Skinning with Dual Quaternions

Ladislav Kavan1,2, Steven Collins1, Jiří Žára2, Carol O’Sullivan1

1Trinity College Dublin, 2Czech Technical University in Prague

Abstract

Skinning of skeletally deformable models is extensively used for real-time animation of characters, creatures and similar objects. The standard solution, linear blend skinning, has some serious drawbacks that require artist intervention. Therefore, a number of alternatives have been proposed in recent years. All of them successfully combat some of the artifacts, but none challenge the simplicity and efficiency of linear blend skinning. As a result, linear blend skinning is still the number one choice for the majority of developers. In this paper, we present a novel GPU-friendly skinning algorithm based on dual quaternions. We show that this approach solves the artifacts of linear blend skinning at minimal additional cost. Upgrading an existing animation system (e.g., in a videogame) from linear to dual quaternion skinning is very easy and has negligible impact on run-time performance.


Keywords: skinning, rigid transformations, blending, dual quaternions, linear combinations

1 Introduction

Skinning and skeletal animation is the technology behind character animation in many applications. In some situations, physically accurate skin deformation, which supports muscle bulging and dynamic effects, is desirable. In other situations, however, a fast algorithm capable of skinning multiple models interactively is needed—especially in the context of videogames.

The standard algorithm for low-cost skinning is known by many names: linear blend skinning, vertex blending, skeletal subspace deformation or enveloping. It is sometimes used not only for skin deformation (as the name suggests) but also to animate cloth, because it is considerably faster than physically based cloth simulation [Cordier and Magnenet-Thalmann 2005]. The basic principle is that skinning transformations are represented by matrices and blended linearly. It is very well known that the direct linear combination of matrices is a troublesome way of blending transformations. This produces artifacts in the deformed skin, even if we restrict the skinning transformations to rigid ones (i.e., composition of a rotation and translation). In spite of these shortcomings, linear blending is still a very popular skinning method, but perhaps only because there is no simple alternative.

Recent previous work suggests converting rigid transformation matrices to ((quaternion, translation)) pairs and blending them instead of matrices [Hejl 2004; Kavan and Zara 2005]. This works, but at a cost: Hejl’s algorithm [2004] imposes constraints on the model’s rigging (specifically, a vertex can only be influenced by neighbouring bones, otherwise artifacts can occur). This could be inconvenient, because linear blending has no such constraints (and it is exploited for many 3D models). Kavan and Zara’s method [2005] does not have this restriction, but uses a complex and computationally expensive Singular Value Decomposition scheme. Obviously, the elegance of linear blend skinning is lost in both cases.

The representation of rigid transformations by matrices or ((quaternion, translation)) pairs illustrates just two possible parameterizations of the group of rigid transformations. Nothing prevents us from blending, for example, 3-tuples ((axis, angle, translation)) or pairs ((axis, translation.sin(\text{angle}))). Even if we restrict ourselves to blending via linear combinations, we can construct infinitely many different blending methods just by considering different parameterizations of rigid transformations.

A natural question arises: which parameterization of rigid transformations is the best one for skinning? First of all, it is necessary to clarify what “best” means in this context. We summarize the mathematical properties that an ideal blending method for skinning should possess in Section 3.1. Next, we show that these properties are satisfied by a representation of rigid transformations known as dual quaternions—a generalization of regular quaternions invented in the nineteenth century [Clifford 1882].

From a practical point of view, dual quaternions offer a very simple yet efficient skinning method. Due to the properties of dual
quaternion blending, none of the skin collapsing effects exhibited by linear blend skinning will manifest themselves. Blending of dual quaternions can be elegantly computed in a vertex shader with complexity comparable to standard linear blending. Dual quaternions are more memory efficient, requiring only 8 floats per transformation (essentially, two regular quaternions), instead of the 12 required by matrices. In an existing application, it is extremely easy to replace a linear blend skinning implementation by a dual quaternion one. All that is necessary is a slight modification of the vertex shader and conversion of the matrices to dual quaternions before passing them to the shader. The model files as well as the internal data structures do not need any change at all – the only difference is in the transformation blending.

