Computing Dense Displacement Fields With Confidence Measures In Scenes Containing Occlusion

P. Anandan

COINS Technical Report 84-32*

December 1984

Abstract

Matching successive frames of a dynamic image sequence using area correlation has been studied for many years by researchers in machine vision. Most of these efforts have gone into improving the speed and the accuracy of correlation matching algorithms. Yet, the displacement fields produced by these algorithms are often incorrect in homogeneous areas of the image and in areas which are visible in one frame, but are occluded in the succeeding frames. Further, these displacement fields are often incorrect even at non-occluded areas that border occlusion boundaries. In this paper, we present a confidence measure which indicates the reliability of each displacement vector computed by a specific type of correlation matching algorithm. We also provide an improved matching algorithm which performs particularly well near occlusion boundaries. We demonstrate these with experiments performed on real image sequences taken in our robotics labaratory.

---

* The report is sponsored by Allen R. Hanson and Edward M. Riseman.

1. INTRODUCTION

One of the powerful techniques that have been studied by researchers in image processing and computer vision for the purpose of matching images is area correlation [Agga81a, Bard80, Hann74, Mora81, Genn80, Burt83, Glaz83, Wong78, Lawt84]. Much of this work has addressed issues in choosing a useful match measure, increasing the accuracy of the match, and in reducing the computational complexity of the matching algorithms. However, most of the current techniques produce false matches when applied to scenes containing occlusion, i.e., where a certain area of the image which is visible in one frame is hidden by other moving areas in the succeeding frames.

Researchers have also attempted to match other image features, such as tokens representing prominent image points [Prag79, Yach81], edges and other intensity contours [Agga81b, Tsuj80, Hayn81], regions [Nage77, Radi81, Roac79], and structural descriptions of edges and regions [Jaco80]. These techniques often provide more accuarate results than correlation matching techniques on scenes containing occlusion. However, the selection of such image features is itself a complex problem. In addition these only provide sparse displacement fields, which makes it difficult to use them for techniques which apply differentiation operators to the flow fields for the computation of motion parameters of the camera and the objects in the scene [Thom83, Praz80].

Our aim in this study is to isolate the situations where correlation matching algorithms fail and to find methods to overcome these failures. We first make a list of the major sources of difficulties for the correlation matching techniques.

- The amount of search required can be large. If the image displacements are n pixels, then the simple correlation technique needs to search over an nxn area. However, many researchers have developed techniques that significantly reduce the amount of search [Glaz83, Mora81, Wong78, Burt83]. These techniques use hierarchical search strategies which usually consists of generating sparse approximate displacement estimates and refining them to produce a dense, precise displacement field.

- The spatial *resolution* of the displacement field produced by correlation matching be-

comes poorer as the size of the sample window increases [Gen80]. Here, *resolution* refers to the ability to faithfully represent sharp discontinuities in the actual displacement field. This loss of resolution occurs because the sample window corresponding to neighbouring pixels overlap each other. As the window size increases, the fraction of the overlapped area also increases. Although this may be overcome to some extent by symbolic feature matching, the problem still remains because, when the displacements are large, the use of features based on small local image phenomena can lead to false matches. Therefore, in this case, the symbolic features employed have to be based on large image phenomena, which also leads to loss of resolution in the resulting displacement field.

- Correlation matching tends to produce false matches in areas of the image where the variation is low. Although this problem has been significantly reduced through matching band-pass filtered images [Burt83, Glaz83], false matches are still produced in large homogeneous areas. Further, when applied along a long straight line, these algorithms often provide incorrect estimates of the component of the displacement vector that is parallel to the line.

- Correlation matching fails most miserably at occlusion points. Points that are present in one image and are occluded in the next should have no match; but correlation matching techniques usually do not have any way of recognizing this. In the case of symbolic matching techniques, static segmentation and boundary extraction processes usually attempt to recognize potential occlusion points.

- Correlation matching also has problems matching non-occluded areas that border occluded areas. The image structure in the windows centered around such points changes between the images, thus making it difficult to match them. This can be frustrating for later processes that specifically require reliable matches at occlusion boundaries in order to segment the image into moving areas, etc. Symbolic feature matching may work well at these points.

As we have noted above, some of the hierarchical search strategies are useful in reducing the amount of search required. However, the problems in processing scenes containing

occlusion still remain. In this study, we use the hierarchical matching algorithm of Glazer, Reynolds, and Anandan [Glaz83] as our basis. We isolate the situations where this matching algorithm fails by computing a confidence measure which estimates the reliability of each displacement vector that is computed by the matching algorithm. We then modify the search strategy to improve the results, especially near occlusion boundaries. The result is a computationally efficient matching algorithm which provides a dense displacement field with estimates of reliability of each displacement vector.

Section 2 of this paper describes the various types of correlation techniques that have been investigated by researchers. Section 3 describes some of the work done by other researchers for finding a confidence measure, and describes the measure chosen for this work. Section 4 describes our modifications to the search strategy. Section 5 describes some applications and the possible future directions of this research.

2. TYPES OF CORRELATION MEASURES AND ALGORITHMS

A variety of correlation measures and associated search strategies have been studied by researchers in the field of image-matching. In this section we briefly review some of these measures and some of these search strategies. The particular choice of the measure and the search strategy is not always independent of each other. In our discussion below, we point out such dependencies where they occur.

2.1 Types of Correlation

Some of the typical correlation match measures that are used by researchers are described in [Hann74]. These include (i) direct correlation, in which the image intensity values in the matched windows are multiplied pointwise and summed, (ii) mean normalised correlation, in which the mean of each window is subtracted from the value at each pixel before the multiplication and summing, (iii) variance normalised correlation, in which the correlation sum is divided by the product of the variances of the two images, (iv) sum-of-squared-differences (SSD), in which the sum of square of differences between corresponding pixels is used as the match measure, and (v) sum of magnitude of the differences, which is similar to SSD, but the absolute values of the differences are used instead of their squares.

A comparative study of some of these match measures can be found in [Burt82]. In his study Burt also suggests the use of Laplacian-filtered images for matching. Although this filtering process can be used in combination with any of the above match measures, his study includes only the Laplacian-filtered direct correlation. The other measures included in this study are direct, mean-normalised, and variance-normalised correlations computed using the unfiltered image. Burt shows that the most reliable results are consistently obtained by choosing correlation with both mean and variance normalisations. However, this process (especially, variance normalisation) is computationally expensive. Therefore, he recommends the computationally efficient Laplacian-filtered direct correlation which appears relatively insensitive to both mean and constrast changes between the images, although this measure performs poorly in the presence of high frequency noise.

The reason for the success of Laplacian-filtered correlation process is that the mean

value of a laplacian-filtered image tends toward zero as the sample window size increases. Thus the filtering process has the effect of mean-normalizing the correlation values. However, we found that when the window sizes are smaller than 8×8, the mean of a sample window, though small, was not nearly zero. In such cases using Laplacian-filtered SSD provides more accurate results than Laplacian-filtered correlation. This is because the SSD measure is sensitive to the difference of the means of the two areas that are compared. For smaller windows, although the means are not zero their differences tend to be small. Our own experiments suggest (see Table 1) that for a fixed size of the matching window, the performance of Laplacian-filtered-SSD exceeds that of the Laplacian-filtered correlation using the same sized window.

	WINDOW SIZE					
	3×3		5×5		8×8	
noise %	corr	ssd	corr	ssd	corr	ssd
0	36.93	66.67	64.07	78.26	79.38	82.03
5	30.53	51.80	60.30	74.10	78.86	81.61
10	21.19	30.44	51.73	60.33	75.21	76.67

Table1 : Comparing correlation and SSD

The tests were conducted using the Mandrill images described in [Glaz83]. The second frame was produced by shifting the first frame by 3 rows down and 5 columns to the right and adding Gaussian noise of standard deviation 0, 5, and 10 % of the intensity range of the images. Square correlation windows of width 3, 5, and 8 pixels were used. The table entries show the % of pixels with the correct displacement estimate.

Hannah [Hann74] points out that the performance of SSD degrades in the presence of mean and contrast changes between the images. Although the filtering process may eliminate problems due to mean differences, contrast changes can still be a problem. Another restriction inherent in using the Laplacian-filtered correlation is the fact that the search area must be small in order to ensure accurate matches. This is so because the high pass filtering process removes image variations below a certain frequency, thus causing the image

structures to repeat beyond a distance corresponding to the filter cut-off wavelength. As we note in the following section, the search strategy used can also be helpful in alleviating some of these difficulties.

2.2 Search Strategies

There are only a limited number of search strategies that have been employed by various researchers. The most obvious strategy is to search the whole area within the expected maximum displacement. Hannah [Hann74] and Gennery [Genn80] use this technique, although Gennery mentions a number of ways to cut down the computational cost of the search, and mentions the possibility of using global techniques to get approximate estimates which can then be improved using local searches. Lawton [Lawt84] uses a search technique that is most suitable for the case of pure translational motion of the camera. In this case a global search is performed for the focus of expansion (FOE), which is the intersection of the translational axis and the image plane. Specific values of the FOE are evaluated and for each, the local searches for optimal feature matches are constrained to lie along radial lines emanating from the assumed FOE.

Wong and Hall [Wong78], Glazer et. al. [Glaz83], Burt et. al. [Burt83], and Moravec [Mora81] all use a multi-resolution coarse-fine strategy, but there are important differences among them. Among these differences, we are interested in the fact that Wong and Hall and Moravec used low-pass filtered images, whereas Glazer et. al., and Burt et. al. used band-pass filtered images. Burt et. al. used a "flow-through" strategy where the searches at the different levels of resolution operate independently of each other. This results in low-frequency coarse resolution searches detecting large displacements and higher-frequency finer resolution searches detecting smaller displacements in the image. Glazer et. al. [Glaz83] use a strategy which utilizes the approximate estimate given by the low-frequency, coarse resolution searches as a starting value to define the search in the higher frequency images, thereby resulting in a more precise displacement estimate.

For large displacements, the use of band-pass filtered images and the coarse-fine strategy for matching is a natural generalisation of the Laplacian-filtered matching technique. In this approach, at any given level of resolution the band-pass filtered image corresponds

to Laplacian-filtering the low-pass filtered image that is faithfully representable (according to Nyquist criterion) at that resolution. The 3x3 search area used both by Burt et. al. and Glazer et. al. effectively limits the search to less than half-the wavelength of the highest frequency information available at each level. This restriction also helps reduce the effect of false matches that may arise due to contrast variation between the images.

The band-pass filtered, coarse-fine search strategy tends to introduces some problems of its own. At occlusion boundaries, where there is a discontinuity in the displacement field, the coarse-resolution processing, errors usually occur due to sample windows which overlap across the boundaries and due to effects of the low-pass filtering process which smooths the image across the boundaries. Since each pixel at a coarse level tends to cover a large area at the finest levels, these coarse-level errors tend to cause incorrect initial estimates to be used at the fine-level pixels, thus leading to a search in areas which do not include the correct match. Typically this creates a large area near the occlusion boundary with incorrect displacement field. Since these errors are primarily due to the hierarchical search strategy, it may be possible to eliminate some of these by using a single level search strategy. However, we believe a better approach would be to maintain the coarse-fine search strategy, but try to recognize such errors as they happen. Our confidence measure is, in fact, an attempt to do precisely that.

