Multiresolution Motion Estimation/Segmentation Incorporating Feature Correspondence and Optical Flow.

P.D.James and M.Spann,

School of Electronic and Electrical Engineering,

The University of Birmingham.

Abstract.

This paper is concerned with the segmentation of scene objects on the basis of their unique uniform motions. A number of previous approaches have been founded upon greyscale spatio-temporal gradient based estimation of the optic flow; these have shown some success. However optical flow only permits a limited range of recoverable motion displacements and exhibits a relatively low robustness to noise. Multiresolution image data can be used to increase the range of allowed motion displacements but the correct resolution at which to compute motion estimates is difficult to determine. It is postulated that with *a priori* knowledge of the elementary motions arising from the dynamic scene, the resolution level of a multiresolution support can be automatically set. These elementary motions may be used to increase noise robustness by permitting a relative rather than absolute classification of motion. We present a multi-stage algorithm in which feature correspondences are used to create a dictionary of elementary motions arising from the scene. The scene is initially segmented into small primitive regions using a maximum *a posteriori* (MAP) criterion in conjunction with a Markov random field (MRF) model and the motion dictionary. An affine motion model and maximum likelihood (ML) criterion are then used to fuse primitive regions of coherent motion into the full scene segmentation. Results for both synthetic and real imagery are given which demonstrate that scene segmentation may be performed across a wide range of motion displacements and at high levels of additive noise.

1. Introduction.

Motion has long been considered a powerful cue for segmentation. Optical flow leads to dense displacement estimates suitable for segmentation. However only small displacements can be recovered and it is not robust to noise. This severely limits segmentation performance on real world imagery. Feature correspondence, although able to yield accurate motion estimates over a wide range of displacements, provides only sparse displacement estimates. Correspondences are not guaranteed to lie on object boundaries and therefore only a limited segmentation can be achieved. In this paper it

BMVC 1995 doi:10.5244/C.9.59

is intended to show how by incorporating feature correspondences into an optical flow based scheme, a co-operative algorithm may be derived which benefits from the strengths of both.

Classical optical flow based motion estimation **(Horn and Schunk 1981)** relies on the computation of local spatio-temporal greyscale gradients within a finite local processing window (typically a cube of dimension 2x2 pixels x 2 frames) and so only small inter-frame differences are recoverable. Multiresolution versions of the image data in the form of a tapering pyramid can be used to increase the recoverable range. Most recently this approach has been pursued by **(Odobez and Bouthemy 1994)** and **(Meyer and Bouthemy 1994)** for motion estimation and **(Goh and Martin 1994)** for motion segmentation. Each pyramidal level consists of an image which is half the spatial resolution of the previous level and is produced by filtering (for example using quadtree sampling) the previous level. The local processing window may be successfully applied at some pyramid level which ensures a displacement less than ≈ 2 pixels per frame. However selection of an appropriate level at which to begin processing is problematic. A pyramid level that is too low may lead to aliasing. However a pyramid level that is too high may also lead to inaccurate motion estimates. At coarser resolutions, fine texturing may have been lost and so gives rise to motion singularities.

Real world noise is characterised by (as in **(Haralick 1994)**) a small random perturbation component modelled as additive Gaussian noise affecting every site and a larger perturbation component affecting only a small number of sites, such as, motion singularities. Low noise robustness to random perturbations is due mainly to the greyscale gradient-based estimation. Furthermore, in areas of low texture, motion estimates may not be able to be computed due to the aperture effect. A model based approach is able to reduce the effect of noise by incorporating a number of greyscale gradient calculations over a large support region. However, enlarging the image support increases the possibility of it straddling two or more motions. This leads to an ill-conditioned model over which motion estimates cannot be recovered. If the elementary motions in the scene are known, a relative measure of motion fit can be used. That is, if the dominant motion of a site can be recovered, a site may be classified on the basis of its separation in displacement (direction and magnitude) from the known scene motion. A schema for the proposed algorithm is shown in Figure 1.