2 Related Work

Historically, the idea of skin deformation by an underlying skeleton is credited to [Mangenat-Thalmann et al. 1988]. Since then, several different approaches to skeletal animation have emerged. Our approach – dual quaternion skinning – falls into the category of geometric methods.

Physically based methods. A logical approach to character animation is to simulate the internal structure of the body: bones, muscles and fat tissues. This can work either with explicit anatomy knowledge [Scheepers et al. 1997; Aubel and Thalmann 2000; Teran et al. 2005], or without [Capell et al. 2002; Guo and Wong 2005; Pratscher et al. 2005]. Physically based methods generally obtain a high level of realism (delivering also dynamic effects and muscle bulges), but at high computational costs.

Capturing real subjects. Several methods successfully exploit modern motion capture and/or 3D scanning devices to capture skin deformation of real people [Allen et al. 2002; Anguelov et al. 2005; Park and Hodgins 2006; Allen et al. 2006]. These approaches are highly accurate, but require expensive hardware and are of course limited only to existing subjects.

Example based techniques. Multiple input meshes can be used both to resolve the artifacts of linear blending and to add additional effects like muscle bulging. Example based methods use either direct interpolation between example meshes [Lewis et al. 2000; Sloan et al. 2001], correction by principal components of example deformations [Kry et al. 2002] or fit the linear blending parameters to match the provided examples [Wang and Phillips 2002; Mohr and Gleicher 2003]. Recently, a more accurate (yet more complex) example interpolation method has been proposed [Kurthara and Miyata 2004] and augmented with an innovative GPU approach [Rhee et al. 2006]. Generally, this class of methods offers a level of realism limited only by the number of input examples. However, the production of examples can be costly, requiring a lot of memory to store them and the animator’s labour.

Geometric methods. In this case, only one input mesh is given (designed in a reference pose). The skeleton-to-skin binding is defined in a direct, geometrical way. The most popular way, established with linear blend skinning, is to bind each vertex to one or more joints. In the latter case, the weight (amount of influence) of each influencing joint must be specified. Advanced blending methods, e.g., direct quaternion blending [Hejl 2004], log-matrix blending [Cordier and Magnenat-Thalmann 2005] and spherical blending [Kavan and Zara 2005] also use this rigging structure. Even though these techniques remove some artifacts, they still fall short of delivering natural skin deformation in all postures (see Figure 1 and Section 3.1).

Some researchers propose combatting skinning artifacts by implementing a different rigging method, e.g., based on sweep surfaces [Hyun et al. 2005] or auxiliary curved skeletons [Yang et al. 2006; Forstmann and Ohya 2006]. In some cases, this also allows advanced effects to be animated, such as muscle bulging. The disadvantages include complexity of the GPU implementation (even the very efficient Forstmann’s method [2006] is about 3 times slower than linear blend skinning) and inconsistency with the established rigging standard: new rigging tools and data formats are needed. In this paper, we argue that the problems of linear blending do not stem from incorrect rigging, but from incorrect blending. With dual quaternion skinning, it is therefore neither necessary to change the rigging structures nor to update existing models.

Dual quaternions. The first reference to dual quaternions (historically called bi-quaternions) appears in [Clifford 1882], along with the more general concept of geometric algebras. These algebras naturally contain not only vectors and quaternions, but also k-dimensional subspaces [Wareham et al. 2005]. This leads to very elegant and dimension-independent expressions of geometric properties, but sometimes unfortunately also to an increase in time and memory complexity of the resulting implementation [Fontijne and Dorst 2003]. Dual quaternions, in turn, are not so general, but are more compact and faster to manipulate. For computer graphics practitioners, the big advantage of dual quaternions is that they are based on regular quaternions – a well-known tool in computer graphics [Shoemake 1985].

Blending vs interpolation. A vast amount of literature has been devoted to the problem of transformation interpolation [Barr et al. 1992; Juttler 1994; Marthinsen 1999; Bela and Kumar 2002; Hofer and Pottmann 2004; Li and Hao 2006]. This is not surprising, because the construction of interpolation curves for given key transformations (e.g., camera orientations) is a fundamental problem in computer animation. Unfortunately, in skinning, we face a different problem: the blending of rigid transformations, i.e., their weighted average (confusingly, in some literature this is also called interpolation). The weighted averages can be used to construct interpolation curves (see [Buss and Fillmore 2001]) but not vice versa.

