GRADIENT-BASED FLOW SEGMENTATION AND LOCATION OF THE FOCUS OF EXPANSION.

Ian Overington
Sowerby Research Centre
British Aerospace PLC
FPC 267, PO Box 5
Filton, Bristol BS12 7QW

ABSTRACT

In previous papers it has been shown how it is possible to sense fragmentary motion and local edge orientation to a high accuracy, directly from illuminance gradient data present in pairs of frames from a time sequence. It has also been suggested how such data may be used to generate optic flow maps. The present paper looks at the practicability of generating such optic flow maps and at the possibilities of scene segmentation and determination of focus of expansion from such maps. It is shown that flow field scene segmentation is easily realisable, and that the focus of expansion can be determined to a high accuracy (typically better than 1 pixel).

I INTRODUCTION.

In past papers it has been shown how it is possible to compute the apparent motion orthogonal to edges, together with the local orientation of the edge fragments, to a high accuracy directly from gradient-based data contained in pairs of frames in a time sequence (Overington 1985a, Overington & Greenway 1986). It has also been discussed in theory how it should be possible, starting from such fragmentary motion data, to compute, to a similar level of accuracy, the motion of local objects in the scene or the global motion of the scene, subject to certain constraints (Overington 1984a, Overington 1985b). The purposes of this paper are threefold. Firstly, the practical feasibility of the latter claim will be demonstrated. Secondly, the practicability of a simple and direct scene segmentation based solely on optical flow will be illustrated. Finally, a method of direct computation of the focus of expansion for the time period between the pair of frames used for the analysis will be discussed.

II BACKGROUND

For some time we have considered that it should be possible to produce computer vision algorithms which have similar capabilities to those of human vision, using the sampling interval as a characteristic unit of distance (Overington 1983, Overington 1984b). Also, human vision has an apparently instant capability to sense motion with great ease, whilst some efforts to achieve motion sensing by computer vision systems have been slow, indirect and imprecise. For instance, some have suffered from the so-called "correspondence" problem of correctly associating features in two frames (Marr 1982, Marr & Poggio 1979). Equally others have been limited by the "aperture" problem, since at any point on an edge (apart from its ends) it is only possible to measure the component of local motion orthogonal to the edge (Hildreth 1983).

It is acknowledged by theorists studying optic flow that a very important attribute of flow in real dynamic viewing of 3-D scenes is the existence of flow field discontinuities. These can provide some of the recovery data for separation of translational and rotational components of optic flow and thereby can provide a means of determining 3-D depth from flow (e.g. Koenderink & Van Doorn 1976, Rieger & Lawton 1985). An equally important parameter in 3-D dynamic optic flow is the instantaneous focus of expansion, this being the basic parameter necessary to define the sensor heading relative to the image frame of reference. It is evident that human perception achieves both flow field segmentation and focus of expansion estimation instantly and with apparent ease. It would therefore seem reasonable to expect that it should be possible to find simple computer processes to carry out both these tasks.

In a previous paper (Overington & Greenway 1986) it was shown that, by carefully addressing the preprocessing methods, it was possible to compute, at each pixel on an edge feature, both the local orthogonal motion between two time samples, to a small fraction of a pixel, and the local orientation of the edge, to better than 1 degree. Such computation was shown to be possible directly from the local average and difference of the two frames, without any correspondence problem. It was subsequently shown in theory that, by modelling the organisation of the orientation sensitive neurones in the visual cortex in primates, it should be possible to use region analysis to provide adequate data to solve the aperture problem (Overington 1985b). All the practical capabilities to be demonstrated in this paper depend on this same theoretical premise, which will be shown to be exceedingly robust in practice.

III THEORY

A. Progressive Window Analysis.

For an image-based coordinate system each edge fragment will have two potential components of motion, one due to the combined effects of XY translational motion of the sensor plus sensor rotation, and one due to Z translational motion. The latter of these will be proportional to the radial distance from a chosen reference point. If the reference point is taken as the optical axis, then the XY and Z flow components relate to the sensor motion. Alternatively, if the reference point is taken as the approximate centroid of a local moving object then the XY and Z flow components relate to the motion of that object relative to the sensor. In many practical situations these are the two most important aspects of optic flow which require to be studied.

* This report describes work done at Sowerby Research Centre as part of Alvey IKBS Project "Spatio-temporal processing and optical flow for computer vision" (Alvey Ref. No. 013).

AVC 1987 doi:10.5244/C.1.23

In the previous theoretical paper (Overington 1985b) it was suggested that if one could localise a moving object then it was theoretically easy to compute the translational motion of the object from the fragments of orthogonal local apparent motion. This is because for any moving rigid body the sum of fragmentary motion components from approximately aligned edge fragments, with due allowance for radial distances of individual fragments from the reference centre, will describe a sinusoid. However, in that paper no explanation was given as to how one was to delineate the object to start with. In fact, such a task is itself difficult in practice. It has therefore been conjectured that it might be possible to derive adequately interpretable data by applying an ordered 2-D window analysis, with various sizes of window, to the fragmentary motion data in the image plane. This assumes that within any given window the true motion would be sufficiently consistent.

Such an assumption might on first sight appear to be very crude. It would seem obvious that it could not hold true for differentially moving objects much smaller than a given window, since for such objects there would be a danger of a great deal of contamination of the data due to differentially moving edge fragments from the surroundings of the object in question. Equally one might expect serious contamination near motion disruption boundaries, where conflicting data from both sides of the boundary would fall within the analysis window. However, these apparent shortcomings are analagous to one of the properties of early human vision which has been exploited in B.Ae. computer modelling and has been demonstrated to be necessary for extraction of high accuracy edge data (Overington 1985c) - that is, effectively high blur (or low resolution) of the input stimulus data relative to the spatial sampling interval. In this case the local optic flow map is effectively blurred by the sampling window. Hence by choosing a suitable spatial interval between window centres a good approximation to the ideal blur / sampling relationship can be achieved. For the present studies we have chosen to use spatial stepping of the window centre for progressive spatial analysis by one quarter of the linear dimension of the currently operating window in each of two orthogonal dimensions. Such a relationship between window size and spatial sampling step is considered to be a reasonable approximation to the ideal blur / sampling relationship discussed previously (Overington 1985c), whilst having the virtue of very simple and economical computation of window motion statistics. A better approximation would theoretically be a Gaussian-weighted window analysis. Such an analysis is presently being studied, but it is much more computationally expensive and will not be further considered here.