3. A CONFIDENCE MEASURE

As noted in the previous sections most correlation matching algorithms generate false matches in homogeneous areas, i.e., where there is a lack of any significant image structure, and around occlusion regions. In particular, there has been a noticeable lack of success in dealing with occlusion areas. Most of the small number of researchers who have directly attacked this problem have usually relied on being able to first detect object boundaries (or edges and junctions of edges that may potentially constitute object boundaries) in some manner [Roac79, Thom80]. A slightly more robust approach is to attempt to enhance a static segmentation process with dynamic information [Radi81].

Previous work that has attempted to provide smoothly varying dense displacement fields (Horn80, Glaz81, Nage83, Hild83) usually propagates displacement information from image areas with significant intensity variations to homogeneous areas. When occlusion is also present, it is usually the case that there are significant intensity variations (due to discontinuities in texture) at occlusion boundaries. However, these boundary estimates are often incorrect and using them for initial estimates tends to confound the errors.

The problem of homogeneous areas and occlusion boundaries would not be so serious if there were methods to recognize such errors and suppress the displacement estimates at such points. However, most algorithms simply produce false matches at these points with no way of recognizing them. The primary focus of this study is to provide a confidence measure, which would help in recognising such false matches.

There have been earlier efforts by Hannah [Hann74], Gennery [Genn80] and Burt [Burt83] to understand this problem. Hannah observes that both the sharpness of the correlation surface at the point of best match and the similarity between the shape of the auto-correlation and the cross-correlation surfaces can be used to decide about the reliablility of the match. However, she does not provide any concrete technique for using this information. Gennery's measure requires a model of the camera noise and scaling effects between the images to be matched. This requires calibration of the camera set-up. Although this appears to be a robust measure, it is often the case that such calibration is difficult due to changes in illumination, surface reflectance, etc. We believe that it is

possible to provide a confidence measure which does not depend on an a-priori model of the camera and the image noise. Burt provides a confidence measure, which in many ways is similar to our own. We describe Burt's measure in greater detail in section 3.4 and compare it with ours.

3.1 Properties of the SSD surface

We define an *SSD surface* as the surface formed by considering the Laplacian-filtered SSD values corresponding to different candidate displacements as the elevation at that displacement. This surface appears to contain a wealth of information about the nature of the image structures at the point being matched. Intuitively, it is clear that where there are significant intensity variations in the image, the match is likely to be reliable and unique, whereas at points in a homogeneous area, this is not so. This fact is noticeable in the shape of the SSD surface corresponding to such points. Usually, the SSD surface corresponding to a point with distinct image structure tends to have a sharp valley centered at the best match value, whereas at a homogeneous point the SSD surface is rather flat.

We conducted an empirical study of the behaviour of the *SSD surface*. For this study, We created a pair of synthetic images by digitally "cutting and pasting" pieces from two images photographed in our robotics lab [Elli84]. Gaussian noise of standard deviation of 10 (20 percent of the standard deviation of the intensity values in image2) was added to the second image. These images are displayed in Figures 3.1 and 3.2. The Folgers coffee can in the center part of the image has been displaced by 14 pixels to the right and 4 pixels down. The boundaries of the displaced segment in the two images are shown in Figure 3.3. The occluded area is shown cross-hatched.

A few points in image 3.1 were chosen to illustrate the behaviour of the SSD surface. These points are highlighted in Fig 3.1. Figures 3.4 through 3.8 display the SSD surface at these points. Each of these consists of two surface displays: One of these, referred to as the *auto-SSD surface*, is generated when a 5×5 sample window centered at the point of interest in the first image was matched with similar windows in the same image centered at all points in an 8×8 area around the point of interest. The other, referred to as the *cross-SSD surface*, is generated when the 5×5 sample window centered at the point of

Figure 3.1 The first frame of the synthetic image pair used to illustrate the behaviour of the SSD surfaces.

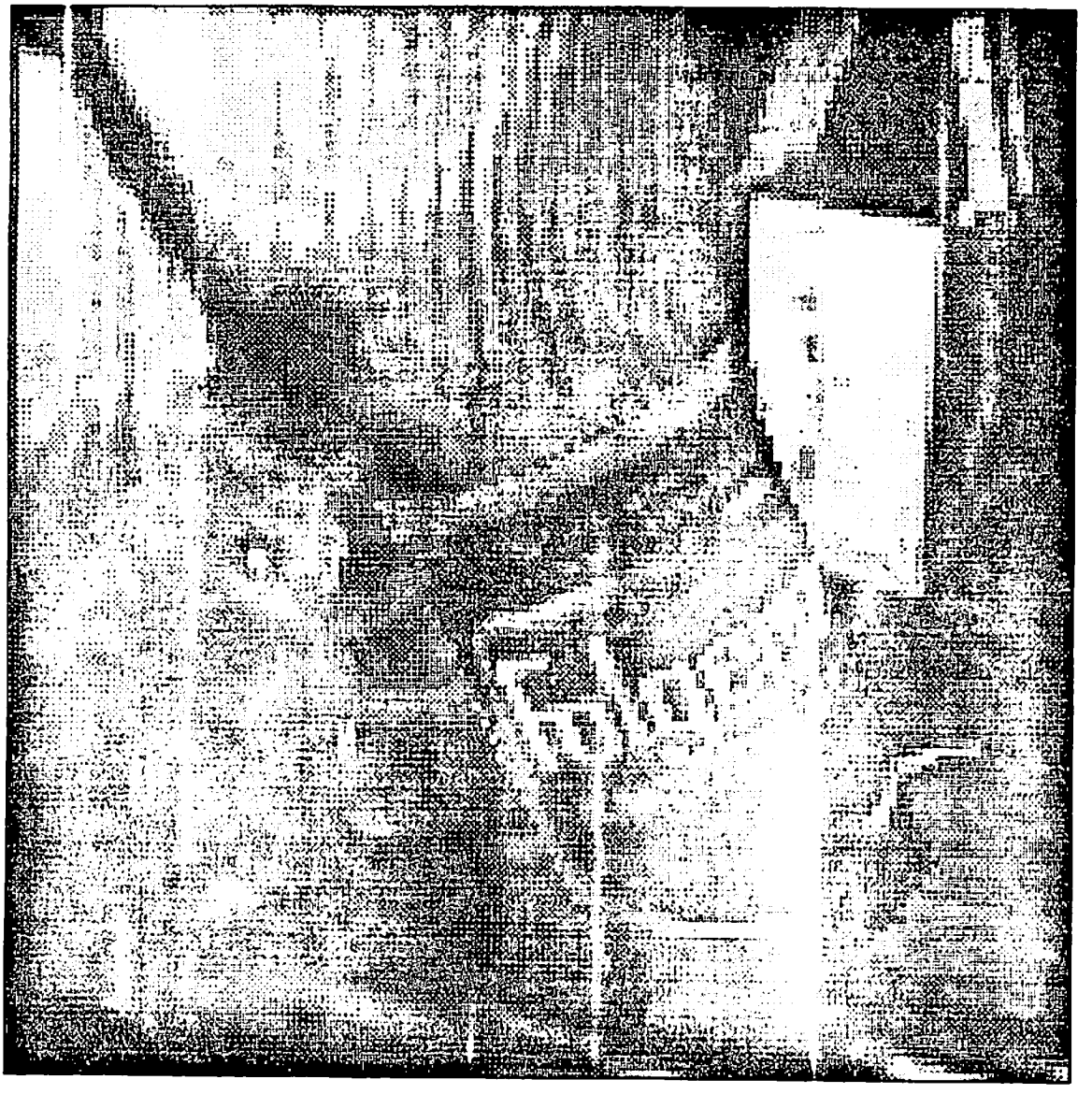


Figure 3.2 The second frame of the synthetic image pair.

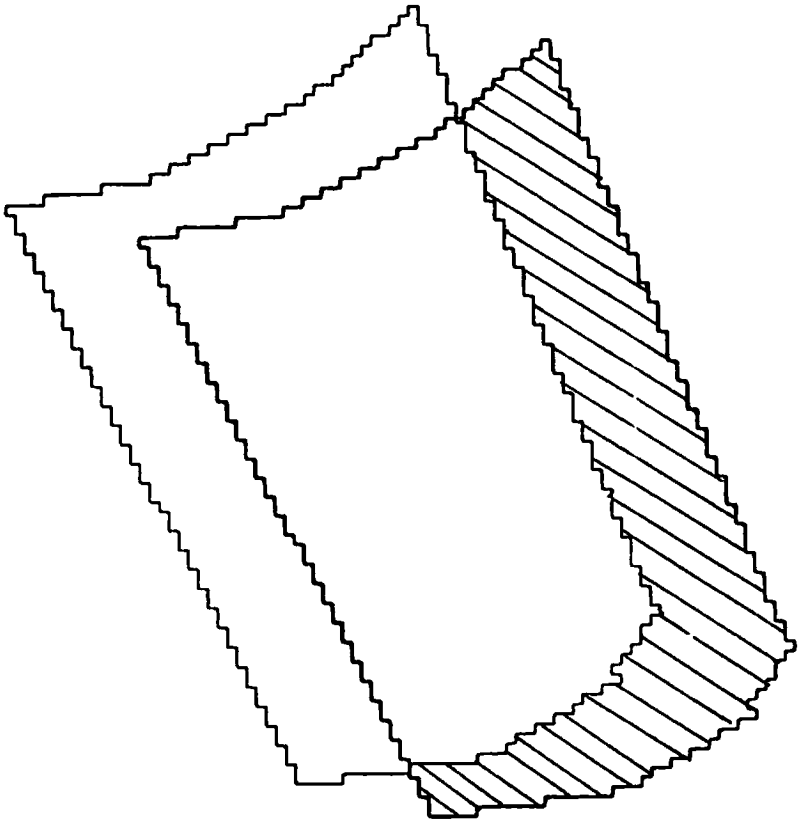


Figure 3.3 The boundaries of the displacement segment in the two Figs. 3.1 and 3.2.

interest is matched with similar windows centered at all points in an 8×8 area around the correct-match point in the second image. In both cases the match measure was the Laplacian-filtered SSD performed at the same level of resolution as the image. Note that for occlusion regions, there is no true match point in the second image; however, we have used the true displacement estimate of the background surface of which they are a part (in our case this is zero displacement, since the background is stationary).

In order to enhance visibility, the surfaces are shown inverted, i.e., the minimum SSD values are shown as peaks rather than as valleys. In each surface display, we have marked the true-match point with an "X" and the point of minimum SSD value with an "O". Also marked are the minimum and maximum SSD values on each of the surfaces, as well as the view-angle (rotation and elevation). All the surface displays are generated by a perspective projection of the surfaces.

Corner point Fig 3.4a and Fig 3.4b displays the values of the *auto-* and *cross-* SSD surfaces at an intensity corner. These correspond to the point a in Fig 3.1. Note the sharp peak in both the *auto-* and *cross- SSD* surfaces (in actuality a sharp valley since these displays have been turned upside-down) centered at the true displacement value. This indicates that the match is reliable in all directions. Further, note that the shape of the surface is well preserved even though the *cross-SSD* surface was generated using images containing significant amount of noise.