3 Dual Quaternion Skinning

In this section, we first quickly review previous geometric skinning algorithms (Section 3.1), identifying their strengths and weaknesses and deriving properties that an ideal blending for skinning should possess. Then we present a short tutorial on dual quaternions (Section 3.2) and discuss dual quaternion blending (Section 3.3). The final dual quaternion skinning algorithm is summarized in Section 3.4.

Conventions. We denote scalars by lower-case letters, vectors and quaternions in bold and matrices by capital letters. Dual quantities are distinguished from non-dual by a caret, e.g., ā denotes a dual number and q a dual quaternion. The i-th component of vector v is written as vi, thus also v = (v1, ..., vn). The dot product of vectors v and w is denoted as ⟨v, w⟩ and ∥v∥ is the usual vector norm. Cross product is denoted as v × w.

3.1 Background on Geometric Skinning

In the following section, we focus only on geometric skinning methods with linear blend skinning-style rigging structure (which is a de facto standard in the videogames industry). A 3D object conforming to this standard consists of skin, a skeleton and vertex weights. The skin is a 3D triangular mesh with no assumed topology or connectivity and the skeleton is a rooted tree (both designed in a reference pose). The nodes of the skeleton represent joints and the edges can be interpreted as bones. However, each node (except the root node) can be easily identified with the edge leading to it, so the difference between joints and bones is rather moot in our case.
The transformations relating joints in the hierarchy are assumed to be rigid (we do not consider scale or shear). The vertex weights describe the skin-to-skeleton binding, i.e., the amount of influence of individual joints on each vertex.

Let us assume that there are \( p \) joints in our model. We usually store the joints in an array, with every joint referenced by a number \( 1, \ldots, p \). In the reference (rest) pose, each joint has an associated local coordinate system. The transformation from the rest-pose of joint \( j \) to its actual position in the animated posture can be expressed by a rigid transformation matrix – let us denote this matrix as \( C_j \).

We assume that vertex \( v \) is attached to joints \( j_1, \ldots, j_n \) with weights \( w = (w_1, \ldots, w_n) \). The indices \( j_1, \ldots, j_n \) are integers referring to the joints that influence a given vertex – in other words, they are indices into the array of joints. There is usually a fixed upper bound on \( n \) (the number of influencing joints), typically 4, following from graphics hardware considerations. The weights are normally assumed to be convex, i.e., \( w_i \geq 0 \) and \( \sum_{i=1}^{n} w_i = 1 \) (however, this non-negativity is not exploited in our algorithm). The \( i \)-th component of vector \( w, w_i \), represents the amount of influence of joint \( j_i \) on vertex \( v \).

The vertex position in the mesh deformed by linear blend skinning is then computed as:

\[
v' = \sum_{i=1}^{n} w_i C_{j_i} v = \left( \sum_{i=1}^{n} w_i C_{j_i} \right) v
\]

that is, taking a linear combination of joint transformation matrices. This explains the skin collapsing artifacts (see Figure 2) – the blended matrix \( \sum_{i=1}^{n} w_i C_{j_i} \) is no longer a rigid transformation, but a general affine one (potentially containing scale and shear factors).

Formally speaking, the skin collapsing effects visualize the fact that the set of orthonormal matrices is not closed under addition.

A blending method that avoids such skin-collapsing defects must therefore deliver a rigid transformation in all cases. The blending of matrix logarithms [Alexa 2002] satisfies this condition and has been implemented for skinning by [Cordier and Magnenat-Thalmann 2005]. Unfortunately, log-matrix blending has a different problem: in some cases, it picks a longer trajectory than necessary when interpolating rotation [Bloom et al. 2004]. This means that instead of a direct, straight rotation, a diversion is sometimes taken, which can result in unsightly artifacts (see Figure 3). Our desired blending should therefore always interpolate rigid transformations along the shortest path, i.e., along a geodesic in the manifold of rigid transformations [Hofer and Pottmann 2004]. Note that this can be achieved for log-matrix blending in the case of two transformations [Li and Hao 2006], but this method does not generalize to \( n \) transformations.