The concept of multiple scale window analysis has further analogy in early vision - that of playing off noise against resolution by multiple scale processing (Baker & Sullivan 1980). For a "busy" scene, with lots of edge fragments at many orientations, one can use a relatively small window for local flow vector analysis, thereby being able to separate out the motion of small objects in the scene. Alternatively, by increasing the window size one can guarantee at least some analysis for scene areas with a relative sparsity of edge fragments, but at the expense of failing to analyse motion for small objects. In the limit, by using a window which is approximately the same size as the complete image one can readily extract a measure of the mean global XY translational motion for the time interval between the two frames considered.

B. Scene Segmentation based on Motion.

The motion vector map which is generated by a two-dimensional progressive window analysis using relatively small windows may be thought of as a pixellated energy distribution map possessing both magnitude and orientation. In other words, the effect of processing with small windows is to smooth and interpolate the optic flow data onto a regular grid of lower resolution. One may consider further processing such data in a number of ways: e.g. taking just the scalar magnitude of the local vectors, or just the orientation data. Alternatively, one can consider the possibility of trying to handle both the magnitude and orientation of the data. In many practical situations it is the magnitudes of the motion vectors which are of by far the most importance. For instance, if one has a local moving object against a static background the only information one requires in order to segment the object from the scene is the magnitude of the motion. Equally, for a flow disruption boundary due to differential depth planes in an image it is rare that there are two equal motion magnitudes with a difference of orientation. There are, however, some situations where one can have roughly equal magnitude and an orientation difference - e.g. for a local moving object and a moving sensor platform or when panning with a moving sensor. Also there is evidence that human vision is equally adept at sensing orientation discontinuities as it is in sensing magnitude discontinuities, so it is considered that it would be unwise to dismiss orientation as unimportant.

For the purposes of the present preliminary study of the possibilities of segmentation based on motion we have chosen to explore the capabilities of further analysis of the map of motion vector magnitudes. However, in order to permit segmentation of regions where the motion may be of similar magnitude but of grossly different direction, we have the option to carry out the secondary processing on distributions of $A\sin\theta$ and $A\cos\theta$, where A is the magnitude of the local motion vector and θ is its orientation. This is found to be a simple but effective way of yielding an adequate representation of all local motion magnitudes for segmentation purposes, whilst separating effects of vectors in widely differing directions. In such a representation all vector magnitudes will be represented in one or other of the two maps as not less than $A\sin 45°$ or $A\cos 45°$ ($=0.707A$). Such distortions of the true amplitude distribution are believed to be quite satisfactory for approximate segmentation purposes: the true data are still available, if required.

As already stated in Section III.A, the data available to us in the vector magnitude maps are essentially a blurred and smoothed representation of local motion, sampled approximately optimally for information transfer. Thus regions of roughly constant flow are represented as constant "intensities", very local regions of flow are represented as blurred "point sources" and regions close to flow disruption boundaries are represented as regions of strong "intensity" gradient. We therefore arrange to treat these data as a new image input to the basic B.Ae. preprocessor for edge extraction (e.g. Overington 1984b, Overington & Greenway 1986). The data are then reprocessed as for the original image data, using a displaced zero grey level of 128 for cosine and sine transform maps. The output yields, relative to the window centre sampling steps, the same vernier position and orientation data as for standard processing of luminance edges data, except that "edge fragments" are now fragments of flow disruption boundaries. Now, vernier position can typically be sensed to approximately 0.03 pixels for low noise scenes. The equivalent sensing of flow disruption boundaries, under ideal conditions, can therefore be expected to be to better than 0.03 sampling steps, which is less than one original pixel for sampling windows up to at least 64 x 64 pixels (i.e. window centre steps of at least 16 pixels). Also the *local orientation of flow disruption boundaries may* be expected to be to 1 or 2 degrees (within the limits of the blurred representation of the flow field).

C. Focus of Expansion (FOE) Computation.

A second powerful potential use to which the motion vector map may be put is in the computation of the focus of expansion for a moving sensor, effectively at the time midway between the two sampled frames. For any local area of the scene, providing the scene itself is static, the computed motion vector from a local window analysis may be considered to be a pure XY translation vector for that small portion of the scene (provided that there is no appreciable sensor roll between the two time samples). It is argued that the fact that this one vector may usually contain components due to XY and Z translational motion plus sensor yaw and pitch is irrelevant for focus of expansion analysis. The confounding of XY translation with yaw and pitch is contained in the magnitude of these local motion vectors, whereas the vector orientation data will always point to the instantaneous focus of expansion in the absence of significant sensor roll. Therefore whatever the distribution of magnitude of motion vectors may be as a result of 3-D scene structure, the orientations of the motion vectors for static parts of the imaged scene will all point to the focus of expansion (provided that there is not any significant sensor roll during the period between the frames). Provided that one is considering sensor platform motion, with only small *local regions of the scene having their own* independent motion, there is the theoretical possibility of computing one estimate of the focus of expansion for each pair of motion vectors, simply by computing their point of intersection. Obviously for motion vectors having very similar orientations such a computed estimate will be subject to gross errors due to inevitable noise in the images and consequent uncertainties in absolute orientation of the individual vectors. There will also be potentially large errors on any estimate based on one or both vectors from regions where there is local scene motion.

