

Joint Detection, Interpolation, Motion and Parameter estimation for Image Sequences with Missing Data

Anil C.Kokaram, Simon J. Godsill

Signal Processing and Communications Group,
Department of Engineering, University of Cambridge.
email: ack@eng.cam.ac.uk, sjg@eng.cam.ac.uk

Abstract

This paper presents methods for detection and reconstruction of 'missing' data in image sequences which can be modelled using 3-dimensional autoregressive (3D-AR) models. The interpolation of missing data is important in many areas of image processing, including the restoration of degraded motion pictures, reconstruction of drop-outs in digital video and automatic 're-touching' of old photographs. Here a probabilistic Bayesian framework is adopted. The method assumes no prior knowledge of the motion field or 3D-AR model parameters as these are estimated jointly with the missing image pixels. Incorporating a degradation model into the framework allows detection to proceed jointly with interpolation.

1. Introduction

Missing data is a common form of degradation in real film and video data. It manifests as dropout in the digital stream in the case of digital broadcasting. In film the problem is caused by abrasion of the film material and the effect is called "Dirt and Sparkle". Previous work has considered the removal of this artefact as a two stage process, first detect the missing locations [1] and then reconstruct the underlying image data [2] using a spatiotemporal image sequence interpolation process. The reconstruction stage may be further specified as a motion reconstruction followed by an image reconstruction stage [3]. The motion interpolation stage is a crucial step in generating useful interpolated data since in the regions of missing data motion estimates are completely unreliable. It is possible to pose the motion reconstruction and image interpolation process as a joint problem [4], but this is an interim step towards a full specification of the problem under one framework. This paper addresses the problem of missing data interpolation for image sequences in perhaps a more coherent manner than has been done in the past. The discussion begins with a specification for the degradation and image model in Section 2 and then employs Bayesian inference to

design an algorithm that addresses the detection and reconstruction problem simultaneously. Section 3 illustrates the broad concepts, introducing the various priors employed for the unknowns and Section 4 then addresses the particular issue of stochastic solution of the system equations.

2. The Models

The observed degraded image sequence, $g(\mathbf{x})$ is assumed to be formed by

$$g(\mathbf{x}) = I(\mathbf{x})(1 - b(\mathbf{x})) + b(\mathbf{x})c(\mathbf{x}) \quad (1)$$

where location $\mathbf{x} = (x, y, n)$ (i.e. co-ordinate (x, y) in the n th frame of the sequence), $I(\mathbf{x})$ is the pixel intensity in the uncorrupted true original, $b(\mathbf{x})$ is a detection variable which is set to 1 when at locations of corruption and 0 otherwise, $c(\mathbf{x})$ is the observed intensity in the corrupted region. It is straightforward to extend the model to include additive noise degradation, and this is left for future work.

It is assumed that the underlying, clean original image sequence can be well modelled by a three dimensional autoregressive model. This model (see, e.g. [5]) with coefficients a_k (for $k = 1 \dots P$) can be represented as

$$I(\mathbf{x}) = \sum_{k=1}^P a_k I(\mathbf{x} - \mathbf{q}_k - \mathbf{d}_{n,n_k}(\mathbf{x})) + \epsilon(\mathbf{x}) \quad (2)$$

$\mathbf{q}_k = (qx_k, qy_k, qn_k)$ is the offset corresponding to coefficient a_k and $\mathbf{d}_{n,n_k}(\mathbf{x}) = (dx_k, dy_k, 0)$ is the motion offset between frames $n - qn_k$ and n at location \mathbf{x} . $\epsilon(\mathbf{x})$ is assumed to be white and Gaussian with variance σ_ϵ^2 .

3. Joint solution

Bayes theorem allows a framework for joint estimation of the unknowns θ and \mathbf{I}_n by writing

$$p(\theta, \mathbf{I}_n | \mathbf{g}_n) \propto p(\mathbf{g}_n | \theta, \mathbf{I}_n) p(\mathbf{I}_n | \theta) p(\theta) \quad (3)$$

where \mathbf{g}_n represents the degraded image frames $n - 1, n, n + 1$, and \mathbf{I}_n is the true data in frames $n -$

1, n , $n + 1$. It is necessary to find expressions for the likelihood $p(\mathbf{g}_n|\theta, \mathbf{I}_n)$, the data model probability $p(\mathbf{I}_n|\theta)$ and prior probability density $p(\theta)$. The first distribution is the likelihood expression, $p(\mathbf{g}_n|\theta, \mathbf{I}_n)$ and is simply $\delta(g(\mathbf{x}) - I(\mathbf{x})(1 - b(\mathbf{x})) - b(\mathbf{x})c(\mathbf{x}))$. This degenerate distribution is a necessary consequence of the ‘‘replacement’’ or ‘‘switching’’ behaviour of the degradation process.

