TWO LAYER SEGMENTATION FOR HANDLING PATHOLOGICAL MOTION IN DEGRADED POST PRODUCTION MEDIA

†*Ben Kent* **Anil Kokaram* †*Bill Collis* *Simon Robinson*†

†The Foundry, London UK
www.thefoundry.co.uk

*Electronic and Electrical Engineering Department,
University of Dublin, Trinity College, Ireland

ABSTRACT

This paper presents a mechanism for dealing with incorrect motion estimation in degraded post production image sequences. This tends to be caused by *Pathological Motion* (a combination of motion blur, complex foreground motion, shadows etc). We describe a method for estimating where such regions are likely to occur by segmenting sequences into foreground and background motion. We show it is possible to produce a conservative matte of which regions in a sequence are foreground, and that blotch detection can be adapted accordingly in such areas.

1. INTRODUCTION

The treatment of missing data in film and video sequences has been of increasing interest in recent years [1] with the advent of consumer digital visual media. The demand for good quality images has increased the requirement for automatic techniques for detecting and removing artefacts like Dirt and Sparkle, Missing Frames and so on from archived video and film media. It is interesting to note that in the digital post-production of film it is routinely necessary to remove relatively small Dirt and Sparkle defects from *modern* film that has been scanned for special effect editing. The level of degradation in this environment is obviously lower than is the case with archived material purely because of the length of time the film material has had to degrade. Nevertheless, the problem of missing data treatment is therefore not restricted purely to the archived media processing domain. Missing data manifesting as defects in an image sequence e.g. Dirt/Sparkle will be called *blotches* in the rest of this paper.

A great deal of effort has been concentrated on developing automated blotch detect/remove tools [1, 2]. However, recent work[2, 3] has highlighted that problems always occur when processing hours rather than seconds of video. This is because much of the work in video processing has assumed that it is possible to write the video model as follows.

$$I_n(\mathbf{x}) = I_{n-1}(\mathbf{f}(\mathbf{x})) + e(\mathbf{x}) \tag{1}$$

where the intensity of the pixel at position $\mathbf{x}$ in frame n is given as $I_n(\mathbf{x})$ and $e(\mathbf{x})$ is a Gaussian distributed error. The model relies on the underlying assumption that it is somehow always possible to build any frame from an image sequence by warping and perhaps cutting parts from other frames (in this case the previous only) and rearranging them. Unfortunately, this is hardly always the case. Fast motion of objects causes blurring, many interesting objects, e.g. clothing, are not rigid and shadows/lighting effects due to self occlusion cause additional complications. This means that in some parts of any sequence it will be impossible to model the behaviour regardless of whether esoteric 3D models could be used. These portions of the image sequence are said to undergo *Pathological Motion (PM)*. Figure 1 shows three frames (I_{n-1}, I_n, I_{n+1}) of a man running up some stairs and the motion compensated frames (I'_{n-1}, I'_{n+1}). An optic flow type method was used for motion estimation here [1]. Note how the motion estimation fails to correctly model the complicated, blurred motion of the man.

Fig. 1. Top row: Frames $n-1, n, n+1$ of a sequence showing PM; note legs and arms show heavy blur and are simply not well correlated between frames. Bottom row: motion compensated frames $n-1, n+1$ showing poor picture reproduction in the region of PM.

All the missing data techniques currently employed assume that missing data manifests as temporal discontinuities. These detect the presence of PM as a temporal discontinuity, and hence remove what is a legitimate part of the sequence. Although it tends to be difficult for viewers to see restoration defects in areas of PM, post-production artists demand that no part of the image is degraded by the blotch removal process. Therefore, no processing of PM regions is often preferred.

