Wei. Knot data analysis using multiscale gauss link integral. *Proceedings of the National Academy of Sciences*, 121(42):e2408431121, 2024. doi:[10.1073/pnas.2408431121](https://doi.org/10.1073/pnas.2408431121).
- [286] H. Sheng, J. Wilder, and D. B. Walther. Where to draw the line? *PLOS ONE*, 16(11):1–16, 11 2021. doi:[10.1371/journal.pone.0258376](https://doi.org/10.1371/journal.pone.0258376).
- [287] R. Short, A. Hylton, R. Cardona, R. Green, G. Bainbridge, M. Moy, and J. Cleveland. Towards sheaf theoretic analyses for delay tolerant networking. In *2021 IEEE Aerospace Conference (50100)*, pages 1–9, 2021. doi:[10.1109/AERO50100.2021.9438167](https://doi.org/10.1109/AERO50100.2021.9438167).
- [288] A. Singer and H.-T. Wu. Vector diffusion maps and the connection laplacian. *Communications on Pure and Applied Mathematics*, 65(8):1067–1144, 2012. doi:[10.1002/cpa.21395](https://doi.org/10.1002/cpa.21395).
- [289] J. J. Singer, R. M. Cichy, and M. N. Hebart. The spatiotemporal neural dynamics of object recognition for natural images and line drawings. *Journal of Neuroscience*, 43(3):484–500, 2023. ISSN 0270-6474. doi:[10.1523/JNEUROSCI.1546-22.2022](https://doi.org/10.1523/JNEUROSCI.1546-22.2022).
- [290] G. Singh, F. Memoli, and G. Carlsson. Topological methods for the analysis of high dimensional data sets and 3d object recognition. In *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007. ISBN 978-3-905673-51-7. doi:[10.2312/SPBG/SPBG07/091-100](https://doi.org/10.2312/SPBG/SPBG07/091-100).
- [291] C. S. Siri. Enhancing cartoon recognition in real time: Comparative analysis of cnn, resnet50, and vgg16 deep learning models. In *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, pages 114–121, 2023. doi:[10.1109/ICAISS58487.2023.10250463](https://doi.org/10.1109/ICAISS58487.2023.10250463).
- [292] L. Siyao, T. Gu, W. Xiao, H. Ding, Z. Liu, and C. C. Loy. Deep geometrized cartoon line inbetweening. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7257–7266, 2023. doi:[10.1109/ICCV51070.2023.00670](https://doi.org/10.1109/ICCV51070.2023.00670).
- [293] H. J. Smith, Q. Zheng, Y. Li, S. Jain, and J. K. Hodgins. A method for animating children’s drawings of the human figure. *ACM Trans. Graph.*, 42(3), June 2023. ISSN 0730-0301. doi:[10.1145/3592788](https://doi.org/10.1145/3592788).
- [294] C. Solomon. The animated world of norman mclaren. *Los Angeles Times*, February 1987. URL <https://www.latimes.com/archives/la-xpm-1987-02-08-ca-1055-story.html>.
- [295] Z. Song, J. Yu, C. Zhou, and M. Wang. Automatic cartoon matching in computer-assisted animation production. *Neurocomputing*, 120:397–403, 2013. ISSN 0925-2312. doi:[10.1016/j.neucom.2012.08.051](https://doi.org/10.1016/j.neucom.2012.08.051). Image Feature Detection and Description.
- [296] D. Spielman. *Spectral and Algebraic Graph Theory Incomplete Draft, dated February 2, 2025*. Yale, New Haven, 2025. URL <http://cs-www.cs.yale.edu/homes/spielman/sagt/sagt.pdf>.
- [297] B. J. Stankiewicz and J. E. Hummel. Categorical relations in shape perception. *Spatial Vision*, 10 (3):201 – 236, 1996. doi:[10.1163/156856896X00141](https://doi.org/10.1163/156856896X00141).
- [298] Z. Su, Y. Tong, and G. Wei. Hodge decomposition of vector fields in cartesian grids. In *SIGGRAPH Asia 2024 Conference Papers*, SA '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400711312. doi:[10.1145/3680528.3687602](https://doi.org/10.1145/3680528.3687602).
- [299] Z. Su, X. Liu, L. B. Hamdan, V. Maroulas, J. Wu, G. Carlsson, and G.-W. Wei. Topological data analysis and topological deep learning beyond persistent homology – a review. 2025. URL <https://arxiv.org/abs/2507.19504>.

- [300] K. Sun, B. Xiao, D. Liu, and J. Wang. Deep high-resolution representation learning for human pose estimation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5686–5696, 2019. doi:[10.1109/CVPR.2019.00584](https://doi.org/10.1109/CVPR.2019.00584).
- [301] V. Surazhsky and C. Gotsman. High quality compatible triangulations. *Engineering with Computers*, 20(2):147–156, July 2004. doi:[10.1007/s00366-004-0282-6](https://doi.org/10.1007/s00366-004-0282-6).
- [302] D. Šýkora, J. Dingliana, and S. Collins. As-rigid-as-possible image registration for hand-drawn cartoon animations. In *Proceedings of the 7th International Symposium on Non-Photorealistic Animation and Rendering*, NPAR '09, page 25–33, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605586045. doi:[10.1145/1572614.1572619](https://doi.org/10.1145/1572614.1572619).
- [303] D. Sykora, D. Sedlacek, S. Jinchao, J. Dingliana, and S. Collins. Adding depth to cartoons using sparse depth (in)equalities. *Computer Graphics Forum*, 2010. ISSN 1467-8659. doi:[10.1111/j.1467-8659.2009.01631.x](https://doi.org/10.1111/j.1467-8659.2009.01631.x).
- [304] D. Šýkora, M. Ben-Chen, M. Čadík, B. Whited, and M. Simmons. Textoons: practical texture mapping for hand-drawn cartoon animations. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*, NPAR '11, page 75–84, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450309073. doi:[10.1145/2024676.2024689](https://doi.org/10.1145/2024676.2024689).
- [305] D. Šýkora, L. Kavan, M. Čadík, O. Jamriška, A. Jacobson, B. Whited, M. Simmons, and O. Sorkine-Hornung. Ink-and-ray: Bas-relief meshes for adding global illumination effects to hand-drawn characters. *ACM Transaction on Graphics*, 33(2):16, 2014. doi:[10.1145/2591011](https://doi.org/10.1145/2591011).
- [306] D. Šýkora, J. Dingliana, and S. Collins. Lazybrush: Flexible painting tool for hand-drawn cartoons. *Computer Graphics Forum*, 28(2):599–608, 2009. doi:[10.1111/j.1467-8659.2009.01400.x](https://doi.org/10.1111/j.1467-8659.2009.01400.x).
- [307] F. M. Taiwo, U. Islambekov, and C. G. Akcora. Explaining the power of topological data analysis in graph machine learning. 2024. URL <https://arxiv.org/abs/2401.04250>.
- [308] W. Tang, P. Yu, and Y. Wu. Deeply learned compositional models for human pose estimation. In *Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part III*, page 197–214, Berlin, Heidelberg, 2018. Springer-Verlag. ISBN 978-3-030-01218-2. doi:[10.1007/978-3-030-01219-9_12](https://doi.org/10.1007/978-3-030-01219-9_12).
- [309] Y. Tang, J. Guo, P. Liu, Z. Wang, H. Hua, J.-X. Zhong, Y. Xiao, C. Huang, L. Song, S. Liang, Y. Song, L. He, J. Bi, M. Feng, X. Li, Z. Zhang, and C. Xu. Generative ai for cel-animation: A survey. 2025. URL <https://arxiv.org/abs/2501.06250>.
- [310] L. Teichmann, G. Edwards, and C. I. Baker. Resolving visual motion through perceptual gaps. *Trends in Cognitive Sciences*, 25(11):978–991, 2021. ISSN 1364-6613. doi:[10.1016/j.tics.2021.07.017](https://doi.org/10.1016/j.tics.2021.07.017).
- [311] F. Thomas and O. Johnston. *The Illusion of Life: Disney Animation*. Hyperion, New York, 1995.
- [312] J. Tompson, A. Jain, Y. LeCun, and C. Bregler. Joint training of a convolutional network and a graphical model for human pose estimation. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1, NIPS'14*, page 1799–1807, Cambridge, MA, USA, 2014. MIT Press. doi:[10.5555/2968826.2969027](https://doi.org/10.5555/2968826.2969027).
- [313] A. Tozzi and J. Peters. Topology of human perception. 03 2019. doi:[10.13140/RG.2.2.24396.05764](https://doi.org/10.13140/RG.2.2.24396.05764).
- [314] K. Turner, S. Mukherjee, and D. M. Boyer. Persistent homology transform for modeling shapes and surfaces. *Information and Inference: A Journal of the IMA*, 3(4):310–344, 12 2014. ISSN 2049-8764. doi:[10.1093/imaiai/iau011](https://doi.org/10.1093/imaiai/iau011).
- [315] G. Unlu, M. Sayed, and G. Brostow. Interactive sketching of mannequin poses. In *2022 International Conference on 3D Vision (3DV)*, pages 700–710, Los Alamitos, CA, USA, Sept. 2022. IEEE Computer

