

Application of rigid motion geometry to historical film restoration.

S. Boukir and D. Suter
Dept. Elect. and Comp. Syst. Eng.
Monash University,
P.O. Box 35, Clayton 3800
Australia

Abstract

Historical film restoration involves locating the position of artifacts and replacing the "missing" portion of the film (obscured by the artifact) with pixels that had been lost. Computer vision research has recently developed many techniques for constraining and predicting parts of a scene based upon the assumption of rigid motion. In this paper, we show how the constraints can help identify artifacts as well as how the prediction can be used to replace the artifact with natural looking portions of the scene. These techniques can be superior, when the rigid motion assumption is valid, to other techniques for historical film restoration.

1 Introduction

The motion picture industry is now over 100 years old. Throughout the world, there are large stocks of motion pictures in a variety of formats: for example, acetate, nitrate or celluloid based film, and magnetic video tape in various formats. Time, usage, and poor storage have taken their toll on much of this material which may be of high cultural, historical, or commercial value.

Because of the ease of reproduction, and of the relative stability of some digital formats, there are projects around the world to digitize the film stock of higher value. There is also much interest in restoring the degraded stock.

Physical methods of restoration have a relatively long history (see, for example, [10]). However, physical means of restoration are labour intensive, costly, and can potentially harm the original stock (at least in the process of making copies for the purposes of restoration).

Digital restoration of historical film has a relatively short history (for example, there is to date, only one monograph on the subject [8]). This is partly because the huge storage and computational cost, of digitization and restoration has meant that large scale digital restoration has not been viable. There are signs that reduced cost/increasing capacity of digital systems may now mean that large scale digital restoration is viable, or will shortly be so.

Restoration can address many defects in the degraded film. For example, colour or intensity degradations, frame jitter (mis-registration), or various "pixel" defects (portions of a frame where the image has been corrupted by dust,

scratches, mould, lifting of the recording medium etc.). It is the latter class of defects we address here: there is a spatial nature in the defect.

The defects can be repaired if we can effectively predict the "true" values of the corrupted pixels. In certain situations this prediction can take a very simple form: if, for example, the corrupted portion of the scene is stationary, we could look to the same region in the previous or subsequent frames to perform a very simple prediction. In portions where the scene may be moving, a simple "block search" can be surprisingly effective [?]. In this approach, one simply searches in the previous frame(s) and in the subsequent frame(s) for blocks that match a possibly corrupted block. If good matches can be found, the block is probably not corrupted. If not, it may possibly be replaced by predictions based upon surrounding block matches (temporally and spatially surrounding).

The block matching approach is just one of several approaches that assume a (locally) very simple motion model. However, pure translation or affine motion models cannot accurately model all of the motion in a realistic scene. For example, the motion on the image plane produced by a rigidly moving object, cannot always be well approximated by such simple motion models. Such motion can be constrained accurately (up to the accuracy of some underlying assumptions concerning the camera model) by a series of geometric constraints: the epipolar constraint between two images or the trifocal constraint amongst three, for example. Ignoring lighting affects and object occlusions, it is actually possible to predict the appearance of a rigidly moving object. This is the essence of the approach investigated in this paper.

We should emphasize that we are not promoting the approaches investigated herein as the "complete and final solution" to the problem of pixel prediction for restoration. As pointed out in the previous paragraph, though the methods presented here can be more precise than methods based upon simple translation or affine motion, they have their limitations that need to be overcome in future work (as discussed in the Conclusion).

The fundamental matrix F is the algebraic representation of the epipolar geometry between two images [6]. This matrix is characterized only in terms of corresponding image points without reference to the camera matrices. This enables F to be computed from image correspondence alone. The trifocal tensor which links point correspondences across three images generalizes the formulation of the fundamental matrix [2][12] [6]. In this paper we restrict our attention to exploiting the fundamental matrix. The fundamental matrix satisfies *the correspondence condition* or *the epipolar constraint*. This condition states that for any corresponding pair of points (m, m') in the two images:

$$m'^t F m = 0 \quad (1)$$

Thus, the point m' lies on the epipolar line $l' = F m$ corresponding to the point m . Similarly, $l = F^t m'$ represents the epipolar line corresponding to the point m' .

