Computing Optical Flow From Two Frames of an Image Sequence

P. Anandan

COINS Technical Report 86-16**

April, 1986

Abstract

A framework is presented for the computation of optical flow. In this framework, a hierarchical, parallel matching approach is employed to solve the correspondence problem between image pairs with large displacements of points and independent object motion. The major elements of this framework are (i) the use of spatial frequency channels, (ii) computation of a confidence measure for each displacement vector, and (iii) the use of a smoothness constraint which propagates reliable displacement estimates to their neighboring areas with less reliable estimates.

A clear motivation for this approach is derived by considering the following characteristics of the displacement field computation process: the nature of the input, computational costs, and the requirements on the output. Although some of the aspects of this framework have been incorporated in various techniques described in the motion literature, there has been no unified treatment of all of these which provides a detailed computational framework. This paper also describes an efficient computational algorithm consistent with this framework and provides the results of its application to a set of real images.

** The report is sponsored by Allen R. Hanson and Edward M. Riseman. This research was supported by DARPA under grant N00014-82-K-0464.

1 INTRODUCTION

The problem of obtaining 3-d structure and motion from a sequence of digital images is usually divided into two parts, the computation of *optical flow* followed by its interpretation. Optical flow is usually defined to be the *velocity field* of image points in the case of a continuous stream of images, and as an image *displacement field* in the case of a discrete image-sequence [40]. The interpretation of optical flow [41,2] involves the determination of the camera motion relative to the environment, the identification of any independently moving objects, and the determination of the 3-d structure of the stationary environment and the independently moving objects. This paper describes a framework for the computation of dense reliable displacement fields from a pair of images containing camera motion as well as independently moving objects. An algorithm consistent with the framework and the results of its application to a pair of real images are also included.

1.1 Computing optical flow – approaches and difficulties

Since most practical situations involve a discrete image-sequence, the focus of this paper is the computation of image displacement fields. Given a pair of image frames, this involves determining the correspondence of image events in the two frames; hence, it is usually called the correspondence or the matching problem [4,40].

The most common approaches used to solve the correspondence problem are intensity-correlation techniques and symbolic token matching techniques. Intensity correlation techniques use the intensity values in an area around an image point as a template for finding a match for that point [4,10,36]. These techniques involve simple computations and provide dense displacement fields. They encounter three major difficulties – errors at areas with insufficient local intensity variations for matching, problems due to large displacements, and errors at areas of the first frame that are occluded in the second frame. The first two of these have been addressed by restricting the computation to "interesting points" in the image (see [10]) and by using hierarchical approaches (see

[14,20,46]) respectively. However, the detection of occlusion has so far proved to be more difficult and remains largely unsolved.

Token matching techniques derive a symbolic description of the underlying intensity structure at a point and search for matches on the basis of that description [32,33,41]. If the tokens are rich (i.e., contain a variety of information to make them uniquely identifiable) and are simple to extract this can lead to robust matches and an efficient technique. However, extracting rich tokens is a complex process, and the criteria for selecting useful features (for matching purposes) is currently not well understood. Since these techniques usually constrain the matching problem by reducing the density of tokens, they provide only a sparse displacement field.

There are also a class of techniques called *gradient based* techniques [16,19,23,24,25] which attempt to determine velocity fields from a continuous stream of images. If the image displacements are small – i.e., less than a pixel, these techniques can be used with discrete image-sequences. However, in most practical situations the displacements are much larger than one pixel; hence, these techniques are not applicable.

In summary, although some of the current techniques address specific difficulties in matching, none seems to have combined all of the ideas to provide a class of robust algorithms. Thus far, a coherent and complete computational framework for this process has not emerged. The approach taken here is to examine the goals of the displacement field computation process and based on these, develop a computational framework.

1.2 Goals of the displacement field computation process

The goals of the displacement field computation process are determined by three major factors: the nature of its input, the requirements on its output, and computational efficiency considerations. The input is a pair of images which contain displacements due to motion of the camera as well as of independently moving objects. In typical video sequences, the image displacements of points

are usually larger than one pixel. Large displacements are also necessary in order to obtain robust results from the interpretation process [3,18].

Since the reason to compute the displacement field is to obtain the 3-d structure and motion, a natural place to look for output-requirements is the interpretation process. This process is usually based on an analysis of the geometric properties of displacement fields [2,5,18,35,39,42]. These analyses indicate that it is useful to obtain a dense displacement field with an indication of the reliability of the displacements.

The need for a dense displacement field is noted in [18,35,39] for scenes with arbitrary camera rotation and translation. Adiv [3] derives the same requirement for scenes containing independently moving objects. Other techniques which depend on the differential properties of the displacement field [26,34,42] also need a dense field.

An indication of the reliability of the displacements seems necessary because the interpretation process is highly sensitive to errors in the displacement field (see [3,6]); therefore, eliminating incorrect matches from consideration will make this process robust. In addition, points that are occluded in the second frame should be indicated as such, rather than given false displacement estimates.

Finally, in order to efficiently compute a dense displacement field, the technique should be suitable for an image-pixel parallel implementation. The computation in each parallel unit should be simple and use local image information, so that a simple parallel architecture with local connectivity can be used.