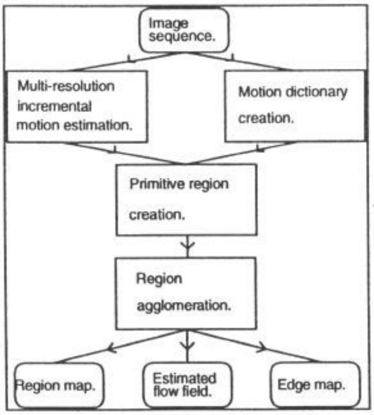
Image sequence.
Multi-resolution incremental motion estimation.
Motion dictionary creation.
Primitive region creation.
Region agglomeration.
Region map.
Estimated flow field.
Edge map.

Figure 1. Algorithm schema.

The image sequence consists of a 2-D projection of the 3-D scene. The image sequence is used to (i) calculate a dense field of motion estimates using an optical flow based multiresolution incremental motion estimator, (ii) create the motion dictionary from the correspondence of feature points between two sequence images. A maximum *a posteriori* (MAP) criterion and Markov random field (MRF) interpretation are used to provide a mathematical framework in which sites within the motion field are associated with entries in the dictionary to produce small 2-D planar patches. To extract more complex motion an affine optical flow model fuses patches of coherent motion together by using a maximum likelihood (ML) criterion.

The organisation of this paper is as follows. Section 2 gives a full description of the motion estimator employed. Section 3 describes the motion dictionary. Section 4 presents the formulation of the planar patch segmentation problem using the MAP criterion. Section 5 gives a full description of the segmentation process. Preliminary qualitative results for both synthetic and real scenes are presented in section 6. Finally, conclusions and possible algorithm extensions are discussed in section 7.

2. Multi-resolution Motion Estimator.

Given the brightness function $I(\boldsymbol{x},t)$ the fundamental constraint equation **(Horn & Schunk 1981)** is given by :

$$\nabla I(\boldsymbol{x},t)^T \boldsymbol{u}(\boldsymbol{x},t) + I_t(\boldsymbol{x},t) = 0 \qquad (1)$$

where $\nabla(I(\boldsymbol{x},t)) = (I_x(\boldsymbol{x},t), I_y(\boldsymbol{x},t))^T$ is the spatial gradient vector of $I(\boldsymbol{x},t)$ and $\boldsymbol{u} = (u,v)^T$ is the flow vector with x and y components u and v and $\boldsymbol{x} = (x,y)^T$.

Consider a model based flow $\boldsymbol{u}_{A,\boldsymbol{b}}(\boldsymbol{x})$ using an affine parameterisation of the flow field as follows :

$$\boldsymbol{u}_{A,\boldsymbol{b}}(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b} \qquad (2)$$

where $A = \begin{pmatrix} m_0 & m_2 \\ m_1 & m_3 \end{pmatrix}$ is the model matrix and $\boldsymbol{b} = (u_0, v_0)^T$. The time dependence has been dropped for simplicity of notation. The affine model is able to describe a range of complex motions including expansion/contraction, rotation and shear **(Campani and Verri 1992)**. Similar modelling approaches have also been adopted for optical flow estimation **(Chou & Chen 1993)** and for dynamic scene segmentation/analysis **(Bouthemy & Francois 1993)**. Both use affine models to characterise the apparent motion. Further, an affine motion model has been used to carry out long range tracking of regions in image sequences **(Meyer & Bouthemy 1994)**.

Let $\boldsymbol{e}_{A,\boldsymbol{b}}(\boldsymbol{x})$ be defined as the error between the true and model based flow. Substituting from Equation (2) :

$$\nabla I(\boldsymbol{x})^T \boldsymbol{u}_{A,\boldsymbol{b}}(\boldsymbol{x}) + I_t(\boldsymbol{x}) = n(\boldsymbol{x}) \qquad (3)$$

where : $n(\boldsymbol{x}) = \nabla I(\boldsymbol{x})^T \boldsymbol{e}_{A,\boldsymbol{b}}(\boldsymbol{x})$ (4)

White Gaussian noise is used to model the random field $n(\boldsymbol{x})$ and is considered to be a zero mean process with variance σ^2. Substituting Equation (2) into Equation (3), the final motion model is given by :

$$\nabla I(\boldsymbol{x})^T (A\boldsymbol{x}+\boldsymbol{b})+I_t(\boldsymbol{x})=n(\boldsymbol{x}) \qquad (5)$$

