Techniques for Disparity Measurement

MICHAEL R. M. JENKIN

Department of Computer Science, York University, Downsview, Ontario, Canada M3J 1P3

AND

ALLEN D. JEPSON AND JOHN K. TSOTSOS

Department of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S 1A1

Many different approaches have been suggested for the measurement of structure in space from spatially separated cameras. In this report we critically examine some of these techniques. Through a series of examples we show that none of the current mechanisms of disparity measurement are particularly robust. By considering some of the implications of disparity in the frequency domain, we present a new definition of disparity that is tied to the interocular phase difference in bandpass versions of the monocular images. Finally, we present a new technique for measuring disparity as the local phase difference between bandpass versions of the two images, and we show how this technique surmounts some of the difficulties encountered by current disparity detection mechanisms. © 1991 Academic Press, Inc.

1. INTRODUCTION

The localization of objects in the real world from a stereo presentation has been one of the major goals of both computer vision researchers and photogrammetric engineers. Two very different computational approaches have been suggested: The underlying computation has been reduced to one either of *correspondence* or of *correlation*. Although both approaches gave rise to initial successes, more recent research has not given rise to the expected improvements in performance.

We begin by examining the classical techniques for measuring disparity in both the computer vision and photogrammetric domains. We construct a number of simple one-dimensional stimuli to investigate the behavior of the two major classes of disparity detectors, and present specific examples which pose difficulties for the various algorithms. The difficulties exhibited by current algorithms are a result of the restrictive nature of the underlying definition of disparity used by the various algorithms. By investigating the notion of disparity in a bandpassed spatial-frequency tuned channel, we relate the local disparity

to the interocular phase difference. We develop a local computational technique that measures this phase difference and thus the local disparity. This new disparity measurement technique is then applied to the stimuli which presented difficulties to the classic disparity detection techniques. In this paper we do not address many of the properties of this new method for disparity detection. Additional work is required to determine the effect of noise on the response of the new detector, and the detector must be placed within a complete stereopsis algorithm in order to examine its abilities more fully. Some preliminary results have been obtained, and the interested reader is directed to Jenkin [14].

In this paper we address only the problem of extracting (or measuring) the initial disparity estimates. We acknowledge that there must exist complex processes that operate on these primitive measurements to arrive at a final three-dimensional description of the environment. However, regardless of the complexity of these later processes, they can only produce results based on the nature of the initial disparity measurements.

2. CLASSIC DISPARITY DETECTORS

2.1. Correlation

Since the mid-1950s correlation-based techniques have been applied to the task of recovering ground height from stereo parallax measurements. (See Wood [37] for a general discussion of the history of automatic correlation techniques.) Two cameras take pictures of the same scene with the cameras' positions separated by a known baseline, or (as in the case of aerial photogrammetry) a single camera is used to take the picture, but the camera is moved between exposures. The computational task is the automatic recovery of surface height from the two images. The classic approach has been to present the task as one of correlation. Fundamentally, the assumption is

14

made that the left and right images are simply shifted versions of each other, and that the measurement of this shift is equivalent to determining the stereo parallax. (Note that this assumption includes assumptions of frontoparallel surface, orthographic projection, and identical lighting conditions as viewed by the two eyes.) The assumption is made that the correct shift is the shift that sets the integral of the square of the difference between the two images to zero. This sum has a minimum when the images are shifted the correct amount with respect to each other. In a corresponding way, the correlation (the integral of their product) has a maximum. We can generalize this idea to measure the shift between two regions in the image. The disparity or shift that will be assigned to the point x can be found as

$$\max_{\text{disp}} \int W(z)\text{left}(x + \text{disp} + z)\text{right}(x + z)dz, \quad (1)$$

where W is a windowing function used to localize the disparity measure, and left and right are the left and right images. The performance of the algorithm in the presence of noisy images can be greatly improved by choosing a windowing function W that has a blurring effect.

Unfortunately, the left and right views of a scene are not simply shifted views of each other. Examine a surface with a regularly repeating pattern (such as ruled note-paper) with your two eyes (see Fig. 1). Hold the surface so that the pattern is vertical, and rotate the surface so that it is almost lined up with the line of sight of one of your eyes. Rotate the surface about the vertical axis. You should note that the spacing of the pattern is different in the left and right eyes, and that the effect of this spacing change (spatial frequency) is dependent on the interocular surface tilt. A simple correlation function applied to this *tilted* surface would have grave difficulties as the left and right views are not particularly similar.

Indeed, in the limit, the spatial-frequency properties of the left and right views can be made arbitrarily different.

It is also possible to explicitly consider nonzero surface tilts in the design of a stereopsis algorithm. A simple measure of differing interocular patterns has been given by Clark and Lawrence [5]. Suppose that the retinal disparity is a linear function of retinal position; then the disparity $d(x)$ has the form

$$d(x_1) = x_r - x_1 = \beta_0 + \beta_1 x_1. \quad (2)$$

If inter-ocular lighting differences are ignored, then I_1' and I_r' (the left and right observed intensities) are related as

$$I_1'(x_1) = I_r'(x_r)$$
$$= I_r'(\beta_0 + (1 + \beta_1)x_1). \quad (3)$$

Even with this simple measure, the left and right eyes do not see simple shifted versions of each other, and for surfaces with a *disparity gradient* (β_1) different from 0, correlation based schemes are likely to fail.

In practice, correlation is not able to deal with scenes that are distorted by perspective projection or by interocular differences in illumination (Horn [12]). The same surface patch will appear with different brightness when viewed from two different directions. Specular reflection in particular will pose a difficult problem. A number of problems have been noted by researchers even when the geometry and illumination are ignored. The task of determining a unique minimum can be very difficult when there are large regions of constant intensity. The converse case in which there is a great deal of variation in image intensity also leads to problems. As Crombie notes

Correlation is an inadequate measure when there is less than, or more than, a moderate amount of image structure at and around points selected for image matching.[1]

For more details on correlation-based methods for stereopsis, and the surface models used, the reader is directed to Förstner [9].

Given that the simple correlation task will not uniquely determine interocular disparities at all points in an image, complex later processes have been proposed to produce a final depth map. Not only must multiple depth measurements at a given image point be resolved, but areas of constant intensity (zero, for example) may have no depth measure assigned, and some technique must be presented to "fill in" these missing values. Many commercial systems utilize a human operator in this final step. For example, the *Gestalt Photomapping System* uses a

FIG. 1. Nonfrontoparallel surface. A piece of paper with vertical lines which is rotated from the fronto-parallel plane (A) will have different spatial frequencies as seen by the two eyes. If we place the paper so that it is rotated with respect to one eye (B), but is perpendicular to the line of sight of the other eye (C) then the two eyes see drastically different structures. The effect of surface deformation under perspective projection has been emphasized for clarity of presentation.

---

[1] Crombie [6, p. 529].

FIG. 2. Left and right intensity profiles of bar target. Note that these profiles are of intensity and are of a one-dimensional target.

human operator to set critical disparities before and during image processing (Kelly [17]).

Consider the application of a correlation algorithm to a simple surface, such as a bar that is tilted in depth (it does not lie completely within the frontoparallel plane). The bar has a zero disparity at one end and a nonzero disparity at the other (See Fig. 2.)

Figure 3 shows the result of applying the correlation algorithm to a 1D slice through a bar. The figure shows the correlation strength for all possible (x, d) positions. Note that no decision has been made to determine a unique measure of disparity for a given x position; rather all possible responses are represented. Classic correlation techniques choose peaks in the disparity for a given x position as the *true* disparity.