1.3 An overview of the framework

In the approach described here, the input images are decomposed into their spatial frequency components by using a set of *spatial frequency channels*. The images in the low-frequency channels provide rough displacement estimates over a large range, whereas the images in the high-frequency

channels provide more accurate estimates. A confidence measure is determined for each displacement vector simultaneously with the displacement computation. A smoothness constraint is then employed to propagate the displacement estimates at high confidence areas and modify those at low confidence areas. Thus, the major elements of the framework are, (i) spatial frequency channels, (ii) the confidence measure and (iii) the smoothness constraint. All computations within each channel are image pixel parallel and based on local information.

In the remainder of this paper, this framework is described in detail and the outline of an algorithm consistent with it is provided. Results of applying this algorithm to a pair of real images is also included. Finally, the limitations of this framework are discussed and directions for future research are outlined.

2 THE COMPUTATIONAL FRAMEWORK

This section contains a description of a framework for the computation of dense displacement fields. The framework is general in the sense that there are a variety of ways of implementing its various components. What is attempted here is to identify those components that must be contained in any algorithm intended to solve the correspondence problem. As noted before, all computations within the framework are suitable for parallel implementation.

The framework is explained by providing a detailed description of its three major elements that were identified in section 1. This is followed by a discussion of its relationship to some other approaches described in the literature.

2.1 Spatial frequency channels

Perhaps the most fundamental idea in the framework is processing based on the spatial frequency decomposition of the image. Such a decomposition can be achieved by using a set of *spatial frequency channels*. Each channel uses information in a specific range of spatial frequencies of the

input images.

The motivation for this decomposition arises from the need to obtain dense and accurate displacement estimates over a large range. Although small local image structures can accurately determine the displacements over a short range, they cannot be used for matching over large range because these structures may repeat. This leads to ambiguities in matching. Therefore, it is clear that in order to process large displacements, large image structures must be used to avoid duplicate matches. However, when matching is based on large spatial structures, there will be a significant overlap of the image areas that determine these structures for neighboring pixels. Therefore, a dense displacement field computed on this basis will vary slowly over the images – i.e., the accuracy of the displacements will be low.

These observations suggest that the structures at different scales should be separated and each used to measure displacements to suitable accuracy over a suitable range. The spatial frequency channels provide a convenient way of achieving such a separation and displacement computation. Similar considerations lead Marr and Poggio [27,29] to suggest a spatial frequency decomposition of the image for stereopsis. There also appears to be psychophysical evidence [1,17,45] to support the existence of spatial frequency channels in the human visual system.

Given that the images are decomposed using the spatial frequency channels, there are three stages to the computation of a displacement field: the spatial frequency decomposition, the matching process within each channel, and a control strategy for combining displacement information from the different channels. The general criteria for designing these stages are discussed below.

Spatial frequency decomposition – The spatial frequency decomposition involves the design of a set of filters and a scheme for representing the output of these filters. The primary requirements of this process are, (i) the filters should be suitable for parallel implementation with computations based on local spatial image information, and (ii) the support regions of the filters in the frequency domain should be narrow and together they should cover the range of image frequencies. The

family of $\nabla^2 G$ filters meet both these criteria and have been studied extensively in computer vision [28,29]. The related family of difference-of-Gaussians (DOG) filters also meet these requirements and can be implemented more efficiently [28]. Psychophysical studies [45] indicate that the spatial frequency channels in the human visual system are similar to the DOG filters.

Match criterion – The need to maintain a simple and uniform processing scheme suggests that the computations within the different channels should be similar to each other. Therefore, it is sufficient to describe the matching process within a single channel. The match criterion within a channel involves choosing either a type of correlation or a type of symbolic token matching scheme – this framework allows both. The primary considerations are efficient computation and generation of a dense displacement field.

Control strategy – The control strategy determines the way in which the displacement information from the different spatial-frequency channels is recombined. For this, a *spectral continuity* constraint (similar to the one suggested for stereopsis in [21,30,43]) should be used. This constraint is described below.

Usually, it can be assumed that the projections of points on different environmental surfaces do not overlap in the image. Hence, the displacement estimates at corresponding image locations in the different channels are due to relative motion between the camera and the same surface; therefore, they must be similar. This means that at any image location, among the duplicate displacement estimates computed in the high frequency channels, the correct one is that which is consistent with the estimates from the low-frequency channel at the corresponding image location. The assumption of no overlap is violated in the case of transparent and fence-like surfaces. At present, this framework does not address such situations.

2.2 Confidence measures

It was noted in section 2.2 that there are usually areas in images with insufficient local information for determining displacements. This is also true in matching by spatial frequency channels. In order to identify such areas, a confidence measure for the displacement estimate should be computed for each pixel. This section describes the properties required of such a measure.

1. The confidence measure should be direction sensitive to indicate the reliability of the different directional components of a displacement vector. The motivation for this stems from the fact that the local structure of an image can be broadly classified into 3 types: homogeneous areas, points along straight edges or lines, and points of high curvature along contours. The matching process will show a different type of performance in each of these situations.

   In a homogeneous area of the image, the displacement estimate will be unreliable in all directions. At a point along an edge the displacement component perpendicular the orientation of edge will be reliable, whereas the component parallel to the edge will not be reliable. Finally, at a point of high curvature, the displacement estimate will be reliable in all directions. In order to discriminate between these situations, a direction sensitive vector measure is needed.

2. The confidence measure should be low at areas of the first image that occluded in the second image. This places a constraint on when the confidence measure can be computed. Since the information from both images is needed recognize occlusion, it cannot be computed before the matching process. However, a tentative measure can be computed before matching and completed during the matching process.