A mechanism for detecting PM would therefore assist any missing data treatment process. Rares and Bornard et al [3, 2] present different schemes for doing this. Rares relies on explicit identification of different types of picture content that could be an

*anil.kokaram@tcd.ie www.mee.tcd.ie/∼ack
†ben@thefoundry.co.uk,bill@thefoundry.co.uk,simon@thefoundry.co.uk

indicator of PM. Although thorough, that technique is computationally expensive and relies heavily on a classification step. The technique presented by Bornard is more implicit and exploits the constraint that missing data in degraded image sequences does not normally cause temporal discontinuities in the same place in several frames. Thus detection of the occurrence of consecutive discontinuities is a direct indication of PM. Bornard's technique uses an MRF model for discontinuity states first presented in [4] and relies on a temporal window of 5 frames.

The previous techniques attempt to express the PM problem in the image space in a pixel flow type framework. What is needed is a mechanism for identifying objects in the sequence and then exploiting that knowledge to discriminate between legitimate object surfaces and defects. Such a process is difficult for archived motion picture film and video because the heavy level of degradation can seriously affect object estimation. However, in the post production case, several observations encourage a different approach to Dirt detection. First of all, the level of degradation is lower than archival material. Missing regions tend to be of sizes up to 20×20 pixels on 2048×1556 film scans (a typical post production resolution). Secondly, there is more legitimacy for a conservative treatment process since, as stated earlier, the film editors always prefer leaving defects rather than damaging film even to the slightest level. Finally, it is common in post-production to use *render farms* for improving computation speed. Thus the same software is applied automatically to different frames on different CPUs. Unfortunately this implies that a history of information about discontinuities may not always be available and there is no guarantee that frames will be dealt with in sequential order. The inter CPU communication necessary to be able to facilitate this increases the complexity of the software substantially.

The next section presents the essence of an idea to overcome these issues. Pictures then illustrate the usefulness of the process with difficult film material.

2. PATHOLOGICAL MOTION AND BLOTCHES

As it is not possible to remove large blotches using spatial information alone, it is necessary to detect blotches by exploiting the assumption that they do not occur in consecutive frames at the same position [1]. Therefore they are sites of temporal discontinuity. Thus the main difficulty in successfully detecting blotches is in accurately estimating the motion between subsequent frames. Hence, the consequence of PM is the incorrect detection of defects where there are none (false positives) and the failure to detect dirt where it exists (false negatives). Figures 2 and 3 show blotch detection performance in the presence of PM using an MRF based blotch detector presented in [4].

The idea presented in [4] is to configure discontinuity fields between frames where a discontinuity variable $O_{n,n-1}(\mathbf{x})$ at each pixel is set to 1 if the pixel $I_n(\mathbf{x})$ is occluded in frame $n-1$, and is zero otherwise. The same situation exists for the forward direction $O_{n,n+1}$. A blotch is indicated if *both* $O_{n,n-1}$ and $O_{n,n+1}$ are set to 1. It is educational to re-express this model in a different manner. Pixels are more appropriately configured as occupying four states, $s(\mathbf{x}) = \{0, 1, 2, 3\}$. In state 0, the pixel is trackable, while in states $1, 2$ the pixel is occluded in the previous or next frames respectively and in state 3 the pixel is occluded in both directions. Hence $s = 3$ indicates a blotch at that site. Collecting the previous, current and next frames into image vector $\mathbf{I}$, motion information into $\mathbf{D}$, the neighbourhood configuration of s into S, and dropping the position $\mathbf{x}$ for clarity, manipulating $p(s|\mathbf{I}, \mathbf{D}, S)$ at each site is the key to the estimation problem. Proceeding in a Bayesian fashion

$$p(s|\mathbf{I}, \mathbf{D}, S) = p(\mathbf{I}|s, \mathbf{D})p(s|S) \tag{2}$$

$p(s|S)$ should express a smoothness constraint on the discontinuity state s, and encourages like states to collect together. A typical Gibbs energy prior is employed here as follows.

$$p(s|S) \propto \exp -\left(\Lambda \sum_{v} \lambda|s \neq s(v)|\right) \tag{3}$$

where v indexes the eight nearest pixel sites, and $\Lambda = 2.0$ here, while $\lambda = 1$ for vertical and horizontal neighbours and $1\sqrt{2}$ otherwise.