However, in a typical scene there may be several hundred motion vectors. These will involve a wide variety of orientations for FOE's within or near the field of view. Alternatively they will exhibit near parallelism for FOE's near infinity. This large number of vectors in turn leads to a very large number of possible pairings, typically of the order of several hundred thousand. Hence it is possible to reject pairings which involve pairs of vectors of closely similar orientation relative to the total spread of orientations and still be left with a large number of contributions to a statistical estimate of the focus of expansion. Provided that the areas of local scene motion are only a small percentage of the total scene, the spurious estimates of the focus of expansion from such local motion will not seriously degrade the mean position estimated from the whole scene.

IV PROGRAMMING

In order to explore the theoretical concepts discussed in Section III and in previous papers (Overington 1984a & Overington 1985b), a computer program has been written which permits a progressive windowed analysis of any digitised 2-D image, followed by a further analysis of the computed motion vectors either as a pseudo-energy map or for estimation of the focus of expansion. Any size of window may be selected, or a succession of window sizes in progressive ratio of 2:1 may be explored automatically. For the moment the program requires *that there be at least one fragment of orthogonal* edge motion contributing to each of the 12 basic 30 degree orientation channels (Overington 1983, Overington 1985c) within a given window for the analysis of that window to proceed. Otherwise the window location is bypassed, with a record being made of inadequate data. As discussed in pevious papers (Overington 1984a & Overington 1985b), it is possible, by what might be called "graceful degradation", to derive some flow data from a local area when a much less complete fragmentary flow is present. However, this requires considerably more programming and more running time for analysis, so we have elected initially to see what can be accomplished with the most rigorous restrictions. Some work has already started on relaxation of the restrictions to the case where fragmentary data should be available from at least one of each pair of opposite orientation channels rather than from each channel. This work is, however, only at an early stage.

To improve the stability of the outputs an option is provided to test the local consistency of the magnitude and/or the orientation of the computed motion vectors in local 3 x 3 groups of output data. By this means it is possible to suppress some of the effects of temporal noise and thereby restrict the outputs mainly to regions of realistic local consistent motion.

At this stage an option is provided to carry out a focus of expansion analysis. Basically this analysis involves taking the orientations of any pair of window centre vectors and computing the *intersection point for these two vectors. The mean* of all the intersection points and the standard deviation on the mean in X and Y are then derived as the statistics of the focus of expansion. If this process is permitted to run on all possible data then typically several hundred thousand computations are carried out. Many of these are of considerable uncertainty owing to the vector orientations being very similar. An option is therefore provided to select a range of orientation differences between individual pairs of vectors within which computations will be permitted. Presently this range is selected intuitively for each situation. It is considered that such range limits should be relatable to the range of orientations present in a given output map and this is under investigation.

The motion vector map is output in two forms. Firstly an image file is produced in which the data are stored in essentially analogue form - thereby suitable for plotting purposes and for subsequent higher level analysis. Secondly a byte-packed image file is produced of the sine and cosine components of the vector data which is in a format compatible with normal digital images stored from video records. This is then immediately suitable for re-entry into a form of B.Ae. preprocessor for sensing boundaries of consistent flow regions. For local flow it is also suitable for re-entry into a modified version of the B.Ae. preprocessor dedicated to sensing of local concentrations of energy, thereby providing outputs only at the centres of local flow regions. This latter output may be then used for search cueing or tracking purposes.

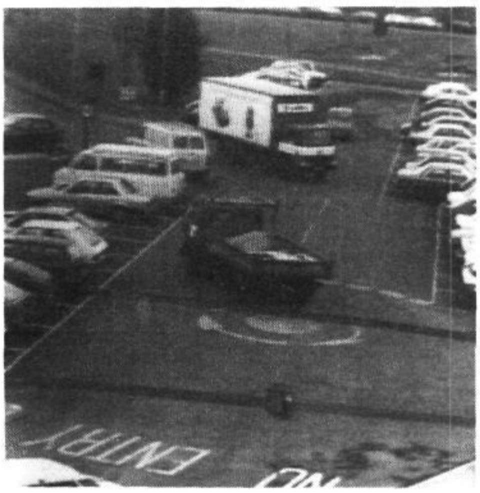
V PRACTICAL TESTS

A variety of images have been processed, including realistic scenes containing local and global motion, pairs of random-dot kinematograms and scenes containing typical growth (expansion) due to sensor platform motion.

The results show a consistent capability to extract both global and local XY translational motion vectors from both realistic scenes and random dot kinematograms. They also show a powerful capability for approximate scene segmentation based on local flow vectors, both for realistic scenes and for random-dot kinematograms. In particular this includes a capability to extract flow-field discontinuities due to both local moving objects and natural occlusions in 3-D input scenes. Finally it has been shown that focus of expansion estimates for natural scene image pairs are subject to very small uncertainties (typically fractions of a pixel).

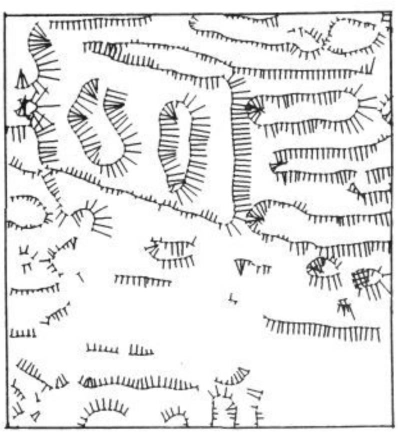
A. Local Motion - Real Scenes.