Decomposing rigid transformation matrices into \((\text{quaternion}, \text{translation})\) pairs and blending them linearly has none of the above-mentioned problems: it always produces rigid transformations and selects the shortest path of the interpolated rotation [Kavan and Zara 2005]. Unfortunately, the blending of quaternions and translation vectors independently has another pitfall – dependence on the body-space coordinate system. In practice, this means that the model’s vertices rotate around the origin of the body-space. The origin is user-defined and usually coincides with the center of mass. In many applications, rotation around the center of mass is exactly what we need. However, this is not the case for skinning.

Rotation of model vertices around the center of mass produces unacceptable skin deformations. Intuitively, vertices influenced by both the lower and upper arm should naturally rotate about a point near the elbow. This is the principle of [Hejl 2004]: the vertices rotate around the nearest joint. In cases similar to that of the elbow (two bones connected by a joint), this works perfectly. However, for more complex joint influences, such as those usually occurring around the armpit or dorsum, artifacts result from the fact that the center of rotation is fixed and does not adapt to the actual posture. In character models, these artifacts can be usually avoided by careful rigging. However, they are inevitable in other deformable objects, such as our simple cloth model shown in Figure 4.

The limitations of Hejl’s method can be resolved as described in [Kavan and Zara 2005], which presents an accurate method to compute the rotation centers. It works by minimizing the translation of the resulting blended transformation (c.f. with Hejl’s method which works correctly only when this translation is zero). Unfortunately, the cost of this approach is significant: it requires a least-squares solution of a set of equations, carried out using Singular Value Decomposition (SVD). Since it is not affordable to run SVD per vertex, the rotation centers are cached and reused for clusters of vertices. This trick enables real-time performance of spherical blend skinning to be achieved. Unfortunately, the recycling of rotation centers may cause discontinuities of the deformed skin in areas where the clusters concur (see Figure 5). These artifacts are however much less frequent (and apparent) than with direct quaternion blending.

Note that it was not necessary (and not even possible) to choose a rotation center in linear or log-matrix blending. This is because both of these approaches are coordinate invariant, i.e., it does not matter if we switch our transformations to a new coordinate system before or after the blending. This is easy to ver-
ify: if matrix $M$ represents a change of coordinates, then matrices $C_{j_1}, \ldots, C_{j_n}$ expressed with respect to the new coordinate system become $MC_{j_1} M^{-1}, \ldots, MC_{j_n} M^{-1}$. Linear blending with these matrices gives:

$$
(\sum_{j=1}^{n} \omega_j MC_{j} M^{-1}) = M \left( \sum_{j=1}^{n} \omega_j C_{j} \right) M^{-1}
$$

which follows from the distributive property of matrix multiplication. We see that the resulting blended matrix is really the same as that for matrices $C_{j_1}, \ldots, C_{j_n}$, just with respect to the new coordinate system (of course). The same reasoning can be repeated for log-matrix blending if we recall the following properties of matrix exponential and logarithm: $\exp(MC_j M^{-1}) = M \exp(C_j) M^{-1}$ and $\log(MC_j M^{-1}) = M \log(C_j) M^{-1}$ (see [Moakher 2002]). This shows that log-matrix blending is coordinate-invariant.

Why does this not work for \textit{(quaternion,translation)} pairs? The reason is that if we define an algebra of \textit{(quaternion,translation)} pairs, with multiplication corresponding to composition of transformations, then this algebra is not distributive. Specifically, it is not distributive with respect to right multiplication [Kavan and Zara 2005] (it is with respect to left multiplication). As right multiplication corresponds to the change of body-space coordinates (and left multiplication to the change of world-space coordinates), this just reflects the fact that blending \textit{(quaternion,translation)} pairs rotates about the origin of the body-space.