The second distribution, $p(\mathbf{I}_n|\theta)$, is the image model. Employing the 3D-AR model here allows the Gaussian nature of the residual sequence to be exploited. Considering a volume of pixel intensities suitably motion compensated and scanned into a vector \mathbf{i} , and a matrix of coefficients \mathbf{A} arranged such that $\mathbf{e} = \mathbf{A}\mathbf{i}$ is a vector of residuals, $\epsilon(\mathbf{x})$; then

$$p(\mathbf{I}_n|\theta) \propto \exp\left(-\frac{\mathbf{i}^T \mathbf{A}^T \mathbf{A} \mathbf{i}}{2\sigma_\epsilon^2}\right) \quad (4)$$

To complete the description of the posterior distribution in equation 3, priors need to be assigned to the parameters. Suitable ‘non-informative’ [6] priors can be assigned to $\mathbf{a}, \sigma_\epsilon^2$. In this case $p(\sigma_\epsilon^2) \propto 1/\sigma_\epsilon^2$. The priors for motion and the detection and corruption fields are discussed next.

4. The motion prior

To encourage smoothness in the local motion field, a Gibbs Energy Prior is assigned to the motion employing an 8 nearest neighbour support. This effectively allows for implicit motion field interpolation in the missing region. To reduce the complexity of the final solution the motion field is block based, with one motion vector being employed for each specified block in the image.

In the manner of e.g. [7, 8] the prior for $\mathbf{d}_{n,n-1}(\mathbf{x})$, the motion vector mapping the pixel at \mathbf{x} in frame n into frame $n - 1$, is as follows.

$$p(\mathbf{d}_{n,n-1}(\mathbf{x})|\mathbf{S}_n(\mathbf{x})) \propto \exp\left(-\left(\sum_{\mathbf{v} \in \mathbf{S}_n(\mathbf{x})} \lambda(\mathbf{v})[\mathbf{d}_{n,n-1}(\mathbf{x}) - \mathbf{v}]^2\right)\right) \quad (5)$$

where \mathbf{v} is each vector in the neighborhood represented by $\mathbf{S}_n(\mathbf{x})$, and $\lambda(\mathbf{v})$ is the weight associated with each clique. In order to discourage ‘smoothness’ over too large a range, $\lambda(\mathbf{v})$ is defined as $\lambda(\mathbf{v}) = 1/|\mathbf{X}(\mathbf{v}) - \mathbf{x}|$ where $\mathbf{X}(\mathbf{v})$ is the location of the block providing the neighborhood vector \mathbf{v} . This location is measured in terms of blocks. This paper does not address the issue of incorporating discontinuities into the motion field both temporally and spatially.

5. Priors for corruption and detection

In practice, each region of missing data tends to have fairly constant intensity (see figure 1), therefore it is reasonable to place a similar energy prior on both the binary field $b(\mathbf{x})$ and the blotch value field $c(\mathbf{x})$.

It is found that acknowledging discontinuities in these fields leads to much better behaviour. Note that edges in the c and b fields must correspond to edges in the image since the corrupted areas will be generally well delineated from their surrounding by a marked grey scale transition. Thus a simple zero crossing edge detector employed on the degraded image will enable the rough configuration of an edge field that can be used subsequently to define the priors on the c and b fields. This data dependent prior must be regarded as an approximation employed solely for improving convergence of the algorithm.

The priors are therefore defined as follows

$$p_c(c(\mathbf{x})|C) \propto \exp\left(-\sum_{z \in C} \lambda_z(\mathbf{x})(1 - u(z, \mathbf{x}))|c(\mathbf{x}) - z|\right)$$

$$p_b(b(\mathbf{x})|B) \propto \exp\left(-\sum_{z \in B} (1 - u(z, \mathbf{x}))|b(\mathbf{x}) - z|\right)$$

where z is one of the values of the variables d or b in the 8 connected neighborhood of the location \mathbf{x} , denoted C , B . $u(z, \mathbf{x})$ is set to 1 if there is a significant zero crossing between the location \mathbf{x} in the image and location from which z is extracted. Thus the smoothness constraint is turned off across significant zero crossings in the image. Note that these priors are defined on the *pixel* resolution image grid, whereas the motion prior discussed previously is defined on a *block* basis.