The correlation-based algorithm produces a large broad peak over a wide range of disparities. The windowing function used was chosen to be a Gaussian with a half-height cut-off which was considerably smaller than the bar width (approximately one-half the width of the

bar). Although a Gaussian was chosen, windowing functions with support regions smaller than the bar width will produce similar responses. The large plateau in the correlation space is due to the relative lack of features in the image, and thus peak finding algorithms will fail.

An additional problem with the correlation-based technique comes from the fact that the background of the stimulus does not have an intensity of zero. Thus for any x position away from the actual bar, there are peaks in the correlation space that arise simply from the correlation of a higher intensity region (the bar) with lower intensity region (the background). Peaks in the correlation surface do not necessarily correspond to true surface structure.

This effect of a nonzero background resulting in a false correlation surface suggests that some sort of initial bandpass filtering should be applied to the raw images to reduce areas of constant or slowly changing intensity to zero. This would require that the correlation surface be computed at a number of different scales to be a true measure of objects in the environment. Figure 4 shows the application of the correlation detector to the bar (Fig. 2) which had previously been bandpass filtered. For this example, the windowing function was chosen to be particularly tightly localized in space in order to highlight problems inherent in the general approach. As there is no formal theory behind how the low-pass filter should be chosen, many filters would be appropriate. Filters with a lower high-frequency cut off will produce better results; however, there is no theory to guide the choice of such a filter.

A number of effects are apparent. The correlation surface is no longer constant over the width of the bar; rather it peaks (in x) near the edges of the bar. This is to
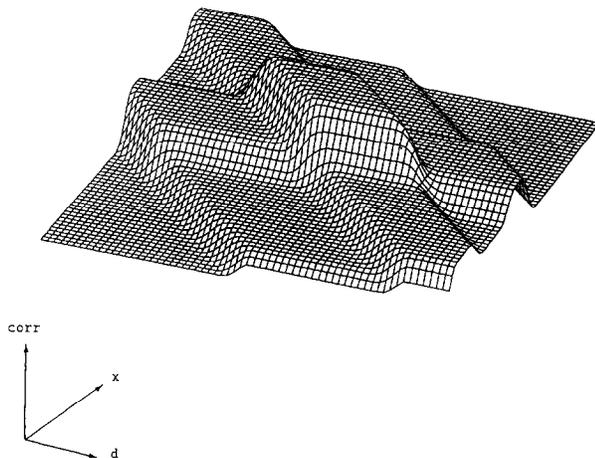
FIG. 3. Correlation (x, d) surface for bar target. Disparity is measured in terms of the left eye's view. Commonly, peaks in slices in d (horizontal axis) are considered as the true disparity for the given x coordinate.
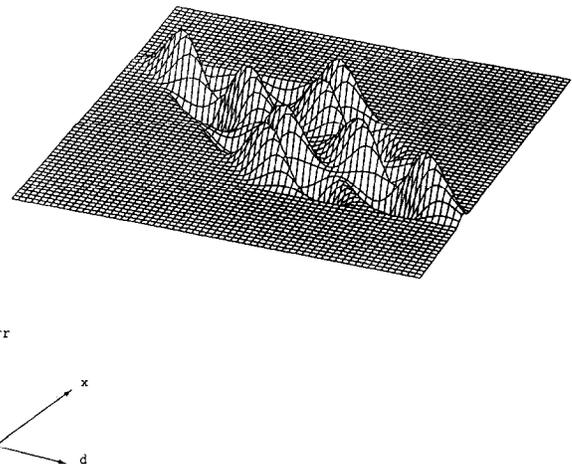
FIG. 4. Correlation surface after initial bandpass filtering. Note that the broad shoulders in Fig. 3 have been successfully removed, but that a number of smaller peaks remain.

be expected as bandpass filtering will tend to emphasize the bar's edges and reduce the weight of the bar's interior. In addition, there are a number of spurious peaks which might be reduced by the application of a low-pass final filter. The blind application of correlation to the output of bandpass filtered versions of the raw images does not appear to produce a useful response.

Much more sophisticated correlation-based techniques are available. Normalized cross-correlation, for example, replaces the simple correlation of (1) with the correlation of zero-mean signals, normalized by their variances. (See Witkin et al. [36] for an application of normalized cross-correlation.) Although more sophisticated techniques are available for correlation-based processing, fundamental problems remain. In particular, the notion that the disparity of a signal can be captured as a single process operating over all scales can be easily confounded by images in which the processes that generate structure operate only locally in scale. Rather than there being one single disparity accounting for all of the structure at a given location, a number of disparities exist, each accounting for a different amount of the structure at a given point in an image. Witkin et al. for example, apply normalized cross correlation to different blurred versions of the image in order to overcome this basic problem.

As a technique for measuring disparity, correlation has been shown to have a number of problems. It has difficulties in dealing with

- perspective projection.
- interocular illumination differences.
- disparity gradients.
- too much structure in an image.
- too little structure in an image.

By adding preconvolution steps, or blurring windowing functions, a number of problems may be surmounted. However, the process of adding these bells and whistles becomes rather ad hoc, and the underlying theory becomes lost in the use of image and correlation space operators whose functions are at best of a "heuristic" nature. Rather than continuing with the addition of unsupported features to correlation algorithms, let us turn our attention to the correspondence algorithms proposed by computer vision researchers.

2.2. Correspondence

Current correspondence-based theories of stereopsis involve three distinct stages (Arnold [1], Barnard and Thompson [2], Grimson [11], Lowrie [19], Marr [20], Marr and Poggio [21, 22], Mayhew and Frisby [24], Ohta and Kanade [27], Richards, Nishihara, and Dawson [31], Pollard, Mayhew, and Frisby [29]). First, the two images of a stereo pair are processed separately to extract monocular features. One common choice of feature is the

presence of a zero-crossing in a bandpassed version of the image (often referred to as an edge). Second, the monocular features in one image are matched with corresponding features found in the other image. In practice this second stage cannot be expected to produce only the correct matches, and a third stage must be considered in order to remove the incorrect matches ("false targets"). There are therefore three main issues in the design of a traditional correspondence based algorithm for stereopsis, namely (i) the choice of image features; (ii) the choice of matching criteria; and (iii) the way false targets are avoided or eliminated.

There are several factors involved in the choice of suitable image primitives. We consider these factors from the traditional viewpoint of extracting symbolic monocular features. Of primary importance is that an extracted feature can be expected to correspond to a particular property of a surface in the scene being viewed, and that this surface property is likely to produce the same type of feature in both images. In other words, we wish to use features that can be expected to produce reliable information about surface structure once they have been correctly matched. It is also important that matches occur sufficiently often to provide a fairly dense description of the disparity. While algorithms are available for filling in expected values of disparity given only sparse data (Terzopoulos [34]), it is obviously preferable to have denser data. Of course, the density of any measurement will be limited by the surface texture. Smooth surfaces will not give rise to as rich a description as surfaces with complex textures.

Finally, the choice of image features strongly affects the options available for obtaining matches and eliminating false targets. This has been expressed very clearly by Marr:

> The basic problem to be overcome in binocular fusion is the elimination or avoidance of false targets, and its difficulty is determined by two factors: the abundance of matchable features in an image and the disparity range over which the matches are sought. If a feature occurs only rarely in an image, the search for a match can cover quite a large disparity range before false targets are encountered, but if the feature is a common one or the criteria for a match are loose, false targets can occur within quite small disparities.[2]

In brief, the features (i) should correspond to surface properties, (ii) should produce matches that are fairly dense, and (iii) should have distributions over a typical image such that false matches can be relatively easily avoided or eliminated.