- Society. doi:[10.1109/3DV57658.2022.00080](https://doi.org/10.1109/3DV57658.2022.00080).
- [316] C. Van Meel, A. Baeck, C. R. Gillebert, J. Wagemans, and H. P. Op de Beeck. The representation of symmetry in multi-voxel response patterns and functional connectivity throughout the ventral visual stream. *NeuroImage*, 191:216–224, 2019. ISSN 1053-8119. doi:[10.1016/j.neuroimage.2019.02.030](https://doi.org/10.1016/j.neuroimage.2019.02.030).
 - [317] C. Walker, E. Winner, L. Hetland, S. Simmons III, and L. Goldsmith. Visual thinking: Art students have an advantage in geometric reasoning. *Creative Education*, 2:22–26, 01 2011. doi:[10.4236/ce.2011.21004](https://doi.org/10.4236/ce.2011.21004).
 - [318] D. B. Walther and D. Shen. Nonaccidental properties underlie human categorization of complex natural scenes. *Psychological Science*, 25(4):851–860, 2014. doi:[10.1177/0956797613512662](https://doi.org/10.1177/0956797613512662).
 - [319] D. B. Walther, B. Chai, E. Caddigan, D. M. Beck, and L. Fei-Fei. Simple line drawings suffice for functional mri decoding of natural scene categories. *Proceedings of the National Academy of Sciences*, 108(23):9661–9666, 2011. doi:[10.1073/pnas.1015666108](https://doi.org/10.1073/pnas.1015666108).
 - [320] B. Wang, T. Sudijono, H. Kirveslahti, T. Gao, D. M. Boyer, S. Mukherjee, and L. Crawford. A statistical pipeline for identifying physical features that differentiate classes of 3d shapes. *The Annals of Applied Statistics*, 15(2):638–661, 2021. doi:[10.1214/20-AOAS1430](https://doi.org/10.1214/20-AOAS1430).
 - [321] F. Wang, K. Tang, H. Wu, B. Zhao, H. Cai, and T. Zhou. Sketchbodynet: A sketch-driven multi-faceted decoder network for 3d human reconstruction. In *31th Pacific Conference on Computer Graphics and Applications, Pacific Graphics 2023*, 10 2023. doi:[10.2312/pg.20231266](https://doi.org/10.2312/pg.20231266).
 - [322] J. Wang and J. Yu. Correspondence construction for cartoon animation via sparse coding. In *Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service, ICIMCS '13*, page 277–282, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450322522. doi:[10.1145/2499788.2499861](https://doi.org/10.1145/2499788.2499861).
 - [323] K. Wang, J. Zheng, and H. S. Seah. Progressive sketching with instant previewing. *Computers & Graphics*, 81:9–19, 2019. ISSN 0097-8493. doi:[10.1016/j.cag.2019.03.017](https://doi.org/10.1016/j.cag.2019.03.017).
 - [324] X. Wang, W. Yang, H. Peng, and G. Wang. Shape-aware skeletal deformation for 2d characters. *The Visual Computer*, 29, 06 2013. doi:[10.1007/s00371-013-0817-1](https://doi.org/10.1007/s00371-013-0817-1).
 - [325] X. Wang, W. Yang, W. Kou, B. Yang, and G. Wang. Topology-aware moving least square deformation for 2d characters. *Computer Animation and Virtual Worlds*, 27(6):546–555, 2016. doi:[10.1002/cav.1675](https://doi.org/10.1002/cav.1675).
 - [326] D. M. Watson, T. Hartley, and T. J. Andrews. Patterns of response to scrambled scenes reveal the importance of visual properties in the organization of scene-selective cortex. *Cortex*, 92:162–174, 2017. ISSN 0010-9452. doi:[10.1016/j.cortex.2017.04.011](https://doi.org/10.1016/j.cortex.2017.04.011).
 - [327] O. Weber and D. Zorin. Locally injective parametrization with arbitrary fixed boundaries. *ACM Trans. Graph.*, 33(4), July 2014. ISSN 0730-0301. doi:[10.1145/2601097.2601227](https://doi.org/10.1145/2601097.2601227).
 - [328] J. Wee and K. Xia. Persistent spectral based ensemble learning (perspect-el) for protein–protein binding affinity prediction. *Briefings in Bioinformatics*, 23(2):bbac024, 02 2022. ISSN 1477-4054. doi:[10.1093/bib/bbac024](https://doi.org/10.1093/bib/bbac024).
 - [329] X. Wei and G.-W. Wei. Persistent topological laplacians — a survey. *Mathematics*, 13(2), 2025. ISSN 2227-7390. doi:[10.3390/math13020208](https://doi.org/10.3390/math13020208).
 - [330] B. Whited, G. Noris, M. Simmons, R. W. Sumner, M. Gross, and J. Rossignac. Betweenit: An interactive tool for tight inbetweening. *Computer Graphics Forum*, 29(2):605–614, 2010. doi:[10.1111/j.1467-8659.2009.01630.x](https://doi.org/10.1111/j.1467-8659.2009.01630.x).
 - [331] J. Wilder, M. Rezanejad, S. Dickinson, K. Siddiqi, A. Jepson, and D. Walther. Lo-

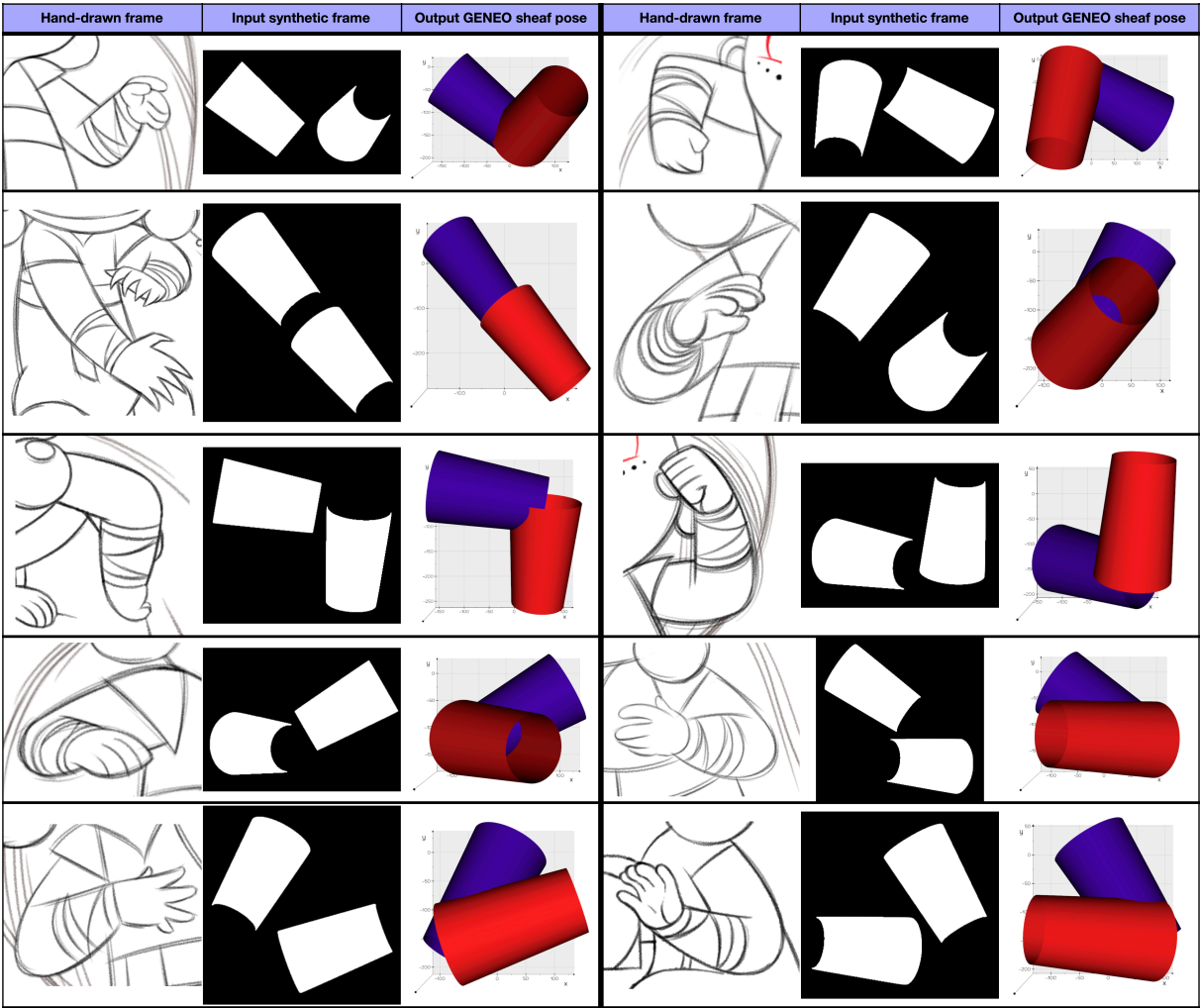
- cal contour symmetry facilitates scene categorization. *Cognition*, 182:307–317, 01 2019. doi:[10.1016/j.cognition.2018.09.014](https://doi.org/10.1016/j.cognition.2018.09.014).
- [332] J. D. Wilder, W. J. Adams, and R. F. Murray. Shape from shading under inconsistent illumination. *Journal of Vision*, 19(6):2–2, 06 2019. ISSN 1534-7362. doi:[10.1167/19.6.2](https://doi.org/10.1167/19.6.2).
- [333] R. Williams. *The Animator’s Survival Kit—Revised Edition: A Manual of Methods, Principles and Formulas for Classical, Computer, Games, Stop Motion and Internet Animators*. Faber & Faber, Inc., 2009. ISBN 0571238343.
- [334] A. Wolf, J. Fan, and A. Monod. Generalized persistent laplacians and their spectral properties. 2025. URL <https://arxiv.org/abs/2509.20220>.
- [335] L. Xian, H. Adams, C. M. Topaz, and L. Ziegelmeier. Capturing dynamics of time-varying data via topology. *Foundations of Data Science*, 4(1):1–36, 2022. doi:[10.3934/fods.2021033](https://doi.org/10.3934/fods.2021033).
- [336] J. Xing, L.-Y. Wei, T. Shiratori, and K. Yatani. Autocomplete hand-drawn animations. *ACM Trans. Graph.*, 34(6), Nov. 2015. ISSN 0730-0301. doi:[10.1145/2816795.2818079](https://doi.org/10.1145/2816795.2818079).
- [337] B. Xu, W. Chang, A. Sheffer, A. Bousseau, J. McCrae, and K. Singh. True2form: 3d curve networks from 2d sketches via selective regularization. *ACM Trans. Graph.*, 33(4), July 2014. ISSN 0730-0301. doi:[10.1145/2601097.2601128](https://doi.org/10.1145/2601097.2601128).
- [338] P. Xu, T. M. Hospedales, Q. Yin, Y.-Z. Song, T. Xiang, and L. Wang. Deep learning for free-hand sketch: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):285–312, 2023. doi:[10.1109/TPAMI.2022.3148853](https://doi.org/10.1109/TPAMI.2022.3148853).
- [339] X. Xu, Q. Zou, and X. Lin. Alleviating human-level shift: A robust domain adaptation method for multi-person pose estimation. In *Proceedings of the 28th ACM International Conference on Multimedia*, MM ’20, page 2326–2335, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450379885. doi:[10.1145/3394171.3414040](https://doi.org/10.1145/3394171.3414040).
- [340] X. Xu, Q. Zou, and X. Lin. Adaptive hypergraph neural network for multi-person pose estimation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(3):2955–2963, Jun. 2022. doi:[10.1609/aaai.v36i3.20201](https://doi.org/10.1609/aaai.v36i3.20201).
- [341] X. Xu, Q. Zou, and X. Lin. Structure-enriched topology learning for cross-domain multi-person pose estimation. *IEEE Transactions on Multimedia*, 25:6272–6284, 2023. doi:[10.1109/TMM.2022.3207578](https://doi.org/10.1109/TMM.2022.3207578).
- [342] C. Yan, D. Vanderhaeghe, and Y. Gingold. A benchmark for rough sketch cleanup. *ACM Trans. Graph.*, 39(6), Nov. 2020. ISSN 0730-0301. doi:[10.1145/3414685.3417784](https://doi.org/10.1145/3414685.3417784).
- [343] L. Yan, T. B. Masood, R. Sridharamurthy, F. Rasheed, V. Natarajan, I. Hotz, and B. Wang. Scalar field comparison with topological descriptors: Properties and applications for scientific visualization. *Computer Graphics Forum*, 40(3):599–633, 2021. doi:[10.1111/cgf.14331](https://doi.org/10.1111/cgf.14331).
- [344] W. Yang. Context-aware computer aided inbetweening. *IEEE Transactions on Visualization and Computer Graphics*, 24(2):1049–1062, 2018. doi:[10.1109/TVCG.2017.2657511](https://doi.org/10.1109/TVCG.2017.2657511).
- [345] W. Yang, J. Feng, and X. Wang. Structure preserving manipulation and interpolation for multi-element 2d shapes. *Computer Graphics Forum*, 31(7):2249–2258, 2012. doi:[10.1111/j.1467-8659.2012.03218.x](https://doi.org/10.1111/j.1467-8659.2012.03218.x).
- [346] W. Yang, H.-S. Seah, Q. Chen, H.-Z. Liew, and D. Šýkora. Ftp-sc: Fuzzy topology preserving stroke correspondence. *Computer Graphics Forum*, 37(8):125–135, 2018. doi:[10.1111/cgf.13518](https://doi.org/10.1111/cgf.13518).
- [347] W. Yang, J. Hua, and K.-Y. Yao. Cr-morph: Controllable rigid morphing for 2d animation. *Journal of Computer Science and Technology*, 34:1109–1122, 09 2019. doi:[10.1007/s11390-019-1963-3](https://doi.org/10.1007/s11390-019-1963-3).