The epipole is the image in one view of the camera center of the other view. For any point x in the first image, the epipolar line $l' = F x$ contains the epipole e' which satisfies [6]:

$$F^t e' = 0 \quad (2)$$

Similarly, $F e = 0$. Thus, the pencil of epipolar lines in the first image (respectively the second image) intersect at the epipole e (respectively e').

3 Image prediction algorithm

The main steps for predicting a new image are:

- Search for matching points [13]
- Robust recovery of the fundamental matrix [5][6]
- Reprojection [1]
- Suppression of Moire effects

3.1 Matching process

The method we use is the KLT (Kanade-Lucas-Tomasi) feature tracker [13]. Reliable features are located by examining the minimum eigenvalue of each 2×2 gradient matrix

$$\begin{pmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{pmatrix} \quad (3)$$

where I_x, I_y are the first derivatives in the x and y direction, respectively. The gradients are summed within a 7×7 window centered around each pixel of the input image. This algorithm also ensures that all the found features are far enough apart. Features are then tracked using a Newton-Raphson method of minimizing the difference between the two windows.

3.2 Robust recovery of the fundamental matrix 2

The previous stage produces *good points* which are then normalized to ensure good numerical solution [5][6]. This normalization consists of applying a similarity transformation so that the transformed points are centered at the origin and the mean distance from the origin is $\sqrt{2}$. The normalized *good points* are used for computing the fundamental matrix in a robust statistical manner [14] to avoid the effects of outliers. The RANSAC robust estimation is done by a repetitive random choice of a subset of seven *good points*, recovering the fundamental matrix, and verifying the quality of the fundamental matrix on the remaining set of *good points*. The number of samples N is set to 588 according to (4). This ensures with a probability $p = 0.99$ that at least one of the random samples of $s = 7$ points is free from outliers for the worse case of $\epsilon = 50\%$ of outliers in the input data [6]:

$$N = \log(1 - p) / \log(1 - (1 - \epsilon)^s) \quad (4)$$

The quality of the fundamental matrix is measured by the distance of the points in the second image from their epipolar line. Only points with error of less than a specified threshold (typically, one pixel) are considered. The best fundamental matrix is the one with the largest number of supporting points, and it is computed again in a usual least-squared manner, using all the points that supported it.

3.3 Reprojection mechanism

The reprojection process is performed every time we wish to generate a novel view. In [3], image points in an image are predicted from image points in two other images and the epipolar geometry between three views, i.e. the knowledge of two fundamental matrices. More precisely, the position of image points are retrieved using the intersection of two epipolar lines. More recently, [1] used the trifocal tensor to generate a novel view from two other views. The limitations of these techniques are mainly concerned with the correspondence problem. Indeed a full correspondence between the two views is necessary to reconstruct the third one. In the following we derive a method without this shortcoming.

Given a pixel $m = (u \ v \ 1)^t$ (projective coordinates) in an image I , the question is *where should its corresponding position $m' = (u' \ v' \ 1)^t$ be in a second image I'* ? The answer is quite simple and only exploits the epipolar geometry between the two views which is entirely encapsulated in the fundamental matrix. So, if at least 7 reliable point-correspondences are available to recover the fundamental matrix, the problem can be solved by the resolution of a linear system.

For notation convenience, let us consider the dual problem of predicting the point m given its correspondent m' .

Draft of paper to appear in ICPR2003
 First, m' must belong to the epipolar line $l' = Fm$:

$$(l') : Ax + By + C = 0 \quad (5)$$

with

$$\begin{cases} A = f_{11}u + f_{12}v + f_{13} \\ B = f_{21}u + f_{22}v + f_{23} \\ C = f_{31}u + f_{32}v + f_{33} \end{cases} \quad (6)$$

where $f_{ij}, i = \overline{1,3}, j = \overline{1,3}$, represent the coefficients of the 3×3 fundamental matrix F .