Along a Straight Edge Figures 3.5a and 3.5b illustrate the two surfaces at a point along a straight edge in the image (point b in Fig 3.1). A ridge like structure along the direction of the edge is clearly visible in both the surfaces. However, note that the peak of the *cross-SSD* surface is shifted away from the correct match-point (i.e., the center of the surface) along the ridge. This indicates that the match estimate is reliable only in the direction perpendicular to the ridge (or the edge) and that we have no reliable information parallel to the ridge.

Homogeneous point Figures 3.6a and 3.6b illustrate the SSD surfaces at a homogeneous point (point c in Fig. 3.1). In this case the SSD surface is rather flat, especially around the center, i.e., the point of best match. Again, the peak of the *cross-SSD* surfrace does not

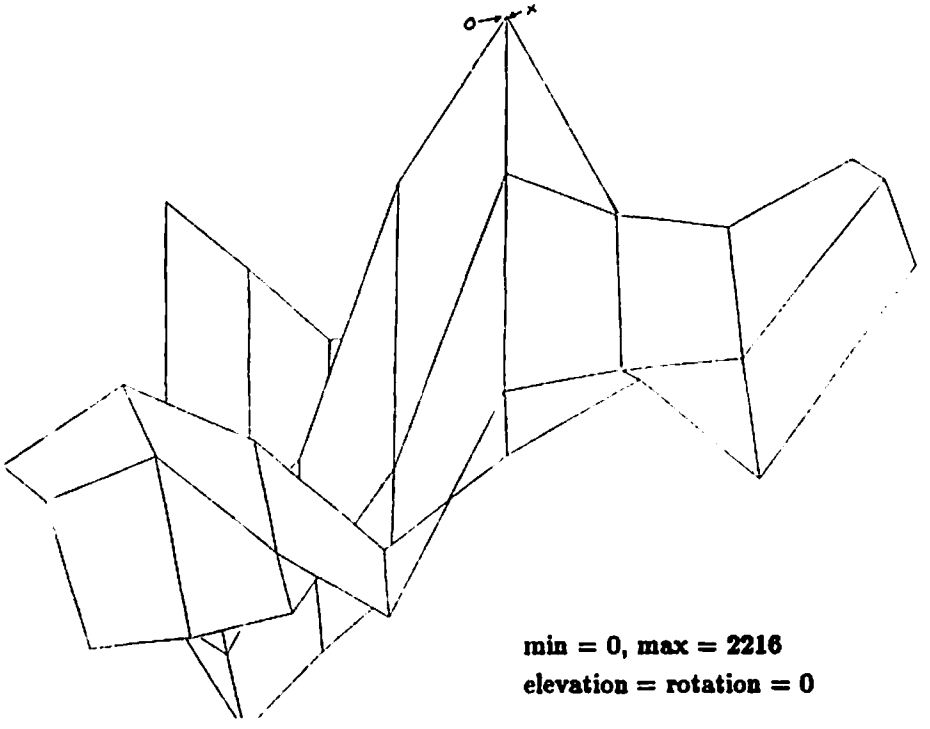


Figure 3.4a The auto-SSD surface at a corner point (point a in the first frame).

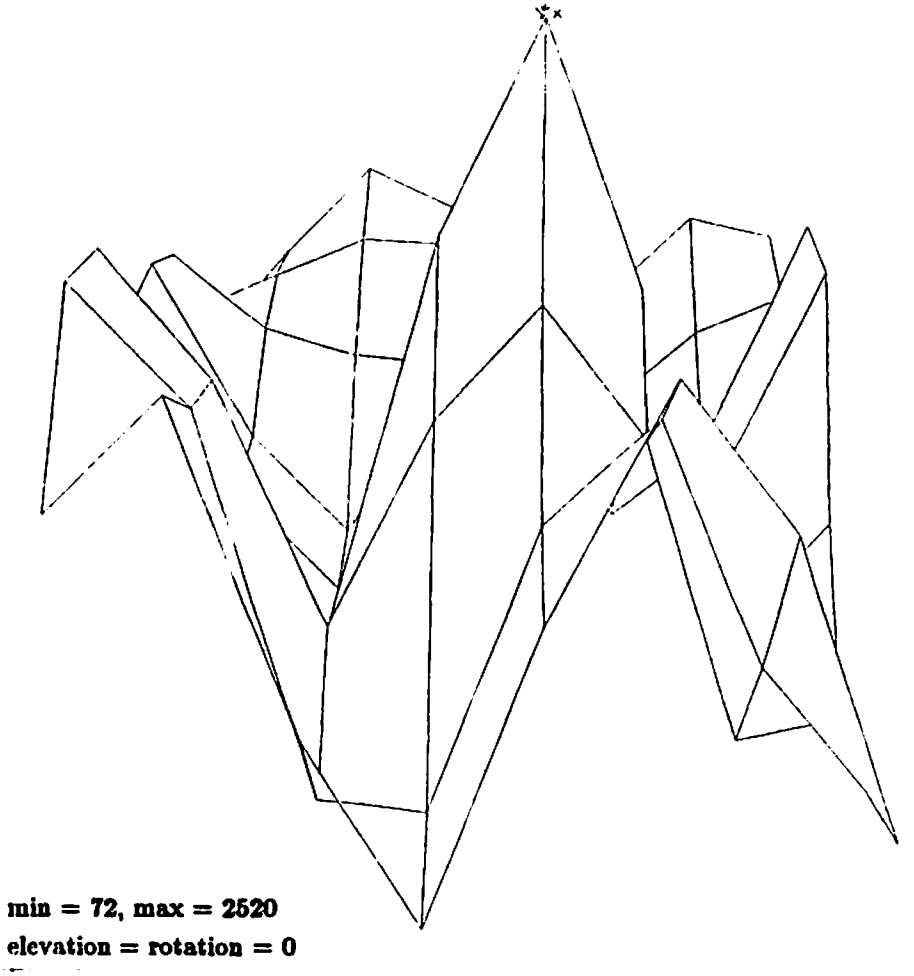


Figure 3.4b The cross-SSD surface at point a.

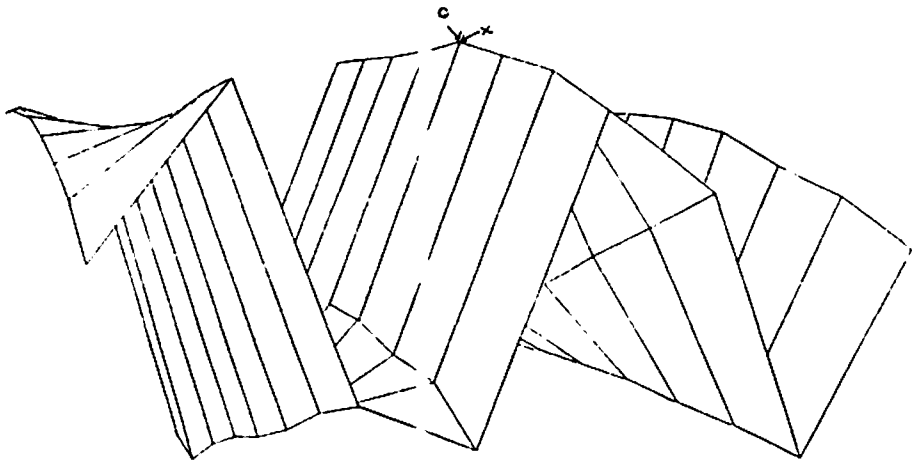


Figure 3.5a The auto-SSD surface at a point along a vertical edge. (point *b* in the first frame).

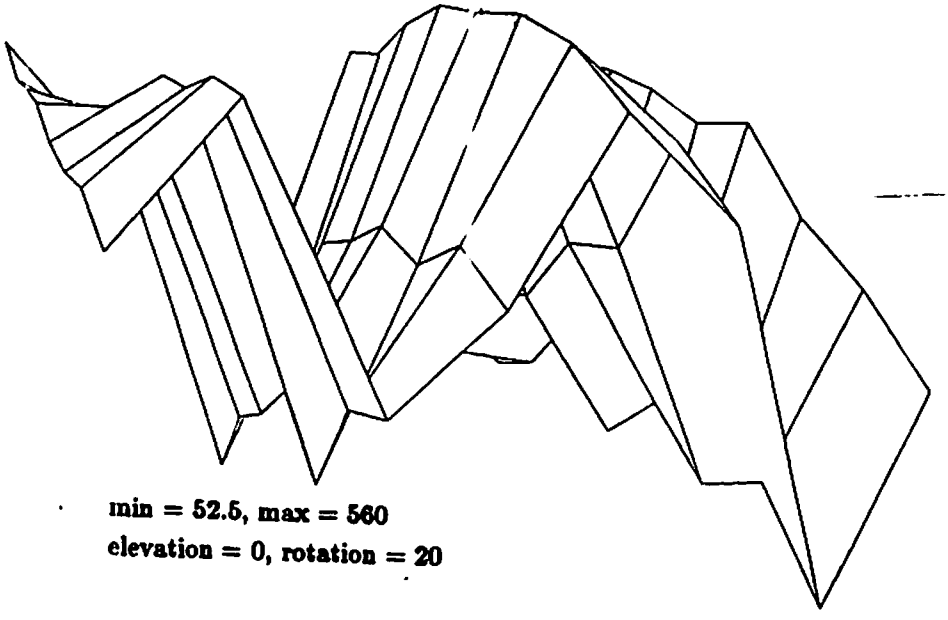
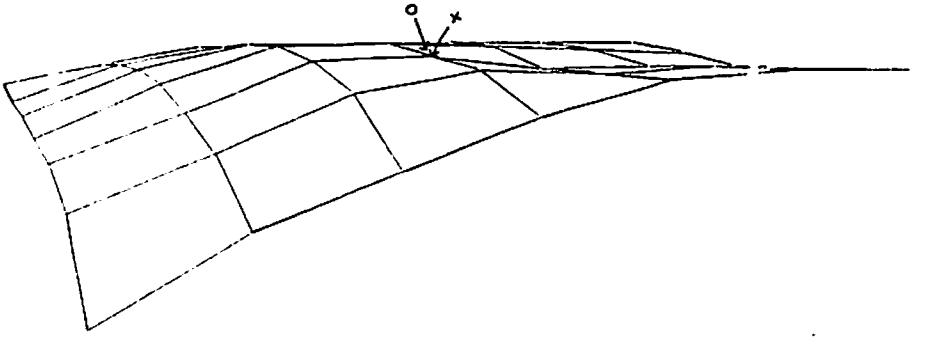
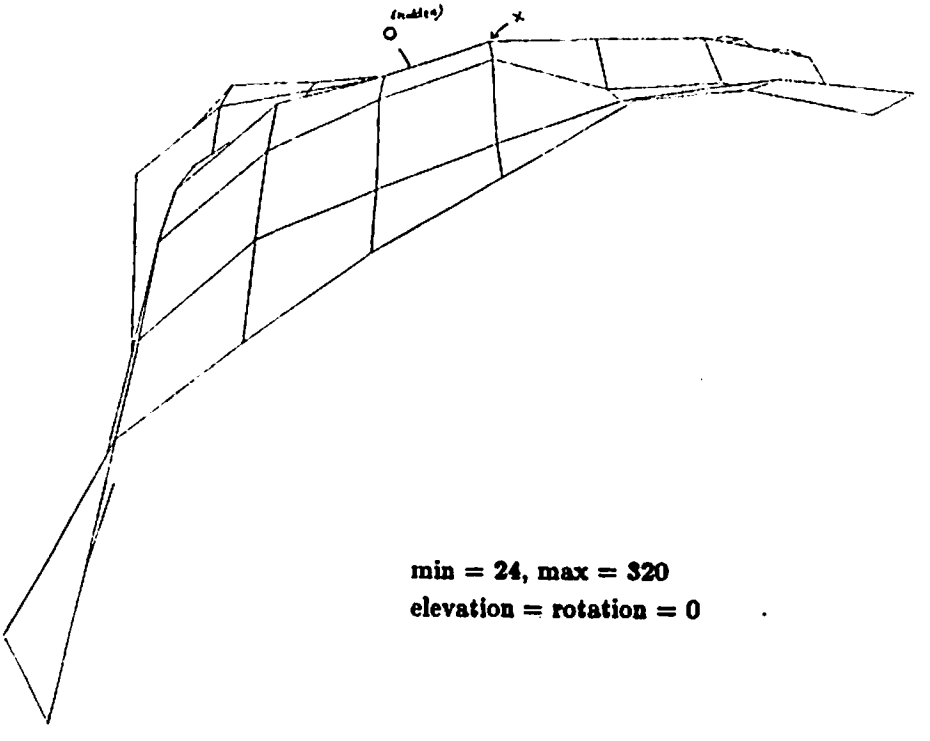