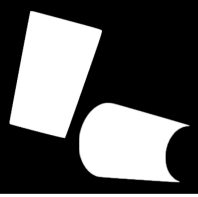
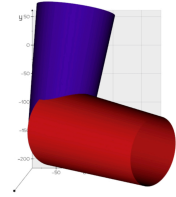
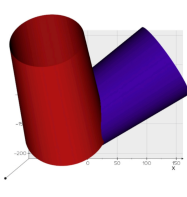
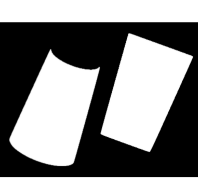
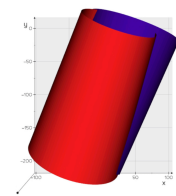
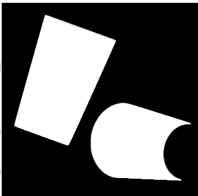
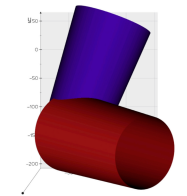
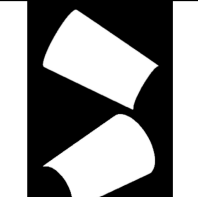
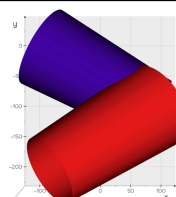
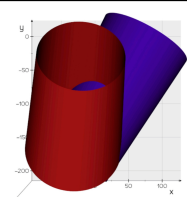
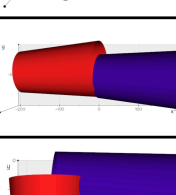
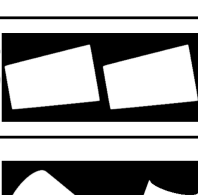
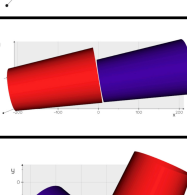
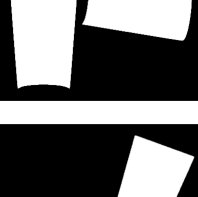
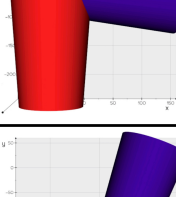
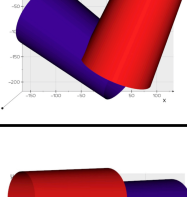
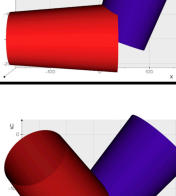
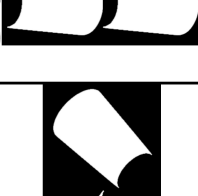
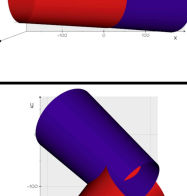
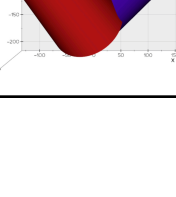
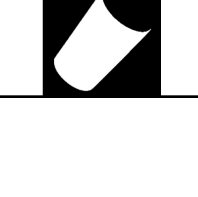
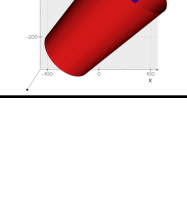
- [348] X. Yang, X. Zhao, and Z. Shen. Are heterogeneous graph neural networks truly effective? a causal perspective. 2025. URL <https://arxiv.org/abs/2510.05750>.
- [349] C.-K. Yeh, P. Jayaraman, X. Liu, C.-W. Fu, and T.-Y. Lee. 2.5d cartoon hair modeling and manipulation. *Visualization and Computer Graphics, IEEE Transactions on*, 21:304–314, 03 2015. doi:[10.1109/TVCG.2014.2360406](https://doi.org/10.1109/TVCG.2014.2360406).
- [350] J. Yin, C. Liu, R. Lin, N. Vining, H. Rhodin, and A. Sheffer. Detecting viewer-perceived intended vector sketch connectivity. *ACM Trans. Graph.*, 41(4), July 2022. ISSN 0730-0301. doi:[10.1145/3528223.3530097](https://doi.org/10.1145/3528223.3530097).
- [351] C. Yu, B. Xiao, C. Gao, L. Yuan, L. Zhang, N. Sang, and J. Wang. Lite-HRNet: A Lightweight High-Resolution Network . In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10435–10445, Los Alamitos, CA, USA, June 2021. IEEE Computer Society. doi:[10.1109/CVPR46437.2021.01030](https://doi.org/10.1109/CVPR46437.2021.01030).
- [352] J. Yu, D. Liu, D. Tao, and H. S. Seah. Complex object correspondence construction in two-dimensional animation. *IEEE Transactions on Image Processing*, 20(11):3257–3269, 2011. doi:[10.1109/TIP.2011.2158225](https://doi.org/10.1109/TIP.2011.2158225).
- [353] J. Yu, W. Bian, M. Song, J. Cheng, and D. Tao. Graph based transductive learning for cartoon correspondence construction. *Neurocomputing*, 79:105–114, 2012. ISSN 0925-2312. doi:[10.1016/j.neucom.2011.10.003](https://doi.org/10.1016/j.neucom.2011.10.003).
- [354] A. L. Yuille and C. Liu. Deep nets: What have they ever done for vision? *International Journal of Computer Vision*, 129(3):781–802, Mar. 2021. doi:[10.1007/s11263-020-01405-z](https://doi.org/10.1007/s11263-020-01405-z).
- [355] O. Zaghen, A. Longa, S. Azzolin, L. Telyatnikov, A. Passerini, and P. Lio. Sheaf diffusion goes nonlinear: Enhancing GNNs with adaptive sheaf laplacians. In *ICML 2024 Workshop on Geometry-grounded Representation Learning and Generative Modeling*, 2024. URL <https://openreview.net/forum?id=MGQtGV5gP0>.
- [356] C. Zhang, L. Yang, N. Chen, N. Vining, A. Sheffer, F. C. Lau, G. Wang, and W. Wang. Creatureshop: Interactive 3d character modeling and texturing from a single color drawing. *IEEE Transactions on Visualization and Computer Graphics*, 29(12):4874–4890, Aug. 2022. ISSN 1077-2626. doi:[10.1109/TVCG.2022.3197560](https://doi.org/10.1109/TVCG.2022.3197560).
- [357] H. Zhang, H. Ouyang, S. Liu, X. Qi, X. Shen, R. Yang, and J. Jia. Human pose estimation with spatial contextual information. 2019. URL <https://arxiv.org/abs/1901.01760>.
- [358] H. Zhao, D. Wang, and L. Liu. Neuromodulatory developmental learning of the mobile robots corresponding to the unexpected obstacles. *Cogn. Syst. Res.*, 88(C), Dec. 2024. ISSN 1389-0417. doi:[10.1016/j.cogsys.2024.101296](https://doi.org/10.1016/j.cogsys.2024.101296).
- [359] L. Zhao, X. Peng, Y. Tian, M. Kapadia, and D. N. Metaxas. Semantic graph convolutional networks for 3d human pose regression. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3420–3430, 2019. doi:[10.1109/CVPR.2019.00354](https://doi.org/10.1109/CVPR.2019.00354).
- [360] R. Zhao, M. Desbrun, G.-W. Wei, and Y. Tong. 3d hodge decompositions of edge- and face-based vector fields. *ACM Trans. Graph.*, 38(6), Nov. 2019. ISSN 0730-0301. doi:[10.1145/3355089.3356546](https://doi.org/10.1145/3355089.3356546).
- [361] Y. Zhao, D. Ren, Y. Chen, W. Jia, R. Wang, and X. Liu. Cartoon image processing: A survey. *Int. J. Comput. Vision*, 130(11):2733–2769, Nov. 2022. ISSN 0920-5691. doi:[10.1007/s11263-022-01645-1](https://doi.org/10.1007/s11263-022-01645-1).
- [362] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, and M. Shah. Deep learning-based human pose estimation: A survey. *ACM Comput. Surv.*, 56(1), Aug. 2023. ISSN 0360-0300. doi:[10.1145/3603618](https://doi.org/10.1145/3603618).