Second, the epipolar line l' is obtained as the line joining m' to the epipole $e' = (u'_e \ v'_e \ 1)^t$:

$$(e'm') : (v'_e - v')x - (u'_e - u')y + u'_e v' - v'_e u' = 0 \quad (7)$$

The coordinates of the epipole e' are computed from (2) which we solve in a least-square manner. Given any image point m' in image I' , we can easily extract the epipolar line parameters as follows. The line $e'm'$ is collinear to the line Fm , both representing the epipolar line l' . Therefore, from equations (6) and (7) we can derive the following constraint:

$$\frac{A}{v'_e - v'} = \frac{B}{u'_e - u'} = \frac{C}{u'_e v' - v'_e u'} = \lambda, \lambda \in \mathcal{R} \quad (8)$$

Substituting (8) in (6) leads to the following linear system of three unknowns u, v, λ :

$$\begin{cases} f_{11}u + f_{12}v + f_{13} = \lambda(v'_e - v') \\ f_{21}u + f_{22}v + f_{23} = -\lambda(u'_e - u') \\ f_{31}u + f_{32}v + f_{33} = \lambda(u'_e v' - v'_e u') \end{cases} \quad (9)$$

In general, λ is a function of the 3D point position (or, alternatively, of (u, v, u', v')). If λ is fixed, then it is easy to see, from the first two equations, that the image plane to image plane mapping is affine. However, it is not "any old affine mapping" but one consistent with a particular subset of 3D motions viewed with perspective cameras. Space does not permit, but a rather straightforward derivation can be carried out, starting from the expression for the Fundamental matrix in terms of an Homography (e.g., [6] p. 223), to show that one such set of circumstances is where the images points are approximately coplanar and the homography associated with that plane has zeros as the first two entries of the last row.

Although the assumption of λ constant is not general, in the preliminary study we made that assumption and found it to still be applicable to enough sequences to make the derived technique useful. Future work will extend the framework towards handling more general situations.

To estimate the parameter λ , we rely on the *good points* $m_i = (u_i \ v_i \ 1)^t, i = \overline{1, K} (K \geq 7)$, previously used for the estimation of the fundamental matrix, e.g.,:

$$\lambda_A = \text{Median} \left(\frac{A_i}{v'_e - v'_i}, i = \overline{1, K} \right) \quad (10)$$

where $A_i, i = \overline{1, K}$, are computed from (6) given (u_i, v_i) .

This simple robust scheme leads to stable results ($\lambda_A \approx \lambda_B \approx \lambda_C$) as confirmed by our experiments. A substitution of $\lambda_A, \lambda_B, \lambda_C$ in respectively the 1st, 2nd and 3rd equation of system (9) eliminates the unknown λ . A least-square technique permits then to estimate the remaining unknowns u and v . These predicted fractional coordinates are mapped into the closest integer coordinate neighbor.

Notice that the computation of λ is performed once only from the set of *good points*. This is very fast to compute.

The simplest scheme for gray-level interpolation is based on a nearest neighbor approach [4]. Thus, in our case, one could simply assign the gray-level $I(u', v')$ of point m' in image I' to the nearest integer coordinate neighbor (u, v) of its computed prediction (\hat{u}, \hat{v}) in image I . Although this interpolation is simple to implement, this method often has the drawback of producing undesirable artifacts, such as distortion of straight edges in images of fine resolution. Smoother results can be obtained by using more sophisticated techniques [4]. Therefore, to obtain a smooth estimate of the gray-level at any predicted point, we rely on a bilinear approach that uses the gray-level of the four nearest neighbors.

3.4 Suppression of Moire effects

The generated image may have some black points because the reconstruction process cannot always fill each point (discretization errors). These well known Moire effects are easily removed on the given image using intensity information of neighbor points. More exactly, a 3×3 median filter is locally used at each black point.

4 Application to film restoration

Our image prediction algorithm is the core of our film restoration method which aims at removing not only small impulsive artifacts (dust spots, hair,...) but large impulsive defects (blotches,...) as well. We have previously developed spatio-temporal restoration algorithms using block-based motion compensation. However, these techniques work well only in case of translational motion and fail in presence of significant rotational component.

Our restoration approach involves accurate detection of the artifacts followed by the reconstruction of damaged areas. This detection/correction scheme has been adopted by several works related to digital restoration [15][9][8].

