COMBINED AFFINE AND TRANSLATIONAL MOTION COMPENSATION SCHEME USING TRIANGULAR TESSELLATIONS

D.B.Bradshaw, N.G.Kingsbury

Signal Processing and Communications Group,
Department of Engineering,University of Cambridge.
email:dbb21@eng.cam.ac.uk, ngk@eng.cam.ac.uk

ABSTRACT

The technique outlined in this paper extends the ability of current warping motion estimation schemes to allow occlusion and uncovering effects to be modelled. Current methods, [1], [2], [4], use a *continuous* rubber sheet approach, consisting of non-overlapping polygons. The new technique estimates and compensates affine and then translational motion within the scene. The latter is achieved by allowing polygons to overlap through the introduction of rips into the sheet which are located in areas where occlusion or uncovering is thought to occur. Results show a reduction in the prediction error when compared to both traditional block-based methods and recently developed warping schemes (without rips).

1. INTRODUCTION

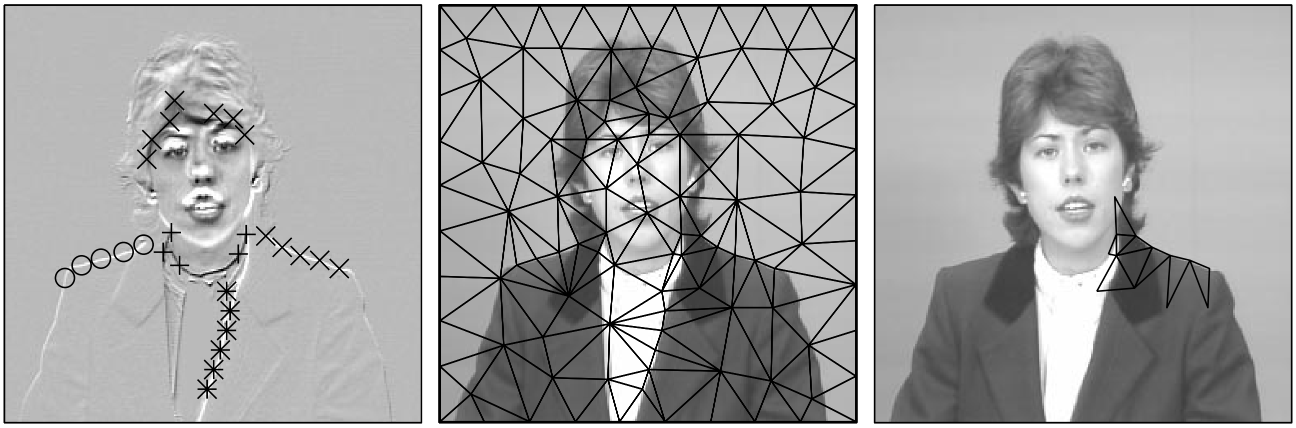
Traditional block-based approaches to motion compensation rely on approximating general motion within an image sequence through the use of a purely translational motion model. To overcome this limitation, recent techniques, known variously as control grid interpolation, warping or affine motion estimation schemes, have been suggested, [1], [2], [4]. These all use a 'rubber sheet' approach whereby the image domain is defined as a set of non-overlapping polygons (generally triangles or quadrilaterals). Motion compensation is then achieved by 'deforming' this rubber sheet. Results using these techniques are promising but are inherently limited because of the 1:1 mapping nature of the warping process. This implies that phenomena such as occlusion or uncovering cannot be correctly modelled. Moreover, translational motion is not directly modelled as only one vertex is altered at any one instance. To overcome these problems a block-matching algorithm is often used prior to applying the algorithm that deforms the rubber sheet. The method outlined in this paper integrates a translational motion model directly into the rubber-sheet algorithm. This is achieved by relaxing the rubber sheet approach by allowing the sheet to rip (by letting adjacent polygons overlap) in areas where occlusion, uncovering or large translational motion is thought to occur.

The technique proposed is a three stage process. Firstly, a triangulation defining the rubber sheet is generated and areas of occlusion or uncovering are flagged. Next affine motion between adjacent frames is estimated and then compensated. Finally, translational motion within the partially compensated frame is estimated and compensated. These stages are described below.