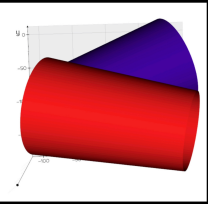
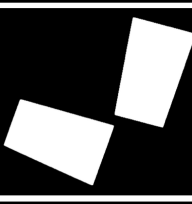
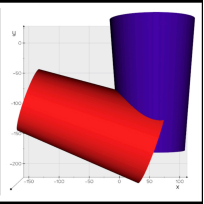
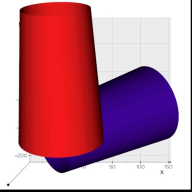
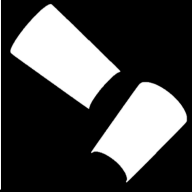
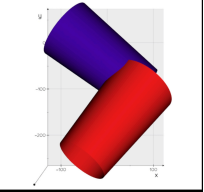
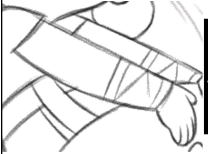
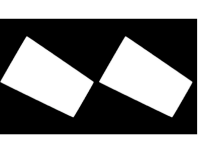
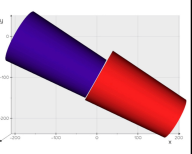
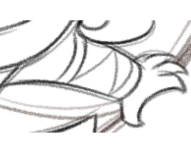
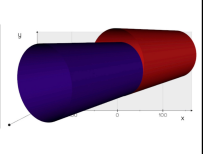
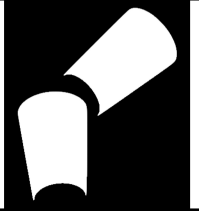
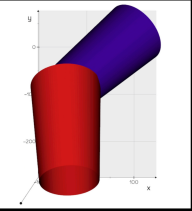
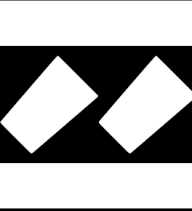
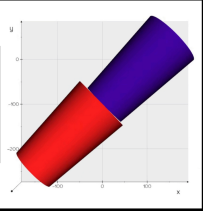
- [363] K. Zhou, J. Huang, J. Snyder, X. Liu, H. Bao, B. Guo, and H.-Y. Shum. Large mesh deformation using the volumetric graph laplacian. *ACM Trans. Graph.*, 24(3):496–503, July 2005. ISSN 0730-0301. doi:[10.1145/1073204.1073219](https://doi.org/10.1145/1073204.1073219).
- [364] Q. Zhou. Pymesh-geometry processing library for pytho. 2020. URL <https://github.com/PyMesh/PyMesh>.
- [365] H. Zhu, X. Liu, T.-T. Wong, and P.-A. Heng. Globally optimal toon tracking. *ACM Trans. Graph.*, 35(4), July 2016. ISSN 0730-0301. doi:[10.1145/2897824.2925872](https://doi.org/10.1145/2897824.2925872).
- [366] T. Zhu, W. Shang, D. Ren, and W. Zuo. Thin-plate spline-based interpolation for animation line inbetweening. 2024. URL <https://arxiv.org/abs/2408.09131>.
- [367] Y. Zhu, J. Popović, R. Bridson, and D. M. Kaufman. Planar interpolation with extreme deformation, topology change and dynamics. *ACM Trans. Graph.*, 36(6), Nov. 2017. ISSN 0730-0301. doi:[10.1145/3130800.3130820](https://doi.org/10.1145/3130800.3130820).

Appendix A

RFQ GENEO Sheaf Implementation
Results



Hand-drawn frame	Input synthetic frame	Output GENEO sheaf pose	Hand-drawn frame	Input synthetic frame	Output GENEO sheaf pose
					
					
					
					
					
					
					

Hand-drawn frame	Input synthetic frame	Output GENEO sheaf pose	Hand-drawn frame	Input synthetic frame	Output GENEO sheaf pose
					
					
					
					

Appendix B

Interviews with Hand-drawn Animation Experts

B.1 Glen Keane Interview Transcript

Glen Keane is a master Disney animator, executive producer, and director whose credits include The Little Mermaid, Beauty and the Beast, Aladdin, and Tangled.

This interview has been edited for clarity.

Glen Keane: When we went to Google [to make the hand-drawn animated short *Duet*], I thought, ‘Oh, this is going to be a wonderful thing. They’re going to have everything I need.’ They had nothing. They gave us a cardboard box with wires, some patch cords, and a couple monitors. They said, “What do you need? Do you need a mouse?”

My son, Max, is a bridge person. Bridge people are those that connect people who draw or do something [where] the technology is something very simple, like a pencil, to math. He’s not a mathematician, but he’s understood how to translate drawing into that. He was my bridge. And bridge people are so, so essential. They are the translators from one medium to the other.

When we, at Disney, were moving to computer animation, it was a horrible couple years of distrust. You’d get on an elevator, one side of the elevator were people working on coding and the effects, but they were the mathematicians, and on the other side were the animators. And they didn’t know how to relate.

[Max] just kept saying yes. He was constantly being asked, can you animate this? You know, add some CG [(computer graphics, i.e., 3D animation)] or some kind of after effects or whatever. And he had no idea how. He went to Ballyfermot [Animation College], but quit that pretty quick and played in a punk band in Dublin at the Amen Dorms. But he had this attitude of like, ‘I will make it. I will survive.’ And so he would say, “Yes, I can do that.” And then he’d go home in a panic, and download tutorials and figure out how to do it. Even now, he says, “I’m always just barely staying ahead of the wolves.”

Rachael Schwartz: As you draw, what is your mental process for conveying a character’s volumetric form and movement while also staying on-model [(an animation term for preserving model sheet character design)]? You know, that tension between: you’re expressing a physical system that the viewer has to understand instantaneously, but you also have this design that’s so stylistic and so grounded in the fact that it’s two-dimensional with the way the shapes are connected and their configuration. What is that

balancing and negotiation process like for you?

GK: That's a great question. There was one point where I really was confronted with that, and it was on [the Disney film] *The Fox and the Hound*. The emotions that I was trying to get, no matter what I did with the model of this fox with the muzzle coming out, I couldn't get. It just felt like I was taking something that was very stiff and trying to bend it, and forcing it, and it never had the emotion. So I thought, 'Okay, I will not try to use the perfect three-dimensional design; I will go with the symbol.' So I just went with a circle and very broad, expressive expressions that communicated instantly. You know, from a mile away, you knew exactly that that fox was disappointed or panicking, and it was funny and it had this energy. So I'd lay that down, and then I'd take this three-dimensional knowledge that I had of the design, and put it on that. And that three-dimensional design was subservient to the symbol. And so I would force and shape that, and it was like, 'Okay, that's the fox, but he feels the way I want him to feel.' And that became the priority. It doesn't have to look like that, exactly like that, it just has to feel like that character. So I really allow the symbols to guide me in not getting too tied down.

But I've also done sculpting, and form means so much. Like one of the things that Ollie said to me was, "Every line you draw is related to another line." And he could tell, I didn't understand what he was talking about. He said, "Someday, you'll get that." But it's like, if you push this, this is a line, but then it creates another line here, and this is the form. It's defining the top and the bottom of the form. Every form is expanding and touching. And when you are sculpting, there's this wonderful tool with teeth on it, and you scrape it over the form, and it has these lines. And when I think that way of the solidity and the beauty of form, and respect it with this symbol of emotion, putting the two of them together, it's the most satisfying thing when it all fits together and I haven't lost either.

RS: I do think a lot about every line relating to another line.

GK: Well, there's another word for it – and you probably don't equate it that way, but maybe you do – it's this word 'appeal.' There's a show on Netflix called *Voir*, and one of their episodes is talking about appeal, and they focus on me talking about that. But appeal is this kind of this mathematical harmony, like music, where everything hits the perfect note. And it's just, 'Oh! Yeah.' And you get it off, and no, it doesn't work.

But I could tell that the way I started thinking about this was my wife has this sense of harmony in design, three-dimensional design. The fact that these three elements [on the table in front of me] are equidistant – that astray, that umbrella, and the phone – would be really bothering her. She wants to create a little tension. These two are together, that one's further away. So we had the salt and pepper shakers and a sugar bowl on the table. And so I thought, 'I'm going to try this out.' Linda left to do something, and then came back. But while she was gone, I slid that [salt shaker], I made it all equidistant. She sat down, and we're talking, and she just kind of [slid the shaker back to its original position]. I said, "Why'd you do that?" [She said,] "Why'd I do what?" [I said,] "You moved the salt next to the pepper there." She said, "I don't know. It just feels better."

But the harmonics of the mathematical harmonics in music, the sounds, there is the same truth in terms of form. And Freddie Moore had this sense of rhythm in his movement and the form. It just makes me so happy. That little drawing I did online where you could see a Francois Boucher drawing, and then you could see this French curve... that's what it is for me. This harmony, this melody.

RS: But the line is still your communication mechanism.

GK: Yeah.

RS: If you could magically create technologies to address challenges in hand-drawn animation workflows, what would they be? What information or technical capabilities do you wish you could have at your fingertips while you draw?

GK: Well, I know for me, the Holy Grail has always been this one thing. And no one's done it. But there's no reason why you can't. And that is: you take a sculpted figure, three-dimensional, beautiful sculpture, you work with it, you get it so it's like, 'Gosh, this is gorgeous, this feels right.' Which I would do with [Disney sculptor] Kent Melton all the time in designing. And then I'd animate. But I'm seeing dimension there, and I'm playing with the forms, like jazz. You're stretching, and you're making something feel straighter just for the energy of it. The form of a thigh, maybe, just feels heavier. That's all in there. And I just want the CG model to snap to the edge [of my hand-drawn line], so that there's a perfect invisible form that I can throw light on, I can throw fabric on, but the line still remains exactly the way I drew it. The rough line. But it has this perfect three-dimensional information on the inside. I would love to be able to just get that.