The algorithm described in section 3 provides a vector valued confidence measure that meets the above requirements. These confidence measures are not only useful for the interpretation process, but are also important for using the smoothness constraint described in the next section. In

addition, when the spectral continuity constraint is applied, these measures can be used to select the most reliable estimates from different channels.

2.3 Smoothness constraint

When an area of the image lacks sufficient local structure for the unique determination of its displacement, the correspondence problem can be regarded as being ill-posed. A traditional approach to solve such problems is to introduce a *regularization principle* which makes the problem well-posed [31]. The regularization constraint usually concerns some global property of the quantities being computed. In this framework, such a constraint can be used separately within each channel to enhance the displacement field computation in that channel.

In the case of the correspondence problem, the regularization principle that has been traditionally used [23,24] is based on the assumption that the environmental surfaces can be regarded to be smooth almost everywhere. This leads to a smoothness constraint on the displacement field [24]. Similar smoothness assumptions can also be found in [10,32].

In its most rigorous form, the smoothness assumption leads to a variational problem [24,38]. This can be solved using a local relaxation algorithm, which can be implemented in parallel. However, the smoothness assumption is violated at flow discontinuity boundaries; hence, the constraint should not be applied across such boundaries. Possible approaches to detecting discontinuities during the smoothing process have been discussed by Terzopoulos [37]. However, it is often necessary to have some *a priori* indication of their location.

A related assumption is that the environmental surfaces are continuous almost everywhere. This leads to a constraint (called the *flow-analyticity* constraint) that the displacement field is an analytic function of image coordinates almost everywhere. Although this constraint has been used in [42] for grouping and segmenting optical flow, it has not been used for obtaining a dense displacement field. A similar approach, called the *figural continuity constraint* is often used in

stereopsis ([21,30]).

The variational formulation of the smoothness constraint has been chosen for this framework. The use of this constraint requires four things: a determination of where the local computation is not reliable (this can be provided by the confidence measures), a precise formulation of the constraint, a method for detecting discontinuities, and an algorithm that uses all of these to obtain a dense and more reliable displacement field.

2.4 Discussion

The foregoing sections described each of the major elements of the framework and how they fit together. This section discusses related approaches to stereopsis and motion analysis, and addresses some possible criticisms of this framework.

A number of similarities between this framework and the computational model for stereopsis that has evolved starting from Marr and Poggio's original theory [21,27,30,43] were noted above. In fact, the stereopsis model has been a source of inspiration for the development of this framework. However, besides the difference between the problems of interest (stereo-correspondence can be viewed as a restricted case of motion-correspondence), there are also other major distinctions between the two models. The Marr and Poggio stereopsis model uses the zero crossing contours of the band-pass filtered images as the basis for matching; hence the spectral continuity and the figural continuity constraints are specific to contour-based approaches. The framework presented here is more general and is therefore applicable to both correlation and token based matching approaches. This increased generality also means that the specific form of the smoothness constraint and the spectral continuity constraint will vary depending on the type of matching chosen.

The approaches described in [14,20,46] for displacement field computation use spatial frequency channels and include parallel computation based on local information and hence bear some similarity to the framework described above. The major differences are the following: Wong and Hall

[46] perform correlation-matching based on low-pass filters whereas this framework uses band-pass filters. Low-pass filtered correlation is usually less reliable as a match measure (see [13] for a discussion). Burt, Yen, and Xu [14] use band-pass filters, but do not incorporate the spectral continuity constraint. Glazer, Reynolds, and Anandan [20] use the same framework, however they as well as the others do not include a smoothness constraint or provide any confidence measures.

A major concern is whether spatial frequency decomposition is the appropriate way of separating large and small scale spatial structures. The distortion caused by the filtering process (e.g., rounded corners, merged contours), in the low frequency channels does not create a serious problem for motion analysis for the following reason. In order to correctly match image events, what is necessary is that these events remain similar between the frames that are matched. This will be the case (barring effects of noise), except near flow discontinuity boundaries. At these boundaries, problems arise because of smoothing across the boundaries. However, these can be significantly reduced by a careful formulation of the strategy used to implement the spectral continuity constraint. An example of such a formulation is given in the algorithm described in the next section.

3 HIERARCHICAL CORRELATION ALGORITHM

An important way of determining the utility of any computational framework is by using an algorithm that is consistent with the framework and by analyzing the performance of that algorithm on real images. One such algorithm is sketched in this section. The algorithm described here utilizes a pyramid representation of the information [47] and uses a hierarchical coarse-to-fine matching strategy. It is ideally suited for a processing cone type architecture [22]. A related algorithm suited for a mesh-connected computer is described in [44].

Since the focus of this paper is on the computational framework, the algorithm is only briefly described. It is convenient to divide the algorithm into five components, each of which is derived from one or more of the aspects described above. A detailed description of these components can

be found in [7,8,12,20] and a complete description of the algorithm can be found in [9].

The description of the algorithm is augmented by a demonstration of its application to the pair of real images shown in figures 1 and 2. These images are two frames from the *road scene* image sequence available at the University of Massachusetts Vision laboratory. The images are at 128 × 128 pixel resolution and the movement between them is a pure camera translation. The image displacements range from 0 to 8 pixels.