$p(\mathbf{I}|s)$ is the likelihood and expresses the lack of correlation in image data along a motion trajectory in the presence of discontinuity. Assuming a translational model of motion such that $\mathbf{d}_{n,n-1}$ is the two component displacement mapping a pixel at $\mathbf{x}$ into position $\mathbf{x} + \mathbf{d}_{n,n-1}$ in frame $n-1$, the likelihood is expressed as follows.

$$p(I_n(\mathbf{x})|s) \propto \begin{cases} \exp -\left(\frac{\Delta_b^2+\Delta_f^2}{2\sigma_e^2}\right) & s=0 \\ \exp -\left(\alpha + \frac{\Delta_f^2}{2\sigma_e^2}\right) & s=1 \\ \exp -\left(\alpha + \frac{\Delta_b^2}{2\sigma_e^2}\right) & s=2 \\ \exp -\left(2\alpha\right) & s=3 \end{cases}$$

where $\Delta_{f,b}$ are forward and backward Displaced Frame Differences (DFDs) as follows,

$$\begin{aligned} \Delta_b &= I_n(\mathbf{x}) - I_{n-1}(\mathbf{x} + \mathbf{d}_{n,n-1}) \\ \Delta_f &= I_n(\mathbf{x}) - I_{n+1}(\mathbf{x} + \mathbf{d}_{n,n+1}) \end{aligned} \tag{4}$$

α is a penalty preventing the degenerate maximisation of the likelihood by setting $s = 3$ everywhere. $\alpha = 2.76^2$ in our experiments as this corresponds to a 99% level of significance given the Gaussian assumption placed on the distribution for the DFDs.

2.1. A practical algorithm

The ICM (Iterated Conditional Modes) algorithm is employed to solve for the states s at each pixel site. At each site the probability $p(s|\cdot)$ is maximised directly by substituting $s = 0, 1, 2, 3$ in turn into the expressions above. Using the multiresolution MRF approach of Heitz et al [5] (to improve convergence and speed), the initial configuration of the field s at the lowest resolution is set to state 0. Iterations proceed at each level of the pyramid until no further change in states is observed. This typically occurs after up to around 20 iterations at the lowest resolution (quarter res) down to only around 3 iterations at full scale.

2.2. Diagnosing PM

As observed by Bornard et al [2] a good indicator of PM is the presence of discontinuities in similar locations in successive frames. In the new expression of blotch detection above, this implies that observations of states values $s = 1, 2$ at similar pixel locations indicate PM. The image can be divided into blocks and the discontinuitiy activity for each block is measured. If the activity in a block is above a threshold over a specified temporal window, that block is said to indicate PM. Dirt detected in these blocks can then be ignored before the repair stage, or the detection can be done again treating these areas differently.

Unfortunately, there are a few problems with this approach. Firstly, it requires the pathological region to be relatively slow moving (the speed determined by the size of the blocks used), as detection requires the resultant high activity to be in the same block in a number of subsequent frames. Therefore, a fast moving pathological object (for example a bird flying across a screen) may not be flagged as pathological. Secondly, if the dirt in the regions flagged as pathological is ignored, any dirt correctly detected in these areas will also be ignored. Alternatively, should you wish to detect the dirt treating the pathological areas differently, it may become necessary to process the sequence twice (a costly solution when a render farm is being employed as data cannot usually be stored from one pass to the next). Finally, pathological areas will not necessarily display the required amount of discontinuity activity over consecutive frames to be flagged as such.

3. EXPLOITING GLOBAL MOTION FOR SIMPLE SEGMENTATION

A different, more compact solution to the detection of PM can be proposed for the post-production application. Assume that pathological regions are only likely to occur in foreground objects. This is valid as most of the time the motion for the background (due, for example, to a camera move) can be easily determined. Therefore, segmenting the sequence into foreground and background would indicate in which areas to be more cautious when detecting dirt. The segmentation, detection and repair could then be accomplished in a single pass of the sequence.

