Detection and Interpolation of Replacement Noise in Motion Picture Sequences using 3D Autoregressive Modelling.

A. Kokaram and P. Rayner,
Department of Engineering, Cambridge University, Trumpington St.,
Cambridge CB2 1PZ.
Tel : (+44) (0)223 332767 *email* : ack@cam.eng.ac.uk

ABSTRACT

Perhaps the most common distortion in degraded film or video material is the occurrence of Blotches of varying colour randomly dispersed in each frame. The Blotches are caused by the abrasion of the film or collection of dirt on the film as it passes through the projection apparatus. They manifest as flashes of bright and dark areas. The Blotches represent regions of missing data (ie they *replace* the original data), and the paper discusses a model based technique for suppressing this artefact.

INTRODUCTION

It is typical to remove impulsive distortion of this kind by applying a Median filter to each pixel in the degraded image sequence. Since Blotches tend to be large in practice ($> 10 \times 10$ pixels) spatio–temporal operations are most successful [8, 2]. However, the median operation tends to blur fine detail and it would be better to restrict the processing to Blotched sites only. Therefore the first part of the paper discusses methods for detecting Blotches. This approach was used in [8], in which a motion compensated detector was used to engage a motion compensated 3D MMF (ML3Dex) at the Blotch sites.

The Median filter is not an interpolator however, and though it is effective in practice, it would be beneficial to develop techniques for interpolation that preserve image detail more effectively. This can be done using the 3D Autoregressive (AR) model to incorporate more spatio–temporal information in a coherent manner.

Because the techniques discussed involve 3D (space/time) operations, it is necessary to direct the operation of the algorithm along motion trajectories. For this paper it is sufficient to note that a Multiresolution Block Matching technique was used that combines ideas from [4, 3] to achieve integer accurate motion estimates. The 3D AR model acts along motion trajectories and implicitly interpolates image information when fractional motion is observed.

THE 3D AR MODEL

This model has been used previously by Strobach [10] for image segmentation and Efstratiadis et al. [6] for motion estimation. The model equation for image frame n is therefore

$$I(i,j,n) = \sum_{k=1}^{N} a_k I(i + q_k^x, j + q_k^y, n + q_k^z) + e(i,j,n) \quad (1)$$

where the grey value at location (i,j) in frame n is given as $I(i,j,n)$; the N model coefficients are defined as a_k $(k = 1 \ldots N)$, and the corresponding location of the kth offset support pixel is given by the support vector $\mathbf{q}_k = [q_k^x, q_k^y, q_k^z]$. It is assumed that the data has already been compensated for motion and this displacement is incorporated into the offsets required for the model support.

DETECTION

Blotches can be characterized as spatio–temporal discontinuities in grey level along a motion trajectory. Assume that the image is corrupted according to the following equation.

$$
\begin{aligned}
g(\mathbf{x}) &= I(\mathbf{x}) + b(\mathbf{x}) \\
\text{where } b(\mathbf{x}) &= \left\{ \begin{array}{ll} 0 & \text{With probability } (1 - P_s) \\ s & \text{With probability } P_s \end{array} \right.
\end{aligned} \quad (2)
$$

Here, s is a randomly distributed grey level representing a Blotch or impulse, occurring at random instants with probability P_s. It is required to detect the occurrences of $b(\mathbf{x}) \neq 0$ in order to isolate the distortion for further examination. The solution is to assume that the undistorted image $I(i,j,n)$, because of its temporal (and to some extent spatial) smoothness, obeys the AR model whereas $b(\mathbf{x})$ does not.

Suppose that the model coefficients for a particular image sequence were known. The prediction error could then be considered to be noise with some correlation structure (See [7]). If $g(\mathbf{x})$ was filtered with the model prediction filter the output could be written as (using eq. 2)

$$\epsilon'(\mathbf{x}) \;=\; g(\mathbf{x}) - \sum_{k=1}^{N} a_k g(\mathbf{x} + \mathbf{q}_k)$$

$$= \; \epsilon(\mathbf{x}) + b(\mathbf{x}) - \sum_{k=1}^{N} a_k b(\mathbf{x} + \mathbf{q}_k) \qquad (3)$$

where $\mathbf{x}$ is the position vector $[i,\ j,\ k]$. Equation 3 shows that the undistorted samples in the degraded signal are reduced to the scale of the residual sequence. The distorted samples can be reduced in amplitude if there are other distorted samples nearby but this does not occur often. Therefore the prediction error generated at a distorted pixel site would be larger than that at uncorrupted sites.