The constraints on the density and on the ease of eliminating false targets are in direct opposition. In particular, the number of possible matches in a given region increases polynomially with the density of a given symbolic

2 Marr [20, p. 127].
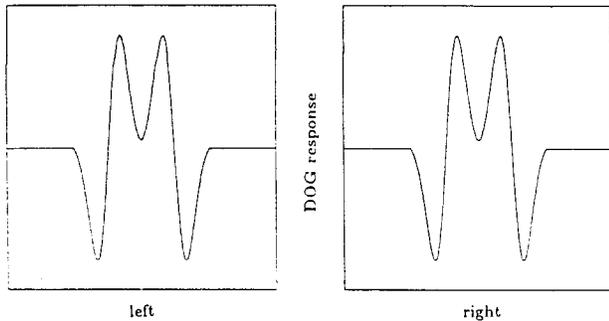
FIG. 5. DOG response for bar target.

feature. Therefore the problem of finding the correct match can be expected to rapidly become more difficult as the density grows.

Many different features have been chosen as a basis for disparity measurement. Perhaps the simplest is the image intensity itself. This makes the technique very similar to the correlation process with the window function W replaced by the delta function. This choice of monocular feature is quite attractive (if for no other reason than its simplicity of computation) for restrictive domains such as random dot stereograms (e.g., Marr and Poggio [21]). For more complex images, it suffers from exactly the same problems as does the correlation method, with the added burden that the use of the delta function for windowing exaggerates the problems.

The use of feature points as monocular primitives has been a favorite approach of computer vision researchers, and provided that the monocular features are appropriately chosen (by human intervention, for example), this approach can be very successful. Automated selection of feature points, such as points of high contrast (Barnard and Thompson [2], Moravec [25]) or corner points (Nagel and Enkelman [26]) have also been used. Unfortunately feature points have been found to be quite unstable (see Thorpe [35] for a comparison of feature point operators). More recent results with feature detectors have shown promise with corner-based detectors (Förstner [10]). However, fundamental problems with features remain. Classification as to feature type can easily change, the feature may not exist at all in a particular image pair (e.g., a corner-based detector might be applied to an image lacking corners), or a feature may not be detectable in one image at all due to interocular differences in image intensities or geometry.

Marr and Poggio [22] were the first to suggest the use of zero-crossings detected in the second derivative of the image as monocular features. Later work by Arnold [1], Grimson [11], Lowrie [19], Mayhew and Frisby [24], and Richards, Nishihara, and Dawson [31] have expanded upon this notion by modifying the types of filtering and the features extracted. However, the basic computa-

tional paradigm has remain unchanged: A spatial-frequency bandpass filter is applied to the image, and a number of features are extracted from the result. Current algorithms select either zero-crossings, or zero-crossings and peaks, at a number of different spatial frequencies as their monocular primitives.

Spatial-frequency tuned channels have been chosen for a number of reasons. Many computer vision algorithms claim biological relevance; that is, they draw support for specific computational processes from findings in physiology and psychology. There is a large body of literature supporting the notion of independent stereo processing from spatial-frequency tuned channels (Hubel and Wiesel [13]; Blakemore, Fiorentini, and Maffei [3]; Poggio and Fischer [28]). From a purely computational point of view structure is often localized in frequency space. As Marr noted,

The spatial organization of a surface's reflectance function is often generated by a number of different processes, each operating at a different scale.[3]

The notion of extracting primitives from bandpass filtered images gives rise to a symbolic pattern matching task as opposed to numerical correlation techniques. Problems that are encountered in correlation techniques due to interocular lighting differences may be reduced if tokens such as zero-crossings from bandpass channels are used. Unfortunately, token-based correspondences give rise to their own problems. (See Jenkin and Kolers [15] for a review of problems with correspondence.) Features may not be extracted from both eyes' views with equal ease, and the features themselves may be highly sensitive to small local perturbations in image intensity. The task of designing matching rules that are robust in the presence of unstable features, while not simply considering all interocular correspondences, is a very difficult problem.

The use of bandpass channels localizes structure in frequency space. Tilted surfaces can be detected so long as the left and right views of the local structure fall within the same channel. Note, however, that if the disparity gradient becomes too large, or if the spatial-frequency tuning becomes too narrow, the correspondence algorithm will be unable to respond to the surface.

Consider the application of a correspondence algorithm to the bar shown earlier in Fig. 2. Figure 5 shows the matching strengths for all possible (x, d) positions. The correspondence-based algorithm correctly identifies one end of the bar target as having zero disparity, while the other end of the target has a nonzero disparity, but makes no inference whatsoever as to the disparity of the intervening region. Although the bandpass filter was chosen so that the zero-crossing found would correspond to

---

[3] Marr [20, p. 46].
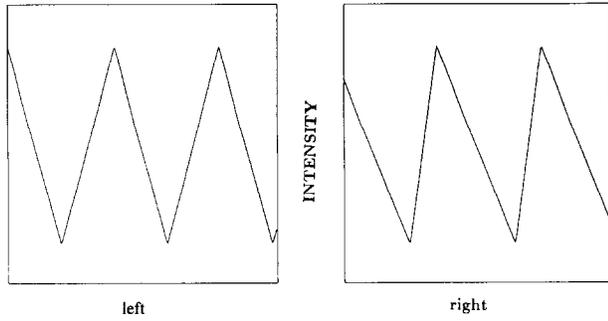
FIG 6. Left and right views of saw-tooth surface. The plot is of intensity as a function of position. Note that the eyes see surfaces of similar frequency but with differing profiles.
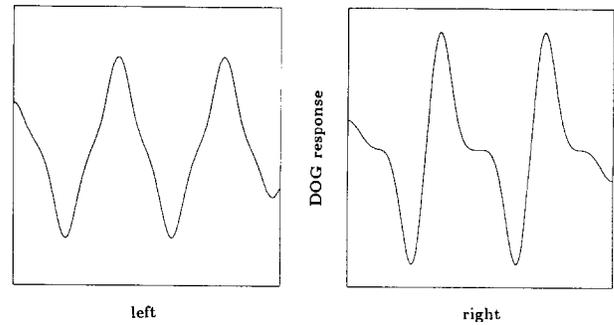
FIG. 8. DOG applied to left and right views.

the edges of the bar, lower spatial frequencies produce responses that have little to do with the actual edges of the bar. In addition, zero-crossings will be located within the structure of the bar due to the frequency of the band-pass filtering. Thus feature matches will occur between zero-crossings that correspond to the edges of the bar and zero-crossings detected within the bar. Figure 5 has zero-crossings at each end of the bar at the correct disparities. A different choice of spatial frequency tuning for the DOG operator may result in the formation of additional zero-crossings.

Mayhew and Frisby [24] reported that surfaces in which intensity changes were coupled to changes in disparity posed difficult problems for zero-crossing based algorithms. Figure 6 is such a surface. The left and right saw-tooth waveforms are of the same frequency but with slightly different profiles. (Note that the left and right images are plots of *intensity*, not depth.) The true depth of the surface is saw-tooth in nature. The "correct" surface disparity is shown in Fig. 7. If standard difference of Gaussians (DOG) filters are used, a simple symmetry argument shows that the exact zero-crossing always occurs halfway between the peaks. The matching algorithm therefore produces only a constant disparity response; the variation in depth is not detected. Moreover, the slope of the zero-crossing in higher frequency channels is

extremely shallow. This is exhibited in Fig. 8, where the left and right images are plotted after being passed through a DOG filter whose peak frequency is set at twice the fundamental. The shallow slope indicates that the positions of the zero-crossings are extremely sensitive to discretization errors and noise. In the left eye's view the zero-crossings will be very difficult to detect, and the slope of the response is almost zero through all zero crossings. In the right eye's view, 50% of the zero-crossings will be easy to detect, while the remaining 50% again suffer from a near-zero slope through the zero-crossings. As matches must be made between these two views, it is unlikely that good binocular matches can be constructed.