3.1 Spatial-frequency channels

A suitable method for the spatial frequency decomposition is provided by the *Laplacian-pyramid transform* proposed by Burt [12]. Briefly, the input to this process is a single digitized image at some resolution (usually a power of two, so $2^l \times 2^l$). The output is a set of images at resolutions $2^i \times 2^i, i = 0, \ldots, l$, represented in a pyramid data structure. There are two stages to this computation: the creation of the *Gaussian pyramid* and the creation of the *Laplacian pyramid*. The Gaussian pyramid is computed by successive 5×5 convolutions of image at one level combined with a reduction in the resolution to obtain the image at the adjacent coarser level. This is equivalent to a set of Gaussian convolutions whose standard deviations are successively doubled, while the corresponding image-resolutions are halved (according to the Nyquist criterion). The Laplacian pyramid is computed as the difference of the images from the adjacent levels of the Gaussian pyramid. Each output image can be regarded as the result of a difference-of-Gaussian (DOG) filter applied to the input image. In the matching algorithm described here, one pyramid is created corresponding to each of the two frames being matched. Figures 3 and 4 show the images from four successive levels of the Gaussian and Laplacian pyramids computed from the image in figure 1.

3.2 Match primitive

The matching within each channel is based on a type of intensity correlation. A 5×5 sample window is chosen around a pixel in the first image, and the intensities in it are compared with the corresponding intensities in same size windows around every possible match candidate in the second frame. These candidates are determined according to the control strategy described in the next section. The *sum of the squared differences* (SSD) of the corresponding pixels is chosen as a measure of the match between the two pixels. This was chosen primarily because it is easy to compute and is never negative – a fact that is used in normalizing the confidence measures described in section 3.4. The details of the rationale are discussed in detail in [7]. Among the various candidates in the second image, the one that minimizes this measure is chosen as the match (within a pixel accuracy).

3.3 Control strategy

A sequential coarse-to-fine strategy is employed; this is explained in greater detail in [20] and [7]. The processing begins at a sufficiently coarse resolution where the displacements are within one pixel at that resolution. At this level, the candidate matches are restricted to pixels in a 3×3 area centered around the corresponding pixel location in the second image.

The process proceeds sequentially from coarse to fine levels. At all levels except the coarsest level, an initial set of displacements for a pixel are obtained by projecting the displacements from the adjacent coarser level. The overlapped pyramid projection scheme described by Burt in [11]. was used in order reduce errors due to smoothing across flow discontinuity boundaries (see [7]). In this scheme, each pixel at a finer level image has four potential parent pixels at the coarse level. All the estimates from the four parents are potentially correct initial match estimates. The union of the pixels in the four 3×3 areas centered around each of the potential initial estimates are considered the candidate matches.

Figure 5 shows the displacement fields computed at different levels of the pyramid for the road

scene images. The processing began at the 16×16 resolution. Figure 6 shows in greater detail the final results computed at the resolution of the input images. In all cases, only a subset of the displacements have been shown in order to enhance visibility.

3.4 Confidence measures

The confidence measure chosen here is based on the shape of the *SSD surface* around the best match location. For each pixel in the first image, this surface is defined as follows: its height corresponding to any displacement is the SSD value corresponding to that displacement. Obviously, such a discrete set of values do not define a unique surface. A local quadratic fit is made to the SSD values corresponding to the nine pixels in the 3×3 area around the best match location .

An empirical study of the SSD surfaces [8] suggests that this surface reflects the local intensity structure of the underlying image. The *principal curvatures* [15] of this surface are used to determine the confidence measures. The components of the confidence measures (called c_{max} and c_{min}) along the two principal axes are normalized values of the corresponding principal curvatures. A detailed description of the vector measure and different types of normalization can be found in [8]. A qualitative consideration of these measures indicates that they satisfy the directional sensitivity properties described in section 3.2. The sensitivity to occlusion is achieved as a part of the normalization process.

Figure 7 displays c_{max} and c_{min} at the two finest levels of the pyramid. The unit vectors superimposed on c_{max} indicate the direction of the maximum principal axes. Note that usually at points along straight edges in the image, these vectors are perpendicular to the edge.

3.5 Smoothness constraint

The smoothness assumption used in this algorithm is similar that employed by Horn and Schunck [24]. The problem of finding a smooth displacement field which approximates the estimated dis-

placements on a discrete subset is formulated as a variational problem. That is, the goal is to find a vector field $U = (u, v)$ which minimizes a quadratic functional $E(U)$ where $E(U) = E_{smooth} + E_{approx}$. The functional E_{smooth} measures the spatial variation of U. The functional E_{approx} measures how well U approximates the initial displacement estimates D given at a set of grid points (x, y). The formulation of E_{approx} uses the confidence measures described above.

The mathematical form of the minimization problem, its relationship to other techniques, and the necessary and sufficient conditions for the existence of a solution are discussed in [8]. In particular, it is shown that this formulation is a generalization of the Horn and Schunck formulation and takes into account more local information wherever available. The solution to the functional minimization problem can be achieved by using the finite-element approach (again, see [8] for details). This approach is based on the one used by Terzopoulos [37] for stereopsis.

The embedding of the smoothness constraint in the hierarchical framework implies that it is not necessary to achieve perfect smoothing (or convergence of the relaxation process) at each level. As long as the coarse level estimates are modified to have approximately the correct magnitude and orientation, the refinement at the finer levels can make the necessary adjustments. This means that at each level, a small number of iterations (usually less than 10) of the relaxation process are sufficient.

The results of applying this smoothness constraint to the road scene images can be seen in figures 8 and 9. In particular, note the improvement of the displacement estimates along the white line on the road and in the homogeneous areas of the sky and the road.