To summarize, we see that an optimal rigid transformation blending method for skinning should:

- always return a valid rigid transformation (to prevent skin-collapsing effects)
- be coordinate-invariant (to avoid the problems with rotation center)
- interpolate two rigid transformations along the shortest path (to mimic natural skin behaviour and avoid excessive stretching)

Note that because we work with all types of rigid transformations, the shortest path property actually embraces two properties: shortest path rotation and translation. Shortest path rigid transformations correspond to \textit{screws} (see next section). Besides the above stated properties, the blending should of course be efficiently computable, typically on a GPU. As we will show in the remainder of this paper, our proposed dual quaternion blending satisfies all of these requirements.

### 3.2 Dual Quaternions

Dual quaternions are not a standard tool in computer graphics, in contrast to regular quaternions. This section provides a brief tutorial; for further details, see [McCarthy 1990; Kavan et al. 2006].
corresponds to multiplication of dual quaternions. If the rotation is described by unit quaternion $q_0$ and the translation by unit dual quaternion $1 + \frac{c}{2} (t_0 i + t_1 j + t_2 k)$ as before, then their composition is

$$(1 + \frac{c}{2} (t_0 i + t_1 j + t_2 k)) q_0 = q_0 + \frac{c}{2} (t_0 i + t_1 j + t_2 k) q_0$$

We can verify, by direct computation, that the result is always a unit dual quaternion and, conversely, that any unit dual quaternion can always be written in this form. Let us assume that we already have a routine for conversion between a $3 \times 3$ rotation matrix and a unit quaternion, as well as a routine for quaternion multiplication. Equation (4) then shows how to convert a $4 \times 4$ rigid transformation matrix to a unit dual quaternion. The opposite conversion, from a unit dual quaternion $q_0 + \epsilon q_0$ to a matrix is also straightforward. The rotation is just a matrix representation of $q_0$ and the translation is given by the vector part of $2 q_0 q_0^*$.

Every unit dual quaternion $q$ can be written as

$$\hat{q} = \cos \frac{\hat{\theta}}{2} + \hat{\epsilon} \sin \frac{\hat{\theta}}{2}$$

where $\hat{s}$ is a unit dual vector with zero scalar part, see [McCarthy 1990] or [Daniilidis 1999]. Note that this looks like the formula for regular quaternions, just employing the dual angle $\hat{\theta}$ and unit dual vector $\hat{s} = s_0 + \epsilon s_0$. The geometric interpretation of those quantities is related to screw motion, that is a rotation and translation about the same axis. Chasle’s theorem [Daniilidis 1999] states that any rigid transformation can be described by a screw motion (see Figure 6). Angle $\hat{\theta}/2$ is the angle of rotation, and unit vector $s_0$ represents the direction of the axis of rotation. $\hat{s}/2$ is the amount of translation along vector $s_0$, and $\hat{s}$ is the moment of the axis. Moment is an unambiguous description of the position of an axis in space. It is given by equation $s_0 = p \times s_0$, where $p$ is a vector pointing from the origin to an arbitrary point on the axis. Which point we choose is not important, because for any other point of the axis, say $p + c s_0$ (where $c$ is an arbitrary scalar), we obtain the same moment: $(p + c s_0) \times s_0 = p \times s_0$. In other words, we can say that while classical quaternions can represent only rotations whose axes pass through the origin, dual quaternions can represent rotations with arbitrary axes.

Dual quaternions exhibit the so-called antipodal property of classical quaternions, i.e., the fact that both $\hat{q}$ and $-\hat{q}$ represent the same rigid transformation. Therefore, when converting matrices to dual quaternions, we must choose an appropriate sign. This is however the same problem as in the case of regular quaternions, but we can apply the same methods as in [Hej 2004] or [Kavan and Zara 2005], i.e., consistently select signs so that the resulting quaternions lie in the same hemisphere.