Of course, the applications of the algorithms presented here are not limited to motion pictures. Similar treatment may be used for any degraded image sequence.

4.1 Detection of impulsive defects

Dust spots and blotches are the most frequent artifacts damaging old films. They typically can be found on one

frame only and will therefore present a strong temporal discontinuity in image brightness. Our defect detector relies on both image prediction, and grey-level conservation. Recall that our image prediction algorithm only needs 7 reliable point-correspondences between the image to restore and any frame of the sequence and not a full correspondence. Thus, we avoid the difficult (especially in case of noisy image sequences) and time-consuming motion compensation step followed by other work in this field [11][8].

The localization of impulsive defects becomes straightforward after image prediction and a simple inter-frame difference may be used for the detection process. However, this scheme is valid only if all the different moving regions are isolated in each frame. This difficult segmentation problem has not been investigated yet. Of course, the former scheme can be used in the simplest situation of one relative motion only (e.g., fig. 5). To handle also the more complex case of one moving object against a static background (e.g., fig. 1), we use a slightly different detector, \mathcal{D} , based on two difference images D_1 and D_2 rather than one only:

$$\mathcal{D}(x, y) = \text{MIN}(D_1(x, y), D_2(x, y)), \forall (x, y) \in \text{image } I \quad (11)$$

where D_1 represents the difference between the current image I and its prediction from another frame of the sequence; and D_2 the difference between I and a successive (or preceding) frame of the sequence. The frame used in D_2 should be different from the one used in D_1 to prevent the risk of detecting also the defects of this frame.

The difference D_1 will highlight the defects of the dominant moving object but also false detections due to the global transformation of the whole image. Thus, background details are likely to be displaced on the predicted image, and will therefore be seen as defects. These false detections are simply removed using the difference D_2 , in which actual defects will also be present while the effects of global transformation (rather than local ones in case of a correctly segmented scene) will disappear.

Finally, we use the double evidence provided by AND-ing the results of two detections. Common spurious points from the two independent detection processes are selected as deteriorations.

4.2 Removal of impulsive defects

This stage consists in post-processing the detection field. The reconstruction of the damaged areas is usually achieved using an interpolation technique [9][7][8]. Our image prediction algorithm provides intact spatio-temporal information to fill the damaged detected areas of a given image. This allows a straightforward restoration avoiding a costly interpolation and, more importantly, does not need any further step for the incorporation of noise (high

frequencies of the image). Indeed, in the context of film restoration, the retrieval of the original noise (film grain) is the most difficult issue which usually causes the failure of a restoration technique. Our restoration algorithm consists of the following steps:

- Dilation (typically of one pixel) of the detection masks. Indeed, the actual border of the defects is usually hardly distinguishable from its surrounding intact area. The resulting oversized masks significantly improve the restoration quality especially in case of large alterations.
- Estimation of the global brightness variation between the current frame and its prediction. Indeed, old films are usually subject to global (more rarely local) intensity instabilities. This luminance variation is computed as the median of the intensity variations of a subset of randomly chosen pixels.
- Assignment of each corrupted pixel with the *adjusted* brightness of its prediction. The brightness modification consists merely in adding to each pixel a fixed amount computed in the previous step.

5 Experiments

Our restoration algorithm has been implemented on old motion pictures. Two examples are shown here, one with a predominant rotation and another with a translation. Figure 1 shows an image pair of the *Mask* sequence, the left image being damaged with a dust spot. The images are 100×100 each and represent a sample of the actual video sequence (720×576). Indeed, the whole sequence exhibits different motions and we do not handle yet the segmentation issue. The moving object is a head, wearing a mask, rotating around itself, thus satisfying the rigidity constraint.

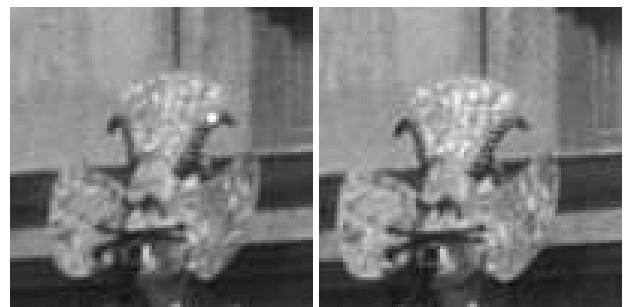


Figure 1: Frames 230 (left) and 240 (right) of the *Mask* sequence

The damaged frame is first predicted using any other image of the sequence (right frame of figure 1). The result of

the prediction is shown in fig. 2. We can notice that the predicted image fits well the original image despite the significant change in shape between the two considered original frames of the sequence (see fig. 1).