2. TRIANGULATION AND RIP LOCATION

An adaptive method based on a Delaunay triangulation is used to generate the mesh, [3]. The size of triangles across the image domain can be varied and points at any position within the image domain can be included in the final triangulation. This allows a high degree of flexibility in defining the final characteristics of the triangulation. The procedures that achieve this are described below.

Smaller triangles are placed in areas of the image domain where large compensated frame difference errors can occur. This is achieved through the use of a background matrix which is derived as follows. The magnitude of the frame difference is lowpass filtered using a large Gaussian kernel. This result is then inverted such that areas of low 'activity' in the frame difference image correspond to large values in the final background matrix and vice-versa. This matrix is then indexed by the triangulation routine to test whether a triangle at a given location is of the correct size. The test compares the triangle's circumscribed radius with the value of the background matrix. If the radius is smaller than the indexed value the triangle



(a) (b) (c)

Figure 1: (a) Rip locations for prediction of Frame 5 from 'Claire' (overlaid on top of the frame difference)
(b) Final triangulation for prediction of Frame 5 from 'Claire'
(c) Translated triangles associated with one rip for prediction of Frame 5 from 'Claire'

is accepted, otherwise it is rejected.

Rips are located on motion boundaries by the following method. A connected set image is derived by passing the magnitude of the frame difference into an algorithm that performs a band-pass 8-neighbourhood connected set function. Next, an edge image is determined by multiplying the current frame with a simple 3×3 filter kernel that approximates a Laplacian operator. The 'rip image' is derived through a Boolean AND operation between the connected set image and the edge image. This final 'rip image' contains a number of regions, each of which corresponds to a separate rip. A number of points are derived for each of the regions. Adjacent points in each region must be no closer than a given threshold (in our simulations, a 3 pixel radius was used) thus limiting the number of generated points. The list of points generated for each region is then included in the final triangulation. The rips and final triangulation for a single frame of the 'Claire' test sequence are shown in Figures 1a and 1b.

3. ESTIMATION OF AFFINE MOTION

A motion vector for each vertex in the triangulation is determined as follows. A cavity is defined which contains all triangles sharing a given vertex. For this given vertex, all points in the corresponding cavity are determined (this defines the search space). The vertex is then displaced to each point and a minimum square error measure for the cavity is evaluated. This measure compares the unwarped pixel values of the current frame with the warped pixel values of the previous frame. This procedure is repeated until every point in the search space has been 'warped to'. Sub-pixel accuracy is achieved by iterating at smaller step sizes in an area local to the initial best match. Non-integer shifts are accommodated by using a bilinear interpolation method. The resulting affine-motion compensated frame is then passed into the translational motion compensation stage of the algorithm.

4. ESTIMATION OF TRANSLATIONAL MOTION

A triangular matching procedure is used to estimate translational motion. A triangle list is associated with each rip and each triangle in a given rip is classified as either moving or stationary by comparing the partially compensated frame data to the current frame data. The triangles labelled as moving are then translated over a given search space (+/-8 pixels) and a best match found. Sub-pixel accuracy is obtained in a similar way to that used in the affine motion stage by searching in an area local to the initial best estimate using sub-pixel step sizes; again non-integer shifts are accommodated through the use of a bilinear interpolation method. Having found the motion vector associated with a rip, adjacent triangles to those classified as moving are tested to see if they also undergo an identical translation. This procedure is repeated for each rip in the image. Figure 1c illustrates those triangles that have been translated for one rip in a single frame of the 'Claire' test sequence.