Therefore thresholding the prediction error is a useful - method of detecting local distortion. In practice, it is best to use a pair of AR models looking forward and backward in time; and observe the two prediction errors at each pixel. This avoids problems with occlusion and uncovering.

Figure 1 compares the receiver operating characteristics when different 3D AR models are used to detect known distortion in an image sequence (WESTERN1) of 60 frames of resolution 256×256. The Blotches were of varying size, shape, and grey level, with a probability of corruption of 0.007. The curves were created by measuring the correct detection and false alarm rates for 20 different thresholds evenly distributed between 0 and 2000.

The models are described by the number of pixels of symmetric support in each frame. A support of 9 pixels therefore refers to a square of 3×3. The models are described in the order *past frame : current frame : next frame*. Therefore a model of the form 1:0:1 yields a prediction equation

$$\hat{I}(i,j,n) = a_{-1}I(i,j,n-1) + a_1 I(i,j,n+1) \quad (4)$$

The detection system uses a forward predictor, N:0, and a backward predictor, 0:N. The two prediction error fields[1] then flag a distortion when

$$(\ [\epsilon_1(\mathbf{x})]^2 \ge t_\epsilon\ ) \ \text{AND} \ (\ [\epsilon_2(\mathbf{x})]^2 \ge t_\epsilon\ ) \qquad (5)$$

Therefore, a Blotch is located when both predictors agree that a match cannot be found in either of the two frames. For figure 1 block sizes of 9×9 were used in generating the prediction error and model coefficients were estimated using the Normal equations.

Because this degradation is artificial, the actual model coefficients can be estimated from the original data prior to degradation. The curves corresponding to these experiments are labelled *Known*, the other model based

detection curves refer to the the case where the model coefficients are estimated from the degraded images. It is clear that the presence of degradation biases the model coefficients adversely resulting in a much worse performance.

Also shown is the result using the SDIa (Spike Detection Index) detector which is a special case of the general AR detector. It uses a 1:0/0:1 system with the model coefficients fixed at 1.0. The result indicates that it performs better in a real situation than the model based approach. However, the performance gained with the 'Known' AR models shows that provided a technique can be found to suppress the effect of distortion in the modelling process [5], there is much to be gained in a spatio–temporal approach. The SDIa can be seen to be similar to a motion compensated version of the detector presented in [9].

INTERPOLATION

The task is now to interpolate information in the blotched regions. This can be done by extending the techniques of [11] to multidimensions. It is assumed that the image function in a block of $M \times M$ pixels is sufficiently stationary for a single model to be used for that block. Further, assume that arbitrary locations within the block are corrupted.

Allowing for a border of pixels at the edge of the $M \times M$ block in the current frame, n say, (so that $(\mathbf{x} + \mathbf{q}_k)$ will never result in a location outside the $M \times M$ block), an equation for the error at every pixel within a centred $B \times B$ block in that frame can be written as below. (Note that a volume of motion compensated data is used.)

$$\mathbf{e} = \mathbf{A}\mathbf{i} \qquad (6)$$

where $\mathbf{i}$ represents an $NM^2 \times 1$ column vector of row ordered pixels from the N $M \times M$ blocks[2], $\mathbf{e}$ is a $B^2 \times 1$ column vector of errors, and $\mathbf{A}$ a matrix of coefficients satisfying the model equation at all the considered points. This coefficient matrix is of size $B^2 \times NM^2$. The vector $\mathbf{i}$ contains intensities of both known and unknown pixels. If this vector is separated into two vectors $\mathbf{i}_u$ (u for unknown) and $\mathbf{i}_k$ (k for known), which represent the known and unknown pixel intensities, then equation 6 can be written as

$$\mathbf{e} = \mathbf{A}_k\mathbf{i}_k + \mathbf{A}_u\mathbf{i}_u \qquad (7)$$

Here, $\mathbf{A}_k$, $\mathbf{A}_u$ are the coefficient matrices corresponding to the known and unknown data vectors. They are submatrices of the $\mathbf{A}$ matrix, made by extracting the relevant columns.