Figure 3.5b The cross-SSD surface at point *b*.



min = 0, max = 273.3
elevation = rotation = 0

Figure 3.6a The auto-SSD surface at a point in a homogeneous area. (point c in the first frame).



min = 24, max = 320
elevation = rotation = 0

Figure 3.6b The cross-SSD surface at point c.

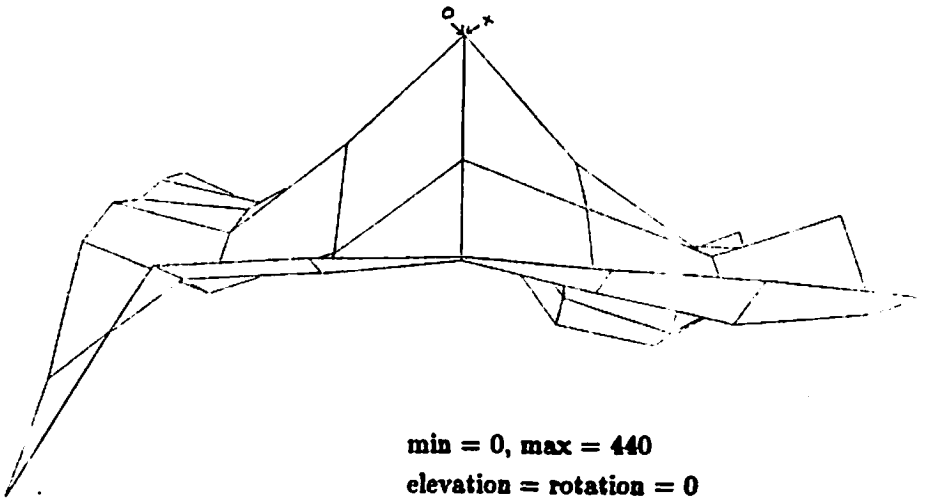


Figure 3.7a The auto-SSD surface at an occluded corner point. (point d in the first frame).

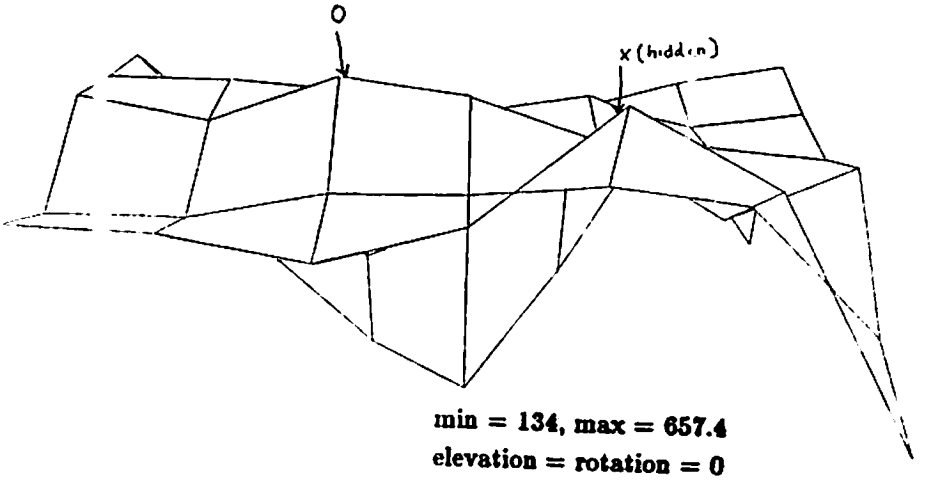
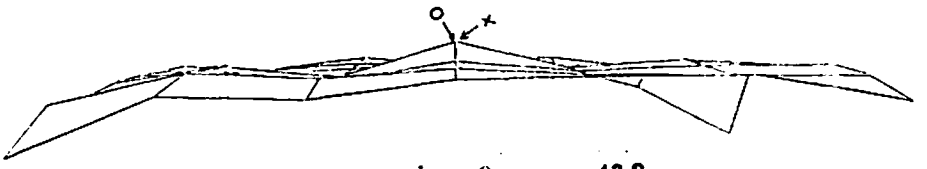
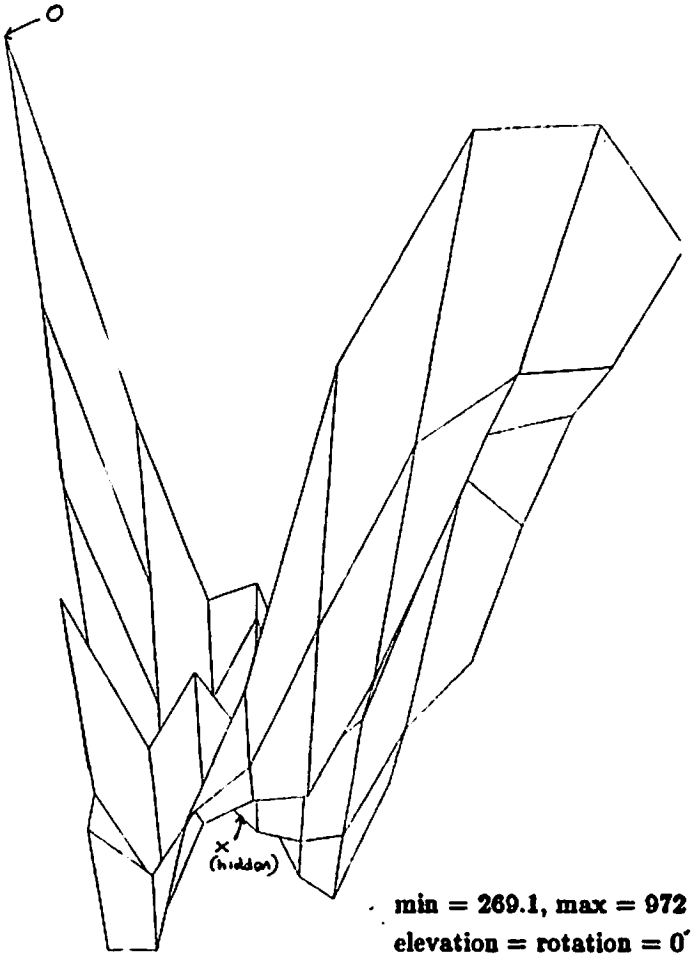


Figure 3.7b The cross-SSD surface at point d.



min = 0, max = 43.2
elevation = rotation = 0

Figure 3.8a The auto-SSD surface at an occluded homogeneous point. (point *e* in the first frame).



min = 269.1, max = 972
elevation = rotation = 0

Figure 3.8b The cross-SSD surface at point *e*.

coincide with the correct match-point. This indicates that the match estimate is unreliable in all directions. Note that in this case the values on the *cross-SSD* surface is much lower than in the case of the sharp corner described above (Fig 3.4a). This behaviour is typical at homogeneous areas of the image in the fine-resolution, higher-frequency bandpass filtered representations, since at such areas most of the image-energy is contained at the lowest spatial frequencies.

An Occluded Corner Point Figures 3.7a and 3.7b illustrate the SSD surfaces at an intensity corner which is in the occluded region (point *d* in Fig. 3.1). In this case, the *auto-SSD* surface display shows a distinct peak, similar to that in Fig. 3.4a. Since the area surrounding the corner point is occluded in the second image, there is no window that properly matches the samplw window surrounding the corner point. Therefore, the *cross-SSD* surface has a somewhat erratic shape.

An Occluded Homogeneous Point Figures 3.8a and 3.8b illustrate the SSD surfaces at a point in a homogeneous area in the occluded region (point *e*). Again, the *cross-SSD* surface shows unpredicatable and erratic behaviour. Note in particular the minimum and maximum values of the *cross-SSD* surface are considerably higher than those of the *auto-SSD* surface.

3.2 The Confidence Measure

The above demonstrations were intended to show how the SSD surface usually captures much of the information about the image structures as well as occlusion effects needed while matching. Although these illustrations capture the behaviour of the SSD surface only at the finest level of resolution, such behaviour is typical also of other levels of resolution. Where a proper match exists (i.e., the non-occluded regions), the SSD value at the point of best match generally seems to be low. At occlusion areas this value seems higher. The curvature of the SSD surface along different directions reflects the degree of variation in the image along those directions, and hence the uniqueness of the match estimate along that direction. This suggests that the confidence in the correctness of the displacement component in any direction should be directly propotional to the curvature of the SSD surface along that direction, and inversely propotional to the SSD value at the point of

best match.

Consider the normalized second derivatives of the SSD surface centered at the point of best match in the four directions 0, 45, 90 and 135 degrees. Each of these can be computed numerically using a 1×3 second-derivative operator oriented in the appropriate direction. These values are normalized in order to maintain the measure between 0 and 1. Such a normalization can be achieved by dividing the curvature measures by a weighted average of the three SSD values used to measure the curvature. This normalization also makes the confidence measure inversely propotional to the SSD value at the point of best match, as observed above.

In the formulae given below, the SSD surface is considered centered at the point of best match. The indexing is relative to that displacement, i.e., the index $(0,0)$ refers to the displacement corresponding to the best match. The figure below explains the indexing scheme.

$$\begin{array}{ccc} S(-1,-1) & S(-1,0) & S(-1,1) \\ S(0,-1) & S(0,0) & S(0,1) \\ S(1,-1) & S(1,0) & S(1,1) \end{array}$$

We compute the four normalized directional second derivatives of the SSD surface as follows:

$$C0 = \frac{S(0,-1) - 2 * S(0,0) + S(0,1)}{S(0,-1) + 2 * S(0,0) + S(0,1)}$$

$$C45 = \frac{S(1,-1) - 2 * S(0,0) + S(-1,1)}{S(1,-1) + 2 * S(0,0) + S(-1,1)}$$

$$C90 = \frac{S(-1,0) - 2 * S(0,0) + S(1,0)}{S(-1,0) + 2 * S(0,0) + S(1,0)}$$

$$C135 = \frac{S(-1,-1) - 2 * S(0,0) + S(1,1)}{S(-1,-1) + 2 * S(0,0) + S(1,1)}$$

where $S(i,j)$ denotes the SSD value at position (i,j) relative to the point of best match.