Foreground/Background segmentation is achieved by exploiting the fact that global motion can generally be readily estimated from most image material e.g. [6]. Regions of the image which do not fit the global motion model are segmented as foreground. Affine global motion is modelled here as follows (for the backward direction)

$$I_n(\mathbf{x}) = I_{n-1}(\mathbf{A}\mathbf{x} + \mathbf{d}_b^g) \tag{5}$$

where $\mathbf{A}$ and $\mathbf{d}^g$ combine to give the global affine transformation.

Segmentation is accomplished by configuring a label field l that is coupled with a discontinuity information as follows

$$l = \begin{cases} 0 & \text{Background pixel, exists in past and next frames} \\ 1 & \text{Background pixel, covered in next frame} \\ 2 & \text{Background pixel, covered in previous frame} \\ 3 & \text{Foreground pixel} \end{cases}$$

The label $l = 3$ indicates foreground, while the other labels labels handle background information. Label configuration $l = 3$ corresponds to outliers in the fit to global motion. These outliers are a collection of two different regions: those undergoing *local* motion and those degraded by dirt. The idea then is to adaptively increase α and Λ in the blotch detection framework in equation 2 above when in regions of foreground detected by $l = 3$.

A practical solution: As can be seen the use of the labels l and s are very similar, with the exception that l relies on global motion information while s relies on total motion information. The ethos of the algorithm used to configure l to achieve a segmentation is therefore the same as that for the blotch detection process outlined above. In this case however, the error signal is derived from the current frame and the globally motion compensated previous and next frames.

$$\Delta_b = I_n(\mathbf{x}) - I_{n-1}(\mathbf{A}_b\mathbf{x} + \mathbf{d}_b^g)$$
$$\Delta_f = I_n(\mathbf{x}) - I_{n+1}(\mathbf{A}_f\mathbf{x} + \mathbf{d}_f^g)$$

Again we assign each pixel a state, but in this case state $l = 3$ is **either** dirt or foreground. The steps are therefore as follows

1. Estimate global affine motion between the current, previous and next frames. Here we use the method outlined in [6].
2. Use this to generate the global motion compensated DFDs (above)
3. Configure l using the multiresolution ICM scheme as for blotches.

The safety of the foreground estimate can be further assured by repeating the detection process using globally motion compensated frames from farther temporal reaches i.e. $n - 2, n, n + 2$. The final foreground region can then be considered the union of all individually detected regions.

It is now necessary to separate those pixels which are denoted as state 3 into dirt and foreground. The label field $l = 3$ is first filtered with a 64×64 box filter to give an *activity* level per pixel. The resulting field is dilated slightly by filtering with an 11 tap gaussian ($\sigma^2 = 3$), and set to 1 when values exceed a threshold of about 1% of the pixels in the block.

This mask indicates the conservative region in which both foreground and dirt could be occuring. To extract the regions in which dirt could be present two strategies are used in combination. The field s is configured using adaptively altered α and Λ that are increased in the regions of foreground. Then the resulting image regions in which labels $s = 3$ indicate the presence of a blotch are post-processed to reject those regions with low spatial contrast. This double check is only conducted within the $l = 3$ i.e. foreground (local motion) regions because it is expected that only high contrast blotches are visible there. In effect the bloth detection is made more conservative in the local motion areas.

Computational Issues: Experiments with film resolution images show that it is not necessary to configure s or l at the highest scale. Thus substantial computational savings are had by operating on subsampled images (factor of 2). Furthermore, the use of a single pass process resolves problems with render farm applications. On a dual processor 2.8GHz Xeon this process takes roughly 15 secs per 2048×1556 film frame (pathological motion detection/ blotch detection/repair).

4. PICTURES

Figure 2 shows results from processing the running man using the method presented here. The local motion mask $l = 3$ correctly

Fig. 2. Top row: Zoom on Original frame and detection using MRF model without PM detection. Middle row: Masks of PM (left [2]) and local motion (using the new technique presented) Bottom: Left degradation caused by dirt repair without regard to PM detection, right: dirt repair after PM detection .

identified the foreground object. Subsequent adaptive processing removed false alarms (shown in the top right image), and allows artefact free reconstruction as shown in the bottom right image. The damage to picture material caused by ignoring PM is shown in the bottom left image where the hand and parts of the arm are erroneously removed.