Multiresolution data is used to extend the recoverable interframe displacement range. For images at t and $t+1$, quadtree pyramids are built. The pyramid level is denoted by p where $p=0$ is the pyramid base. It is possible to estimate motion by directly applying the model given by (5) to image data at an appropriate pyramid level. This is not satisfactory as the maximum resolution of motion estimates that can be recovered is 2^{p+1}. At coarser resolutions (i.e. higher interframe displacements) this is far too coarse to allow differentiation between objects at a common pyramid level. An incremental motion estimate, as proposed by **(Meyer and Bouthemy 1994)**, is therefore employed. The estimator is able to refine a previous (possibly initial) estimate using finer resolution data at the next pyramidal level. Initial motion estimates are made at an appropriate resolution level p by applying (5) in a least squares manner over an 8x8 pixel support to obtain parameters estimates $\left(\hat{A}^P, \hat{b}^P\right)$. This estimate is then refined by projection of parameter estimates down to the receptive field (in this case 16x16 pixels) of the next resolution level. The projection operation creates four separate support regions of 8x8 pixels (see Figure 2). Splitting the receptive field ensures an acceptably small 8x8 support region is maintained. The new support regions are shown enumerated 1-4 in Figure 2.

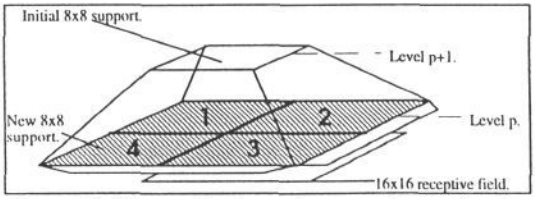

Figure 2. The projection of parameter estimates.

The parameter estimates are used to account for an initial displacement when calculating partial derivatives. Equation (6) is used to obtain least-squares estimates $\left(\Delta\hat{A}^p, \Delta\hat{b}^p\right)$ of $\left(\Delta A^p, \Delta b^p\right)$ at the next level **(Meyer and Bouthemy 1994)**.

$$\nabla I\left(\boldsymbol{x}^p + 2\delta\hat{\boldsymbol{x}}^{p+1}, t+\delta t\right)\cdot\left(\Delta A^p \boldsymbol{x}^p + \Delta b^p\right)\delta t + I\left(\boldsymbol{x}^p + 2\delta\hat{\boldsymbol{x}}^{p+1}, t+\delta t\right) - I\left(\boldsymbol{x}^p, t\right) = 0 \ (6)$$

where $\delta\hat{x}^{p+1}$ is the incremental estimate from the previous level.

These are then added to the overall estimate to form the current estimate at level p:

$$\hat{A}^p = \sum_{i=p}^{P-1} \Delta\hat{A}^i + \hat{A}^P \text{ and } \hat{b}^p = \sum_{i=p}^{P-1} 2^{i-p}\Delta\hat{b}^i + 2^{P-p}\hat{b}^P \qquad (7)$$

3. The Motion Dictionary.

The motion dictionary is composed of a list of motion entries (see Figure 3).

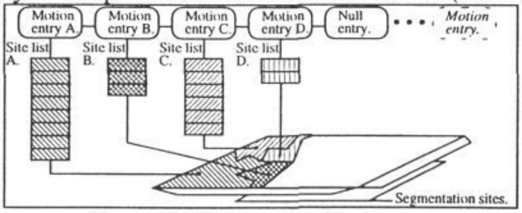

Figure 3. The motion dictionary.

These represent elementary 2-D motions arising from the scene and are derived by application of a feature point correspondence algorithm **(Zhang, Deriche, Faugeras and Long 1994)** to the image sequence. The motions are considered elementary since they are either simple 2-D displacements or part of more complex 3-D motion. We assume that at least one feature correspondence can be recovered per object (more in the case of complex motions). This, in practice, is not difficult to ensure since many feature points are available. A motion entry is ranked to a resolution level according to its interframe displacement which determines the resolution (that is pyramid level) at which to compute optical flow when classifying a site. Each motion entry holds a site list. Membership of a motion entry's site list directly specifies the interpretation assigned to a site. The membership of a site is strictly restricted to one motion entry. Therefore a site list represents one unique region which is defined by the list members. Every instance of the motion dictionary automatically contains a 'null' motion entry. A site at which the optic flow cannot be computed (for example at a singularity) is assigned to this entry and subsequently excluded from the segmentation. A parent/child hierarchy may be defined within the motion dictionary. In this way two motion entries and their motion lists may be merged by simply making one motion entry the child of another (Figure 4).