At this point various possibilities arise. Since each of these four measures provide information specific to the corresponding direction they could all be separately maintained. Alternatively, a conservative measure will be to choose the minimum of these four measures.

We have adopted this latter approach for our study. Hence our confidence measure is,

$$MIN(C0, C45, C90, C135)$$

3.3 The Confidence Measure of Burt, Yen and Xu

Burt, Yen and Xu [Burt83] describe a confidence measure which is very similar in form to the one described here. Their measure uses the Laplacian-filtered correlation surface and takes the four directional second derivatives in a manner similar to ours. Their normalisation technique is also similar to ours. However, the two measures differ in some significant ways.

First, our measure is based on the SSD surface, whereas their measure is based on the correlation surface. While the SSD function remains strictly positive, the correlation function can also have negative values. The presence of negative values can cause the normalized values to go below 0 as well as above 1. Although large negative values around the point of best match indicates a strong peak, in such situations, their confidence measure is below zero.

Second, in the strategy used by Burt, et al. confines the search at each level to a 3×3 area centered around zero displacement. This means that the curvature measures are taken at a point within this window, even though the actual displacement may be large. In band-pass filtered images, the structures can repeat themselves beyond a certain distance, thus causing false matches with high confidence (i.e., a unique match within the search window) to occur.

Finally, Burt's second derivative operators are centered at (0,0), even when the best match point occurs elsewhere within the 3×3 window. The 1×3 operator is too small to be a good approximation of the curvature at any point away from where it is centered. Therefore, if the displacement is not (0,0), Burt's measure may not be a good indication of the directional curvatures at the true match point,

3.4 A Demonstration

Having defined our confidence measure, we now proceed to demonstrate its utility with an example. For this purpose, we chose a pair of real images, called *Folgers images.* The images are shown in figures 3.9 and 3.10 and constitute a stereo pair. There are two prominent surfaces at different depths, viz., the Folgers coffee can and the textured background with the plant. Occlusion at the left side of the can is clearly visible. This area in the first image has no matches in the second image. Figure 3.11a displays the displacement field, and Figure 3.11b displays the confidence measure. The brightness of the confidence measure is proportional to the degree of confidence. The displacement field has been sub-sampled for convenience of display.

These figures reveal some important facts regarding the confidence measure. First, in large homogeneous areas of the image the confidence measure is low and the displacement estimates are often incorrect. For illustration purposes, one such area is marked "A" in Figures 3.9, 3.11a, and 3.11b. Second, the confidence measure is also low along straight edges, although the component of the displacement vectors normal to the edge often seems to correct. One such area is marked "B" in the three figures. Third, the confidence measure is low in occluded areas and around occlusion boundary. One such area is marked "C" in the three images. Finally, although the confidence measure is low both in homogeneous areas and occluded areas, it does not discriminate between the two.

In order to further illustrate the correlation between the confidence measure and the accuracy of the displacement estimates, figure 3.12 displays only those displacement values which have a confidence of 0.3 or more.

Figure 3.9 The folgers image - first frame.

Figure 3.10 The folgers image - second frame.

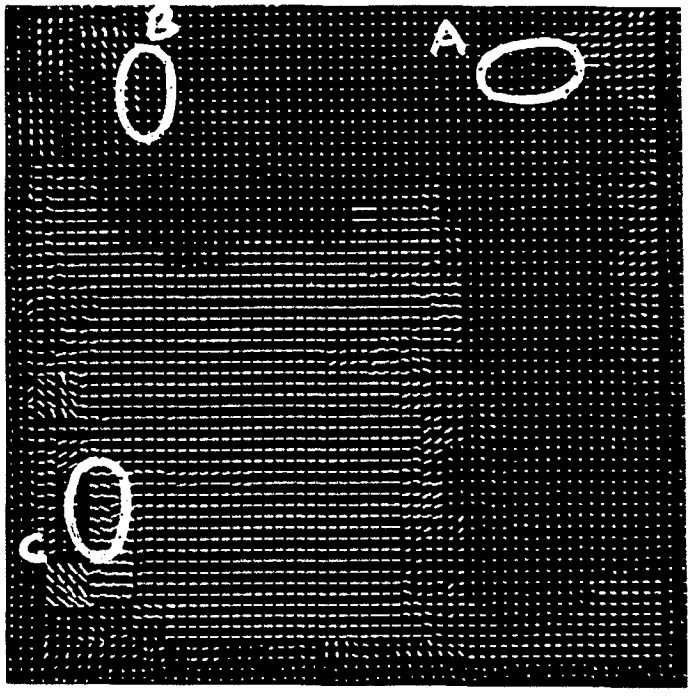


Figure 3.11a The displacement vectors for the folgers image pair.

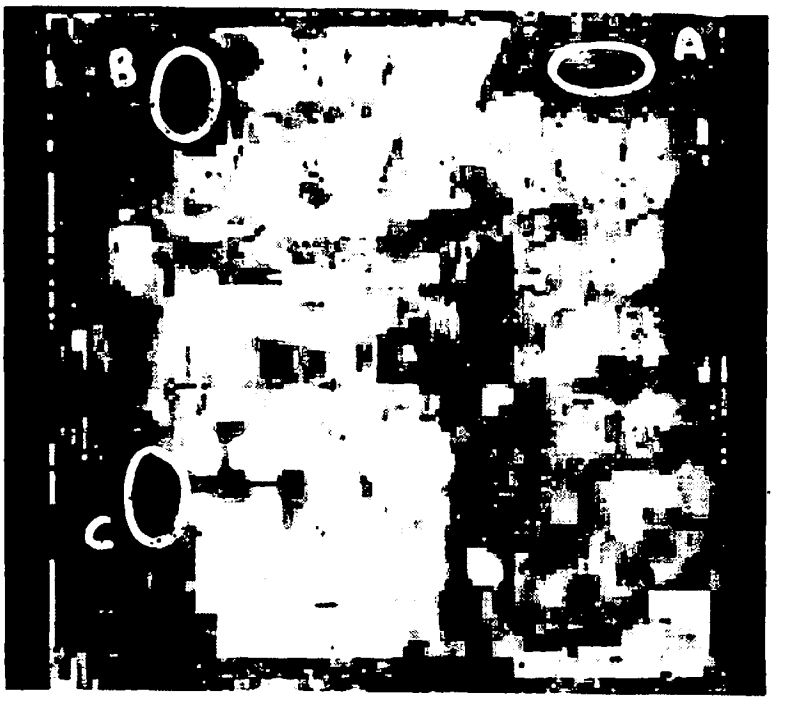


Figure 3.11b The confidence measure for the folgers displacement field.

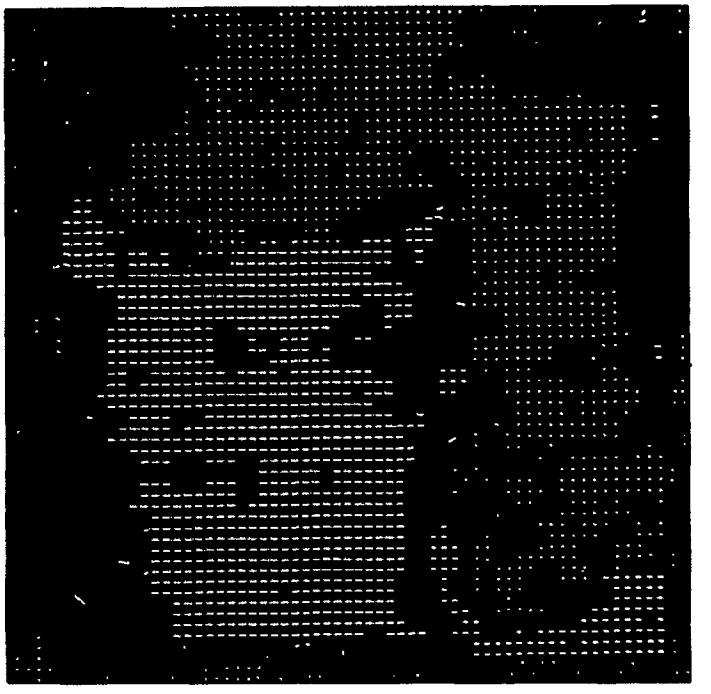


Figure 3.12 The displacement vectors for the folgers pair which have a confidence of at least 0.3.

4. A MODIFIED SEARCH STRATEGY

The confidence measure which we described in the previous chapter is useful in isolating the areas in an image where the displacement estimates are unreliable, and often incorrect. In this chapter we describe two modifications to the search strategy of Glaser, Reynolds and Anandan which significantly reduce the errors in the displacement field, particularly near occlusion boundaries.

We describe briefly the search strategy of Glaser, et. al. in order to familiarize the reader with the terminology involved. The search begins at an appropriately coarse level so that all image displacements are less than one pixel distance at that level. For each pixel in the first frame at the coarse level (say level l), matches are found within a 3×3 window centered around the corresponding pixel in the second frame. Each of these estimates are then projected to the four pixels at level $l+1$ of the pyramid that are directly covered by each pixel at level l. The displacement estimates have to be multiplied by 2 in order to take into account the reduction in the pixel-width between levels l and $l+1$. For each pixel at level $l+1$, the search is conducted in a 3×3 area around these estimates. This process of projection and search is continued downto the finest level of the image pyramid. Figure 1.1 illustrates the projection of a displacement vector and the local 3×3 search at the finer level.

Both our modifications concern the projection of the displacement estimates from a coarse level to the next finer level. First, we restrict the projection of coarse estimates to only those with high confidence. Second, we allow the coarse level estimate at each pixel to be projected in an area larger than the 2×2 area directly covered by that pixel. Both these modifications are explained in greater detail in the following sections

4.1 Restricting Projection to High Confidence Estimates

The restriction of projection to only high confidence estimates is suggested in Glaser, et. al. [Glas83]. The motivation for this idea stems from the fact that when incorrect coarse-level estimates are projected down, the 3×3 searches at the finer levels are conducted in areas of the second frame that do not include the true-match point. This causes incorrect

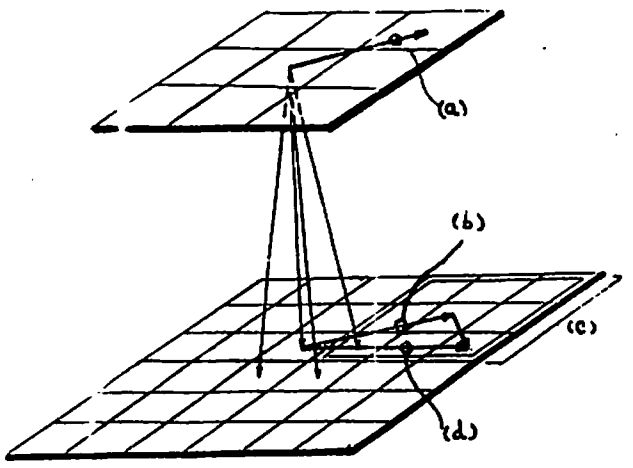


Figure 4.1 The local 3×3 search in [Glas83].