6. Solution using the Gibbs Sampler

Consider that \mathbf{i} contains at least the missing pixels (indicated at $b(\mathbf{x}) = 1$) and their immediate AR support. The missing pixels are denoted \mathbf{i}_U and the remaining pixels as \mathbf{i}_K . The vector of AR parameters a_k is denoted \mathbf{a} . The Gibbs Sampler now operates iteratively, given some starting guess for the unknowns, by drawing random samples from the conditional posterior distribution for each unknown in turn:

$$\mathbf{a} \sim p(\mathbf{a}|\mathbf{i}, \sigma_\epsilon^2, \mathbf{b}, \mathbf{d}, \mathbf{c}, \mathbf{g}); \quad \mathbf{b} \sim p(\mathbf{b}|\mathbf{i}, \sigma_\epsilon^2, \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{g});$$

$$\mathbf{c} \sim p(\mathbf{c}|\mathbf{i}, \sigma_\epsilon^2, \mathbf{a}, \mathbf{d}, \mathbf{b}, \mathbf{g}); \quad \sigma_\epsilon^2 \sim p(\sigma_\epsilon^2|\mathbf{a}, \mathbf{i}, \mathbf{b}, \mathbf{d}, \mathbf{c}, \mathbf{g});$$

$$\mathbf{d} \sim p(\mathbf{d}|\mathbf{a}, \mathbf{i}, \mathbf{b}, \sigma_\epsilon^2, \mathbf{c}, \mathbf{g}); \quad \mathbf{i}_U \sim p(\mathbf{i}_U|\mathbf{a}, \mathbf{i}_K, \sigma_\epsilon^2, \mathbf{b}, \mathbf{d}, \mathbf{c}, \mathbf{g})$$

where \mathbf{c}, \mathbf{b} etc are vectors containing the relevant parameters. These conditionals can be derived by manipulation of the joint posterior given in equation 3. This sampling procedure is repeated until convergence is reached according to some suitable criterion. The

framework allows straightforwardly for minimum mean-square (MMSE), maximum *a posteriori* (MAP) or sampled estimates of \mathbf{i}_U by simple manipulation of the sampled values following convergence.

Assuming a ‘non-informative’ [6] framework of prior distributions for \mathbf{a} , σ_ϵ^2 , the sampling operations for \mathbf{a} , σ_ϵ^2 , \mathbf{i}_U involve simple random draws from well-known distributions. However some practical considerations encourage an altered sampling strategy.

7. Adaptations to the Gibbs Sampler

The convergence of the Gibbs Sampler is generally improved if several unknowns are sampled jointly [9]. This is possible using the method of composition [4]. In this scheme, a random draw from the density $p(\mathbf{a}, \sigma_\epsilon^2, \mathbf{d} | \theta_{-\mathbf{a}, -\mathbf{d}, -\sigma_\epsilon})$, for instance, is made possible by the decomposition

$$p(\mathbf{a}, \sigma_\epsilon^2, \mathbf{d} | \mathbf{b}, \mathbf{c}, \mathbf{i}) = p(\mathbf{a} | \mathbf{d}, \mathbf{b}, \mathbf{c}, \sigma_\epsilon^2, \mathbf{i}) p(\sigma_\epsilon^2 | \mathbf{d}, \mathbf{b}, \mathbf{c}, \mathbf{i}) p(\mathbf{d} | \mathbf{b}, \mathbf{c}, \mathbf{i}) \quad (6)$$

Note that \mathbf{a} , σ_ϵ^2 , \mathbf{d} are conditionally independent of \mathbf{g} . The various composition terms can be derived by successively *integrating out* \mathbf{a} , then σ_ϵ^2 from the posterior distribution.

Random draws from $p(\mathbf{a}, \sigma_\epsilon^2, \mathbf{d} | \mathbf{b}, \mathbf{c}, \mathbf{i})$ can therefore be implemented by drawing from $p(\mathbf{d} | \mathbf{b}, \mathbf{c}, \mathbf{i})$ followed by a draw from $p(\sigma_\epsilon^2 | \mathbf{d}, \mathbf{b}, \mathbf{c}, \mathbf{i})$ using the value of \mathbf{d} drawn previously, then similarly for $p(\mathbf{a} | \dots)$, using the samples just generated for σ_ϵ^2 , \mathbf{d} . In this manner a joint draw for \mathbf{b} , \mathbf{c} , \mathbf{i}_U is also achieved.