3.3 Blending of Dual Quaternions

Using Equation (4), we can convert our skinning transformations to unit dual quaternions $\hat{q}_1, \ldots, \hat{q}_n$. The next task is to compute a blended unit dual quaternion $\hat{q}$ with respect to the given convex weights $w = (w_1, \ldots, w_n)$. This can be done by taking their linear combination followed by a normalization (because we need a unit dual quaternion). We call this Dual quaternion Linear Blending (DLB):

$$DLB(w; \hat{q}_1, \ldots, \hat{q}_n) = \frac{w_1 \hat{q}_1 + \ldots + w_n \hat{q}_n}{\|w_1 \hat{q}_1 + \ldots + w_n \hat{q}_n\|}$$

In the following, we show that DLB has all the properties required in skinning (as summarized in Section 3.1). Firstly, it is clear that DLB always returns a rigid transformation, because DLB computes a unit dual quaternion which can be subsequently converted to a rigid transformation matrix.

Secondly, in order to demonstrate the coordinate-invariance of DLB, we have to verify that for any unit dual quaternion $r$, the following formula is true:

$$DLB(w; \hat{r}_1, \ldots, \hat{r}_n) = r \cdot DLB(w; \hat{q}_1, \ldots, \hat{q}_n)$$

In fact, this breaks down to verifying two similar properties, called left and right invariance. Left invariance requires that

$$DLB(w; \hat{r}_1, \ldots, \hat{r}_n) = \hat{DLB}(w; \hat{q}_1, \ldots, \hat{q}_n)$$

and right invariance, by analogy, requires that:

$$DLB(w; \hat{q}_1, \ldots, \hat{q}_n) = r \cdot DLB(w; \hat{r}_1, \ldots, \hat{r}_n)$$

It should be clear that left and right invariance together imply coordinate-invariance. To prove left invariance, it is sufficient to apply the distributive property of dual quaternions and the multiplicative property of the norm (Section 5.2):

$$DLB(w; \hat{r}_1, \ldots, \hat{r}_n) = \frac{w_1 \hat{r}_1 + \ldots + w_n \hat{r}_n}{\|w_1 \hat{r}_1 + \ldots + w_n \hat{r}_n\|} = \frac{r_1 \hat{q}_1 + \ldots + r_n \hat{q}_n}{\|r_1 \hat{q}_1 + \ldots + r_n \hat{q}_n\|} = r \cdot DLB(w; \hat{q}_1, \ldots, \hat{q}_n)$$

because $\|r\| = 1$, according to our assumption. Demonstration of the right invariance proceeds along the same lines. We see that DLB is indeed coordinate-invariant.

Thirdly, we show that when DLB is applied to two rigid transformations, it interpolates them along the shortest trajectory. Let therefore $\hat{p}, \hat{q}$ be two unit dual quaternions. We will denote their blending as $DLB(t; \hat{p}, \hat{q})$. Thanks to the left invariance of DLB (proven above) we can write

$$DLB(t; \hat{p}, \hat{q}) = \hat{p}^* \hat{DLB}(t; \hat{p}, \hat{q}) = \hat{p} \hat{DLB}(t; 1, \hat{p}^* \hat{q})$$

It is therefore sufficient to show that the path between 1 (the identity transformation) and $\hat{p}^* \hat{q}$ will be the shortest one, i.e., given by the screw corresponding to $\hat{p}^* \hat{q}$. Since $\hat{p}, \hat{q}$ are unit dual quaternions, so is $\hat{p}^* \hat{q}$ and therefore, according to Equation (5), there exists a unit dual quaternion with zero scalar part and $\hat{r}$ (a dual scalar) such that $\hat{p}^* \hat{q} = \cos \frac{\hat{r}}{2} + \hat{r} \sin \frac{\hat{r}}{2}$. Therefore, $DLB(t; 1, \hat{p}^* \hat{q})$ can be re-written as

$$DLB(t; 1, \hat{p}^* \hat{q}) = \frac{1 - t + \hat{r}^* \hat{p}^* \hat{q}}{1 - t + \hat{r}^* \hat{p}^* \hat{q}} = \frac{1 - t + t \cos \left( \frac{\hat{r}}{2} \right) + \hat{r} \sin \left( \frac{\hat{r}}{2} \right)}{1 - t + t \cos \left( \frac{\hat{r}}{2} \right) + \hat{r} \sin \left( \frac{\hat{r}}{2} \right)}$$
which means that the screw axis of $DLB(t; \mathbf{p}, \mathbf{q})$ is the same for all $t \in [0,1]$ and is given by $\tilde{n}$ – the only thing that varies is the angle of rotation and amount of translation, both encoded in $1-i\tau \cos(\frac{2}{3}) 1-i\tau \sin(\frac{2}{3})$ and $\frac{\tau}{1-i\tau \mathbf{p} \mathbf{q}}$. In other words, $DLB(t; \mathbf{p}, \mathbf{q})$ produces a shortest path screw motion. It should be noted that this screw motion does not have a constant speed – in an analogy to linear interpolation of regular quaternions. However, it can be demonstrated that the velocity of this motion is actually not far from constant [Kavan et al. 2006], which explains why this issue does not present any visible drawbacks in skinning.