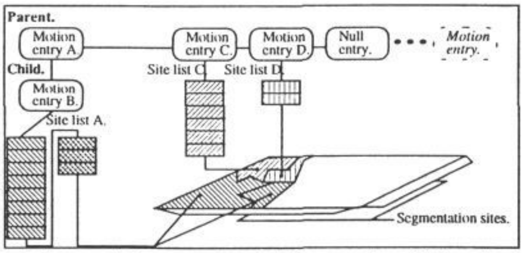

Figure 4. Parent / child hierarchy within the motion dictionary.

4. MAP Formulation of Solution.

An interpretation X of the scene specifies the set of labels given to each lattice site based on the observable image sequence data. The set $DICT = \{d_1, d_2, \ldots d_Z\}$ denotes the motion dictionary and fully specifies the state space for elements of X so that $x_s \in DICT$ where s indexes the set X.

The set $\Omega = \{\omega = (x_1, \ldots, x_M) : x_i \in DICT, 1 \le i \le M\}$, where M is the image lattice size, specifies all possible configurations of the interpretation field of which ω is one possibility. The problem becomes that of finding the most probable interpretation $X = \omega$ given to the scene data D. This may be expressed using the MAP criterion :

$$\mathrm{P}(\mathrm{X}=\omega^{*}\,|\,\mathrm{D})=\max_{\omega\in\Omega}\left(\mathrm{P}(\mathrm{X}=\omega\,|\,\mathrm{D})\right) \tag{8}$$

using Bayes' theorem this maybe written as :

$$\max P(X|D)=\frac{\left(P(D|X=\omega)P(X=\omega)\right)}{P(D)} \tag{9}$$

$P(D|X)$ specifies how well the scene data is explained by the current interpretation. $P(X)$ specifies the extent to which the interpretation meets our prior expectations of a sensibly ordered solution. $P(D)$ is ignored as it is constant with regard to the maximisation.The joint probability $P(X|D)$ is related to the noise distribution. Zero-mean white Gaussian noise of standard deviation σ is assumed. Therefore $P(D|X)$ can be written as:

$$P(D|X)=\left(\frac{1}{2\pi\sigma^{2}}\right)^{N/2.}\cdot\exp\left(-\frac{\|n\|^{2}}{2\sigma^{2}}\right) \tag{10}$$

where $\left\|n^{2}\right\|=\sum_{i=i}^{N}(n_{i})^{2}$ is the total contribution to noise over a region of N points.

As in section 2, a noise corrupted projection of the motion onto the image plane is considered. From (5), $\|n\|^{2}$ is given by

$$\|n\|^{2}=\sum_{s\in\Lambda}\left(\nabla I(\boldsymbol{x}_{s})^{T}(\hat{A}\boldsymbol{x}_{s}+\hat{b})+I_{t}(\boldsymbol{x}_{s})\right)^{2} \tag{11}$$

where the summation is over all pixel sites and $\boldsymbol{x}_s$ is the spatial position of the site s relative to its associated feature correspondence.

The estimated model parameters $(\hat{A},\hat{b})$ are derived by solving the affine model over the region in question. The resolution at which estimation begins is determined by the resolution rank of the motion dictionary entry for the current interpretation.

The expectation of the scene interpretation is modelled as an MRF. This allows a site's interpretation to be determined solely on the basis of some local finite neighbourhood. Furthermore it may be shown that if X is an MRF with respect to some local neighbourhood it is also a Gibbs distribution with respect to that neighbourhood **(Geman and Geman 1984)**. Thus $P(X)$ can be expressed in terms of an energy function $U(\omega)$:

$$P(X=\omega)=\exp\left(-U(\omega)\right)/Z \tag{12}$$

A second order system of the eight nearest neighbours is used and a set of cliques C is defined both spatially and temporally (see Figure 5).