RS: So you want the 3D sculpture, or model, or whatever it is, to be animated by your drawings?

GK: Yes. The drawings become the AI intelligence for it. It would require probably scanning in a lot of my drawings, the way I've been picturing it, just teaching it to understand what these symbols mean. And then it can start to learn. And maybe it'll take six months of that, but it would know me. It would know me.

RS: Imagine a computer could understand hand-drawn characters and volumetric poses, so animators could easily search the entire film – or what portion of the film has been drawn so far – for frames matching certain volumetric poses and movements. What would you think of this kind of technology, if the computer could like follow along with the world of the film and the motion of the characters as you create them? And not just have them siloed into sheets of physical or virtual paper, organized by [the order in which they appear in the film]?

GK: This is what I've been... Yeah. I mean, when we first moved from hand-drawn to computer, my thought was, 'I want those models, the 3D models, to be subservient to my line.' And so the only way we could do it was by an animator, John Ripa, went through the... I did drawings. Well, first I animated it: Ariel, it was a [*The Little Mermaid*] thing. And then they took a CG version of her and snapped – well, it didn't snap to [my hand-drawn lines], but they tried to get as close to the [hand-drawn] pose as possible. And then we worked at them expanding and sculpting that to those lines, which is laborious and stupid for doing any kind of production. But I wanted to see, what does it feel like when you have an actual three-dimensional form there? It was really cool. It wasn't perfect, but you could have it be perfect. And what's nice is that, when you draw a movement like this, there's this little flow. I might just have this wrist and go down to that finger, maybe a couple little slashes, but you don't want everything drawn in. You want just enough. And so I'm trying to push into the area of impressionistic movement, where the computer can help me by throwing light or fabric [onto the character] that I would love to have, and I can freely draw. It will add pattern, like this [complex fabric print] would be wonderful to have. Something that I'm not going to, I don't want to draw all that, but I would love to see it.

[During my Kilkenny Animated Festival talk today,] I showed a little animation of this girl walking, for a future film that I want to do. I want that to be a CG film, but I don't want it to be a CG film. I

want it to be a hand-drawn film, but I want to benefit from the best of both worlds. As a matter of fact, that was a group that I started around 2001, 2002 at Disney, called “Best of Both Worlds,” where I got hand-drawn folks, and I got CG folks, and art directors, and everybody together said, ‘What is the best of both worlds? What would be the ideal?’ And that was what everybody agreed: it would be best if computers could really deliver exactly what you were expressing in your drawing. Or, for people that didn’t draw, to be able to move a CG pose, and it would give those dynamics and the harmony that I was talking about with appeal. It’s like, is the drawing in tune? Is the pose in harmony? Does it have this tilt, rhythm, twist? Does it have a rhythmic movement to it? These are the things that, it’s mathematics, which I don’t do, but I know that I have something in me that knows when it’s right and when it’s not.

RS: So, you’re doing this in 2001. But you sound, 23 years later, like you’re still dissatisfied, and I’m still dissatisfied too. So, what’s wrong with [hand-drawn] animation technology development and research? What’s missing? These breakthroughs and this synergy, why are they not happening?

GK: Well, I think that there’s a satisfaction level that, well, the audience likes it, so we just take it. I think [of] computer animation as a used car salesman. You walk onto a used car lot, and you want to buy a car. You’ve got something in your head, but it’s showing you something that you can afford and that it has on the lot. But it’s not at all what you want, and it’s trying to convince you: CG looks like it’s shaded just right and everything, and it’s like, ‘This is solid, this is real. Buy it.’ And I was saying, ‘Don’t buy it. Don’t buy it. Push it.’

[My talk today] was all about, ‘How do you get technology, how do you get mathematics to solve the problems you need to solve?’ It was with Rapunzel’s hair [on the Disney film *Tangled*], and that was totally impossible. I realized that instead of trying to solve it, what I was hearing was, ‘Can you just have Rapunzel not touch her hair?’ I realized the real problem is to motivate them to want to solve it, to your question. Why isn’t it [happening]? Because people don’t want to. It’s hard. It’s painful. Does anybody really care? Well, if you’re going to do a character like Rapunzel, she cares. She’s real. To me, she is. She’s real. And so I animated this in the morning, just this rough little animation of showing how connected Rapunzel was to her hair. She’s walking, she’s touching it, she’s throwing it, she’s kicking it. And I showed it to all of the coders and the effects [team] and everybody. And they’re just about having a heart attack. But they can tell, ‘I like that girl. She’s wonderful. I want her to touch her hair. I’m going to figure this out.’ And they did. And so that’s the same path I would go. Make them want, show them why, how wonderful the future will be with this.

RS: I suspect that part of the reason why hand-drawn animation has been largely set aside is that 3D animation – and all the technology created for it – dovetails with the past several decades of cultural excitement around technological advancement in the Information Age. This heightened interest in 3D animation has incentivized massive advancements in 3D animation technologies, which in turn push 3D animation forward. But we don’t see comparable advancements in hand-drawn animation technology. How do you feel about people trying to make [technology] to help push hand-drawn animation [forward], in the same way that all of these amazing advancements have pushed 3D animation, as an art form, forward?

GK: I lived in that setting aside of hand-drawn [animation]. And yet, I gotta say, part of it makes me kinda happy, because I’m so dissatisfied with [where hand-drawn animation] was going. It was living in a world that was already limited by technical limitations of painting on cells. Why? We don’t have to do that. Drawing is so important, but... I continually ask myself, ‘If you took Degas or Michelangelo and brought them to today, and did not show them any animated movies but gave them what is possible with technology, with animation, what would they do with it? Where would they take it?’ I’m talking

about specifically Michelangelo and Degas, because they both love sculpting and painting and drawing. So Degas would do a ballet dancer, and he would do a sculpture of her, and then he would do a pastel of her. He would take photos of models, and then he would draw them. He [would be] doing all of it together, but separately. And I wanna do it all at the same time. I want to do what I call ‘sculptural drawing.’ It’s sculptural drawing, and this is the goal for me.

So sculptural drawing, as I’m drawing [a character], I am completely imagining the whole three-dimensional figure that’s in there – I’m just drawing one edge of it. I came to Paris and studied sculpture, and then we did this film *Tarzan*, where I had to draw [such that] a human body was the costume. That was just so wonderful, feeling this form. Usually when I’m animating a character like Ariel [(from Disney’s *The Little Mermaid*)] or the Beast [(from Disney’s *Beauty and the Beast*)], I’m turning them in space to prove to you that they are three-dimensional. So that’s what I want: this perfect world where I can animate freely, and the understanding of a [three-]dimensional figure is being integrated into that drawing, but all that you are seeing is still just the freedom of my line.

The line is this seismograph. I think of it as the seismograph of my soul. There’s an energy to it. I mean, imagine taking a Degas ballerina and the beautiful black line, this charcoal color and this pastel, and just cleaning that up to one perfect line. You would kill it. There’s an energy that has to be preserved. And so, the reason I’m glad that [hand-drawn] 2D [animation] is not what it was any longer, is that it sets me free. [At my talk,] I was showing some of Beethoven’s ninth that I’m animating, and it has this freedom, and dimensionality, and everything that I want. And there’s just one moment in there where it’s... I’m just animating light, on black paper. So, that figure of dimensionality, there is a solid form there. But just showing parts of it that you want the light touching. When I look at that, I think, ‘Well, I don’t want, actually, the computer to be throwing the light. I want to be the one drawing that.’

[Keane plays an audio clip of Hans Zimmer speaking]: ‘We don’t use the orchestras. If we don’t figure out ways of keeping the orchestras relevant, then the orchestras are going to disappear, just like hand-drawn animation disappearing. We will lose something unfathomable in our culture.’

[GK:] “Unfathomable.” I love that.

[Zimmer:] ‘The thing that you’re not supposed to see, and I love it, is the hand of the artist. I love seeing that in every... I love when everything... Every little bit is hand-drawn, and it becomes... It’s just... For me, it’s just beautiful. That’s just like playing music.’

[Keane plays a clip of his in-progress hand-drawn animation, drawn in a loose, rough line style:] I don’t want to clean that up. No, absolutely not. But there’s a form in there; we just put in some shadow that suggests it. My son keeps saying, “Dad you don’t have to draw so much in there. You’re going to leave it a little bit looser.” I had four assistants working with me, doing the in-betweens. A lot of it was all on ones. And it was this moment, right here, I had to just stop, and get up and run.

Those are things that I will keep pushing for, but it’s not going to stop me. I’m going to do it by hand if I don’t have [this technology]. But that’s coming. That is coming.

RS: What if [hand-drawn animation] hadn’t gotten trapped where you left it? What if the limitations that are so inherent to the art form just vanished by some magical technology? How would you feel about its future then?

For example, suppose you’re not Glen Keane, and you’re just an animator working on a film. You are one person, where the form [of the character’s body] that you want to convey exists in your mind, and the feeling you want to convey exists in your mind and your body and your hand. But there are 10-20+ of you [on the animation team] all trying to meld into one expressive force for whatever character you’re

trying to introduce to the world, and put through their life and their story arc and their own emotional journey and physical journey.

The only way that that happens right now is trying to get your minds to fit together into one cohesive whole. And once you're done [drawing the film's frames], there are tens of thousands of them, and they go into a chronological organizational structure. So the world of the film gets trapped to these sheets of paper that are all trying to converge towards telling the same story from all of these different minds. That is a fundamentally overwhelming task for a team of animators to execute properly for a feature length film. And it's an overwhelming data management task. It is so much information, and the information itself is alive, and yet it's trapped in this inaccessible format, partially because there's just like so much of it. You can't possibly look through it and interact with it all at the same time. So what I'm getting at is, those are the limitations I see. I don't know if that's what you see.

GK: When you're saying that, I'm picturing *Wolfwalkers*, for example, and all of the different shots and all of the different movements and all the different angles, and the way they're drawn from behind from the side from above, close-ups, and that all of that is from a perspective. It's a three-dimensional perspective. It may be flat, but it's projecting as if it's seen from here. So then there's another one seen from here. Another one here, here, here, here, here. And if it would take some time, but to be able to put all of that and go, 'okay, this is from this perspective.' And then there's an understanding of what the character is. There's no drawing that actually says all of that, but the truth of it is actually now in the mathematical understanding of all of that.