Note that changing the spatial-frequency tuning of the bandpass filters does not seem to improve the performance of the correspondence algorithm. As Mayhew and Frisby [24] reported, and as we have found, higher spatial-frequency tuned channels do not produce significantly different results. Figure 9 shows the matches of zero-crossings to the saw-tooth surface at different spatial frequencies. As can be seen, information gathered over three octaves does not encode a large amount of the

FIG. 7. True disparity of the saw-tooth surface.
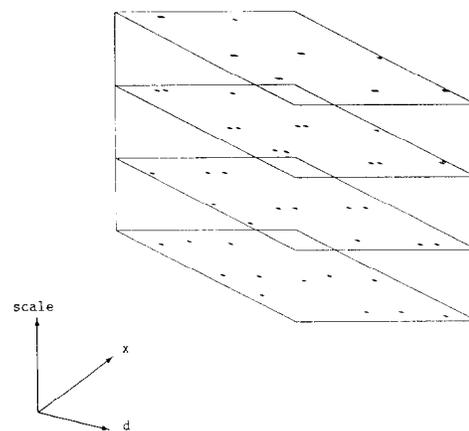
FIG. 9. Zero-crossings at different spatial frequencies. The superimposed planes (x, d) show the locations of zero-crossing across spatial frequencies. The spatial frequency scale is logarithmic.

structure in the saw-tooth surface. It would be difficult to imagine how these zero-crossings could be integrated to recover the true surface structure.

Mayhew and Frisby have reported more successful results using both zero-crossings and peaks from the output of the bandpass filtering. Indeed when this information is added to the correspondence algorithm, more matches are recovered. Unfortunately, peaks are generally more difficult to detect accurately in the presence of noise, and due to the discrete nature of the responses, very few response points are recorded for the underlying disparity surface. The underlying problem is that the extraction of tokens from a bandpass channel severely undersamples the output from a spatial-frequency tuned channel. Any disparity inferred from these tokens will also undersample the surface structure under examination. Although by an appropriate choice of channels tuned to different spatial frequencies it may be possible to recover this structure, it seems a waste to discard useful information from each bandpass channel. Rather than utilize only isolated features from the channel, and thus possibly lose information and at the very least complicate the task of "filling in" missing disparity values, it would be useful to base a disparity measurement primitive on the entire output of the spatial-frequency tuned channel.

2.3. Summary

Many correspondence mechanisms for disparity measurement implicitly assume that surfaces are frontoparallel, and that the imaging system samples using an orthographic projection. These assumptions allow the same surface feature to give rise to identical left and right monocular views of three-dimensional structure. Unfortunately, objects do not in general lie in frontoparallel planes, and imaging systems are not orthographic in nature. The same problem that plagues correlation-based systems (left and right views are not simply shifted versions of each other) also arises in correspondence-based systems. Other algorithms, which allow for tilted surfaces, still assume that the processes that gave rise to a feature in one view are the same processes that gave rise to the feature in the other view. Sophisticated monocular primitives extracted from bandpass channels may be less sensitive to surfaces with nonzero disparity gradients or interocular lighting effects, but these problems will still exist and cannot be completely avoided.

Perhaps the major advantage of correlation techniques is that correlation promises a high density of responses. Theoretically, correlation is capable of providing a depth determination at each and every image point. Techniques which use zero-crossings in a spatial-frequency tuned channel can by definition only respond twice over each wavelength. This lower density of response requires an even more robust process to "fill in" the missing values,

and solving this problem has not proven particularly easy.

Even with the very simple stimuli presented here, neither the correlation nor the correspondence techniques have proven very successful. Correlation responses may have a high density, but they are highly susceptible to either too little or too much structure in an image. Correspondence provides a more robust set of responses, but the responses are sparse, and some later process must be postulated to "fill in" the missing responses. As illustrated here, and as demonstrated by Mayhew and Frisby, surfaces in which intensity changes are tied to disparity changes pose serious problems for zero-crossing based approaches. The practice of extracting sophisticated primitives from bandpass channels as matching tokens seems to create as many problems as it solves.

Both correspondence- and correlation-based schemes have difficulties when applied to certain simple one-dimensional structures. Even without addressing the complex task of determining overall surface structure, current techniques may miss gross surface structure. In terms of human binocular vision, Mayhew and Frisby [24] have shown that saw-tooth surfaces (such as the one used in Fig. 6), are perceived as saw-tooth structure in depth. Thus there is evidence that the correct interpretation can be obtained, but that current algorithms are unable to detect this structure. Some new technique for disparity measurement is required.

Correlation and correspondence approaches to stereopsis are in a sense complementary; correlation produces a dense set of responses but has difficulty with constant or rapidly changing structure and with interocular image differences, while correspondence avoids some of these problems by considering the image at different scales but then fails to obtain a dense set of responses by matching only sparse tokens. It would be desirable both to avoid the problem of structure at different scales and at the same time to produce a dense response. The development of such a measure is the goal of the next section.

3. DISPARITY AS PHASE DIFFERENCE

In this section we present our basic computational approach used to measure local disparity information. To illustrate the proposed technique in its simplest form we consider one-dimensional sinusoidal signals here. We note, however, that only relatively minor modifications of this basic technique will be needed to deal with more general two-dimensional images.

In particular, we let $I_l(x)$ and $I_r(x)$ be the left and right "images," where

$$I_l(x) = A \sin(\omega_l x + \theta_l); \quad A, \omega_l > 0,$$

$$I_r(x) = B \sin(\omega_r x + \theta_r); \quad B, \omega_r > 0. \tag{4}$$

Note that, since different perspectives in the left and right views can alter the spatial frequencies of the observed patterns, we do not assume that $\omega_l = \omega_r$. However, we do assume that $\omega_l \approx \omega_r$ (i.e., $|\omega_l - \omega_r|/(\omega_l + \omega_r) \ll 1$), and in practice we ensure this by considering two bandpassed versions of the left and right raw images.

One way to define a local phase difference between two bandpass signals is to consider matching "features" such as peaks and zero-crossings. For I_l and I_r as above, this basically amounts to matching the arguments of the two sinusoids, modulo 2π. Thus, we define the local phase difference to be

$$\phi(x) = \lfloor(\omega_l - \omega_r)x + (\theta_l - \theta_r)\rfloor \in [-\pi, \pi). \quad (5)$$

Here $\lfloor\theta\rfloor$ denotes the principal part of the angle θ, which is obtained by adding an integral multiple 2π to θ so that the result lies in the interval $[-\pi, \pi)$. Note that this mod operation produces a discontinuous function, $\phi(x)$, whenever $\omega_l \neq \omega_r$. Moreover, the discontinuities correspond to boundaries of intervals in x where the matching of peaks and zero-crossings between the two images is one-to-one.