(a) Displacement vector at level N
(b) Displacement vector projected to its four sons at level N+1 (only one of the four sons is shown)
(c) Search in a 3 x 3 area at level N + 1 (search area shown in double lines)
(d) Updated displacement vector

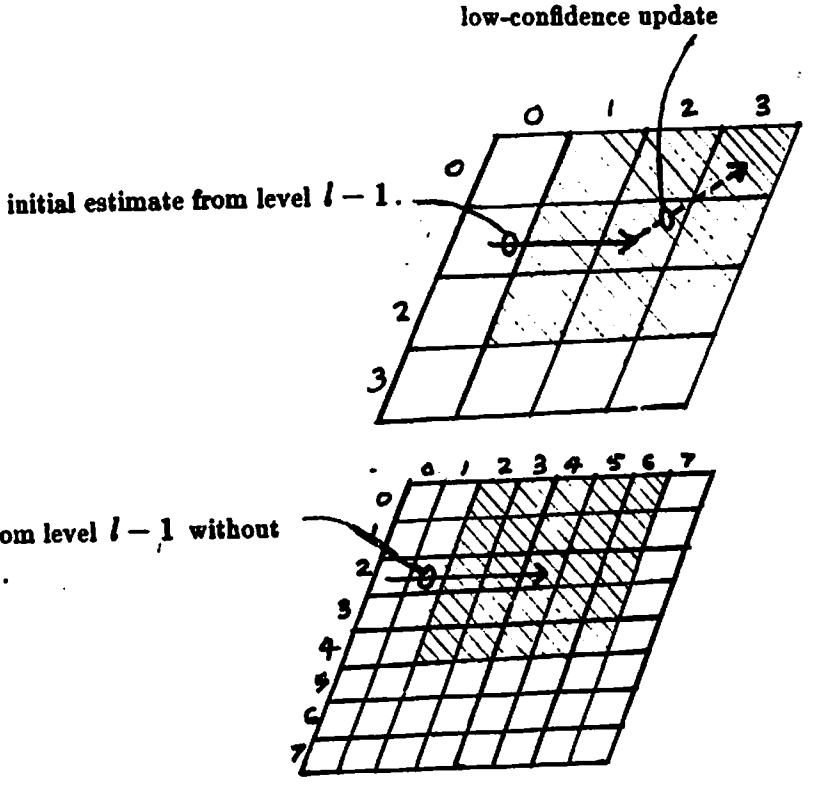


Figure 4.2 The expanded search area

Pixel $(0,2)$ at level l has a low-confidence displacement update. Therefore, its initial estimate from level $l-1$ is passed down to its children pixels at level $l+1$. The expanded search area is shown cross-hatched.

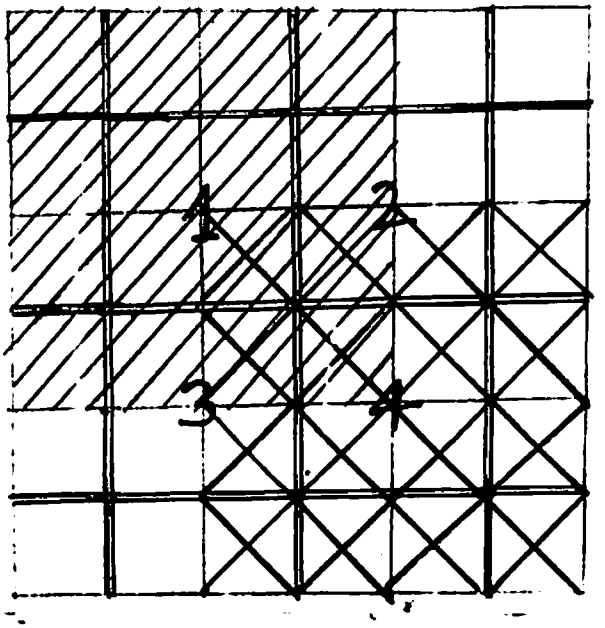


Figure 4.3 The overlapped pyramid projection

The thick double lines show pixel boundaries at level l. The thin lines show pixel boundaries at level $l+1$. The projection area of level l pixels 1 and 4 are shown in the figure. Note that each pixel at level $l+1$ has four parents at level l.

matches in all the subsequent levels. If on the other hand, these incorrect coarse-level estimates are altogether rejected, the finer-level searches can be conducted over larger area than the usual 3×3 windows, and the true-match can perhaps be recovered.

At any level l of the pyramid, the displacement updates at pixels where the confidence measure is low can be suppressed (this can be achieved by a simple threshold on the confidence measure). Wherever such updates are suppressed, we simply pass the displacement estimates from the *parent* pixel at level $l-1$ to the *children* pixels at level $l+1$. Assume that at such a pixel at level l, we are searching over a window of radius r (at level l), centered about the displacement estimate from level $l-1$. This corresponds to a window of radius $2 \times r$ at the next finer level. Hence the lack of update at level l would require that we search over a window of radius $2 \times r$ at the *children* pixels at level $l+1$. If there is still no update at this level the search window radius should be doubled at the next level below, and so on. This is illustrated in Figure 4.2.

4.2 Modified Projection of Coarse Estimates

Our second modification to the search strategy involves the manner in which the coarse level displacement estimates are propagated to the fine level. The strategy of Glazer, et. al. projects the displacement value of a *parent* pixel as the estimate for all of its four *children* at the next finer level. At areas which are near discontinuities in the displacement field (e.g., occlusion boundaries), this approach can cause incorrect estimates to be projected from a coarse-level pixel to the finer level pixels. This occurs because, at a coarse-level of resolution the boundary of discontinuity can be placed only within a coarse accuracy. For fine-level pixels along one side of the boundary, it is then possible that the coarse estimates from the other side of the boundary are projected down. This causes the search at the subsequent levels to find incorrect matches.

We propose a slightly different method of projecting the coarse level displacement estimates to the next finer level. This idea is based on the "overlapping" pyramid idea of Burt, et. al. [Burt80] Each pixel at the finer level $l+1$ is considered to have four potential parents at the coarser level l (see Figure 4.3) We consider all the four estimates as possible initial estimates for the search at level l and conduct searches around each

of these estimates. The displacement corresponding to the best match in this exapanded search area is then chosen as the updated displacement estimate for that pixel at level $l+1$. In this way the pixels along the boundaries of displacement discontinuities are not bound to an incorrect coarse estimate. This allows for more precise placement of the boundaries of displacement discontinuities at the finer level.

It is obvious that we can combine both the modifications into a uniform algorithm. This would involve choosing the appropriate search radius for each of the estimates of the parent pixels according to their confidence value and their search radius at level l.

4.3 A Demonstration

In this section, we describe an experiment to demonstrate the effects of applying both our modifications to the search strategy. Fig 4.4 and 4.5 are two images from the sequence called *poster images*. The displacement estimates based on the old search strategy are shown in Fig 4.6. As in the case of our figures in Section 3, the displacement estimates have been subsampled to enhance visibility. The shaded areas have estimates with confidence below 0.3 and the white areas have estimates with confidence above 0.3. Note the predominance of the low confidence values around the occlusion boundary. Fig 4.7 displays the displacement estimates generated by a search strategy which incorporates only the first of the two modifications, viz., suppressing low confidence estimates. Again, the shaded areas are areas of low confidence (below 0.3). We make the following observations about Figures 4.6 and 4.7. First, the low-confidence areas are rectangular in shape. This is due to the strict 1 to 4 projection used in these search strategies. In both cases if an error is made at a coarse level that is not suppressed, it is possible that the search areas at the finer levels for all the children pixels may no longer include the true match points. Since no information is used from the correct neighbours, these go uncorrected. Second, the restriction of projection of coarse estimates (Fig. 4.7) seems to improve the displacement estimates in some areas (of the background), whereas introduces more errors at others (near the occlusion boundary). This is because some of the low-confidence coarse estimates are actually correct. Eliminating these and expanding the search area can lead to false matches due to repeated features.

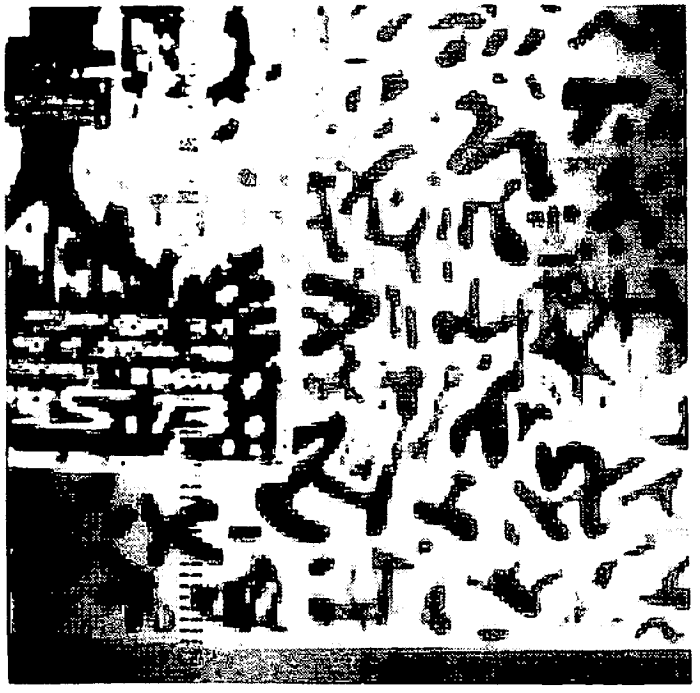


Figure 4.4 The Poster Image - first frame

Figure 4.5 The Poster Image - second frame

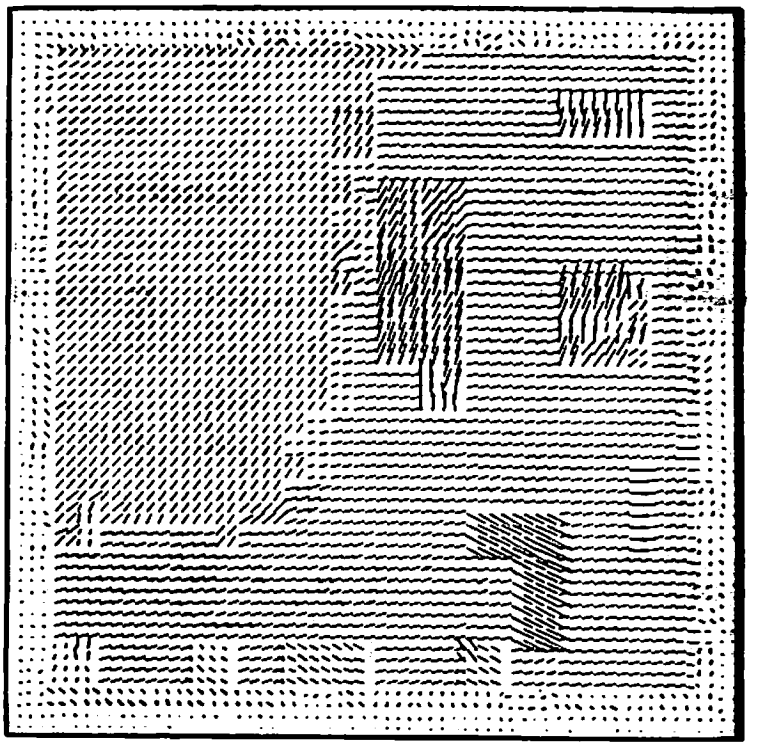


Figure 4.6 Results of the search strategy described in [Glaz83]