4 SUMMARY AND FUTURE RESEARCH

A computational framework for the determination of dense displacement fields and an algorithm consistent with that framework have been described. There are three major issues that have not been fully addressed within this framework: detection of discontinuities, effects of spatial smoothing

of the image across discontinuities, and processing multiple (i.e., more than two) frames. These form the basis for future research directions.

Although the finite element method provides natural ways of incorporating known flow discontinuities (see [37]), the current algorithm does not include any method for identifying them. Possible ways of doing this are described in [7,23,37]. An additional constraint that can be used is the "continuity of discontinuities" assumption suggested by Marr [29].

As noted in section 2, the spatial frequency filtering process can lead to smoothing across flow discontinuity boundaries. Although the overlapped projection strategy reduces the matching errors due to this smoothing, there may still be difficulties when the relative movement between a *small* object and the surrounding background area is *large*. A patch of low confidence measures at the location of the small object will usually indicate such an occurrence. However, the current framework does not include explicit means for addressing such situations.

The importance of extending the two-frame matching process to a sequence of frames need not be stressed. Possible approaches for this can be based on a constraint of *temporal coherence of motion*. These include tracking points and refinement of displacements, using temporal displacement continuity constraint (or a minimum acceleration constraint), and using temporally consistent behavior of areas of low-confidence for the detection and tracking of small objects.

ACKNOWLEDGMENTS

The author wishes to acknowledge major contributions from Deborah Strahman, Philip Kahn, and Prof. Allen Hanson towards the organization and clarity of the paper. Thanks are also due to Lance Williams for his comments and for pointing me to Marr's book at the right times, and to Les Kitchen for his comments. Finally, none of this would have been possible but for the strong commitment and continued advice and support of Prof. Edward Riseman.

References

[1] Adelson E. H. and Movshon J. A. The perception of coherent motion in two-dimensional patterns, *ACM Workshop on Motion*, Toronto, Canada, pp. 11-16, 1983.

[2] Adiv G, Determining 3-d motion and structure from optical flows generated by several moving objects, *IEEE T-PAMI*, 7 (4), pp. 384-401, 1985.

[3] Adiv G., Inherent ambiguities in recovering 3-d motion and structure from a noisy flow field, *Proc. CVPR*, pp. 70-77, 1985.

[4] Aggarwal J. K., Davis L. S., and Martin W. N., Correspondence processes in dynamic scene analysis, *Proc. IEEE*, Vol. 69, No. 5, pp. 562-572, 1981.

[5] Aggarwal J. K., Structure and motion from images, *DARPA IU Workshop Proc.*, pp. 89-95, 1985.

[6] Aggarwal J. K., Structure and motion from images: fact and fiction, *Proc. of the third Workshop on Computer Vision: Representation and Control*, pp. 127-128, 1985.

[7] Anandan P., Computing dense displacement fields with confidence measures in scenes containing occlusion, *SPIE Intelligent Robots and Computer Vision Conference*, Vol. 521, pp 184-194, 1984, also *COINS Technical Report 84-32*, University of Massachusetts, December 1984.

[8] Anandan P. and Weiss R., Introducing a smoothness constraint in a matching approach for the computation of displacement fields, *DARPA IU Workshop Proc.*, pp. 186-196, 1985.

[9] Anandan P., The measurement of image motion, *Ph. D. dissertation*, Computer and Information Science Dept., University of Massachusetts, Amherst, Ma., *in preparation*.

[10] Barnard S. T. and Thompson W. B., Disparity analysis of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-2, Number 4, July 1980, pp. 333-340.

[11] Burt P. J., Hong T. H and Rosenfeld A., Image segmentation and region property computation by cooperative hierarchical computation, *IEEE Trans. Systems, Man, Cybernetics* 11, 1981, pp. 802-809.

[12] Burt P. J., Fast filter transforms for image processing, *Computer graphics and image processing*, 16, pp. 20-51, 1981.

[13] Burt P. J., Yen C. and Xu X., Local correlation measures for motion analysis: A comparative study, *IEEE Proc. PRIP*, pp. 269-274, 1982.

[14] Burt P. J., Yen C. and Xu X., Multi-resolution flow-through motion analysis, *IEEE CVPR Conference Proceedings*, June 1983, pp. 246-252.

[15] do Carmo M. P., *Differential geometry of curves and surfaces*, Prentice-Hall, New Jersey, 1976.

[16] Cornelius N. and Kanade T., Adapting optical flow to measure object motion in reflectance and X-ray image sequences, *Proc. ACM Siggraph/Sigart Interdisciplinary workshop on motion*, Toronto, Canada, pp. 50-58, 1983.

[17] Daugman J. D., Spatial visual channels in the Fourier plane, *Vision Research*, Vol. 24, No. 9, pp. 891-910, 1984.

[18] Fang J. and Huang T. S., Some experiments on estimating the 3-d motion parameters of a rigid body from two consecutive image frames, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, NO. 5, pp 545-554, September, 1984.

[19] Fennema C. L. and Thompson W. B., Velocity determination in scenes containing several moving objects, *Computer Graphics and Image Processing*, 9, pp 301-315, 1979.

[20] Glazer F., Reynolds G. and Anandan P., Scene matching by hierarchical correlation, *IEEE CVPR conference*, June 1983, pp. 432-441.

[21] Grimson W. E. L., Computational experiments with a feature based stereo algorithm, *IEEE T-PAMI*, Vol. PAMI-7, No. 1, pp. 17-34, 1985.