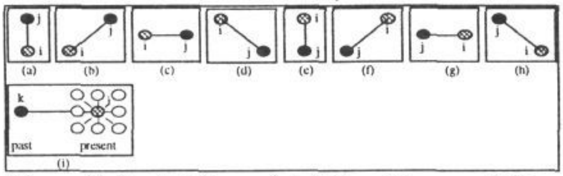

Figure 5. Spatial (items (a)-(h)) and temporal (item (i)) clique set.

Each clique is a subset of the neighbourhood system. The total energy term is given as the sum of these individual cliques :

$$U(\omega)=\sum_{C}V_C(\omega) \tag{13}$$

where V_C represents an individual clique potential.

The spatial cliques are used to support a piecewise constant *a priori* constraint which assumes that sites will have common motion interpretations. However the interpretation $x_s = d_{null}$ may be selected at a fixed cost of α. This allows sites where motion estimates cannot be recovered (primarily singularities and greyscale discontinuities) to be isolated at the cost of accepting the d_{null} state. The energy term is therefore given as:

$$V_{spatial}(ij)=\left\{\begin{matrix}\alpha\ if\, \boldsymbol{x}_i = d_{null} \\ \beta\ \text{ if}\, \boldsymbol{x}_\text{i} \neq \boldsymbol{x}_j \text{ otherwise}\end{matrix}\right\} \tag{14}$$

Cliques which include a d_{null} site are ignored. The penalty term β is determined from the data and is specified as the difference between interpretations x_i and x_j in the motion dictionary. The measure of difference, as defined in **(Barron, Fleet and Beauchemin 1994)** is given by Equation (15).

$$\psi=\arccos(\vec{v}_c\cdot\vec{v}_e) \tag{15}$$

where $\vec{v}\equiv 1/\left(\sqrt{u^2+v^2+1}\right)(u,v,1)$

This represents an angular measure of the difference between the two motion vectors $\vec{v}_c$ and $\vec{v}_e$ both in terms of direction and velocity. The measure is convenient as it handles both large and small speeds. The reference vector $\vec{v}_c$ is supplied by the motion dictionary entry and the comparison vector $\vec{v}_e$ by the estimated motion displacement of the site computed from data at the resolution level specified by the motion dictionary entry.

The temporal clique is used to support a site memory similar to that proposed in **(Murray and Buxton 1987)**. The energy function given in Equation (16) favours site interpretations which remain constant from one iteration to another. The penalty term ζ is gradually increased over the number of iterations a certain interpretation has survived for.

$$V_{temporal}(ik)=\left\{\begin{matrix}-\zeta\ if\, \boldsymbol{x}_i=\boldsymbol{x}_k \\ +\zeta\ \text{ otherwise}\end{matrix}\right\} \tag{16}$$

The total local energy term is specified as :

$$\text{U}_\text{i}(\omega)=\sum_{\text{j}}\text{V}_\text{spatial}(\text{ij})+\sum_{\text{k}}\text{V}_\text{temporal}(\text{ik}) \tag{17}$$

and the global energy is given as :

$$\text{U}(\omega)=\sum_{\text{i}}\left(\sum_{\text{j}}\text{V}_\text{spatial}(\text{ij})+\sum_{\text{k}}\text{V}_\text{temporal}(\text{ik})\right) \tag{18}$$

5. Algorithm Description.

5.1. Interpretation Field Initialisation.

The MRF field is initialised by applying Equation (15) at each site within X to assign an initial interpretation from D. The optical flow is computed at the required resolution determined by the resolution rank of the interpretation. The interpretation which provides the best level of fit is selected.

5.2. Primitive Region Creation.

The optimisation process consists of minimising the global energy function given by Equation (18). This is performed using a deterministic relaxation process based upon the iterated conditional modes (ICM) algorithm **(Besag 1986)**. Under this scheme, sites are visited randomly. At each site the energy function is calculated for each possible member of the motion dictionary. The chosen interpretation is that which maximises the local decrease in energy (Equation (17)). The site is labelled by associating it with the appropriate motion dictionary entry. The process terminates when the global energy cannot be reduced any further by local site changes within one complete iteration. Local minima are avoided by a good initial setting of the interpretation field. The interpretation field then consists of a number of roughly 2-D planar patches or 'primitive' regions which are associated with motion dictionary entries. Those motion dictionary entries not associated with any region are removed.