The local disparity, $d(x)$, is the distance the images must be shifted with respect to each other so that I_l and I_r agree up to a multiplicative constant (B/A in the above example).[4] This definition leads to the following form for the local disparity;

$$d(x) \equiv \frac{1}{\bar\omega} \phi(x), \quad \bar\omega \equiv \frac{1}{2}(\omega_l + \omega_r) \quad (6)$$

With this definition it follows that the left image shifted to the right by $\frac{1}{2}d(x)$, namely $I_l(x - \frac{1}{2}d(x))$, is a constant times the right image shifted by the same amount to the left, namely $I_r(x + \frac{1}{2}d(x))$. Positive disparity, therefore, corresponds to objects that are *farther* from the fixation point of the two eyes.

In order to extract $\phi(x)$ we consider the point-by-point multiplication of the left and right images, that is,

$$I_lI_r = AB \sin(\omega_l x + \theta_l)\sin(\omega_r x + \theta_r)$$

$$= \frac{AB}{2}(\cos((\omega_l - \omega_r)x - (\theta_l - \theta_r))$$

$$- \cos((\omega_l + \omega_r)x + (\theta_l + \theta_r))). \quad (7)$$

By low-pass filtering this product with a filter, L, having a high-frequency cut-off below $\omega_l + \omega_r$, we obtain

---

4 We assume frontoparallel surfaces in our development of the disparity detectors; in a later section we will generalize the result to surfaces with nonzero disparity gradients.

$$P(x) \equiv L * (I_l(x)I_r(x))$$

$$= \frac{ABK}{2} \cos((\omega_l - \omega_r)x + (\theta_l - \theta_r))$$

$$= \kappa \cos(\phi(x)), \quad \kappa \equiv \frac{ABK}{2}. \quad (8)$$

Here $\phi(x)$ is as in (5) and $K = K(\omega_l - \omega_r)$ is the sensitivity of the low-pass filter to the frequency $\omega_l - \omega_r$. This result is encouraging since $P(x)$ depends only on the desired local phase difference and a scale factor involving the product of the amplitudes. However, a *local* technique is needed to disambiguate the amplitude and the relative phase information inherent in $P(x)$. The technique must be local since we cannot expect bandpassed images to be well approximated by (4) over intervals of roughly the length of one wavelength of $P(x)$, that is, over lengths $2\pi/|\omega_l - \omega_r|$. Therefore it is not possible to estimate the scale factor by examining the behavior of $P(x)$ over one of its wavelengths. Instead it is necessary to obtain an estimate for $\phi(x)$ based on information available over no more than a few wavelengths of the base frequency $\bar\omega$ (i.e., $\Delta x \simeq 2\pi/\bar\omega$).

A suitable local technique can be obtained by the inclusion of a specific shift in one (or both) of the images before the pointwise product is calculated. For example, we define

$$P(x, s) \equiv L * \left(I_l\left(x - \frac{1}{2}s\right)I_r\left(x + \frac{1}{2}s\right)\right)$$

$$= \kappa \cos(\bar\omega s - \phi(x)), \quad (9)$$

with $\bar\omega$ as in (6). Now, to obtain an approximation for ϕ near a given point x_0, the value of s is adjusted so that a particular local feature of $P(x, s)$ occurs at $x = x_0$. Suitable local features of P are, of course, features that are independent of the product AB, and include zero-crossings and local extrema. Here we chose to avoid seeking local extrema of $P(x_0, s)$, which can be fairly sensitive to noise. Instead we consider a simple feedback control which provides a value $s_\infty(x_0)$ such that

$$P(x_0, s_\infty(x_0)) = 0,$$

$$\frac{\partial P}{\partial s}(x_0, s_\infty(x_0)) > 0. \quad (10)$$

For the moment assume that such an $s_\infty(x_0)$ is computed. Then we see from (9) that (10) is satisfied if and only if

$$\lfloor\bar\omega s_\infty(x_0) - \phi(x_0)\rfloor = -\frac{\pi}{2}. \quad (11)$$

Therefore the local relative phase difference is given by

$$\phi(x_0) = \left\lfloor \bar{\omega} s_\infty(x_0) + \frac{\pi}{2} \right\rfloor, \qquad (12)$$

which can be computed given $s_\infty(x_0)$ and the mean frequency $\bar{\omega}$.

The basic idea behind the control loop is best illustrated with the following simple scheme. Consider $s(t; x_0)$, which is defined to be the solution of the differential equation

$$\frac{ds}{dt} = -P(x_0, s(t; x_0)), \quad t > 0,$$

$$s(0; x_0) = s_0. \qquad (13)$$

Here s_0 is some initial guess for the local phase difference at x_0. The meaning of (13a) is made clear in Fig. 10. Roughly speaking, Eq. (13) simply states that if $P(x_0, s)$ is positive then s should be decreased, and if $P(x_0, s)$ is negative then s should be increased. Except for the unstable situations where the initial guess $s(0; x_0)$ precisely satisfies (10a) but with the inequality (10b) reversed, the solution $s(t; x_0)$ of (13) converges to a solution of (10) as $t \to \infty$. We refer to this limit as $s_\infty(x_0)$.

If the mean frequency $\bar{\omega}$ is known, then the relative phase can be computed from (12) and the limit $s_\infty(x_0) = \lim_{t \to \infty} s(t; x_0)$. In addition it follows from (5) and (12) that the local disparity $d(x_0)$ satisfies

$$d(x_0) = s_\infty(x_0) + \frac{\pi}{2\bar{\omega}} = s_\infty(x_0) + \frac{\lambda}{4}. \qquad (14)$$

In practice $\bar{\omega}$ might be approximated by ω_0, the peak frequency of the bandpass filter used to compute $I_l(x)$ and $I_r(x)$. The error in the disparity calculated using (14), with $\bar{\omega}$ replaced by ω_0, is then

P(x,s)
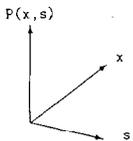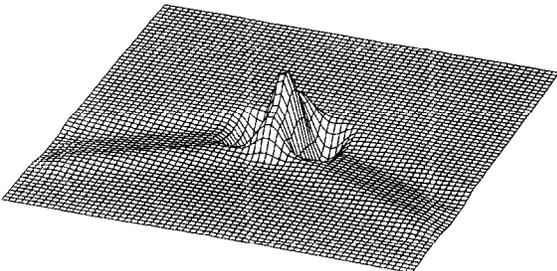
FIG. 10. $P(x, s)$ for a step edge at zero disparity.

$$\Delta d = \left(\frac{1}{\omega_0} - \frac{1}{\bar{\omega}} \right) \simeq \frac{1}{4} \left(\frac{\bar{\omega} - \omega_0}{\omega_0} \right) \lambda_0, \qquad (15)$$

where $\lambda_0 = 2\pi/\omega_0$ is the wavelength corresponding to the peak frequency, and $(\bar{\omega} - \omega_0)/\omega_0$ is the relative displacement of $\bar{\omega}$ from the peak frequency. For example, if the bandwidth of the bandpass filter is less than an octave, then the relative displacement of $\bar{\omega}$ is bounded by $\frac{1}{2}$, and (15) implies

$$|\Delta d| < \frac{1}{8} \lambda_0. \qquad (16)$$

We note, however, that for images in which the power spectrum is relatively smooth near the peak frequency ω_0, the appropriate value of $\bar{\omega}$ will be closer to ω_0 than is indicated above. In this case we can expect to obtain results whose accuracies are well within the bound provided in (16). This completes the description of one technique for measuring the disparity d. We end this section by describing a modification of this technique which provides the disparity to a greater accuracy.