[22] Hanson A.. R and Riseman E. M., Processing cones: A computational structure for image analysis, in: *Structured computer vision*, Tanimoto S. and Klinger A. (Eds.), Academic Press, New York, 1980.

[23] Hildreth E. C., The measurement of visual motion, *PhD dissertation*, Dept. of Electrical Engineering and Computer Science, MIT, Cambridge, Ma., 1983.

[24] Horn B. K. P., and Schunck B. G., Determining optical flow, *Artificial Intelligence*, vol. 17, pp. 185-203.

[25] Limb J. O. and Murphy J. A., Estimating velocity of moving images in television signals, *Computer Graphics and Image Processing*, Vol. 4, pp. 311-327, 1975.

[26] Longuet-Higgins H. C., and Prazdny K., The interpretation of a moving retinal image, *Proc. Roy. Soc. London, B.*, vol. 208, pp. 385-397, 1980.

[27] Marr D. and Poggio T., A computational theory of human stereo vision, *Proc. Roy. Soc. London, Ser. B*, 204, pp 301-308, 1979.

[28] Marr D. and Hildreth E. C., Theory of edge detection, *Proc. Royal Society of London*, B207, pp. 187-217, 1980.

[29] Marr D., *Vision*, Freeman Press, San Francisco, 1982.

[30] Mayhew J. E. W. and Frisby J. P., Psychophysical and computational studies towards a theory of human stereopsis, *Artificial Intelligence*, Vol. 17, pp. 349-385, 1981.

[31] Poggio T. and Torre V., Ill-posed problems and regularization analysis in early vision, *MIT AI Memo 773*, April, 1984.

[32] Prager J. M. and Arbib M. A. Computing the optic flow: The MATCH algorithm and prediction, *Computer Vision, Graphics, and Image Processing*, 24, pp 271-304, 1983.

[33] Radig B., Kraasch R., and Zack W., Matching symbolic descriptors for 3-d reconstruction of simple moving objects, *Proc. of IEEE ICPR*, Miami, Florida, pp. 1081-1084, 1980.

[34] Rieger J. H. and Lawton D. T., Determining the instantaneous axis of translation from optic flow generated by arbitrary sensor motion, *Proc. ACM Siggraph/Sigart Interdisciplinary Workshop on Motion*, Toronto, pp. 33-41, 1983.

[35] Roach J. W. and Aggarwal J. K., Determining the movement of objects from a sequence of images, *IEEE transactions on Pattern Analysis and Machine Intelligence*, Vol. 2, No. 6, pp. 554-562, 1980.

[36] Rosenfeld A. and Kak A. C., *Digital picture processing*, Academic Press, New York, 1976.

[37] Terzopoulos D., Mutiresolution Computation of Visible-Surface Representations, *Ph. D. Dissertation*, Massachusetts Institute of Technology, Jan. 1984.

[38] Terzopoulos D., Image analysis using multi-resolution methods, *IEEE transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No. 2, pp. 129-139, 1986.

[39] Tsai R. Y., and Huang T. S., Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces, *IEEE T-PAMI*, 6, 1984.

[40] Ullman S., Analysis of visual motion by biological and computer systems, *IEEE computer*, pp. 57-69, 1981.

[41] Ullman S. *The Interpretation of Visual Motion*, The MIT Press, Cambridge, Ma., 1979.

[42] Waxman A. and Wohn K., Contour evaluation, neighbourhood deformation and global image flow: planar surfaces in motion, *CS-TR-1394* University of Maryland, April 1984.

[43] Williams, L., Spectral continuity and eye vergence movement, *Proc. of the ninth IJCAI*, pp. 985-987, 1985.

[44] Williams L.R. and Anandan P., A coarse-to-fine control strategy for stereo and motion on a mesh-connected computer, *to appear in IEEE CVPR Conference*, Miami Beach, Florida, 1986.

[45] Wilson H. R. and Bergen J. R., A four mechanism model for spatial vision, *Vision Research*, 17, pp. 1177-1190, 1979.

[46] Wong R. Y. and Hall E. L., Sequential hierarchical scene matching, *IEEE Trans. on Computers*, Vol. 27, No. 4, pp. 359-366, 1978.

[47] Tanimato S. and Pavlidis T., A hierarchical data-structure for picture processing, *Computer Graphics and Image Processing*, Vol. 4, No. 2, pp. 104-119, 1975.

Figure 1: The first image of the road-scene image pair.

Figure 2: The second image of the road-scene image pair.

Figure 3: The images in four levels of the Gaussian pyramid.