The expressions required to perform these joint draws can be derived [4] by forming a vector \mathbf{e} containing all the excitation terms from equation 2 within a block of data and expressing it as $\mathbf{e} = \mathbf{i} - \mathbf{I}\mathbf{a}$. This then leads to the following distributions required in the joint draw $p(\mathbf{a}, \sigma_\epsilon^2, \mathbf{d} | \mathbf{i})$.

$$\begin{aligned} p(\mathbf{a} | \mathbf{d}, \mathbf{i}, \sigma_\epsilon^2) &= \mathcal{N}_P(\hat{\mathbf{a}}, \sigma_\epsilon^2 (\mathbf{I}^T \mathbf{I})^{-1}) \\ p(\sigma_\epsilon^2 | \mathbf{d}, \mathbf{i}) &= \text{IG}((N - P)/2, E(\hat{\mathbf{a}}, \mathbf{i}, \mathbf{d})/2) \\ p(\mathbf{d} | \mathbf{i}) &\propto \frac{E(\hat{\mathbf{a}}, \mathbf{i}, \mathbf{d})^{-(N-P)/2}}{|\mathbf{I}^T \mathbf{I}|^{1/2}} \\ \text{where } \hat{\mathbf{a}} &= (\mathbf{I}^T \mathbf{I})^{-1} \mathbf{I}^T \mathbf{i} \\ \text{and } E(\hat{\mathbf{a}}, \mathbf{i}, \mathbf{d}) &= \sum_{\mathbf{x} \in \text{block}} \epsilon(\mathbf{x})^2 \end{aligned} \quad (7)$$

N is the number of pixels in the image block.

In practice the draws for \mathbf{b} , \mathbf{c} , \mathbf{i} are performed jointly on a pixel by pixel basis, sampling from the expression

$$p(b(\mathbf{x}), c(\mathbf{x}), i_n(\mathbf{x}) | \mathbf{d}, \mathbf{a}, \sigma_\epsilon^2, B, C, \mathbf{i}_n(\bar{\mathbf{x}}), \mathbf{g}) \quad (8)$$

where $i_n(\mathbf{x})$ is the true intensity at \mathbf{x} in frame n , and $i_n(\bar{\mathbf{x}})$ denotes a neighborhood of previously estimated

intensities. This joint draw is unusual because of the degenerate delta function in the likelihood, but since b is a binary field, c is at most an 8 bit field, and this draw is specified on the *pixel* grid a feasible sampling scheme results. The distributions required for the composition sampling can be derived by integrating the posterior.

The utility of the algorithm is increased if there already exist initial estimates for motion in particular. It then becomes possible to draw samples from the numerically evaluated p.d.f. for $\mathbf{d} \sim \mathbf{p}(\mathbf{d} | \mathbf{i}, \sigma_\epsilon^2, \mathbf{a}, \mathbf{b}, \mathbf{c})$ in a local region around the current estimate. In essence, the actual procedure employed for sampling for \mathbf{d} involves proposing 8 candidate vectors from the neighborhood and perturbing these surrounding vectors by ± 1 pixel. The motion sample is then drawn from this set by direct numerical evaluation of the probability distribution assuming that the probability of all other samples is zero. Initial motion estimates can be taken from any number of motion estimators currently available, a multiresolution gradient based technique is employed here [3].

8. Results and Final Comments

The top of figure 1 shows a section of corrupted original from a movie sequence, with the major distortion boxed in white. The Gibbs Sampler was allowed to run for 120 iterations, with $\lambda_c = 0.15$, $\lambda_b = 4.0$, a 5 point temporal AR model, and a 10 sample ‘burn-in’. Figure 1 third from top, shows the locations which were set to 1 more than 100 times during the iterations. These figures show a much better detection behaviour than simple temporal difference thresholding¹ (shown second from top in figure 1), although the false alarms tend to be larger where they do occur. The next two pictures show the MMSE result (i.e. the average of the samples for \mathbf{i}_U) generated from the last 110 samples, and the 120th sample respectively. The reconstructions show good performance, even in the areas of false alarm.

For comparison, the last picture shows the reconstruction result generated from the alternative three stage method [3] using a spatio-temporal median filter as the interpolator. It illustrates that the current result is of better quality. The technique presented here does suffer from the drawback that it is unable to successfully delineate the extremities of a blotch. The remnant blotch boundaries can just be seen in the MMSE reconstruction as compared to the median result where the problem was artificially solved by using a pessimistic estimate of the blotch location. This can be addressed by incorporating a low-pass filter into the

¹Threshold set to 10.0 here

degradation model proposed here, perhaps with detrimental effects on convergence.

The results so far indicate that the proposed technique is potentially quite a powerful process with respect to “hands-off” operation; however further testing is needed to confirm whether this is not offset by the increased computational load.