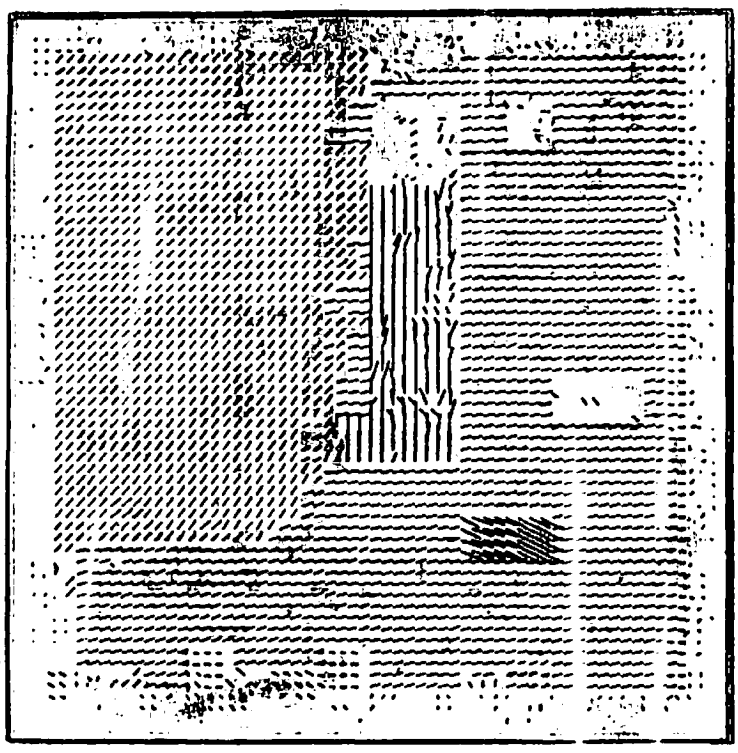


Figure 4.7 Results using restricted projection

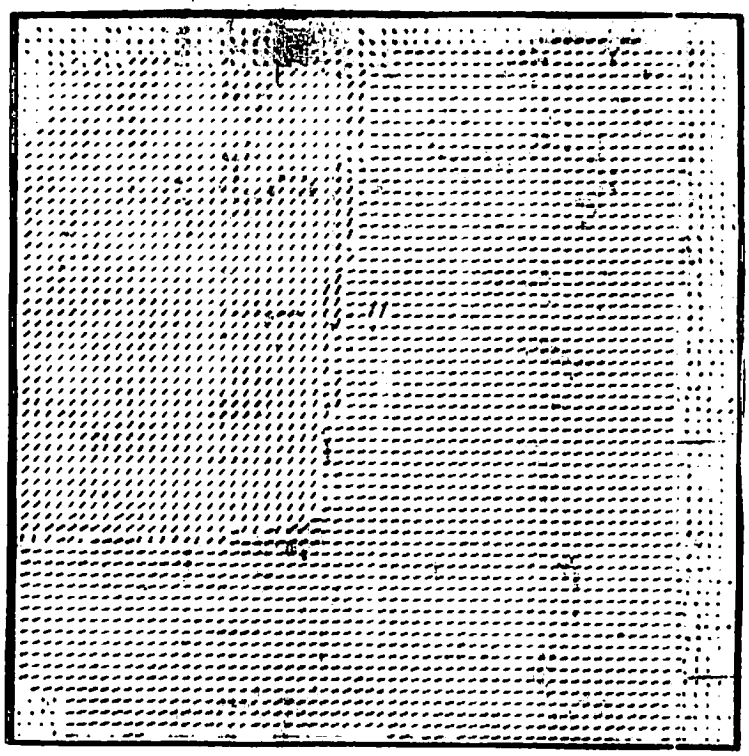


Figure 4.8 Results using overlapped projection

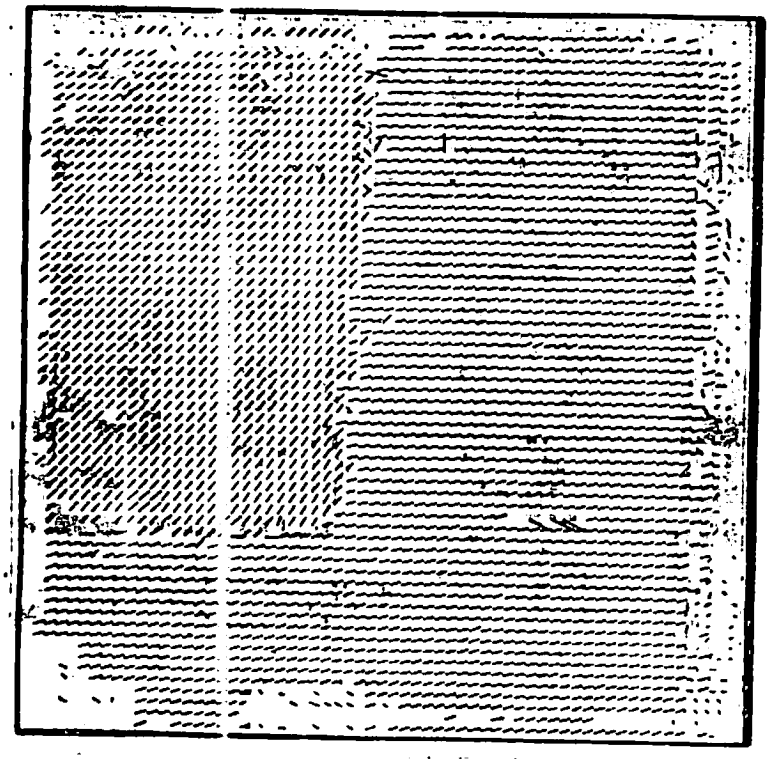


Figure 4.9 Results using both our modifications

Fig 4.8 displays the displacement estimates superimposed on low confidence areas as provided by a search strategy which incorporates only the second modification, viz., the overlapped pyramid projection. Fig 4.9 displays the displacement estimates superimposed on the shaded low confidence areas as provided by a search strategy that incorporates both the modifications mentioned above. In Figures 4.8 and 4.9, is easy to see that a dramatic reduction in the size of the low-confidence areas has been achieved. Figure 4.8 shows that the modified projection strategy provides the major contribution to the improvement in the displacement field. This is so because, in this approach the coarse-estimates are not altogether eliminated. Instead, information is used from neighbours who may have correct estimates.

Later in this paper, we discuss other possible ways of utilising the confidence measure to improve the matching results, all of which are currently under investigation.

5. APPLICATIONS AND FUTURE WORK

Thus far in this report we have described our confidence measure, and demonstrated its use in our modification to the search strategy. This measure can also serve as useful piece of information for techniques that process the flow field. These techniques usually involve using the flow field for the extraction of camera and object motion parameters ([Adiv84],[Rieg84],[Praz80],[Tsai84]). With some modifiations, the confidence measure can also be used by processes that attempt to improve the accuracy of the flow field through the use of various consistency constraints ([Nage83],[Glaz81]). In the following sections, we outline the immediate applications of the confidence measure, and describe the directions in which this study can be extended to include the various modifications necessary for other applications.

5.1 Use by Prameter Computation Algorithms

One of the primary uses of image displacement fields has been to recover camera and object motion parameters and orientation and depth of image surfaces. Typically, techniques that address these problems involve solving a system of equations ([Long81],[Tsai84]) or minimizing an error measure ([Adiv84],[Rieg84],[Praz80]).

The techniques that involve solving a system of equations use the displacement estimates as the known variables. Often these techniques tend to be highly sensitive to errors in the displacement fields (see [Adiv84]). In such cases, eliminating the displacement estimates which have a low confidence measures from consideration can help in reducing errors and enhancing the reliability of the overall results of these processes.

The techniques that minimize an error measure or compute a hough-transform, can use the confidence in two ways. First is the same as above, i.e., to eliminate the displacement estimates with low confidence measures. Second, these techniques usually compute a global error or transform which has contributions from each displacement vector. This confidence can be weighted by the confidence measures. In this way inaccurate displacement vectors (which typically have low confidence), do not contribute to the optimization process, thus enhancing the reliability of that process. As an example of this latter use,

we point to Adiv ([Adiv84]), who attempts to segment the image into regions which have consistent displacement fields within those regions, and compute the 3-d motion paramters corresponding to these fields. His technique is a multi-stage one, and involves transformations from the displacement vectors to affine-transformation parameter space, as well as least-square-error fits of 3-d transformations to the displacement fields. He uses the confidence measure as a weight associated with the contribution of each displacement vector to the affine transformation as well as to the error measure.

5.2 Future Work

Our experiments using real images taken in our robot-lab revealed several facts regarding our confidence measure and our modified search strategy (see Section 3.5). Here we focus on two of those facts. First, along straight edges the confidence measure is low, although the displacement-components perpendicular to the edges seems correct. Second, the confidence measure does not discriminate between occluded areas and homogeneous areas of the image.

In the following sections, we describe the possible approaches to addressing these issues. Although we intend to pursue these directions, the reader should be forewarned that what follows is largely speculative.

Directional Information about Matching

The demonstrations of the behaviour of the SSD surface at typical areas of the image with directional structures (e.g., edges) clearly showed that such directional information was indeed noticeable in the shape of the SSD surfaces. More specifically, we noted earlier that along edges in the images, we see a ridge like SSD surface where the orientation of the ridge corresponds to the orientation of the edge in the image. Hence, the directional confidence measure along the direction of the edge is low, whereas it is high in the direction perpendicular to it. It has also been well recognized by many researchers (to name two, [Glaz81] and [Horn80]) that a directional feature in the image (say, an edge) can provide reliable information about the component of the displacements in the direction perpendicular to that feature, whereas it can provide no information about the component of the

displacements along the direction of that feature.

There are two possible ways in which this directional confidence information can be utilized. The first is to use it in an algorithm similar to the modified search strategy described in section 4. In this case, the search area would not be expanded along the direction where the SSD surface shows significant variations. Instead, we can expand the search area in the direction along the ridge, thus obtaining a somewhat rectangular search area.

The second method is to use these in an algorithm that propogates information between neighbouring pixels, especially along edges and curves in the image in order to bring together reliable information about different directions. One way of doing this is described in [Glaz81]. Each pixel provides a linear constraint equation on the displacement vector at that pixel. In Glaser's approach, a least square-error solution for the system of the constraint equations from neighbouring pixels along an edge is considered to be the true displacement vector for all pixels along the edge. Given that the SSD surface captures the directional information, we have available to us information similar to these constraint equations. An added benefit of using the SSD surface over the constraint equations is the fact that the SSD surfaces also provide information about noise variations and occlusion effects.