A number of sequences of real scenes containing moving cars and lorries have been explored in order to test out the capabilities of the processes reported in Section III.B for local motion extraction and segmentation. A typical sequence of stages of the processes is shown in Figure 1. Figure 1a shows one of an original pair of frames entered into the B.Ae. image preprocessor, this scene containing many static objects (buildings, vegetation, road markings, parked cars) plus two moving lorries. The raw fragmentary apparent motion map resulting from the standard B.Ae paired-frame analysis is shown in Figure 1b. In this figure print resolution precludes from showing the individual apparent motion vectors clearly, but the impression of local motion is created by the regions of apparently thickened or brush-like contours. Figure 1c shows an enlargement of a fragment of Figure 1b including one of the lorries. In this figure can be seen the actual discrete pixel by pixel apparent motion vectors which have tended to merge in Figure 1b. Figure 1d presents, as vector representations, the motion vectors at the centres of the sampling windows after a 32 x 32 pixel window analysis. In this figure the regions where there are no vectors are those regions where there were inadequate input fragments to satisfy the restriction discussed in Section IV. In figure 1e is shown the same vector map after applying a 3 x 3 local neighbourhood orientation consistency test (orientation standard deviation less than 20 degrees). The clusters of vectors in the region of the lorries and the relative absence of any regular

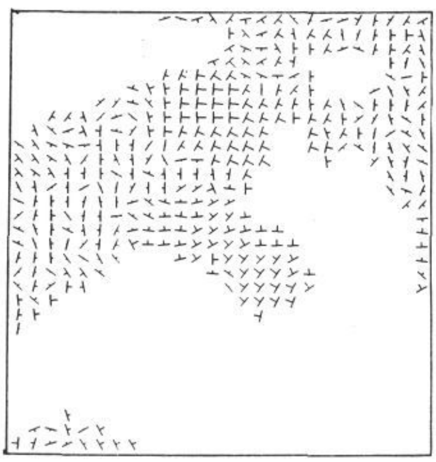


a) One of the input frames.

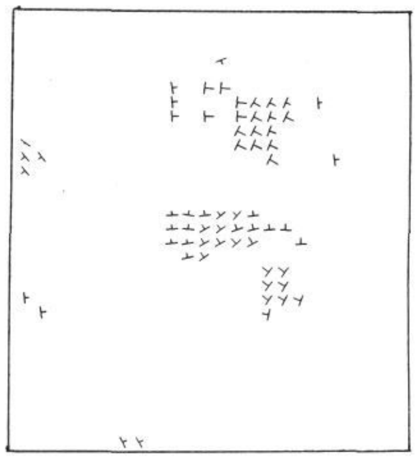
b) Fragmentary edge and motion output from paired-frame analysis. "Thickening" of profiles indicates local flow magnitude.



c) Enlargement of portion of b) to show individual edge and motion fragments. Stems of T's indicate magnitude of motion orthogonal to local edge fragments.



d) Flow vector map from 32 x 32 pixel progressive window analysis (vectors at intervals of 32/4 = 8 pixels).



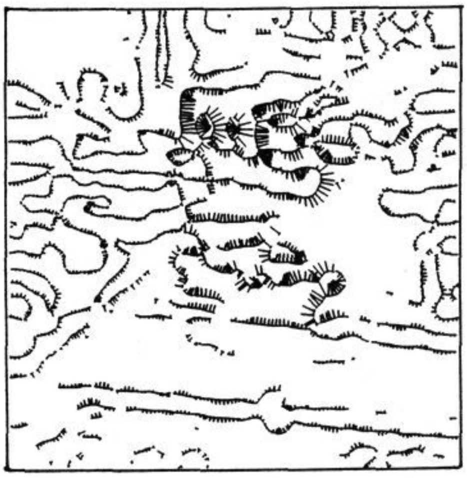
e) As d), but with rejection of vectors where local 3 x 3 patch has orientation uncertainty greater than 20 degrees.

Figure 1. Example of progressive analysis of local flow from a real scene containing two moving vehicles (256 x 256 pixels).

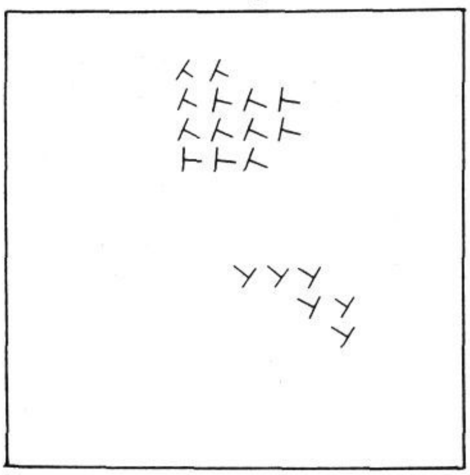
clustering of vectors elsewhere is evident. It will however be noticed that the groups of consistent vectors around the lorries form rather irregular patterns. It has been found that this is largely due to the lorries being too large relative to the analysis window (that is, the image of each lorry occupying too many pixels).

A considerably more controlled localisation of motion is obtained from a reduced scale initial image processing. In such a case the interpretation of the 2nd difference fragmentary motion map is difficult (Figure 2a), there being little semblance of the actual form of the lorries or other scene details in evidence. Yet when a 32 x 32 pixel window analysis is carried out and the local neighbourhood orientation consistency test is applied, it is found that a very clear clustering of motion vectors is obtained (Figure 2b).

The sine and cosine transformed magnitude data from this vector map were taken as a new input to the B.Ae. edge extraction process with the result shown in Figure 2c. As an alternative we can process the same data through a variant of the B.Ae. preprocessor in order to sense the approximate centroid of each local cluster of consistent flow vectors for tracking purposes.



a) Fragmentary edge and motion output from paired-frame analysis.



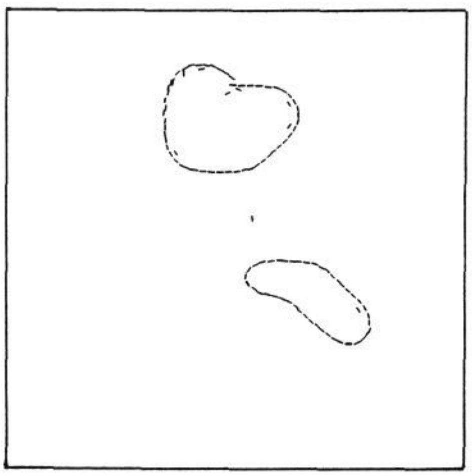
b) Flow vector map from 32 x 32 pixel progressive window analysis with rejection of vectors where local 3 x 3 patch has orientation uncertainty greater than 20 degrees.