Note that the error estimate in (16) is based entirely on the uncertainty in the value of $\bar{\omega}$, which is needed to compute the disparity from the limit, $s_\infty(x_0)$, of the loop filter. The technique illustrated below uses a loop filter whose limit is precisely the local disparity. The value of $\bar{\omega}$ is not needed, and the error produced by its use is thereby eliminated. The key to this approach is an appropriate choice of the initial filtering. In particular, suppose the left and right images I_l and I_r are obtained using a zero-phase and a $-\pi/2$-phase filter, respectively. For example, sine and cosine Gabor filters could be used. (The Gabor pair must be slightly modified so that both the sine and cosine Gabors are bandpass.) Then the right image becomes

$$I_r(x) = A \sin\left(\omega_r x + \theta_r - \frac{\pi}{2} \right). \qquad (17)$$

Forming $P(x, s)$ as in (9), with $I_l(x)$ as in (2.1a), we find

$$P(x, s) = \kappa \sin(\bar{\omega} s - \phi(x)). \qquad (18)$$

Now the control scheme (13) provides $s(t; x_0)$ such that $s(t; x_0) \to s_\infty(x_0)$ as $t \to \infty$, where $s_\infty(x_0)$ satisfies

$$[\bar{\omega} s_\infty(x_0) - \phi(x_0)] = 0. \qquad (19)$$

Finally, we see that if $\bar{\omega} s_\infty(x_0) \in [-\pi, \pi)$ then (4), (5), and (19) imply that $s_\infty(x_0) = d(x_0)$. Therefore the control scheme converges to the local disparity, and the precise value of $\bar{\omega}$ is not needed. We will use Gabor filtering in the following experiments with the disparity detector.

3.1. *Other phase difference methods*

Phase-based methods have been considered previously in the matching of Fourier transforms between image pairs (Kuglin and Hines [18]; Stokes [33]). Both methods are similar to the phase matching proposed here, in that the task of measuring disparity is reduced to the problem of measuring the phase difference between two signals. Stokes has shown that if the phase is measured over a long input signal (about twice the wavelength), very poor values for the phase shift may be reported. He has also demonstrated that noise can seriously affect the results obtained by phase matching over larger signal lengths.

The method proposed here differs from these previous methods in that we measure the phase difference locally. The computation of phase difference in our work has no similarity to Fourier-based schemes, and the density of the sample points that are considered using the phase-based scheme is considerably higher than once every other wavelength.

Sanger [32] has also considered disparity computation using phase measurements obtained from Gabor filters. His technique involves a direct measurement of the phase in the left and right signals, and then a subtraction. This results in a much smaller disparity range than the technique proposed here. We have also considered the use of Gabors and different techniques of disparity measurement (Jepson and Jenkin [16]).

3.2 *Phase-Sensitive Demons*

The general form of the phase measurement scheme discussed above is given in Fig. 11. We summarize this scheme below. The raw images, I_l^0 and I_r^0, are first sent through bandpass filters B_l and B_r, respectively, to produce

$$I_\nu(x) = \int_{-\infty}^{\infty} B_\nu(x - z)I_\nu^0(z)dz, \quad \omega = 1, r. \quad (20)$$

The left and right bandpass image is shifted by $s/2$, and the phase detector,

$$P(x, s) = \int_{-\infty}^{\infty} L(x - z)\left[I_l\left(z - \frac{s}{2}\right)I_r\left(z + \frac{s}{2}\right)\right] dz, \quad (21)$$

is applied. Here L is a low-pass filter. The response of the phase detector is fed into the loop filter, which for now we take to be

$$\frac{ds}{dt} = -P(x, s), \quad t > 0. \quad (22)$$

It is also convenient to introduce an initial guess, say

$$s(0; x) = s_0. \quad (23)$$

The value of s computed in (22) is fed back to the shift module. The loop is said to be *locked* when $P(x, s) = 0$, and in this case we set $s_\infty(x) = s$. We refer to this mechanism as a *phase-sensitive demon*.

We imagine the initial stereo images decomposed into a multiple spatial scale representation, perhaps through the use of a Burt pyramid (Burt [4]) or the DOLP transform (Crowley and Stern [7]). Also there is a similar hierarchical scheme that can be used to construct Gabor-like receptive fields (Fleet and Jepson [8]). The filter B_l and B_r are members of such a family. For each spatial scale there is a demon applied at each sample point x and, like all good demons, they are independent of other demons at different x's or different scales. In particular, we emphasize that the demons are simply collecting *local* disparity information. A globally consistent interpretation of the computed local disparities is the job of either a subsequent level of processing or a network connecting demons to their neighbors.
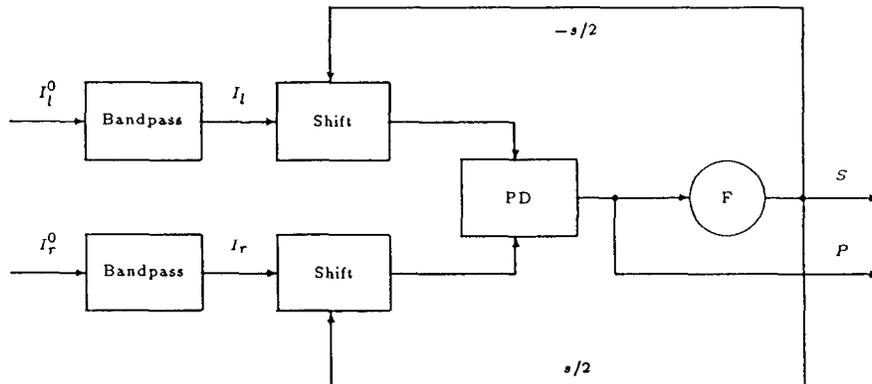
FIG. 11. Feedback computation of P. F is a loop filter which modulates the output of the phase detector PD.

3.3. *Dealing with a disparity gradient*

The disparity demons that we have presented are tuned for a specific spatial frequency, a specific retinal location, and a specific disparity. In addition, these disparity demons are tuned to a particular disparity gradient ($\beta_1 = 0$). As with the correspondence-based algorithms using features extracted from bandpass channels, the use of bandpass signals will allow surfaces with nonzero disparity gradients to be processed. The limit of the disparity gradient will be controlled by the width of the bandpass filter used. In this section we will sketch how the demons can be modified to measure surfaces with a nonzero disparity gradient.

Suppose that a detector is required that will be tuned for a particular disparity gradient β_1. Then from (3), $I'_1(x_1)$ = $I'_r(\beta_0 + (1 + \beta_1)x_1)$. Suppose that I'_1 is a sinusoidal function of x, say with frequency ω; then I'_r is also a sinusoidal function of x, but with a frequency $\omega/(1 + \beta_1)$, and with a phase shift of β_0. In order to apply a disparity measure similar to (9) the left and right images must be bandpass filtered to equal spatial-frequency channels. We can accomplish this by resampling the right (or left) eye's view so that the resampled images have similar central frequencies. Equation (20) is modified so that the right raw image is sent through a bandpass filter which has a central frequency $\omega_r = 1/(\beta_1 + 1)\omega_1$. The resulting bandpass channel is then resampled at $(\beta + 1)$. The result of this construction is that I'_1 and I'_r are bandpass signals with similar spatial-frequency tuning although they do not necessarily encode the same spatial-frequency channels from the raw images. We can then apply (21), (22), and (23) to measure disparity between I'_1 and I'_r. A locked response indicates the presence of a surface with a disparity gradient β_1 at the point in space for which the demon was tuned.