A more general way of using neighbour information for the improvement of the displacement estimates is to use relaxation-smoothing techniques. These are similar to the now classical Horn and Schunk smoothing technique, or perhaps closer to the constrained-smoothing approach described by Nagel [Nage83], and the surface reconstruction approach described by Terzopoulos [Terz84]. These are discussed in the following section.

Creation of Dense Reliable Displacement Fields

The Constrained Smoothing Approach — Nagel [Nage83] describes an approach that performs a form of constrained smoothing of the displacement fields. His approach integrates the intensity gradient constraint of Horn and Schunk with a smoothness constraint on the flow field which is somewhat different from theirs. The smoothness constraint is

modified with weights that are inversely proportional to the intensity gradients at a given pixel. At each pixel, the component of the displacement vector along the direction of the intensity gradient is not allowed vary significantly, whereas the component in the normal direction is allowed to change more freely.

The Surface Reconstruction Approach — Terzopoulos [Terz84] describes an approach for interpolating sparse data for reconstructing a surface. This approach involves formulating the interpolation problem as a problem of minimizing the potential energy of a thin plate, with sparse uncertain data about the location of the surface. The plate is suspended on needles corresponding to the available sparse data. The needles have springs attached to their ends, whose stiffness are proportional to the reliability of the available data at the locations of the needles. The finite element method approach is used to numerically solve the variational problem.

Comparing the Two Approaches — There are three important differences between the approaches of Nagel and Terzopoulos.

First the variational problems are somewhat different. Nagel's smoothness requirement is non-isotropic, i.e., at any point in the image, the smoothness requirement is not the same in all directions. Although there is provision in the approach of Terzopoulos for performing non-isotropic smoothing, his precise formulation of the potential energy is isotropic.

Second, Terzopoulos includes available sparse data and reliability information in his variational problem. Nagel's approach, being a generalisation of Horn and Schunk's work, refers directly to the spatio-temporal intensity variations of the image, and does not include any initial displacement data. However, it is possible to extend this approach to include such data as the intial values for the iterative smoothing process.

Third, Nagel uses the finite-difference approach to numerically solve his variational problem, whereas Terzopoulos uses the finite-element method. As Terzopoulos notes in his thesis, finite element method has several advantages. First, it has a well established mathematical theory which is useful to understand stability and convergence issues. Second, at locations of known discontinuities in the image, it is possible to fracture the surface,

i.e., to prevent the smoothing process from functioning across discontinuity contours in the image. Finally, it allows arbitary shaped grids on the image. The disadvantage is that the finite-element method is a more complex formulation.

Our Problem — Our data consists of a dense displacement field with directional confidence information everywhere. Our goal is to replace this by a dense field by propagationg information from reliable areas to unreliable areas of the image. We would also like to prevent the propagation across occlusion boundaries. Under these conditions it is natural to consider an approach similar to that of Terzopoulos.

Our problem can be seen as a generalization of Terzopoulos' surface reconstruction problem to a vector field. We will use directional information regarding the reliability of the available data.. It also differs from Nagel because the directional information is based on both images rather than a single image. We will locate discontinuities in the displacement field through attempting to identify occlusion (see next section). At this point, we can either use an isotropic (like Terzopoulos), or a non-isotropic (like Nagel) smoothness constraint. One of our future goals is to pursue both approaches and compare them.

Recognition and Processing of Occlusion

Occlusion, although a source of failure and frustration for most algorithms that attempt to produce a dense displacement fields, is very useful for the purposes of segmentation of the image into objects at different depth or with different movements. Therefore any process that detects occlusion very early in the processing can be useful for focus of attention, tracking as well as more accurate computation of the various image properties.

The confidence measure discussed in this study produces low values for both homogeneous areas as well as occlusion areas. However, often it is possible to separate these situations using the information in the SSD surface. In real world images, it is usually the case that where there are occlusion boundaries, there are also discontinuities in the image texture. This suggests that all values on the auto-SSD surface will usually be high at occlusion boundaries, whereas they will be uniformly low at homogeneous areas. This

observation can be useful in the identification of occlusion areas in an image.

This is not as simple as it may seem from the discussion above. For one thing, there are questions about how to translate such real valued numbers as the confidence measure, and the values on the SSD surface into binary decisions about occlusion. Further, it is also possible to encounter situations where there is a compounding of occlusion situations and homogeneous areas, which makes it difficult to identify them as being due to occlusion. The key to these problems may lie in separating the confidence measure into two components. One is a *matchability* measure which could be based on single image information, and would classify points according to their degree of "matchability". For example, an intensity corner will be highly matchable, whereas a point in a homogeneous area of the image will be poorly matchable. Another component is a *match-existence* measure which would identify points in occluded areas, and in areas with changing image structure. This measure could be derived by comparing the shapes of the auto- and cross- SSD surfaces. These questions will be the focus of our own future work in this direction.

Acknowledgements

I wish to thank Profs. Al Hanson and Ed Riseman for their comments on the manuscript. Thanks are also due to Frank Glazer and George Reynolds for the interest they have shown in this work. My belated thanks to Randy Ellis for help in obtaining the images and the PMTs for my figures. This research was supported by DARPA under grant N00014-82-K-0464.

REFERENCES

[Adiv84] Adiv, Gilad, Determining 3-D Motion and Structure from Optical Flow Generated by Several Moving Objects, *COINS Technical Report 84-07* University of Massachusetts, April 1984.

[Agga81a] Aggarwal, J. K., Davis, L. S. and Martin, W. N., Correspondence Processes in Dynamic Scene Analysis, *Proceedings of the IEEE*, Vol. 69, Number 5, May 1981, pp. 562-572.

[Agga81b] Aggarwal, J. K. and Martin, W. N., Analyzing Dynamic Scenes Containing Multiple Moving Objects, in *Image Sequence Analysis* (T. S. Huang, editor), Berlin Hidelberg; Springer Verlag, 1981, pp. 335-380.

[Bard80] 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.

[Burt80] 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, 802-809

[Burt82] Burt, P. J., Yen, C. and Xu, X., Local Correlation Measures for Motion Analysis: A Comparative Study, *IEEE Proceedings of PRIP*, 1982, pp. 269-274. Also *IPL-TR-024*, ECSE Dept., RPI, 1982.

[Burt83] Burt, P. J., Yen C. and Xu X., Multi-Resolution Flow-Through Motion Analysis, *IEEE CVPR Conference Proceedings*, June 1983, pp. 246-252.

[Elli84] Ellis, R., Robot Imaging System, *Laboratory for Perceptual Robotics* Internal Memorandum, no. 6, Jan, 1984, Univ. of Massachusetts at Amherst.

[Genn80] Gennery, D. B., Modelling the Environment of an Exploring Vehicle by means of Stereo Vision *Report No. STAN-CS-80-805*, Dept. of Computer Science, Stanford University, California.

[Glaz81] Glazer F., Computing Optic Flow, *IJCAI-7*, Vancouver B. C., Canada, Aug. 1981, pp. 644-647.

[Glaz83] Glazer, F., Reynolds, G. and Anandan, P., Scene Matching by Hierarchical Correlation, *IEEE CVPR conference*, June 1983, pp. 432-441.

[Hann74] Hannah, M. J., Computer Matching of Areas in Stereo Images, *Stanford A.I. Memo 239*, July 1974.

[Hans80] Hanson, A. and Riseman, E. M., Processing Cones: A Computational Structure for Image Analysis, in *Structured Computer Vision*, (Tanimato, S. and Klinger, A., editors), Academic Press, New York, 1980.

[Hayn81] Haynes, S. M. and Jain, R., Detection of Moving Edges, *CSC-82-004* Intelligent Systems Lab., Dept. of Computer Science, Wayne State Univ., Dec. 1981.

[Hild83] Hildreth, E. C., Computing Velocity Field Along Contours, *ACM Siggraph/Sigart Workshop on Motion* Toronto, Canada, Apr. 1983, pp. 26-32.

[Horn80] Horn, B. K. P. and Shunck, B. G., Determining Optical Flow, MIT A.I. Memo Number 572, April 1980.

[Jaco80] Jacobus, C. J., Chien, R. T. and Selander, J. M., Motion Detection and Analysis by Matching Graphs of Intermediate Level Primitives, *IEEE Transactions on PAMI*, Vol. PAMI-2, Number 6, Nov. 1980, pp. 495-510.

[Lawt84] Lawton D. T. Processing Dynamic Image Sequences From a Moving Sensor, *Ph.D Dissertation and COINS Technical Report 84-05*, University of Massachusetts, 1984.

[Long81] Longuet-Higgins, H. C., A Computer Algorithm for Reconstructing a Scene from Two Projections, *Nature* 293, 1981

[Mora81] Moravec, H. P., Robot Rover Visual Navigation, UMI Research Press, Ann Arbor, Michigan, 1981.

[Nage77] Nagel, H.-H., Analyzing Sequences of TV-Frames, *IJCAI 77*, MIT, Cambridge, Ma.,

Aug. 1977, p. 626.

[Nage83] Nagel, H.-H., Constraints for the Estimation of Displacement Vector Fields from Image Sequences, *IJCAI-83*, Karlsruhe, West Germany, pp. 945-951.

[Prag79] Prager, J. M., Segmentation of Static and Dynamic Scences, *COINS Technical Report 79-07 and Phd Dissertation*, University of Massachusetts, Amherst, 1979.

[Praz80] Prazdny, K., Egomotion and Relative Depth Map from Optical Flow, *Biology and Cybernetics*, Vol. 36, 1980, pp. 87-102.

[Radi81] Radig, B. M., Image Region Extraction of Moving Objects, in *Image Sequence Analysis* (T. S. Huang, Editor), Berlin Heidelberg: Springer-Verlag, 1981, pp. 311-354.

[Roac79] Roach, J. W. and Aggarwal, J. K., Computer Tracking of Objetcs Moving in Space, *IEEE Transactions on PAMI*, vol. PAMI-1, 1979, pp. 127-134.

[Terz84] Terzopoulos, D., Mutiresolution Computation of Visible-Surface Representations, *Phd Dissertation*, Massachusetts Institute of Technology, Jan. 1984.

[Thom80] Thompson, W. B. and Barnard, S. T., Lower-Level Estimation and Interpretation of Visual Motion, *Computer*, Aug. 1981.

[Thom83] Thompson, W. B., Mutch, K. M. and Berzins V. A., Determining Qualitative Spatial Properties from Optical Flow Fields, *Proceedings from the Workshop on Vision, Brain and Cooperative Computation* Univ. of Massachusetts, Amherst, Ma., May 1983.

[Tsuj80] Tsuji, S., Osada, M. and Yachida, M., Tracking and Segmentation of Moving Objects in Dynamic Line Images, *IEEE Transactions on PAMI*, Vol. PAMI-2, No. 6, Nov. 1980, pp. 516-522.

[Wong78] Wong, R. Y. and Hall E. L., Sequential Hierarchical Scene Matching, *IEEE Transactions on Computer*, Vol. 27, No. 4, 1978, pp. 359-366.

[Yach81] Yachida, M., Determining Velocity Map by 3-D Iterative Estimation, *IJCAI-81* Vancouver, B. C., Canada, Aug. 1981, pp. 716-718.