RS: Yes, exactly. If the computer can take all of those drawings and put them together like that, then that's your character. And she lives in three dimensions. She's in the computer. I feel like that would open up so many different avenues.

GK: And Rachael, the thing that I don't think you have to solve is: what are the steps afterwards? Like if you have all of that information, you're going to discover something that you don't even know. You just move forward, just make that happen. And something wonderful is going to be there. And then you go to the next step, and then to the next step. And that's just the way this all works.

RS: What do you think those next steps would be if that was feasible computationally?

GK: Well, that's what I was thinking of when we were talking there. And I realized that if a computer had an imagination, if the computer had a full knowledge of that, it would be basically a sculpture that's living and acting and moving in with everything, from every angle that's been animated there. And there'd be gaps [in poses and view angles], because not everything has been done. You may not have a shot looking up when that character is pointing at the camera or something. But you would gain so much of an expanded knowledge of that character. Model sheets were always horrible.

B.2 Mark Mullery Interview Transcript

Mark Mullery is the Chief Technology Officer of the five-time Oscar-nominated hand-drawn animation studio Cartoon Saloon, whose credits include The Breadwinner and Wolfwalkers.

This interview has been edited for clarity.

Rachael Schwartz: When you use a lattice, you put a grid on an image – or even just part of an image – and deform the image based on the grid. This is different from having a mesh or a click-and-drag shape

deformation handle where the shape itself is parametrized, whereas with lattices, the space in which the shape lives, like the plane, is parametrized. Does lattice deformation look bad, and if so, why? What are the issues with it?

Mark Mullery: It does look bad, yeah. It mostly looks bad. I've never made the distinction between lattices and meshes before, so I'm going to use the terms interchangeably. There's a bunch of cases where we've done some really nice lattice/mesh work. I've used that quite a bit, say, when we do waterfalls. We will frequently get somebody to paint long directional brush strokes [depicting waterfall flow], and then we will fix that into an infinite loop so that it's constantly moving. Then we will take a mesh, bend it, and embed that in [to warp the looping waterfall brush strokes], and then it looks like it's flowing. That kind of stuff works: regular shapes that you're trying to put very normal bends into.

But adjusting meshes has so many problems, like where to begin. I would start by saying it has very poor sophistication of articulation. You just can't articulate within it. You're really only doing XY movements on things. You need Bezier curves, or something like that, in order to make the thing curve at all. Then you have to say when you want it to curve and when you don't, because otherwise you'll just get this angle with a 50 pence piece shape.

RS: Unless there's a smoothing function on it.

MM: You can put a smoothing function in, that's true. One of the biggest problems is that you are actually pushing and pulling pixels around, such that you're taking individual pixels and making one massive, or taking five or six and you're crushing them to the size of one. How does that even interpolate, and what do you do on top of that? So yeah, it has an awful lot of problems. In order to make articulation work, you have to make it a small element, like just the top of somebody's arm, or just the forearm, or something like that. Then you have to put rules on exactly how and where it can bend, and put limits on how much it should bend. It's not inherently a bad process because you can make it work, but there are just so many landmines around you.

I would also make one further point as well. I don't know how relevant this is to what you're saying there, but when you are taking warping and pulling around a fixed descriptive texture, for our business and brand identity, you are immediately [working exclusively] in the digital world because there is no physical analogue to doing that. Let's say you want to do paper cutout animation under the camera. One of cutout animation's defining features is that it does not bend and warp. Shapes rotate, and then are substituted or replaced [to depict volumetric motion]. You can imagine that as this physical looking thing. Once you are bending and warping textured rendered elements, you are in the computer. There is nothing wrong with that, but if you are trying to replicate something physical, the mind just doesn't believe it.

RS: Yes, as soon as you start warping a stroke, even if it's a vector stroke and its quality is preserved, it doesn't look like a human hand created it. Using either lattices or meshes to warp strokes, shapes, contours, etc. always outputs animation where something looks off, unless it's just motion trapped in the plane. That's because interpolating strokes is not equivalent to the volumetric understanding involved in interpolating occluding contours. It doesn't look the same. You just have artist-drawn physical source and target character states, and lines that morph into other lines across those inbetween frames. It's not equivalent to every new frame describing a different artist-drawn physical state of a character's body.

MM: There's a very good quote from Norman McLaren, the animator. It is something along the lines of, "Animation is not the art of drawings that move, it's the art of movements that are drawn." It sounds slightly woo-woo when you first hear it, but then you go, 'No, that is such a serious distinction.' Because

it doesn't look right. The problem with interpolation is attempting to make the drawing move, whereas drawing a movement is about choosing exactly what you put in place. Even if you do lattice or mesh interpolation on an individual stroke, that can work, but you need to use a new one on the next frame. You can't use that one again. If you keep using that one again, you've lost it. If you move between two [hand-drawn strokes via lattice/mesh interpolation], even with some pretty nice algorithms, that's maybe going to be all right for this one basis, for this one line, for this first draft of what somebody used.

I'm not saying it doesn't have value, it clearly does, but it works in small ways to start to get you somewhere. But if you took a whole character drawn that way and a whole character drawn this way and try to [mesh/lattice] in-between them, it doesn't matter how good your algorithm is, it will not be the same thing [as hand-drawn inbetweening].

RS: Even besides the technical issues with reusing deformed strokes, on a visual level, they can easily be deformed beyond the physical arc of what the human hand can draw. Especially with vector graphics. And viewers notice that difference.

MM: Yeah.

RS: The thing that I keep coming back to is the fact that a hand-drawn frame is a representation of a concept that the animator is visually communicating to the viewer. And the quality of representation is, again, artistic choices. So you have functional choices of what is being represented and how that is arising on screen, based on how the body works, etc. You have the artistic choices of what the region shapes are, how things are drawn, how things move, etc. Those are all decisions made by the animator with a lot of intention, and a lot of information moving through their head in very complex ways. And when you start to get to a point where you're telling the computer to fill in the gaps between two key frame representations, without giving the computer any understanding of what the drawing is representing, that's where you start to get problems.

I've seen a recent paper that has an animation of a character extending their arm, which [is a simple, mostly planar motion], but there's something wrong with the way that the character's upper back is shaped as the view changes slightly over intermediate frames.

MM: A lot of what animators focus on is balance and center lines and things like that. The most usual way of doing this is that the head is basically always in line with the foot. And if you are unbalancing [the character's body], unbalancing is good, but unbalancing is good specifically in the middle of a movement. It's really hard to get across to a machine what you're trying to do there. When animators have a sense of their own bodies, or of what a body does, then as they're attempting to move this arm out like this, they're intrinsically going, 'Oh, as I move, the center of my body's line is actually shifting to here. I'd better just twist the torso slightly to accommodate for that.' Or the bone shape is different [as they move]. As you look at something, more volume is revealed, basically. So all of these things, it's not that a machine can't do it. I don't mean that. But it produces things very bluntly that I think we're doing a very good job of imagining before we go to draw.

RS: Going back to the lattice deformation, when you're stretching and compressing pixels and apply smoothing functions, doesn't that create artifacts that can cause issues in DIP (Digital Ink & Paint) software? If you have color-fill up to a specific type of line, and then that type of line isn't preserved by the smoothing function on a pixel basis, that can lead you to need to go back and do corrections.

MM: Yeah. There's a general problem in DIP with micro transparencies, the specific transparency of

lines, of pixels. Anytime you draw a line and then [paint fill up to that line], the order of operation there matters because we're presuming some kind of automatic [micro transparency handling] is happening here. But when the line is re-imagined [by a stroke deformation algorithm] on top of [the initial hand-drawn line], you can end up with a gap or [another type of line degeneration] because this transparency was added or appeared as the result of the smoothing algorithm.

RS: Basically, DIP is more complicated than the line that you can visually see with your eyes.

MM: Yes.

RS: Moving on from lattices, a bunch of tweening algorithms process strokes separately from the plane in which they're drawn, for instance a network of strokes [comprising] the drawing. These methods need to get some kind of stroke correspondence between keyframes. It's much easier to tween keyframes that have equivalent sets of strokes, where there's a one-to-one correspondence between the strokes. There is a Disney paper from 2010 that is specifically designed for tight in-betweening, where the source and target keyframes have almost identical stroke configurations and topology – by that, I mean connectivity and structure. The premise of the paper is that they want to alleviate the workload from manually tweening these tight keyframes, which makes sense.

MM: Yeah, that's totally valid.

RS: Yes. But especially since most animation happens on 2s in our world, it's much more likely to have a stroke topology change across two consecutive keyframes than it is to have these tight inbetweening instances. And when it doesn't have that, the users have to specify down to extreme details, like one-to-one correspondences across keyframes, which is very finicky. Or you can handle mild occlusion with visibility toggles on strokes, like if you're drawing a hand, you would draw the backs of the fingers in frames where you don't see them. And then you would have to toggle [stroke visibility] on and off as the hand unfurls, and operate the toggles on a teeny tiny level.

MM: The reason I react by saying that this is valid is because, in a funny way, it almost feels like you have to exhaust this way of thinking. If anyone was presenting that as a genuine way of working, you'd go, 'Whoa, hang on.' But someone presenting it theoretically, I totally follow.

RS: It's a really valid theoretical argument. My point is that it only works when the theory matches up really well with the drawings. And most of the time that doesn't happen. A lot of the papers that work like this have small user studies that show some percentage of time was gained [when animating using their auto-tweening algorithm], and they expect more time gained with the more time users spend learning to use software. They're basically saying, 'Yeah, you have to do extra work of drawing the occluded fingers and resolving all the mismatched strokes, but it still takes less time than drawing all of the in-between frames.' I'm wondering, despite that technically being perhaps true, if the amount and frequency of tinkering you have to do [to make these auto-tweening algorithms work] may not justify integrating into animators' workflows. If I'm drawing a hand, I'm thinking about what I see. I'm not thinking about what the fingers that I don't see. The extra work doesn't just come from drawing extra strokes, it's also the mental effort of learning to think about and practice drawing in a totally new way.