9. References

- [1] A. Kokaram, R. Morris, W. Fitzgerald, and P. Rayner. Detection of missing data in image sequences.. *IEEE Image Processing*, pages 1496–1508, Nov. 1995.
- [2] A. Kokaram, R. Morris, W. Fitzgerald, and P. Rayner. Interpolation of missing data in image sequences.. *IEEE Image Processing*, pages 1509–1519, Nov. 1995.
- [3] A. Kokaram and S. Godsill. A system for reconstruction of missing data in image sequences using sampled 3D AR models and MRF motion priors. In *European Conference on Computer Vision 1996*, pages 613–624. Springer Verlag, April 1996.
- [4] S. Godsill and A. Kokaram. Joint interpolation, motion and parameter estimation for degraded image sequences with missing data. In *Signal Processing VIII*, volume I, pages 1–4, September 1996.
- [5] P. Strobach. Quadtree-structured linear prediction models for image sequence processing. *IEEE PAMI*, 11:742–747, July 1989.
- [6] G. E. P. Box and G. C. Tiao. *Bayesian Inference in Statistical Analysis*. Addison-Wesley, 1973.
- [7] J. Konrad and E. Dubois. Bayesian estimation of motion vector fields. *IEEE Trans PAMI*, 14(9), September 1992.
- [8] C. Stiller. Motion-estimation for coding of moving video at 8kbit/sec with gibbs modeled vectorfield smoothing. In *SPIE VCIP.*, volume 1360, pages 468–476, 1990.
- [9] J. Liu, W. H. Wong, and A. Kong. Covariance structure of the Gibbs sampler with applications to the comparison of estimators and augmentation schemes. *Biometrika*, 81:27–40, 1994.

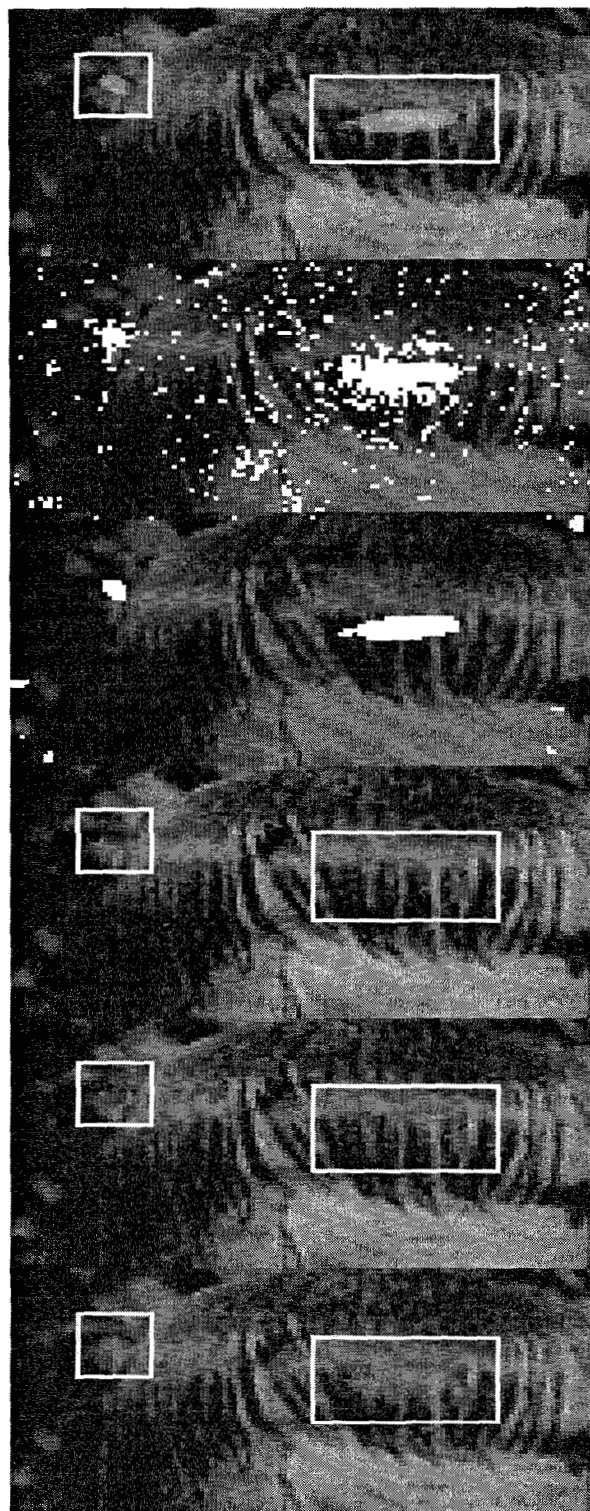


Figure 1: Top to bottom : Degraded original, Detection using SDIA, Detection with the Gibbs Sampler, MMSE Interpolation, Sample from sampler, Interpolation using median.