B. Local Motion - Random-dot Kinematograms.

It is now a well-known fact in psychophysical experimentation that when human vision is presented with pairs of random dot patterns which contain coherently related patches of random dots - either *to the two eyes or in temporal sequence - a strong* impression of the coherent patches is evident (e.g. Julesz 1971, Van Doorn & Koenderink 1984). Such visual capability is usually assumed to be evidence of a very complex perceptual pattern matching processing. Yet it is instant and compelling. We speculated that our simple paired-frame analysis and subsequent window analysis should be capable of sensing such coherent patches, since it did not rely on pattern matching. We therefore produced computer-generated random-dot patterns with coherent displaced patches much as used by Julesz and by Koenderink, and attempted to process them through our flow analysis programs. The results were spectacular. Not only did the patches stand out quite clearly, but the spatial resolution capabilities for bar patterns were satisfyingly similar to those found in psychophysical experiments. In addition, in regions where there should be no motion at all, the fragmentary apparent motion outputs were exactly zero. This verifies that the flow analysis processes do not, in themselves, generate any spurious motion predictions. (Details of this part of the study and a selection of results have been published separately - Overington 1987).

C. Global Motion.

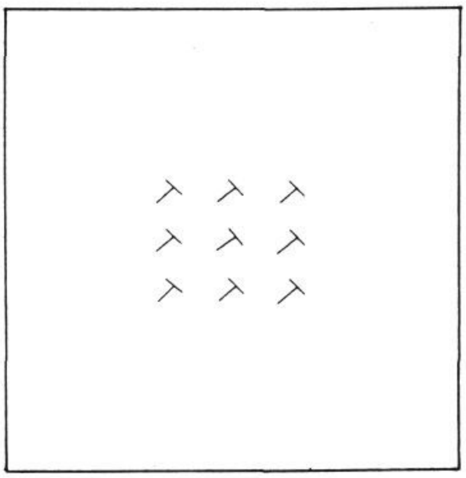
A convenient way of testing our ability to sense global motion within an image was to intentionally displace one frame from a paired-frame sequence by an integer number of pixels horizontally and/or vertically. Fractional pixel shifts were precluded at this stage owing to the complexities of reliable sub-pixel resampling of the original image frames. However, quite large global motions could theoretically be explored by first scaling down an image by n x n pixel blocking. Pairs of displaced frames could then be processed through our flow programs as far as the windowed vector analysis, using a large window size. An estimate of mean global flow could then be taken as the mean of the



c) Segmentation from data shown in 2b by reprocessing through a version of the B.Ae. edge sensing preprocessor.

Figure 2. Example of processing of pair of frames as in figure 1, but with a linear scale reduction of 2:1 (i.e. 128 x 128 pixels).

a) One of the input frames (256 x 256 pixels).



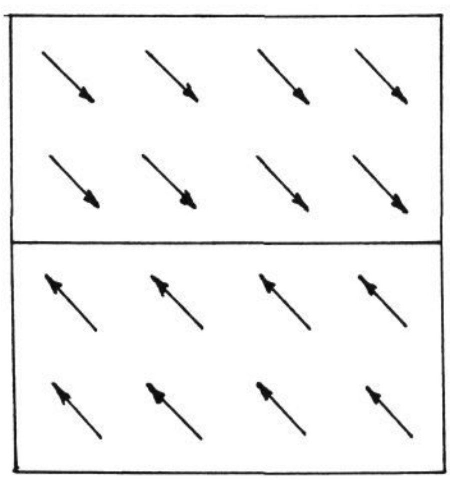
b) Flow vector output map from 32 x 32 pixel window analysis carried out on a 4:1 scale reduced version of a).

Figure 3. Example of measurement of global flow on a real scene.

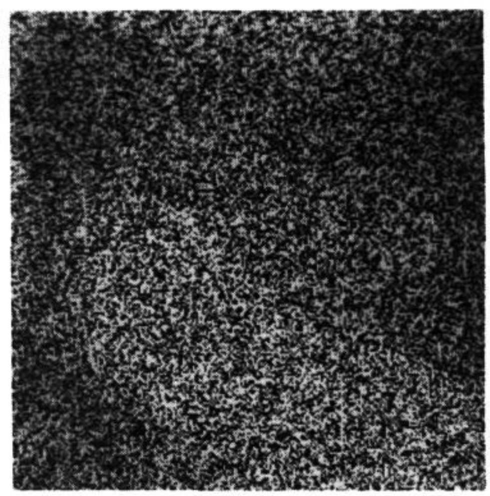
set of outputs from the window analysis. A typical image subjected to this treatment is shown in Figure 3a. The image can be seen to contain a lot of miscellaneous structure. One of the pair of images was displaced by 3 pixels horizontally and 3 pixels vertically. The pair of images were then resampled, using 4 x 4 pixels blocking, in order to reduce the effective motion for analysis to 0.75 x 0.75 pixels. The output from the windowed analysis is shown in Figure 3b. The consistency of the flow vectors is evident visually. The statistical estimate of global motion in this case was 4.04 original pixels at an angle of 136.6 degrees. Since the imposed input displacement was 3 pixels horizontal and 3 pixels vertical the true global motion was 4.24 pixels at an angle of 135 degrees.