5.3. Primitive Region Agglomeration.

The primitive regions define an initial approximation to the complete segmentation. In order to determine the complete region structure, the number of regions and affine motion parameters, a fusion process is required. This is carried out by considering merging entries within the dictionary. The association of sites to a motion dictionary entry defines a region and resolution over which the affine model (given in Equations (5) and (6)) may be computed. The partial structural knowledge improves affine model estimates since the support region is large but does not straddle multiple motions. The possible merging of two motion dictionary entries is based on a ML decision (**James and Spann 1995**).

$$\frac{\left(\hat{\sigma}^2(R_1)\right)^{|R_1|}\left(\hat{\sigma}^2(R_2)\right)^{|R_2|}}{\left(\hat{\sigma}^2(R_1 \cup R_2)\right)^{|(R_1 \cup R_1)|}} > T_{merge} \tag{19}$$

where $\hat{\sigma}^2(R)$ is the affine model error for region R and T_{merge} is a threshold.

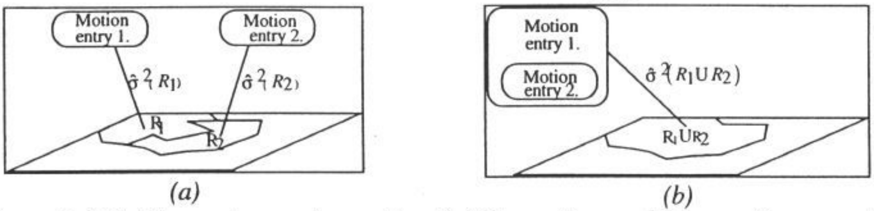

Figure 6. Primitive region agglomeration. (a) Computing model errors for separate regions. (b) Computing model error for combined region.

The merging process is shown in Figures 6(a) and 6(b). If the regions are merged, one motion dictionary entry is removed from the motion dictionary and re-introduced as a child of the other motion dictionary entry. Thus sites with interpretations belonging to one motion dictionary entry automatically assume new interpretations belonging to the other motion entry. A new region is therefore formed from the agglomeration of R_1 and R_2. This process is carried out for each motion entry and iterated until no new merges are accepted on one complete iteration.

6. Results.

Figure 7(a) shows a synthetically produced scene consisting of 3 squares translating at 6 pixels/frame downwards, 1 pixel/frame upwards and left and 3 pixels/frame to the right respectively (top to bottom left to right). Initial dictionary entries totalled 39. Figures 7(b) and 7(c) show the primitive and agglomerated regions respectively.

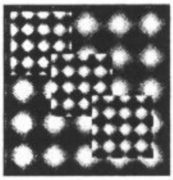

Figure 7(a)

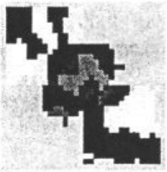

Figure 7(b)

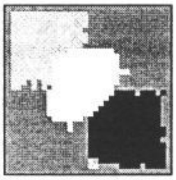

Figure 7(c)

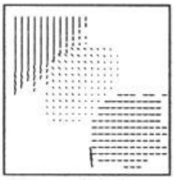

Figure 7(d)

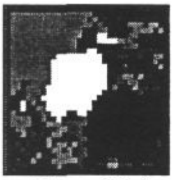

Figure 7(e)

T_{merge} was set to 5000.0 and α set to 10.0. Each square was segmented at a different and correct resolution. The estimated displacement field is shown in Figure 7(d). Figure 7(e) shows the effect to adding white gaussian noise of variance 120.0 to the scene. As can be seen, although small fragments are present each square's structure can be recovered. This noise robustness is gained by using the relative classification criterion.

Figure 8(a) shows a camera translating horizontally across a stationary scene. This produces a parallax effect where closer objects appear to move faster than those farther away. The tree trunk undergoes the greatest motion of ≈ 5 pixels/frame. Initial dictionary entries totalled 64. Figures 8(b) and 8(c) show the primitive and agglomerated regions respectively. T_{merge} was set to 5000.0 and α set to 10.0. The estimated displacement field is shown in Figure 8(d). Figure 8(e) shows the effect to adding noise of variance 120.0 to the scene. As can be seen the segmentation is largely unaffected.