As an example, suppose that we wish to build a detector that is tuned to objects that have a disparity gradient β_1 of 1. Such an example is shown in Fig. 12. Instead of applying filters to the left and right views with the same frequency, we use filters with central frequencies such that the left frequency is twice that of the right. The phase detector requires that its input signals be of the
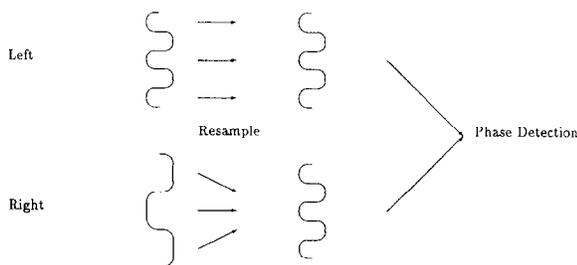
FIG. 12.   Nonfrontoparallel surface.

same frequency, so we must resample one (or both) of these signals. In Fig. 12 we chose to resample the right bandpass signal. We scaled the right signal by a factor of $\frac{1}{2}$ so that after resampling the left and right signals have the same spatial frequency. We can then apply our disparity detector to these two signals. In this way we can construct a detector tuned to a nonzero disparity gradient.

We note that other interocular differences can be measured in a similar fashion. Interocular orientation differences, contrasts, etc., can easily be encoded by a disparity sensitive demon. All that is required is that the signals available to the demon be of similar spatial frequencies along the axis that passes through the two eyes. The demon will operate independently of other information encoded in its response.

For example, to produce a demon which is sensitive to opposite contrasts in the left and right eyes view, we need only modify one of the initial bandpass filters (B_1, say) so that its output is inverted. To produce a demon that is sensitive to structure oriented at 45° in the left eye's view, and to structure oriented at −45° in the right, B_1 and B_r must be modified to be orientation specific to +45° and −45°, respectively. Of course, although the construction of demons tuned to a large number of interocular differences is possible, constraints on the total number of demons will exist in any environment (and in biological systems in particular).

3.4. *Discussion*

The underlying mechanism we have presented for disparity detection is one of measuring the local phase difference between two small windows in a bandpass channel. Computationally, this appears as a modification to the simple correlation scheme presented earlier: The images are cross-correlated, and a feature (in our case, a zero-crossing) is extracted. A number of important differences are apparent:

1. The initial image is not correlated directly, nor is a windowing function used in an attempt to blur the image to reduce noise in the correlation. Bandpass versions of the images are used. Thus we measure structure at different scales in the scene, rather than performing a global (across scales) measure. A correlation system that does not utilize measurements at different scales can easily be confounded by scenes that have conflicting structure at different scales.

2. Classic correlation is based upon the correlation of identical left and right signals. In the phase-based detector, the bandpass filtering in the left and right images need not necessarily be similar. For example, use of the Gabor pair discussed earlier results in the left and right images being convolved with radically different functions. If disparity gradient tuning is required, then left

and right images may not be convolved with filters having similar bandpass properties.

3. The output of the correlation is low-pass filtered. If this step is not performed then the high-frequency additive term in (7) will not be removed, and the resulting P space will not be sinusoidal. The detection of the correct (low-frequency) zero-crossing from within the high-frequency component without the low-pass filtering would be quite difficult, if not impossible. Such as surface is shown in Fig. 4.

4. Our disparity detectors have been constructed to extract the local relative phase difference between two signals. Correlation applied to bandpass signals followed by low pass filtering may produce similar results, but without a formal description of what is being extracted the results are difficult to interpret, and an intelligent choice of the low pass filter is not possible.

4. TESTING THE DISPARITY DETECTOR

Figures 13 and 14 show the response of our disparity detectors to the bar target (Fig. 2) and to the saw-tooth target (Fig. 6) respectively.
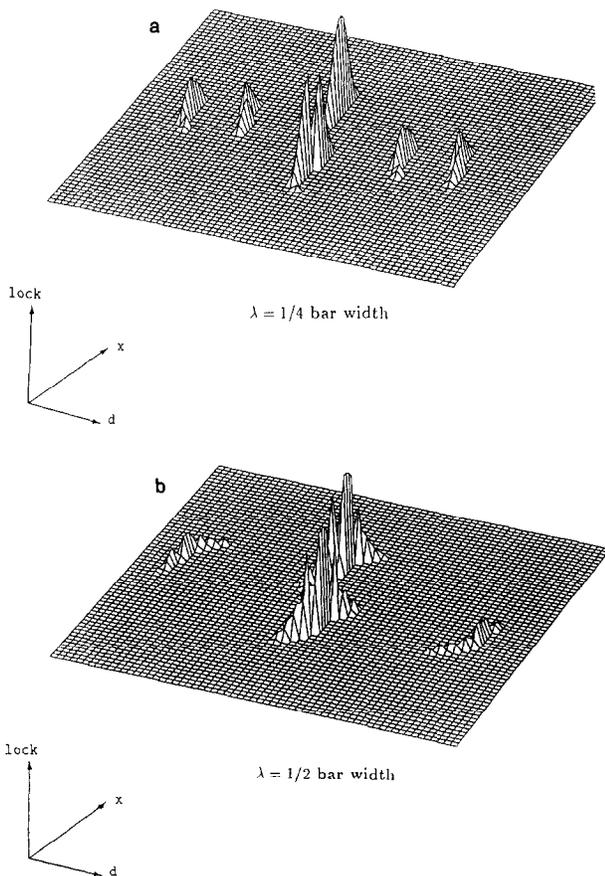
FIG. 13.   Disparity demon response to bar target. The height of the response is proportional to the slope in s as the surfce passes through 0 in $P(x, s)$ and is zero othewise.
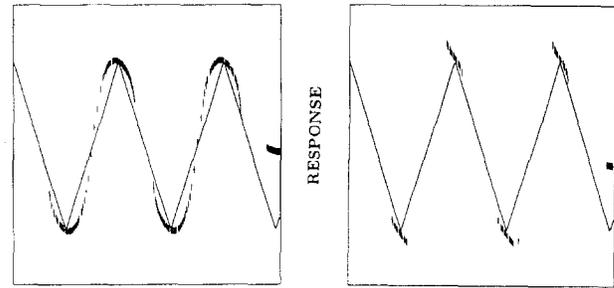
FIG. 14.   Disparity demon response to saw-tooth texture. Responses are shown for filters tuned to 4 and 8 times the base frequency of the saw-tooth function. The results are superimposed over the true surface disparity.

Consider the response to the bar first. The bandpass channels were chosen to be approximately one-half the width of the bar. The detector has responded at zero disparity at one end of the bar, and with nonzero disparity near the other end. Within the bar, the disparity detector has interpolated possible disparity values. In addition, at the ends of the bar, the strength of the disparity response falls off rapidly beyond the bar's edge. The falling off distance can be easily computed a priori from the spatial-frequency properties of the bandpass filtering. The disparity detector begins to address the problem of providing results in a region near discontinuities in intensity, rather than just at the discontinuities.

In addition to the correct zero disparity, responses were also obtained from points in s that were considerably different from zero. These points arise from the discrete nature of the digital implementation and undersampling of the $P(x, s)$ space. Note that these "false targets" could be eliminated by simple thresholding or the task of distinguishing these responses could be left to whatever later process deals with the general problem of false targets.

Figure 14 shows the responses near zero to the saw-tooth surface. Due to the regularly repeated structure, a large number of different interpretations are available. The same response pattern will be repeated in disparity space every multiple of the wavelength of the pattern. The structure near disparity zero, however, consists of a saw-tooth section. By utilizing the full bandpass signal in phase locking, structure can be detected that might otherwise be missed. Such missing structure can only further complicate later processes that might attempt to produce coherent descriptions of global surface structure.