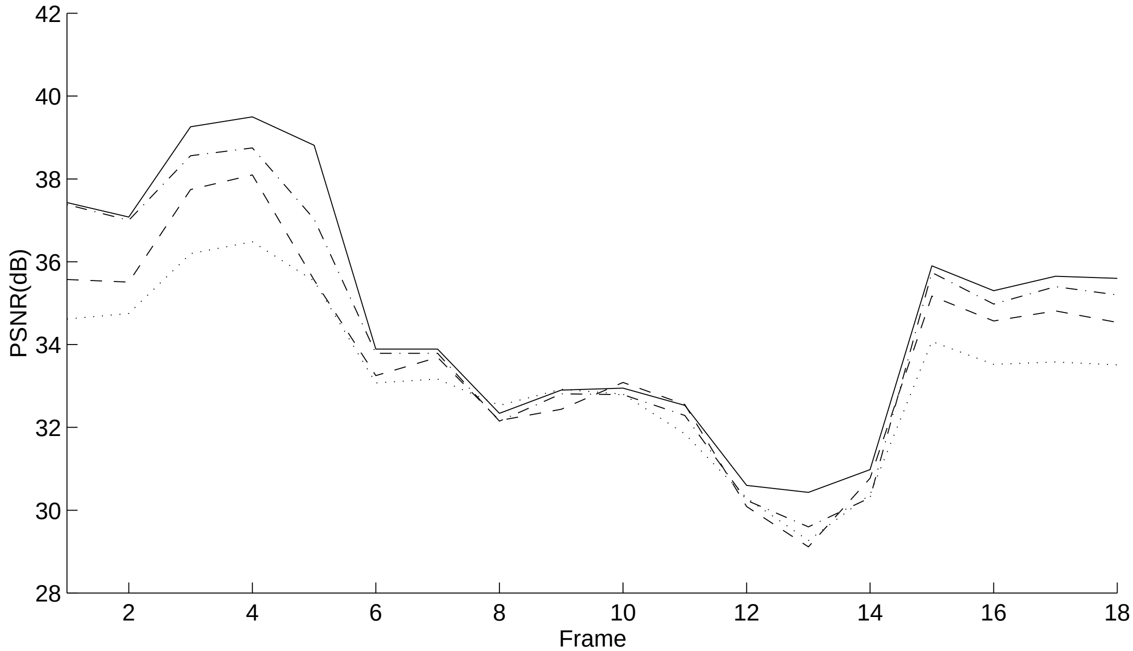


Figure 2: Displaced frame difference errors for the ‘Trevor’ test sequence
———— : Combined translational/affine method
– · – · – : Affine method
– – – – : Overlapping-block gradient-based method
· · · · · · · · · : Sub-pixel block-matching method

5. CODING CONSIDERATIONS

This scheme is primarily aimed at low bit-rate applications, to date all our simulations have used an average of 64 vertices. The frame difference required to generate the adaptive triangulation can be locally constructed at both the encoder and decoder by subtracting the last two previously reconstructed frames. If the rip locations are derived using the current frame they have to be transmitted as extra side information. However, the rips can be derived from the last locally decoded frame thus requiring no coding overhead apart from the motion vectors associated with each of the rips. The results in this paper have used rips derived from the current frame to assess the maximum performance gain available.

6. RESULTS

Figure 2 illustrates the Peak Signal-to-Noise ratio for 18 frames of the ‘Trevor’ sequence. Results using the combined affine and translational compensation schemes are shown, as are results obtained from a purely affine compensation scheme (ie without rips). The translational motion model enhances the performance of the algorithm (the mean increase being +0.41dB). Also shown are results obtained from an overlapping-block gradient-based scheme and a sub-pixel non-overlapping block-matching based scheme. The combined scheme performs as well as, or better than the other schemes throughout the 18 frames. The mean PSNR for each of the schemes is shown in Table 1.

	‘Trevor’
Combined Scheme	34.68dB
Purely Affine Scheme	34.27dB
Gradient-based Scheme	33.84dB
Block-matching Scheme	33.26dB

Table 1: Mean PSNR values obtained from the ‘Trevor’ test sequence

Note particularly Frame 13 from the ‘Trevor’ sequence. In this case, the block based nature of the other schemes results in inaccurate compensation, this is illustrated in Figure 3 where a portion from the displaced frame difference of Frame 13 of the ‘Trevor’ sequence is shown. In the mouth area, and at the top of the head and also on around the shoulders the combined affine/translational scheme compensates more accurately than the other block-based schemes.