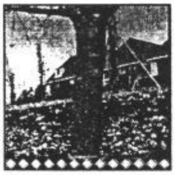

Figure 8(a)

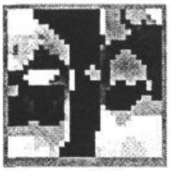

Figure 8(b)

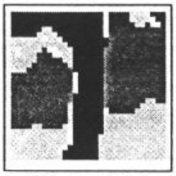

Figure 8(c)

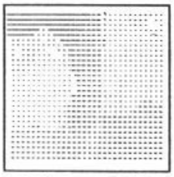

Figure 8(d)

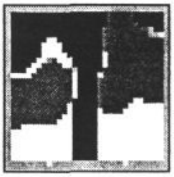

Figure 8(e)

7. Conclusions.

Segmentation from motion using greyscale spatio-temporal gradient based estimation of the optic flow has only a small range of recoverable motion displacements and a low robustness to noise. This paper has addressed these problems by incorporating feature correspondences and multiresolution data. The correspondences provide a dictionary of motions within the scene, allowing a relative classification of a site and a local multiresolution support to be set at the correct resolution. The results show good robustness to noise and recovery of motion over a much wider range (0 - $\approx$7 pixels) of interframe displacements. A number of possible additions to the algorithm are attractive. The spatial location of a feature correspondence may be used to weight the probability of a site's interpretation based on its distance from the spatial position. Also, by including a measure of confidence along with motion dictionary entries the most accurate interpretations may be weighted so as to favour their selection above inferior ones.

References.

[1] Barron JL, Fleet DJ, Beauchemin SS, "Performance of Optical Flow Techniques", Int. Journal of Computer Vision, **12**, 1, 1994, pp 43-77.

[3] Besag J, "On the Statistical Analysis of Dirty pictures", Journal of Royal Statistical Society, **48**, 3, 1986, pp 259-302.

[4] Bouthemy P, Francois E, "Motion Segmentation and Qualitative Dynamic Scene Analysis from an Image Sequence", Int. Journal of Computer Vision, **10**, 2, 1993, pp 157-182.

[5] Campani M, Verri A, "Motion Analysis from First-Order Properties of Optical Flow", CVGIP:Image Understanding, **56**, 1, 1992, pp 90-107.

[6] Chou WS, Chen YC, "Estimation of the Velocity Field of Two Dimensional Deformable Motion", Pattern Recognition, **26**, 2, 1993, pp 351-364.

[7] Geman S, Geman D, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images", IEEE Trans. on Pattern Analysis and Machine Intelligence, **6**, 6, 1984, pp 721-741.

[8] Goh WB, Martin GR, "Model Based Multiresolution Motion Estimation in Noisy Images", CVGIP:Image Understanding, **59**, 3, 1994, pp 307-319.

[9] Haralick RM, "Performance Characterization in Computer Vision", CVGIP:Image Understanding, **60**, 2, 1994, pp 245-249.

[10] Horn BKP, Schunk BG, "Determining Optical Flow", Artificial Intelligence, **17**, 1981, pp 185-203.

[11] James PD, Spann M, "Model-Based Motion Estimation/Segmentation using an Adaptive Pyramidal Approach.", Int. Journal of Computer Vision, submitted for publication.

[12] Meyer FG, Bouthemy P, "Region-Based Tracking Using Affine Motion Models in Long Range Image Sequences", CVGIP:Image Understanding, **60**, 2, 1994, pp 119-140.

[13] Murray DW, Buxton BF, "Scene Segmentation from Visual Motion Using Global Optimisation", IEEE Trans. on Pattern Analysis and Machine Intelligence, **9**, 2, 1987, pp 220-228.

[14] Odobez J, Bouthemy P, "Robust Multiresolution Estimation of Parametric Motion Models Applied to Complex Scenes", INRIA France, Internal Publication 788, 1994.

[15] Zhang ZY, Deriche R, Faugeras O, Long QT, "A Robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry", Technical Report No. 2273, INRIA, France, 1994.