D. 3-D Flow Disruption Boundaries.

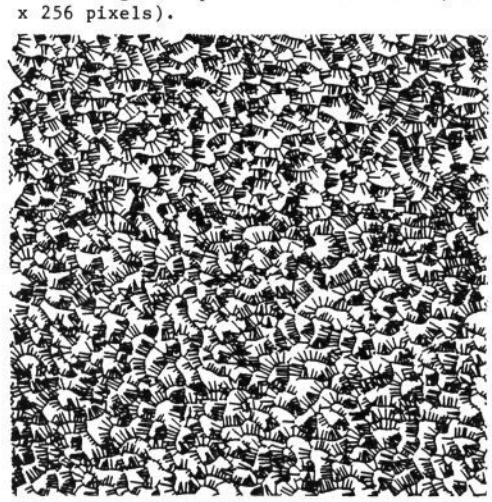
In the absence of suitable real data sequences for exploration of 3-D flow disruption boundary extraction, and since we had already shown the capability of our processes for sensing of discontinuities in random dot patterns, it was decided to demonstrate sensing of 3-D flow disruption boundaries by generating split field random dot pairs where the upper and lower halves



a) Conceptual split flow field input.

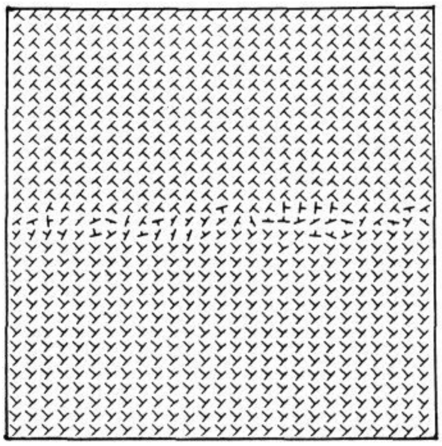


b) One of the random-dot patterns containing the split field flow data (256 x 256 pixels).

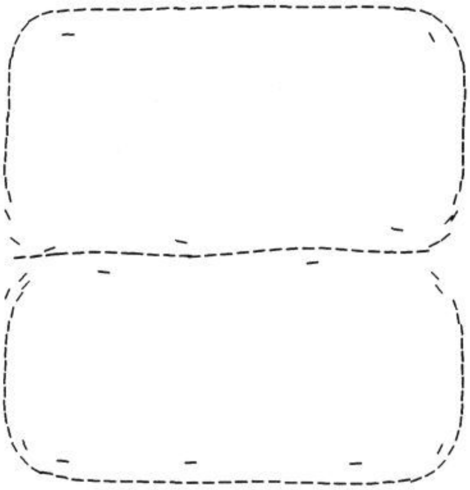


c) A 128 x 128 pixel central portion of the fragmentary edge and motion output from paired-frame analysis of the random-dot patterns.

Figure 4. Example of sensing of major flow field disruptions - synthetic random-dot images.



d) Flow vector output map from 32 x 32 pixel window analysis carried out on data in c).



e) Flow field disruption boundaries from reprocessing of the data in d) through a version of the B.Ae. edge sensing preprocessor. Note that in this case the output segments both the split field and the frame, since the whole frame contains motion.

Figure 4 (continued).

were subject to different motions. Various differential flow conditions were studied, always generating good disruption boundaries. A typical example is given in Figure 4. Figure 4a presents a conceptual split field motion as encoded in random dot pairs. One of the typical random-dot patterns containing the coherent regions of optical flow is shown in Figure 4b. The fragmentary edge and motion output resulting from paired-frame analysis of the pair of random-dot patterns is shown at Figure 4c. It will be seen that this output appears to contain very little intelligence. However, a 32 x 32 pixel window analysis produced the vector map in Figure 4d, where two regions of very clearly aligned motion vectors separated by a narrow band of uncertainty can be observed. A replay of the sine and cosine transforms of this map through the B.Ae. edge sensing preprocessor produced the flow disruption boundaries in Figure 4e (identical maps from sine and cosine transforms due to the motion being at 45 degrees). Here we have both the actual flow disruption boundary and the edges of the frame delineated. Also contained in the output data, of course, are the relative strengths of the changes of flow magnitude at the edges of the frame and across the disruption boundary, which may be utilised for recovery of the actual flow pattern, if desired.

E. Focus of Expansion.

Several image sequences have been studied for extraction of focus of expansion estimates, these including both situations where the FOE was well within the field of view and situations where the FOE was a long way outside the field of view. In all cases the rough position of the FOE is very evident from visual inspection of the windowed flow vector maps. A typical example of a motion vector map for an FOE within the field of view is shown in Figure 5b. In this case the sensor was approaching the scene shown in Figure 5a. The flow vectors resulting from window analysis, shown in Figure 5b, are clearly radiating from a point some two-thirds of the way down the image and horizontally central. A focus of expansion computation carried out on these data gave an error on the mean position of +/- 0.12 pixels. The full image had dimensions of 256 x 256 pixels.

A typical example from an image sequence where the sensor was approaching the same scene, but also translating roughly horizontally from left to right is shown in Figure 5c. Here it can clearly be seen that the flow vectors are pointing to a location well outside the field of view. An FOE computation on these data gave the position as (1175.46,202.54) with errors on the means of +/-5.36 & 0.53, measured from the top lefthand corner of the image (i.e. a horizontal location some 3.6 fields of view to the right of the righthand edge of the image). The larger error on the X mean in this case is, of course, inevitable and totally acceptable, since it is the ratio of error to distance outside the field of view which is important in such situations.

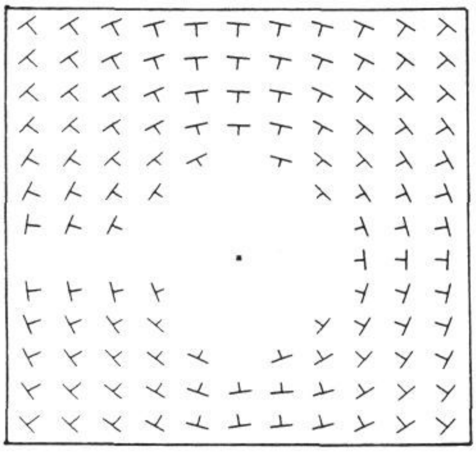
F. Processing of Extended Sequences.