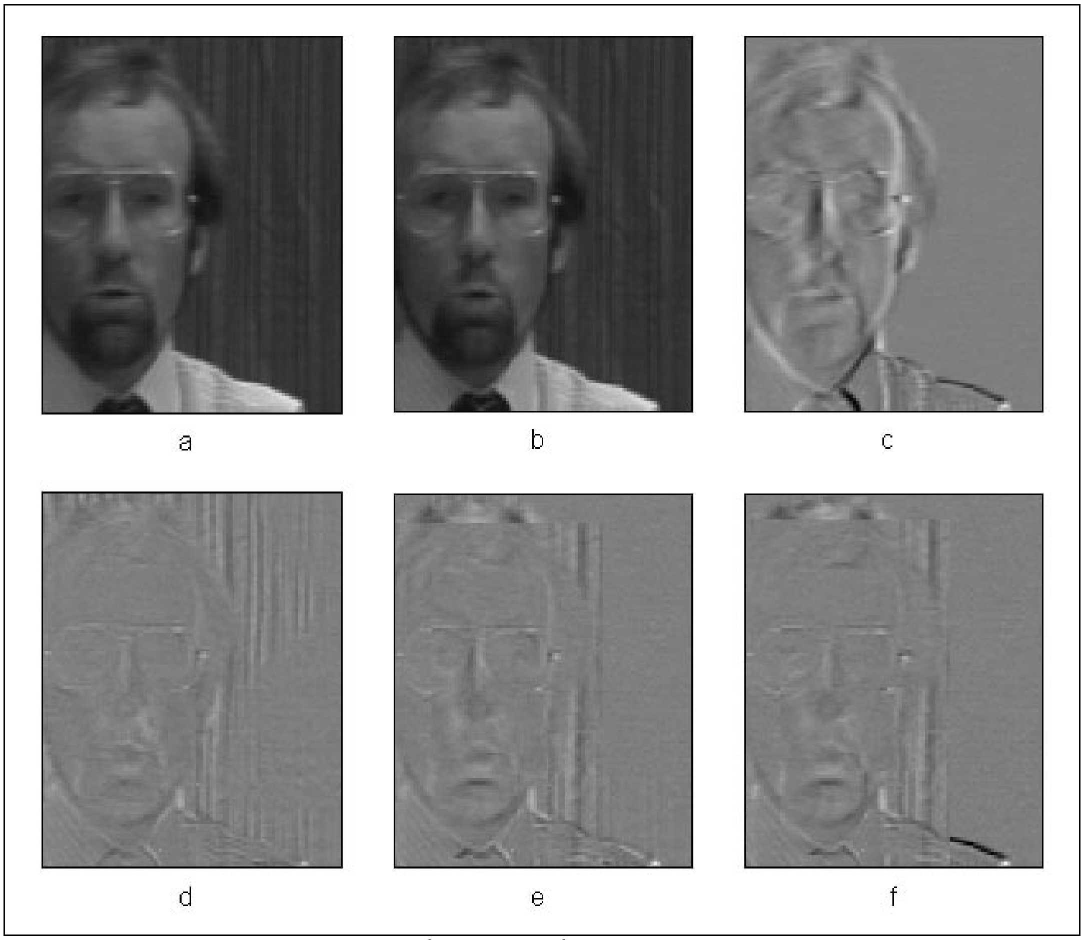


Figure 3: Prediction of Frame 13 from the 'Trevor' test sequence
(a) Frame 12 from the original sequence. (b) Frame 13 from the original sequence. (c) Frame difference (Frame 13 - Frame 12). (d) Prediction error from the affine/translational method. (e) Prediction error from the gradient-based method. (f) Prediction error from the block-matching method.

7. REFERENCES

[1] Y. Nakaya and H. Harashima. Motion Compensation Based on Spatial Transformations. *IEEE Transactions on Circuits and Systems for Video Technology*, 4(3):339–356, June 1994.

[2] J. Nieweglowski, T.G. Campbell, and P. Haavisto. A Novel Video Coding Scheme Based on Temporal Prediction using Digital Image Warping. *IEEE Transactions on Consumer Electronics*, 39(3):141–150, August 1993.

[3] S. Rebay. Efficient Unstructured Mesh Generation by means of Delaunay Triangulation and Bowyer-Watson Algorithm. *Journal of Computational Physics*, 106:125–138, 1993.

[4] G.J. Sullivan and R.L. Baker. Motion Compensation for Video Compression using Control Grid Interpolation. In *Proceedings of ICASSP'91*, pages 2713–2716. M9.1, May 1991.