Figure 3 displays similar results for a sequence showing a woman on roller skates. Here the local motion mask $l = 3$ identifies both the moving foreground object and the pathological area where the shadow crosses the man on the bench.

5. FINAL COMMENTS

This paper has presented a new method for dealing with PM in degraded image sequences. Sequences are intially segmented into foreground and background motion, and blotch detection is applied more cautiously in foreground areas (in which we assume motion estimation is likely to be less reliable). The main novel contribution is the ability to determine pathological regions before the blotch detection process (previous work has used the output of the blotch detector over a temporal window). This yields a more efficient and compact solution, allowing PM detection, blotch detection and repair to be combined into a single stage. The resulting process is more amenable to render farm processing which is key for post-production applications. Examples of the sequences shown above can be seen at www.mee.tcd.ie/∼sigmedia/benicip/.

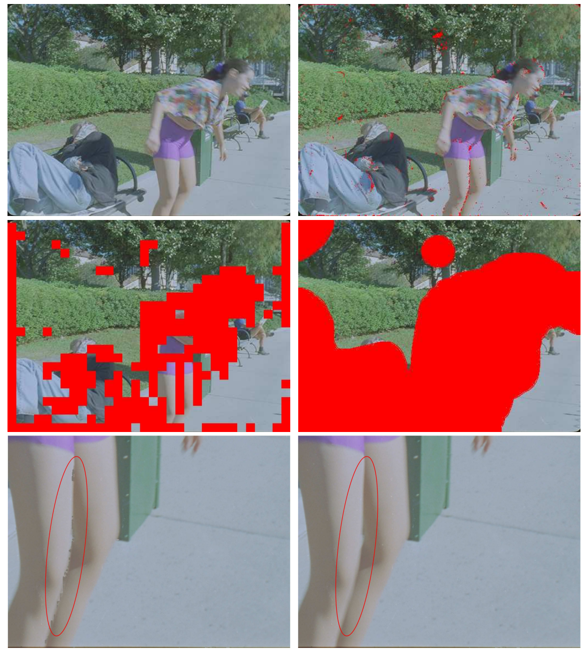

Fig. 3. Top row: Original frame and detection using MRF model without PM detection. Middle row: Masks of PM (left [2]) and local motion (using the new technique presented) Bottom: Left degradation caused by dirt repair without regard to PM detection (especially on the edge of the leg), right: dirt repair after PM detection .

6. REFERENCES

[1] A. C. Kokaram, *Motion Picture Restoration: Digital Algorithms for Artefact Suppression in Degraded Motion Picture Film and Video*, Springer Verlag, ISBN 3-540-76040-7, 1998.

[2] Raphael Bornard, Emmanuelle Lecan, Louis Laborelli, and Jean-Hugues Chenot, “Missing data correction in still images and image sequences,” in *ACM Multimedia*, December 2002.

[3] J. Biemond A. Rares, M. J.T. Reinders, “Complex event classification in degraded image sequences,” in *Proceedings of ICIP 2001 (IEEE), ISBN 0-7803-6727-8*, Thessaloniki, Greece, October 2001.

[4] A. Kokaram, R. Morris, W. Fitzgerald, and P. Rayner, “Detection of missing data in image sequences.,” *IEEE Image Processing*, pp. 1496–1508, November 1995.

[5] F. Heitz, P. Pérez, and P. Bouthemy, “Multiscale minimization of global energy functions in some visual recovery problems,” *CVGIP : Image Understanding*, vol. 59, no. 1, pp. 125–134, January 1994.

[6] F. Dufaux and J. Konrad, “Efficient, robust and fast global motion estimation for video coding,” *IEEE Transactions on Image Processing*, vol. 9, pp. 497–501, 2000.