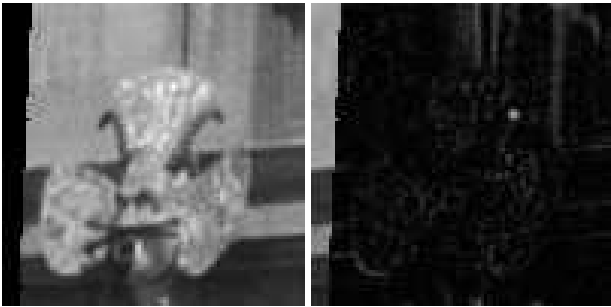


Figure 2: Prediction of frame 230 using frame 240 of the *Mask* sequence (left). Difference between the original frame and its prediction (right)

Another image of the sequence is used to predict the damaged frame of the *Mask* sequence. Both predictions are used as inputs in our detection algorithm leading to the results depicted on the left part (respectively the right part) of fig. 3. The double evidence of these two independent detections provides reliable results, almost free from false detections, as can be seen on the left part of fig. 4.

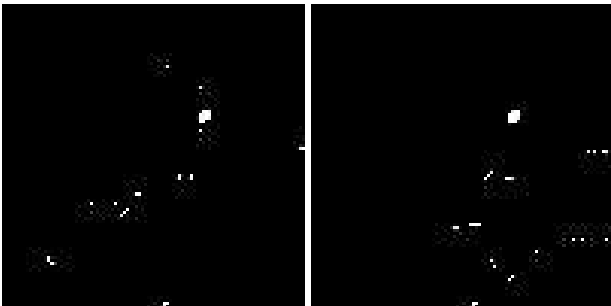


Figure 3: Defect detection based on frame 240 (left) , respectively on frame 228 (right), of the *Mask* sequence

Finally, the detection masks are used to restore the damaged areas. Figure 4 exhibits the restored damaged frame using its prediction from frame 240. Associated movie clips (paper11a, paper11b, and paper11c) show the original, predicted and restored frames of the entire sequence.

Successful restoration results have also been achieved on the *Building* (256×256) sequence whose motion is almost a pure translation (see the original frames on fig. 5 and the illustration of our algorithm steps on figures 6 and 7). Associated movie clips (paper11d, paper11e, and paper11f) show the original, predicted and restored frames of the entire sequence.

Notice that our restoration approach needs only a few parameters to be set. Furthermore, they do not require a

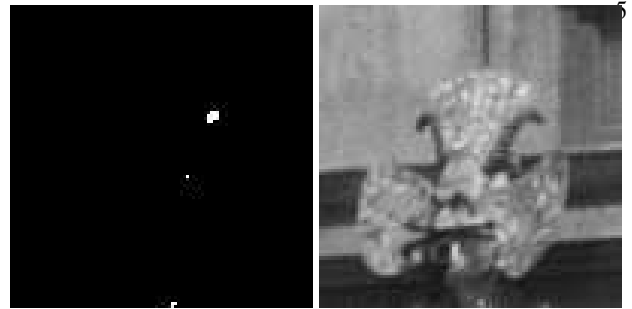


Figure 4: Detection (left) and removal (right) of impulsive defects on frame 230 of the *Mask* sequence



Figure 5: Frames 30 (left) and 26 (right) of the *Building* sequence



Figure 6: Prediction of frame 30 using frame 26 of the *Building* sequence (left). Difference between the original frame and its prediction (right)



Figure 7: Detection (left) and removal (right) of impulsive defects on frame 30 of the *Building* sequence

Draft of paper to appear in ICPR2002
fine tuning. We used the same values for all the sequences we have experimented with: the number of features n to be tracked has been fixed to 30 (the actual number of found matching pairs is $15 \leq n \leq 30$), the number N of samples used in the RANSAC estimation of the fundamental matrix has been set to 588 according to (4) and the detection threshold $S = 25$.