MM: Yeah. It puts me in mind of somebody listening to jazz and saying, 'Well, you have to listen to the notes they're not playing.' But what goes through my mind there is that time spent making little micro decisions is not the same thing as time spent drawing freely. Even if it took [the animator] less time, they

might have expended a lot of calories, literally, by attempting to think all the way through those micro operations. And the second thing I would say is so much drive – not so much from research specifically, but from companies attempting to work in this way – is, either intentionally or unintentionally, subtly changing the profile of the kind of animator you want working for you. The logic goes, ‘Hey, if we just have this incredibly elaborate machine, then all we need are very, very detailed machinists to operate it,’ as opposed to the way [Cartoon Saloon] would think of going, ‘We want artists and we want to provide tools that keep them in the artist flow.’

RS: But you still need the machinists to be artists, or else you’re making crazy workflow trade-offs. And that type of person [who can take on that role] is ultra rare.

MM: Oh yeah, god. How many people, what kind of skill profile is that, that somebody could efficiently be a star artist and a star machinist simultaneously?

RS: For sure. There’s more I want to talk about with the one-to-one stroke correspondence. It’s not just the tightness thing. It’s also that consecutive keyframes don’t usually have the same number of strokes. It’s not like animators are given a set of building blocks of strokes and then puzzle-piece them together into [intermediate frames].

MM: But that’s why we got to where we got to with rigs. Well, that’s not exactly why, but that’s why with rigs, their scaffolding and deformation guarantees are the only way to make them work. That’s why you need to actually build the entire rig in advance: to build in those kinds of strict correspondence guarantees between frames. I have a lot of respect for 2D rig software. I think it’s limiting in an awful lot of ways, but I have a lot of respect for the animators and for the process. But it is not the same as hand-drawn. Whenever you try to apply [rig animation workflow logic] to hand-drawn, you go, ‘Well, you could make this work as long as you were to effectively rig [the character or body motion] up in advance, and then you’ve just departed from hand-drawn.’

RS: You’re asking the animators think about drawing in ways too distantly related to the pursuit of drawing.

MM: Exactly.

RS: Lots of tweening methods, even methods that actually seem pretty viable, require animators to draw a new keyframe every time there’s a topology change [over intermediate frames]. And the type of topology change can depend on the method. Sometimes it’s like stroke connectivity, sometimes it’s two shapes merging into one. [This approach is] interesting because it’s not the worst thing in the world, right? The worst case scenario is that you would just manually draw frames that otherwise you would want to auto-tween.

MM: In principle, it’s still saving you time on the frames you don’t have to draw.

RS: Yeah. And these methods are interesting because in order to work at the level of stroke connectivity, strokes need to be algebraic objects that you can manipulate. From a technical standpoint, they handle topology change by tracking and performing topological operations in individual strokes. But for animators to know when to use [these methods] in practice, they might also need to be tracking strokes, so that they can recognize when stroke topology change occurs. And that might be another drain on their energy, if a small motion involves dozens of different strokes breaking and forming. What are your thoughts on asking

animators to think about their drawing in terms of stroke tracking, and stuff like that?

MM: I was going to say was animators don't think in terms of changing topology, but let me counter that. There is a way in which they do, or in which some animators do. Certainly changes in topology are a concern in rig animation all the time, in a way that they are not in CG (computer-generated) animation.

RS: Right. Because topology change is an emergent property of embedding a volumetric object in a plane.

MM: Perfectly put. But that is not the case when you have a 2D rigged puppet. You get in this thing where if you need to turn [an arm a certain way, animators] have to choose the moment at which that happens. Say, in Moho [animation software], you have animated layer order with the arm parented to the torso, and you can pull [the arm] out from behind [the torso] and but in front, but you have to choose the frame in which it happens. So if you have animators who have worked on both hand-drawn and rigged animation, they actually would be able to think in that way. I have another case as well. I think Clean-up animators do have to think this way. On *Wolfwalkers*, there were two bad cases, where the front and back designs of characters were irreconcilable. With Merlin and the soldiers. You couldn't move between different character design viewpoints without visual inconsistency.

RS: Which is different from just having different character geometry per viewpoint, which is something we always have to do [at Cartoon Saloon].

MM: Yes. To put it simply, one of the reasons the generative AI stuff we see always looks a bit weird and doesn't work is because, in a sense, animation is defined by what changes and what remains the same. The movement and articulation of the body is changing, but the body remains the same. Animation represents the motion of bodies. The problem is that you can't [animate the soldier] without the armor plate changing into a different thing. So the point is when they need to be, certainly when they hit a problem, Clean-up artists who are trying to solve shapes and volumes actually do probably think in topology. They would never describe it that way, but they actually are.

RS: I agree with you, but I think that's a different type of topological thought than what the methods I'm describing are demanding. Clean-up is thinking about where certain shapes are when a character is in a certain pose at a certain viewpoint, like, 'If we want to resolve this pose a certain way, then we need to make sure that we're holding on like a version of that pose where you can see certain design elements of the arm, so that's recognizable to the viewer.' Stuff like that, decisions about like what shapes to show, what shapes get collapsed, what shapes get split by which other shapes. That's a topological way of thought that has to do more with regions.

But these methods involve thinking topologically about drawings as networks of strokes, rather than networks of regions that are connected to each other with respect to the way the character's body is structured. With strokes, you have different topological cases. You have cases where the strokes are just in the 2D plane. You have cases of layers, which is pretty intuitive. There are other cases which work better, but maybe less intuitive to people not used to thinking topologically, like stroke immersion vs. stroke embedding. That's a hard thing [for non-mathematicians] to conceptualize.

MM: You know, it's a hard thing to conceptualize, but I would say that Clean-up animators do think in strokes [in topological ways]. If you watch them work, I mean, it's very funny. They appear slightly insane because they will just be going, 'No... No... No' [as they repeatedly draw and redraw strokes].

RS: But isn't that them looking for the line? They're not looking for the piece of stroke that slots in

somewhere and connects to all the other strokes, necessarily.

MM: No, not necessarily, but kind of. Because when they animate, when they draw each frame, they will be kind of asking themselves, ‘How is this line going to travel across that?’ They will think in that way. And the other thing that’s surprisingly similar is that if they have a stroke criss-cross point [like a topological immersion], they do have to treat those differently in their mind. They have to embody that one stroke is in front of the other. Especially because strokes might parallax, and if they start tracking that criss-cross point, then it will behave differently. It will behave as though it’s this one volumetric object. I think they are drawing those criss-crossed lines consciously separate from each other.

RS: Gotcha. I want to drill down on some other differences from a computational perspective. There’s this one auto-tweening method that’s quite recent, where a drawing is interpreted as sets of strokes. So if you have an eyeball and an eyelid, that’s a set of strokes. You want to automatically tween a blink, which means that the eyelid stroke will have to be in front of the eyeball strokes. The eyeball strokes have to be grouped together so that the lid can slide over them uniformly. And they’ll have to be associated with the lid so that the computer knows that when the lid moves, the lid stroke should be occluding. That all has to be specified, often manually, at a stroke-by-stroke level. And if the strokes change enough [over intermediate frames], like if the eye is turning to the side or something, you have to redefine the grouping, because there’s going to be new types of [stroke] interactions.

MM: I think we are noticing a relationship between how modern Clean-up is done and these ways of thinking in terms of topology and changing in topology. You would dismiss that [relationship] out of hand, but there’s a part of me that goes, ‘Well, don’t for a second.’ You jump in and think why this might or might not work. A lot of that topological thinking would be happening sort of subconsciously, in the way that the animators are visualizing [their drawing tasks], and maybe even in the order of operation of how they go and draw the eyelid. But I think there’s a chasm between that vs. being a willing machinist detailing which strokes are related and how.

RS: Yeah, I think you’re right. I mean, I wouldn’t be professionally and intellectually where I am if I didn’t think there was a deep relationship between thinking topologically and animators’ thought process while they work.

MM: Exactly. I have an intuitive sense that there is a relationship there, but it’s not the same thing as how this software should work.

RS: And it’s not like animators couldn’t overcome these [workflow demands] and learn to think like that [so they could use this type of software]. They would just kind of have to understand how the software works. And once they got that understanding, it would probably automatically tap into a comprehension that they already had. They would just realize that they need to formalize it.

MM: Yeah, or move into a conscious mode of working with it.

RS: I’m still wondering about the potential cost to workflow of tinkering with those types of setups, where you have to redefine them every time there’s a sufficient enough change [between keyframes].

MM: To speak as our business [i.e., Cartoon Saloon], we just don’t want to do that. There was a desire, I think, to get tighter and tighter and more perfect drawings, until we realized machines were the experts at doing tighter and more perfect drawings. And [Cartoon Saloon] switched suddenly to going, ‘We’d rather

just do things loose and imaginative as it is.’ I can’t speak for the business [of hand-drawn animation] across the world. I’m sure there’s lots of companies who want to find solutions to improve efficiency, but it just wouldn’t chime with the way [Cartoon Saloon] wants to work at all.

RS: I still have to answer the question of, ‘Would it actually improve efficiency?’ And that goes back to your point of artists and machinists, because thinking that in way and being able to execute those types of operations is one thing, but to do it over and over again with that level of mental load on top of the mental load of trying to draw really quickly.

MM: And trying to find the character in there.

RS: Yeah, you know, you’re saving time on frames. You don’t have to draw [some of the intermediate frames]. But is that worth the cost of everything that workflow entails? It reminds me of what you’ve described to me before, of why [Cartoon Saloon] doesn’t animate with vector graphics. Like you said, animators will repeatedly draw and redraw lines to get them right. But the difference is that with vector strokes, you might actually not be able to get it right [because it may not be technologically possible to express the exact line you need as a vector stroke]. In theory, unlike with vector strokes, you should be able to get these tweening methods to work with enough time put in. There’s not some kind of mathematical incompatibility, like with the vectors, that completely prevents you from doing what you want to do artistically. So there is a finite end to the tinkering, but it is a lot of tinkering.

MM: My answer is basically that I don’t think [auto-tweening software involving these types of user interventions] would improve efficiency. But my answer differs in degree, depending on what kind of business and what kind of software and what kind of artist we’re talking about. But I essentially come down on thinking, I don’t think I want the animator engaging in this kind of thing.