### 3.4 The Final Algorithm

This section discusses how to efficiently implement dual quaternion skinning. First, we need to convert the skinning matrices $C_1, \ldots, C_p$ (where $p$ is the total number of joints) to dual quaternions $\mathbf{q}_1, \ldots, \mathbf{q}_p$, unless our application works with them already. This will typically be done on a CPU and will not take long, because the conversion to a dual quaternion involves just one quaternion multiplication (see Equation (4)) and the number of joints $p$ is usually quite small. The dual quaternions $\mathbf{q}_1, \ldots, \mathbf{q}_p$ are then sent to the GPU as uniform parameters, each represented by a $2 \times 4$ matrix.

The skin deformation, i.e., the DLB itself and conversion back to the matrix form, takes place in the vertex shader. Since the number of vertices tends to be orders of magnitude higher than the number of joints, the vertex shader code should be optimized. The first step of $DLB(\mathbf{w}; \mathbf{q}_1, \ldots, \mathbf{q}_p)$, i.e., the computation of the linear combination $\mathbf{b} = \sum_{j=1}^{p} w_j \mathbf{q}_j$, is straightforward. However, the following normalizations, i.e., the computation of $\mathbf{b}' = \frac{\mathbf{b}}{\|\mathbf{b}\|}$, and the subsequent conversion of $\mathbf{b}'$ to matrix $M$ can be optimized as follows:

If $\mathbf{b} = \mathbf{b}_0 + \varepsilon \mathbf{b}_e$, then the norm $\|\mathbf{b}\| = \|\mathbf{b}_0\| + \varepsilon \frac{(\mathbf{b}_0 \cdot \mathbf{b}_e)}{\|\mathbf{b}_0\|}$, according to Equation (3). The inverse is given by

$$\frac{1}{\|\mathbf{b}\|} = \frac{1}{\|\mathbf{b}_0\|} - \varepsilon \frac{(\mathbf{b}_0 \cdot \mathbf{b}_e)}{\|\mathbf{b}_0\|^3}$$

according to Equation (2). Therefore,

$$\mathbf{b}' = \frac{\mathbf{b}}{\|\mathbf{b}\|} = \frac{\mathbf{b}_0}{\|\mathbf{b}_0\|} + \varepsilon \left( \frac{\mathbf{b}_e}{\|\mathbf{b}_0\|} - \frac{\mathbf{b}_0 (\mathbf{b}_0 \cdot \mathbf{b}_e)}{\|\mathbf{b}_0\|^3} \right)$$

Now, $\mathbf{b}' = \mathbf{b}_0' + \varepsilon \mathbf{b}_e'$ needs to be converted to a homogeneous matrix $M$. Its rotational part is given simply by $\mathbf{b}_0' = \frac{\mathbf{b}_0}{\|\mathbf{b}_0\|}$. The translation is given by the vector part of $2b_e' (b_0')^*$. This simplifies to

$$2b_e' (b_0')^* = 2 \left( \frac{\mathbf{b}_e}{\|\mathbf{b}_0\|} - \frac{\mathbf{b}_0 (\mathbf{b}_0 \cdot \mathbf{b}_e)}{\|\mathbf{b}_0\|^3} \right) = 2 \left( \frac{\mathbf{b}_e - \mathbf{b}_0 (\mathbf{b}_0 \cdot \mathbf{b}_e)}{\|\mathbf{b}_0\|^3} \right)$$

Since the scalar part of $2b_e' (b_0')^* = 0$ (according to Equation (4)), it means that there is no need to evaluate the vector part of $\mathbf{b}_0 (\mathbf{b}_0 \cdot \mathbf{b}_e)/\|\mathbf{b}_0\|^3$. Therefore, we can compute the translational part of matrix $M$ just by computing the vector part of $2.\mathbf{b}_e/\|\mathbf{b}_0\|$, and its scalar part can be safely ignored. This means that we can avoid the lengthy dual quaternion normalization.