To derive an interpolation, $\mathbf{i}_u$ must be found. Following Vaseghi [11], this is done by minimizing the squared

---

[1] ϵ_1 , ϵ_2

[2] N frames are used

error $\mathbf{e}^T\mathbf{e}$ with respect to $\mathbf{i}_u$. This yields the solution

$$\mathbf{i}_u = -[\mathbf{A}_u^T\mathbf{A}_u]^{-1}\mathbf{A}_u^T\mathbf{A}_k\mathbf{i}_k \qquad (8)$$

This equation implies knowledge of the model coefficients which must be estimated from the corrupt data.

The prediction error at the suspected distorted points may be weighted to zero so that this 'doubtful' data does not affect the estimation process. The general approach is to weight the prediction error by some function $w(\mathbf{x})$ such that the new prediction equation may be written as

$$\epsilon_w(\mathbf{x}) = w(\mathbf{x})\sum_{k=0}^{N} a_k I(\mathbf{x} + \mathbf{q}_k) \qquad (9)$$

Where $a_0 = 1.0$, and $\epsilon_w(\mathbf{x})$ is the weighted error at position $\mathbf{x}$. Minimizing the squared error $[\epsilon_w(\mathbf{x})]^2$ with respect to a_k then yields weighted Normal equations, which can be solved as usual. A weighting function set to 0 for all the Blotch positions and 1 otherwise is found to be effective in practice.

The figure 4 shows a photograph of one frame of the *GO-RN* sequence (fig. 3) artificially corrupted. The images are 256×256 and have been shrunk vertically to fit in the space provided. The sequence consists of 3 frames in which a main figure (the *GORN*) undergoes a small rotation, zoom and camera pan, amounting to a motion of less than 5 pixels per frame. Perfect detection was assumed. Only the second frame of the sequence was corrupted. All the interpolators used the same motion vectors from the same motion estimator.

Figures 5 and 6 show the result using ML3Dex [8] and 3D AR interpolation respectively. The filters presented in [2, 1] are not compared since they cannot remove this size distortion in a non–recursive implementation. The 3D AR algorithm used a 9:8 causal model with equations set up via the prediction error in the current and next frame. A block size of 17×17 pixels was used to generate the model coefficients via weighted estimation. The restorations are virtually identical except for the right eye of the creature. The median operation replaces two blocks there with a fairly uniform intensity, but the AR interpolator restores the texture almost perfectly. It can be seen that ultimately, the median filter is not as accurate as the model based interpolator for textured regions. The MSE for each restoration is 5.2 and 1.86 for the median and 3D AR processes respectively.

Figure 2 compares the performance of the Median Vs the AR interpolators on the corrupted WESTERN1 sequence (See fig. 1.) The motion estimation was done on the degraded sequence. The SDIa detector was used to give 80% Correct and 1% False Alarm rate. The curves show that the model based approach can give consistently better results than the Median operation. The 9:8

model performs the best because it incorporates spatial as well as temporal information. Even though ML3Dex also does this, it is more sensitive to errors in motion estimation since it relies more heavily on spatial information from surrounding frames.

CONCLUSIONS

For detection, the 3D AR model is unlikely to be useful unless a technique is used to remove the effects of outliers [5]. The SDIa detector has been introduced and this performs better than previous techniques. For interpolation, the 3D AR model is able to reconstruct the image with high accuracy, and is better able to handle fine texture than a median operation. The success of the technique for image sequences indicates that the high temporal correlation in the image sequence allows the AR framework to cope with spatial non–stationarities. However, its computational complexity is higher than the Median operation and this restricts its usefulness in real time applications in the short term.

References

[1] Bilge Alp, Petri Haavisto, Tiina Jarske, Kai Öistämö, and Yrjö Neuvo. Median-based algorithms for image sequence processing. In *SPIE Visual Communications and Image Processing*, pages 122–133, 1990.