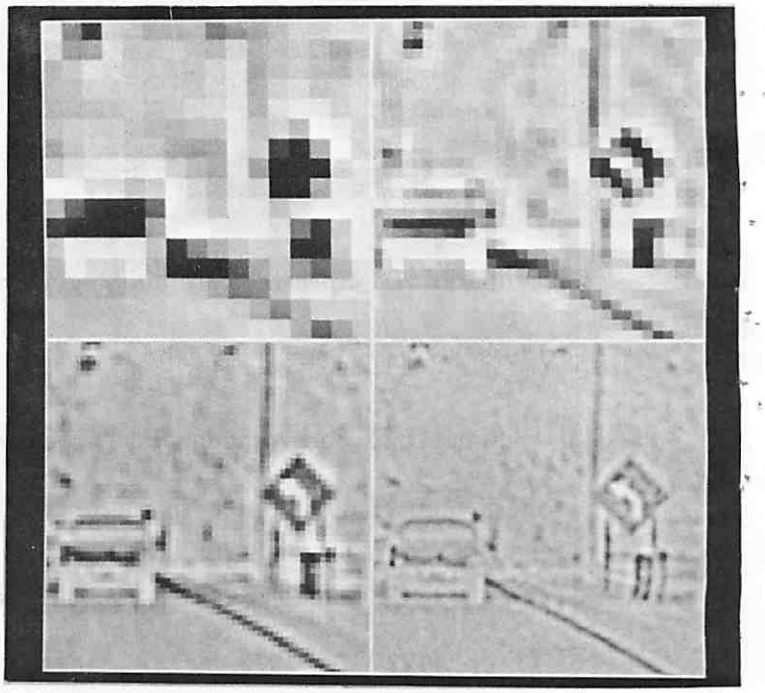

Figure 4: The images in four levels of the Laplacian pyramid.

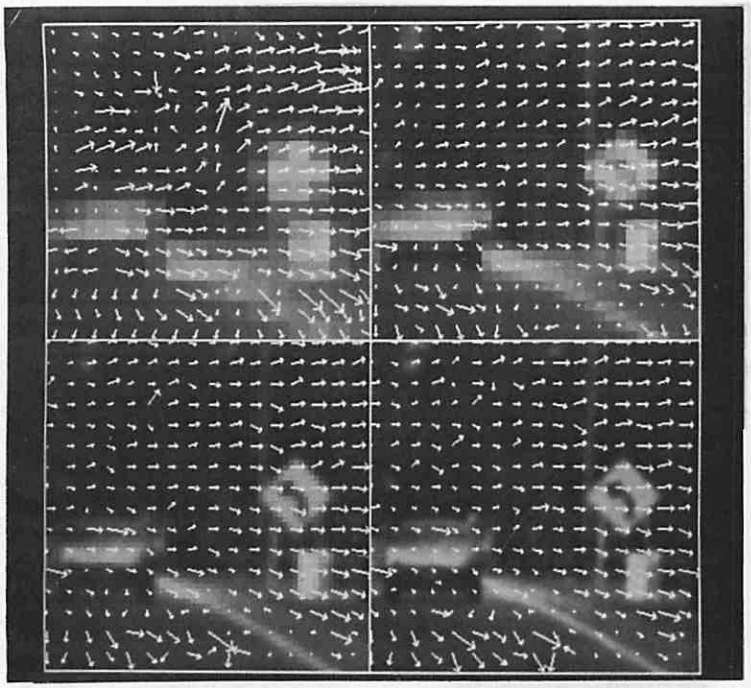

Figure 5: The displacement field at four successive levels. The background images are those from the corresponding levels of the Gaussian pyramid of the first image. The top-left corresponds to level 4 (i.e., 16×16) and the bottom-right image is at level 7 (i.e., 128×128). At all levels, only a subset of 16×16 vectors have been shown in order to enhance visibility.

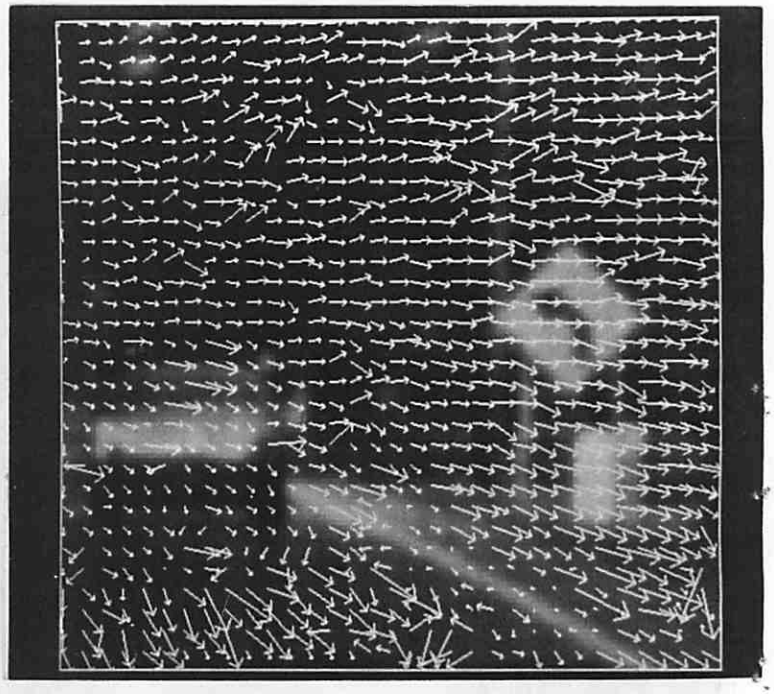

Figure 6: The displacement field computed at the resolution of the input images. Again, only a 32×32 sample of vectors have been shown in order to enhance visibility.