Matrix $M$ is then used to transform the input vertex $\mathbf{v}$. Thanks to the fact that $M$ is always a rigid transformation matrix, there are no complications with the inverse transposition and normalization (as in linear blending [Mohr and Gleicher 2003]) and we can use $M$ directly to transform also the vertex normal $\mathbf{v}_n$. Dual quaternion skinning can thus be summarized in the following pseudocode:

---

**4 Results and Comparison**

In our experiments, we use a female model with 5002 vertices, 9253 triangles and 54 joints. We compare the proposed dual quaternion skinning with linear blending, direct quaternion blending [Hejl 2004], log-matrix blending [Cordier and Magnenat-Thalmann 2005] and spherical blend skinning [Kavan and Zara 2005]. Some artifacts are better visualized on a simple model of cloth (6000 vertices, 12000 triangles and 49 joints). Note that the only variable in our experiments is the transformation blending – the input data (model files and postures) are always the same. The visual results
Figure 8: Comparison of direct quaternion blending (left) and dual quaternion blending (right). Only the latter delivers smooth deformation.

support our hypothesis that our DLB method (which satisfies the properties stated in Section 3.1) is indeed free of artifacts, see Figures 7, 8, 9 and 10.

In order to compare computational performance, we have implemented both CPU and GPU versions of dual quaternion skinning. The average times are reported in Figure 11. The performance of log-matrix vs. spherical blending has been compared only on the CPU, because the GPU implementation of these algorithms has not been provided in the previous work. Note that our implementation of log-matrix blending uses an optimization for rigid transformations based on the Rodrigues formula, as suggested in [Alexa 2002].

From the measurements, we see that dual quaternion, linear and direct quaternion blending [Hejl 2004] have quite similar performance. Although our algorithm is slightly slower than both linear and direct quaternion blending, we believe that this is not a high price to pay for the elimination of artifacts. When compared to log-matrix and spherical blending, we see that dual quaternion skinning is more than twice as fast (and also much easier to implement).

Figure 9: Comparison of log-matrix (left) and dual quaternion blending (right). The shortest-path property of dual quaternion blending guarantees natural skin deformations.

5 Conclusions

In this paper, we propose a novel skinning method based on the blending of dual quaternions. We justify its usefulness not only on practical examples, but also by identifying the mathematical properties important for skinning and showing that dual quaternions satisfy them. Our method is efficient, very simple to implement and does not require the modification of existing models or authoring tools (even though advanced methods exploiting the linearity of linear blend skinning, such as [Mohr and Gleicher 2003], would have to be adapted). We therefore believe that it provides a really practical alternative to the popular but inaccurate linear blend skinning.

The proposed algorithm has several limitations. Dual quaternion skinning simulates skin-to-skeleton binding in a simplistic way and is therefore unable to produce realistic musculature or dynamic effects. Also, dual quaternions are limited only to rigid transformations and are thus not suitable for models whose parts can scale or shear (e.g., some cartoon characters).

From a broader point of view, the blending of dual quaternions has potentially many more applications than just skinning. More generally, it is a method for rigid transformation blending with interesting properties. Therefore, it could be potentially useful in contexts such as motion blending, analysis or compression [Alexa 2002]. The in-
vestigation of other applications of dual quaternions in computer graphics promises to be an interesting area for future work.

6 Acknowledgements

We wish to thank Carlo H. Séquin for early illuminating discussion on the topic and the anonymous reviewers for their helpful comments. We acknowledge the support of the Higher Education Authority of Ireland. This work has been partly supported by the Ministry of Education of the Czech Republic under the research programs LC-06008 (Center for Computer Graphics) and MSM 6840770014.

References