[2] G.R. Arce. Multistage order statistic filters for image sequence processing. *IEEE Transactions on Signal Processing*, 39:1146–1161, May 1991.

[3] M. Bierling. Displacement estimation by heirarchical block matching. In *SPIE VCIP*, pages 942–951, 1988.

[4] J. Boyce. Noise reduction of image sequences using adaptive motion compensated frame averaging. In *IEEE ICASSP*, volume 3, pages 461–464, 1992.

[5] E. DiClaudio, G. Orlandi, F. Piazza, and A. Uncini. Optimal weighted LS AR estimation in presence of impulsive noise. In *IEEE ICASSP.*, volume E3.8, pages 3149–3152, 1991.

[6] S. Efstratiadis and A. Katsagellos. A model based, pel–recursive motion estimation algorithm. In *Proceedings IEEE ICASSP*, pages 1973–1976, 1990.

[7] A.K. Jain. *Fundamentals of Digital Image Processing*. Prentice Hall, 1989.

[8] A. C. Kokaram. *Motion Picture Restoration*. PhD thesis, Cambridge University, England, May 1993.

[9] R. Storey. Electronic detection and concealment of film dirt. *SMPTE Journal*, pages 642–647, June 1985.

[10] P. Strobach. Quadtree–structured linear prediction models for image sequence processing. *IEEE PAMI*, 11:742–747, July 1989.

[11] S. V. Vaseghi. *Algorithms for the restoration of archived gramophone recordings*. PhD thesis, Cambridge University, England, 1988.
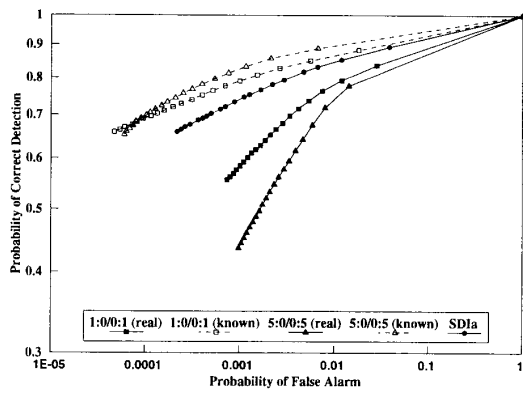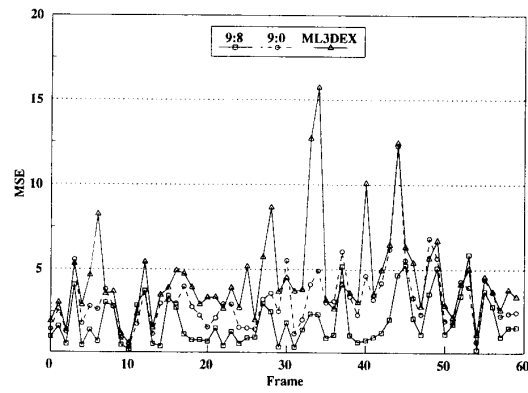
Figure 1: Detection performance.

Figure 2: Error in reconstruction.

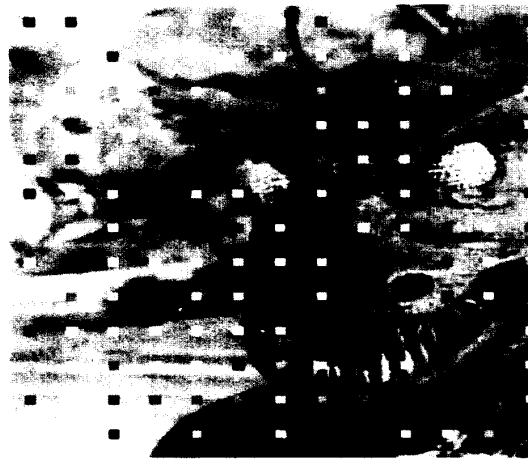Figure 3: Original Frame 2 of *GORN*.

Figure 4: Corrupted Frame 2 of *GORN*. (Missing patches are 5 × 5 large.)

Figure 5: Restored using ML3Dex.

Figure 6: Restored using 9:8 3D AR model.