Thus far the discussion has centred entirely around the extraction of instantaneous flow data from pairs of image frames. In any practical situation, of course, there will be a continuous stream of flow data from which one would wish to determine time histories of various flow parameters. We have therefore addressed the problem of automating our complete sequence of paired-frame analyses, such that they can interrogate and extract various flow parameters from streams of video images. It has proved possible to combine the paired-frame analysis version of our basic fragmentary edge and motion sensing program, the windowed flow analysis program and the secondary "edge" and local centroid processing into a single program which can progressively reject one video frame and replace it by the next but one in a sequence, such that the outputs are a time sequence of various types of flow data.

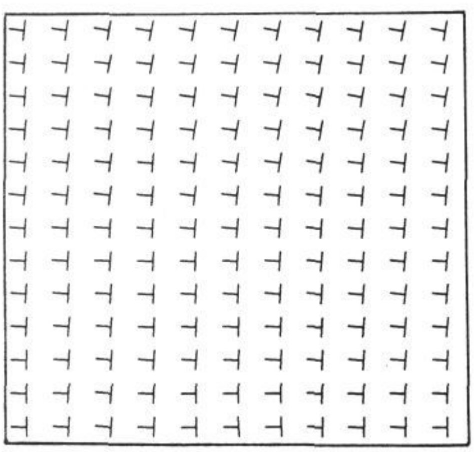
We have processed sequences of frames containing local moving objects and also sequences involving sensor translational motion through this composite program. It is possible to illustrate the time history of motion of local moving objects in terms of sequential estimates of magnitude and direction of computed object motion, whilst it is possible to illustrate directly the time history of FOE computations for sequences involving sensor translation.

In Figure 6a & b are presented the time histories of computed magnitude and direction of motion of the two lorries shown in Figure 1 for a sequence of 15 frames. Since the lorries must be presumed to be moving relatively consistently over the short time interval covered by the frame sequence (some 0.7 seconds) one would expect smooth trends of

a) One input image from a sequence where the sensor was approaching a village scene roughly along the optical axis.

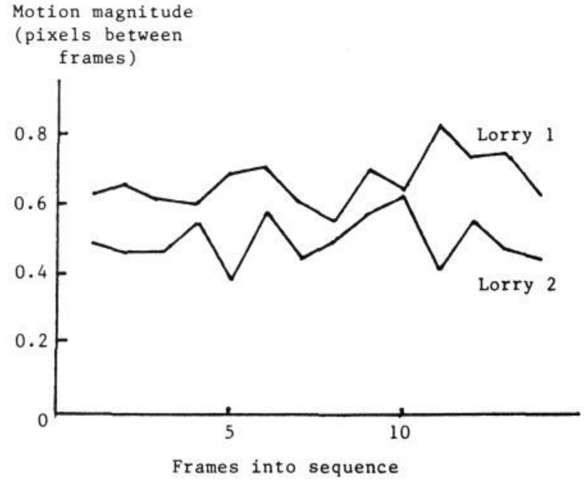


b) Motion vector map from a 64 x 64 pixel window analysis. Motion vectors having a local 3 x 3 orientation uncertainty of greater than 20 degrees are suppressed. The computed focus of expansion is shown by the spot in the central region of the vector map.

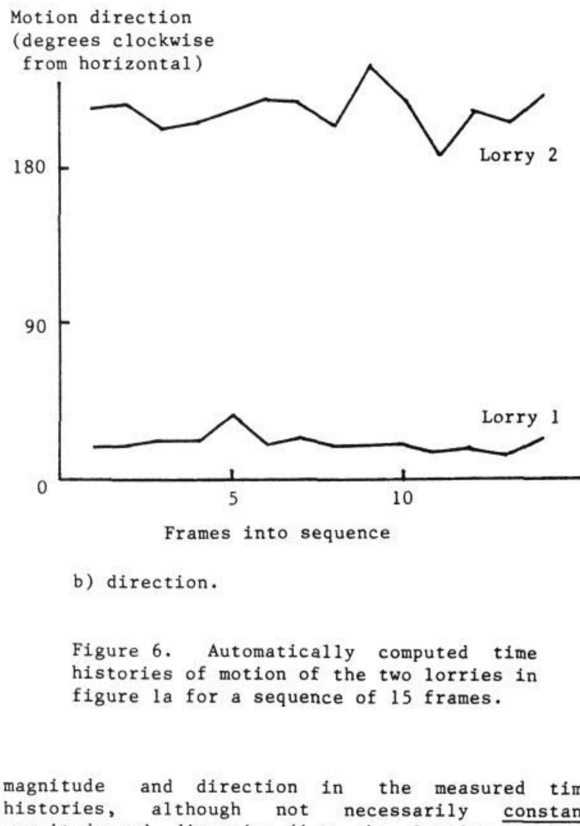


c) Motion vector map from a 64 x 64 pixel window analysis for a situation where the sensor was approaching the village scene of figure 5a with additional large translational motion in X.

Figure 5. Example of sensing of focus of expansion.



a) magnitude.

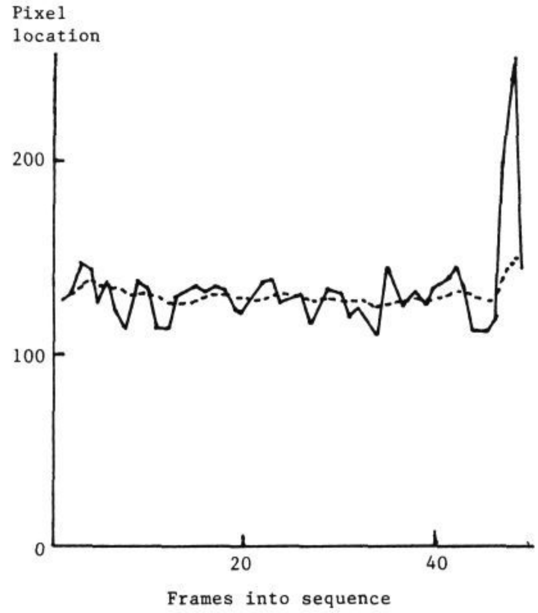


b) direction.