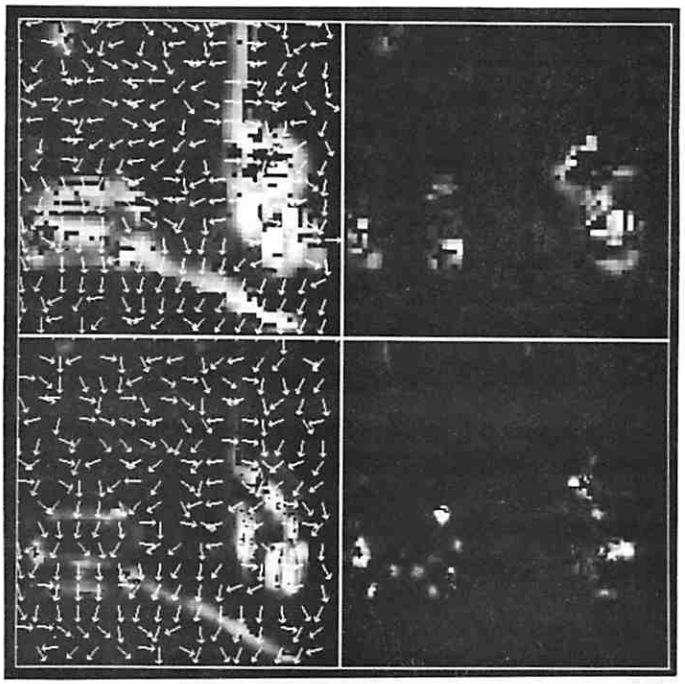

Figure 7: The confidence measures at the two finest levels of processing. The upper two images show the c_{max} (on the left) and c_{min} (on the right) at level 6 (i.e., 64×64) as intensity images. The unit vectors superimposed on c_{max} indicates the direction of e_{max}. The vector e_{min} will always be perpendicular to e_{max} and hence is not shown. Only a 16×16 subset of the unit vectors have been shown in order to enhance visibility. The bottom two images are similar measures computed at the resoution of the input image.

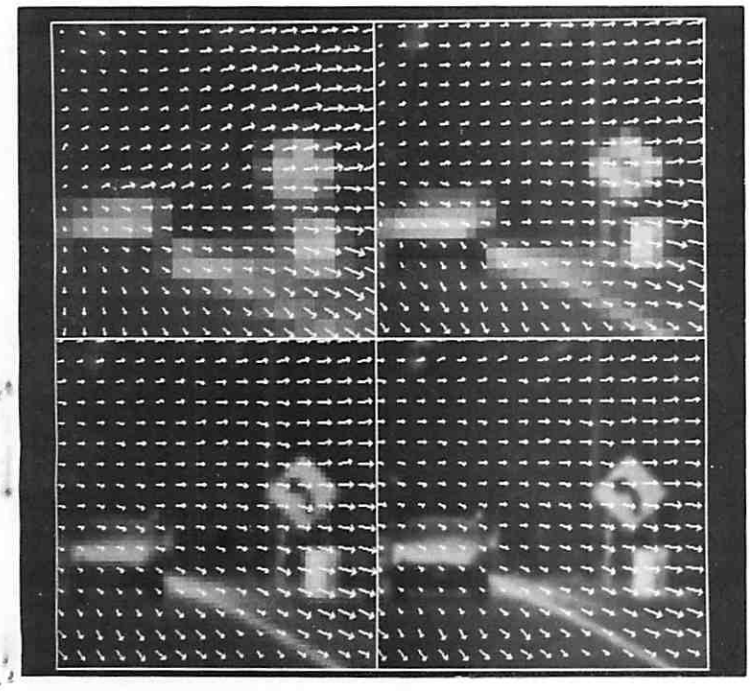

Figure 8: The *smoothed* displacement field at four successive levels. A subset of 16×16 vectors have been shown in order to enhance visibility. Comparing this with the results in figure 5 indicates that the improvements begin at the coarsest level.

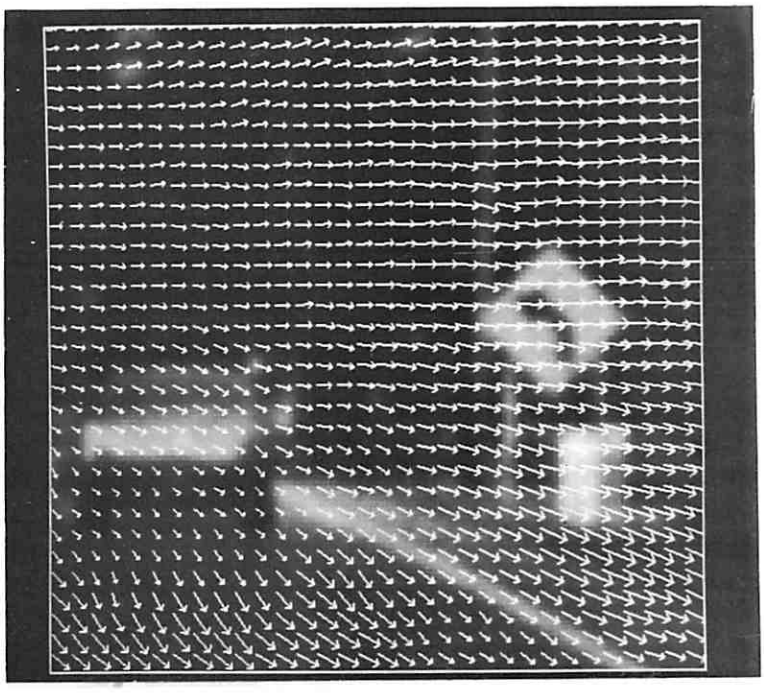

Figure 9: The *smoothed* displacement field computed at the resolution of the input images. Again, only a 32×32 sample of vectors have been shown in order to enhance visibility. Comparing this with the results of the displacement field in figure 6 indicates major improvements along straight lines and homogeneous areas.