RS: Okay, let me give you an example. The algorithm I just described, where you have to hook [strokes] up into foreground-background relationships – that system, computationally, mathematically, fundamentally, cannot handle a case of a ribbon overlapping. You know, like one of the cancer research pins. Or, it can’t handle a soda can with a label on it, where the label is one stroke, but you can’t see the whole stroke. The problem is that in both those cases, strokes that exist on multiple depth levels at a time. And computationally, just because of the way that they designed [this method], in order to make strokes exist on two [depth] levels at the same time, you would have to parse the single stroke into two strokes on different levels. That type of failure case, you know, it’s not a deal breaker. It works great, until you get to someone’s hair or cloak blowing in the wind. Which, again, isn’t the end of the world, cause you can just go back to drawing.

MM: If you limit it down to the cases where it might work well, it might have very little negative impact. My worry is that there’s so many layers to what the animator would have to go through there. Because on one hand, they would have to understand the process. There’d be like a learning curve. They’d have to spend a bunch of time trying and maybe failing to use the tool properly, and only [after that learning curve period] could you work out if they would get fatigued by this kind of workflow. If it works and they can use it, great. But what if this number of conscious operations every hour burns them out by lunch? Even so, it simply can’t be dismissed out of hand. It has too many relationships to how animators already think that it’s worth thinking about.

RS: The last thing I’ll say is that another theme seems to be emerging, that the better that animators

can think along the lines of how the software they're using is representing and processing data, the more successful they'll be with using the software, and the more fluidly they'll be able to understand when and how to use it. And I actually think when and how to use it might be the most worrisome issue with these kinds of methods, because if animators spend time trying and failing to use these tools on cases that the software fundamentally can't handle, like self-overlaps, not knowing why they can't get it to work – that's a lot of work down the drain. And you're only going to know exactly when and when not to use it if you really understand how the underlying algorithm is functioning.

MM: Or if you were forced to work this way, you could develop an intuition over years through brute force. Through just repetitive success and failure, which would take a very long time. It would be a lot better to have a mental model of how [the software] works.

RS: Which, going back to your point, is not actually necessarily prohibitive from a theoretical perspective, because there's such a strong relationship between how animators are already thinking and what these softwares actually do when you drill down into the fundamentals. But [gaining fluency and intuition with the software] is still a big project for any trained artist.

There's one paper that's probably the most flexible of any method. But in user trials, the biggest problem was the really steep learning curve. And the reason why it's so successful is because it turns the drawing very explicitly, much more explicitly than any other method that I'm looking at, into a full topological space. With that, you get guaranteeably good ways of collapsing strokes and regions [over volumetric articulate motion]. As long as you draw the topology change frames, you can animate things like a head turning, and actually lose the eye strokes and the mouth strokes correctly because they'll just collapse down. And because it is so rigorously a mathematical topological space, you get all of the mathematical guarantees of how topology works for free, which allows for these collapses. The number one issue that the authors talk about is that animators have to basically learn topology to overcome the software's learning curve. Which isn't necessarily a bad thing, but it connects back to your point about the artist-machinist hybrid, and of how many [people capable of doing that work] are actually in existence from an employment perspective, not necessarily a theoretical perspective.

MM: You know Anita [Gaughan, Cartoon Saloon animator], her animal action stuff is so excellent. I mean, Tomm [Moore, Cartoon Saloon co-founder/director] basically poached her out of animation school. There's a bunch of theories [explaining how she got so good]. One theory is that she evidently works extremely hard, seems to animate a lot in the evening, seems to come home and keep practicing, keep driving. But let's say she found it particularly hard to think topologically, and so for animal action we replaced her with someone who could. I think we would have lost something we were trying to aim for.

RS: Because the art always has to come first over the ability to interact with the machines. That's a really important point, and that's always what you're going to hire for.

MM: Yeah. That's what [Cartoon Saloon] believes in.

RS: Well, it's what makes [the animation] good. It's what makes people want to watch it.

MM: Yeah. Say there was an animator who really leaned on the machine a lot, and got a lot of animation done. It's probably nowhere near as good as Anita's, but we got it done in X amount of time. With our company specifically, when you start talking about efficiencies and things like that, yes, it matters, but it's low down the list for us.

RS: Correct me if I'm wrong, but I think what you're saying is that you want to balance maximum efficiency with sufficiently maximal preservation of the artistic quality and artistic vision. Quality can volley up and down depending on what type of technology we choose to employ, but as soon as it dips below a certain threshold, it's not worth efficiency maximization anymore.

MM: So that's fairly well put, and I guess I would be saying the same thing. You know, flip it around and suppose we're getting various different kinds of AI and calculation tools and using them for something reasonably complex to compute, like stellar photography or planetary motion. Imagine we started saying, 'These [types of tools] don't really suit mathematicians, so let's just get other people who are good at working with the software.' You might end up [approaching the project] in a certain way. Imagine if the most important skill was the speed you could type, and you pulled mathematicians out of the project to hire speed typists instead. There's part of me that goes, 'I think you need a mathematician in there.' So that's where I come from. When I think about my own job and the whole idea of looking for new animation technologies that will improve efficiencies, my guiding star is that we should take people who are talented artists and then try to get systems and machines that support them in that way.

When we start talking about changing the profile of the person [we want to hire], – saying, 'This computation only works if you're a fast typer, it doesn't matter if you're a good mathematician' – we can talk in pure theory about how this could be more efficient. I'm really open to that, and I'm actually really open even to seeing where there are overlaps [between hand-drawn artistry and software-based efficiency]. But provided we remain a hand-drawn studio, I'm going to take people who are naturally good at hand-drawn animation unaided, and then try to provide aids for them. I'm going to be driven by that, as opposed thinking about this an entirely new way, where animators' own innate senses of movement and artistry don't matter as much as their ability to understand topology and work with those operations without getting fatigued.

RS: I'm thinking about how the people who have the ability to make animation software have been trained in math and computer science and image processing. Often, that's what they know, that's the world they live in, that's what they start with, and that is the place from which they reach out into the world, trying to offer things from their own expertise to potentially solve or alleviate problems that they see [in animation]. That has fundamentally produced an entire field – excluding stuff like internal Pixar researchers collaborating with artists on technology for specific projects – where [the methods that we have been discussing] are how researchers can best conceive of the problem of auto-tweening within the language of computational math and data structures that they're familiar with. They're trying to make it as easy as possible for [animators] to meet them where they are. They're basically saying to animators, 'Your problem looks like our problem. This is our best interpretation of your problem [(auto-tweening)] in the language of our problem [(computational topology)]. And because of how our problem works, we can help you solve your problem in all these different ways. We can't solve every instance of your problem, and you have to approach your problem in a particular way, but that shouldn't be an issue because it's relatively innocuous work.'

That sounds great, and it seems like it should work. But it creates the fundamental issue that you're describing, where the pool of people who are supposed to be using this [software] are not necessarily equipped to use it, or are not equipped to understand how to use it efficiently. And [the hand-drawn animation industry] prioritizes an entirely different skill set [to that required to effectively use the software], and views the technology side of it as something potentially helpful, but not essential.

MM: Yeah. You know, it's funny, the US truckers union is called the Teamsters, right? That's because

they used teams of horses [to deliver goods] when they were founded, but of course they don't anymore. I don't know if the same people who were good at driving teams of horses eventually became good truck drivers or not. But if the technology moved to such a place such that the original horse drivers couldn't adapt, Atlas would shrug a little bit. There's someone who will commit to [learning how to operate trucks]. There is the sort of cold, capitalist way in which all of this has to play out economically. It's like the emergence of puppet-style hand-drawn animation, where a technology gelled with people who were capable of learning it. Suddenly, it takes off.

The fact is, at any point someone could strike gold, theoretically, and have this perfect [auto-tweening software]. Say they struck gold, and [could automate] animation that looked hand-drawn, but the catch was that we couldn't use any of our existing hand-drawn animators. We would have to use this different class of people [who could best operate the software]. Maybe Cartoon Saloon wouldn't do that very quickly, but someone would. Someone would switch and do it.

RS: I mean, I think that's going to be an AI situation in 5-10 years.

MM: It's what everyone's trying. Attempting to leverage AI into making the better mousetrap, into making a good movie. The thing is, everyone will be using it, so there will be a tremendous amount of AI sludge. And the more sludge there is, the more audiences will switch off. And the easier it will be for prestige [of human craftsmanship] to survive in that way. And then, I think companies are going to try to come along and try to do prestige, but they're going to try to use the same methods that laypeople can use to produce games and memes and whatnot. I suspect they'll find it difficult to release that in 500 theaters.

Then, you might find people who are able to make a marketable film with AI, but it just takes a tremendous amount of work on the back end, after it's done, to craft it into something that's usable. Then you have to ask, 'Is this seriously how you want to work, how you want to spend a 25 million production budget? You don't want to work for the ground up? You'd rather have people chisel away at [the AI output] until it looks passable?'

RS: It's the fundamental underlying issue with artistic technologies. I mean, this has much more to do with generative AI than with auto-tweening tools, which are genuinely helpful or trying to be helpful. But for researchers like me who work on this, there's always the possibility of slipping thinking that the technology output can get good enough to justify downscaling the role of human artists in the art production process. It's really easy to start [building technology] from a flawed design premise.

MM: Yeah, exactly. I think ultimately what we're seeing at the moment is generative AI that are essentially toys for people to play with. Toys have a bit of a half-life, and they're a lot of fun. There are two possibilities. Either they're a lot of fun for now, but eventually we don't really use them anymore. Or they completely become the norm for how everyone makes memes, or stuff like that. They become so normal that everyone stops noticing how much they use them. Right now, half the charm of memes is that they're badly put together. Maybe eventually, tastes will change, and everyone will prefer the look of AI memes. But either way, it will be hard to have the AI aesthetic work for toys while also being the aesthetic of prestige. It's going to be really difficult, if not impossible, for those trends to exist simultaneously.

Where AI may appear, where it likely will appear, is much more in the margins. It's going to be much more, for instance, that we're able to use AI to do a first paint-potting pass in DIP. We could lower the DIP team down to two people, who would do the other 80% of the DIP work. It's just not going to sound that sexy to anybody, but it will save you money. And it will be a genuine application of the technology.