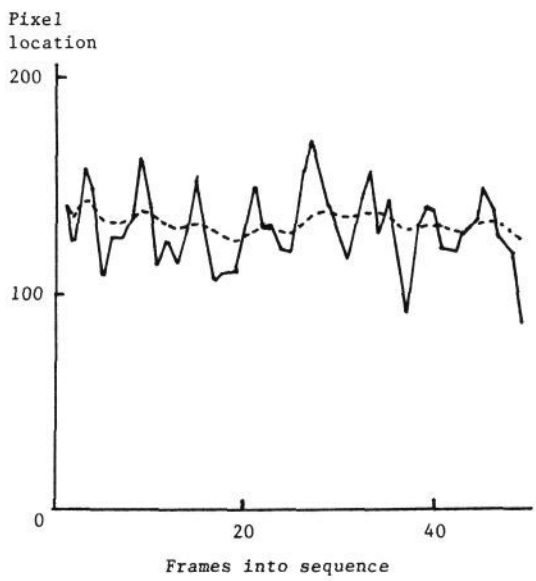
Figure 6. Automatically computed time histories of motion of the two lorries in figure 1a for a sequence of 15 frames.

magnitude and direction in the measured time histories, although not necessarily constant magnitude and direction (i.e. the lorries may have been turning and/or accelerating). Although there is some scatter in the frame to frame outputs, the general consistency is evident.

In Figure 7a & b are presented as solid lines the plots of a time sequence of the X and Y FOE mean values for a series of 50 consecutive frames of the closing sequence of Figure 5b. It will immediately be observed that although there is a satisfying consistency of the trend of these means, there is far more frame to frame scatter than would be expected from the statistical uncertainties on the means as quoted in Section IV.E. We have therefore carefully studied (by visual inspection of pairs of frames presented alternately) the actual movements from frame to frame. It is found that, despite the fact that this image sequence was taken with utmost care using an optical bench, there are significant random lateral and vertical shifts of the image from frame to frame which subjectively tie up with the scatter of the computed FOE's. We therefore have confidence that the majority of the



a) X time history.
——— raw data.
- - - - - - - "smoothed" data with 8 frame exponential decay time constant.



b) Y time history.
——— raw data.
- - - - - - "smoothed" data with 8 frame exponential decay time constant.

Figure 7. Automatically computed time histories of instant to instant focus of expansion coordinates for the sensor approaching the village scene in figure 5a.

scatter in our time plots is genuine frame to frame variations in the input data, and not due to errors introduced during the computations. Thinking about this it has been realised that any practical system (including the human eye/head) will be subject to such instabilities. One should therefore have some form of temporal smoothing on the output in order to yield a satisfactory time history. In natural systems such a smoothing would most likely be achieved by having an exponential decay with a relatively long decay time constant. We therefore applied such a smoothing function, with a time constant of 8 framing intervals. The resulting time histories of the FOE were as shown by the dotted lines in Figure 7a & b. A similar smoothing would, of course, be expected in perception of local object flow as in Figure 6.

VI CONCLUSIONS

A wide variety of techniques for subsequent processing of fragmentary motion data derived from a paired-frame gradient-based primary motion analysis program have been discussed and demonstrated. It is concluded that it is readily possible to carry out many important facets of motion analysis to a high accuracy on both real scenes and random-dot pairs using simple and direct methods of computation, thereby overcoming a number of commonly encountered problems and limitations.

REFERENCES

Baker K. & Sullivan G.D. (1980), 'Multiple band-pass filters in image processing', IEE Proc., 127(Part E), 173.

Hildreth E.C. (1983), 'The computation of the velocity field', MIT Artificial Intelligence Laboratory Memo. 734.

Julesz B. (1971), 'Foundations of Cyclopean Perception', University of Chicago Press.

Koenderink J.J. & Van Doorn A.J. (1976), 'Local structure of movement parallax of the plane', J. Opt. Soc. Am., 66, 717.

Marr D. (1982), 'Vision', W.H.Freeman & Company, San Francisco..

Marr D. & Poggio T. (1979), 'A computational theory of vision', Proc. R. Soc. Lond. B, 204, 301.

Overington I. (1983), 'Computer simulation of preperceptual processing of form', Proc. of the SPIE, 397, 73.

Overington I. (1984a), 'Simple analysis of local and global optical flow for essentially two-dimensional scenes', B.Ae.D. SRC Report JS10204.

Overington I. (1984b), 'VISIVE - a simulation of early visual processing of form', B.Ae.D. SRC Report JS10192.

Overington I. (1985a), 'Direct measurement of fragmentary velocity and stereo depth', Sensor Review, 5, 214.

Overington I. (1985b), 'A simple, direct practical method of sensing local motion and analysing local optical flow', Proc. of the SPIE, 596.

Overington I. (1985c), 'A paradox - high fidelity from poor image quality', Proceedings of Machine Intelligence '85, IFS Publications, Bedford, U.K..

Overington I. (1987), 'Simple analysis of random-dot stereograms and kinematograms', Proc. of the SPIE, 804.

Overington I. & Greenway P. (1986), 'Theoretical and practical super-resolution performance in simple processing of local form, motion and stereo disparity', Proc. of the SPIE, 728.

Rieger J.H. & Lawton D.T. (1985), 'Processing differential image motion', J. Opt. Soc. Am. A, 2, 354.

Van Doorn A.J. & Koenderink J.J. (1984), 'Spatio-temporal integration in the detection of coherent motion', Vision Research, 24, 47.