6 Conclusion

We have demonstrated that one can detect and then correct certain defects in film and video material via methods that exploit the geometry of rigid motion. Where the assumption of a particular subclass of rigid motion is valid, the method ought to be more accurate than methods that use simple heuristics such translation or arbitrary affine motion models to achieve the same ends. Where the motion is essentially pure translation, we have demonstrated that the method introduced here is also effective.

There are several avenues for future work. We are extending the techniques to work with a more general class of rigid motions. Moreover, although there are situations where the entire image moves with a single motion consistent with 3D rigid motion (e.g., static world and moving camera), to be effective the complete system would have to be constructed so as to handle multiple rigid motion within the field of view. Hence, one naturally turns to methods of motion segmentation.

Likewise, there are other situations where the implicit assumptions behind our methods will fail to be met. The method does not handle occlusion - including self-occlusion. Methods that attempt a reconstruction of depth have this potential (at greater cost and risk of error). Occlusion can also be handled by certain "one sided" (temporal) continuity tests. That is, an occluded region is seen until the point of occlusion in a continuous fashion. We also do not model lighting effects - ours is a purely geometric approach.

Finally, although one can handle articulated motion (with each part rigidly moving), one cannot handle, with our approach, arbitrary yet frequently occurring natural motions.

Nonetheless, where one can reasonably detect the presence of rigid motion (in an automated or semi-automated fashion), the methods we have implemented, and the related methods that are sure to follow, should add to the set of useful techniques one can employ to restore film and video material.

Acknowledgements

This work was conducted whilst the first author was on leave from La Rochelle University (France). The authors

gratfully acknowledge the support of the Australian Research Council (IREX award) that made this collaboration possible.

References

- [1] S. Avidan and A. Shashua. Novel view synthesis by cascading trilinear tensors. *IEEE Transactions on Visualisation and Computer Graphics*, 4(4):293–306, October-December 1998.
- [2] O. D. Faugeras and B. Mourrain. On the geometry and algebra of the points and line correspondences between n images. *Proc. of ICCV'95, Int. Conf. on Computer Vision*, pages 951–962, 1995.
- [3] O. D. Faugeras and L. Robert. What can two images tell us about a third one? *Proc. of ECCV'94, European Conf. on Computer Vision*, 1:485–492, May 1994.
- [4] R. C. Gonzalez and R. E. Woods. *Digital image processing*. Addison-Wesley Publishing Company, 1993.
- [5] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580–593, October 1997.
- [6] R. I. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2000.
- [7] S. Kalra, M. N. Chong, and D. Krishnan. A new Auto-Regressive (AR) model-based algorithm for motion picture restoration. *ICASSP'97, International Conference on Acoustics, Speech and Signal Processing*, 4, 1997.
- [8] A. Kokaram. *Motion picture restoration. Digital algorithms for artefact suppression in degraded motion picture film and video*. Springer-Verlag, 1998.
- [9] R. D. Morris. *Image sequence restoration using Gibbs distributions*. PhD thesis, University of Cambridge, May 1995.
- [10] P. Read and M-P Meyer. *Restoration of Motion Picture Film*. Butterworth-Heinemann, London, 2001.
- [11] L. Rosenthaler, A. Wittmann, A. Günzl, and R. Gschwind. Restoration of old movie films by digital image processing. *IMAGE'COM 96, Bordeaux, France*, pages 1–6, May 1996.
- [12] A. Shashua. Trilinear tensor: the fundamental construct of multiple-view geometry and its applications. *International Workshop on Algebraic Frames For The Perception Action Cycle (AFPAC), Kiel Germany*, September 1997.

- ~~Draft of paper to appear in ICPR2002~~
[13] J. Shi and C. Tomasi. Good features to track. *CVPR'94, IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 593–600, 1994.
- [14] P. H. S. Torr and D. W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *Int Journal of Computer Vision*, 24(3):271–300, 1997.
- [15] R. Veldhuis. *Restoration of lost samples in digital signals*. Prentice Hall, 1990.