Figure 15 shows the output of the disparity detector when it is applied to the saw tooth surface at different spatial frequencies. The disparity range in Fig. 15 is larger than the range in Fig. 9, and thus shows some of the false targets that occur due to the regularly repeated structure in the raw images. The output of the disparity
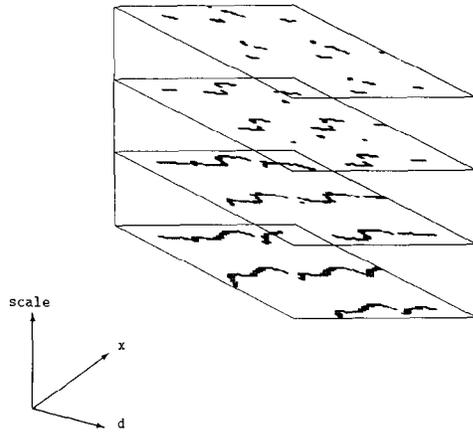
FIG. 15. Disparity demon at different spatial frequencies.

FIG. 17. DOG response for bar target.

detector should be compared with the output of the zero-crossing correspondence algorithm shown in Fig. 9. The disparity detector has recovered surface structure in considerably more detail than is available using sparse correspondence techniques. Higher spatial frequencies tend to encode information about the position of the peaks, while lower spatial-frequency tuned channels encode more of the gross structure of the surface.

The correlation based schemes perform reasonably well on the saw-tooth texture, successfully correlating the peak shapes between the two images. Unfortunately, the strength of the correlation depends on the intensity of the texture. That is, the strength of the response is less when the intensity of the signal is closer to zero.

Finally, as suggested by the reviewers, we consider the effect of noise on the three disparity measurement techniques. As a base case, consider a simple bar target with identical presentations to the left and right cameras (see Fig. 16). The central bar region has a higher intensity than the background, and the background has a nonzero intensity. Figure 17 shows the output of a difference of Gaussians operator to the stimuli shown in Fig. 16. As expected, two sharp zero-crossings can be found at the edges of the bar. Depending on the sensitivity of the zero-

crossing detection process and on the central frequency of the operator used, additional zero-crossings can also be detected. Note the difficulty in rejecting the false zero-crossings within the bar in Fig. 17, but accepting the true zero crosssings between the peaks in Fig. 8.
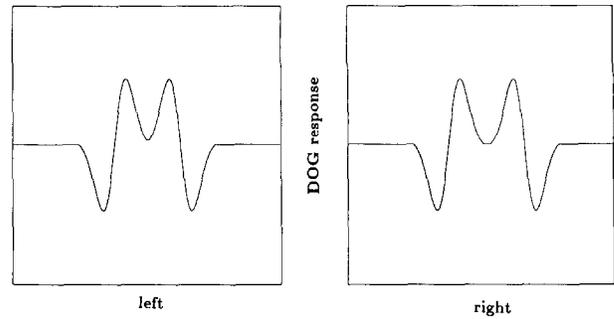
The output of a simple correlation process using a Gaussian windowing function with a support region considerably smaller than the bar width is shown in Fig. 18. As one would expect from the large featureless areas, the task of peak finding in the correlation surface would be quite difficult. Figures 19a and 19b show the positive zero-crossings in the $P(x, s)$ space for the phase difference operator at two different spatial frequencies (roughly 1/4 (a) and 1/2 (b) of the bar width). Strong responses are seen near the edges of the bar with smaller incorrect responses from false zero-crossings at incorrect disparities. These responses are consistent with the responses obtained with the tilted bar target earlier in the paper.

Now consider the effect of adding Gaussian white noise with a standard deviation of 20% of the intensity difference between the bar and the background (see Fig. 20). The general structure of the bar is visible, although

FIG. 16. Left and right intensity profiles of bar target.
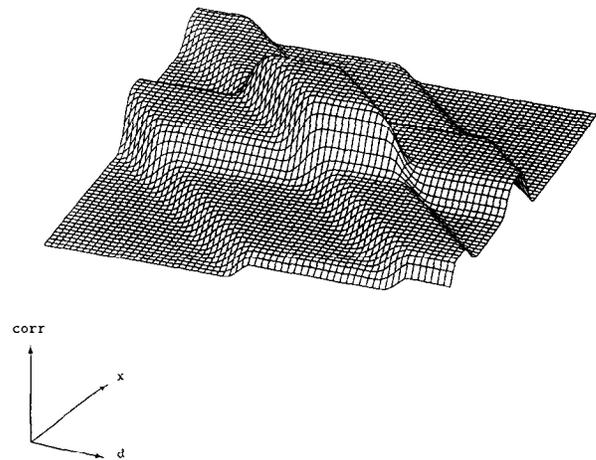
FIG. 18. Correlation surface for bar target.

λ = 1/4 bar width

lock

λ = 1/2 bar width

lock
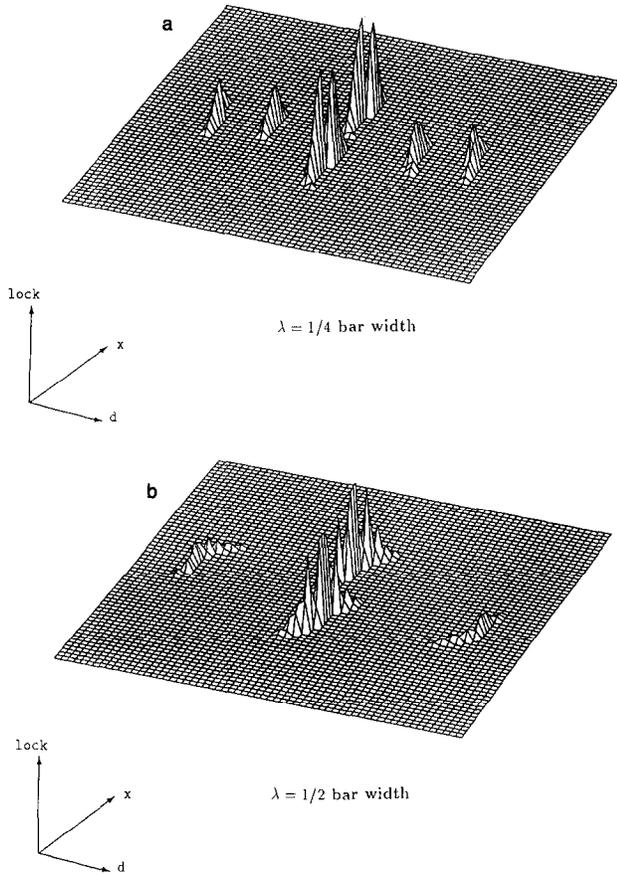
FIG. 19.   Phase difference response for bar target.

FIG. 21.   DOG response for corrupted bar target.

the fine structure is lost within the noise. The application of the DOG operator gives a large number of zero crossings (see Fig. 21) which will provide many possible matches. The task of sorting out the correct matches from the noise response will be quite difficult without the use of very strong high level constraints. The correlation space no longer shows the smooth structure of Fig. 18, but the new peaks introduced are simply the correlation of noise added to the two signals. The peaks do not correspond to the correct disparity of the bar, only to the

maximum correlation of the noise added to the bar (Fig. 22). The addition of random, uncorrelated texture does not simplify the correlation task. The phase difference technique responses are shown in Fig. 23. The noise introduces more false zero-crossings, but strong responses are still seen at the edges of the bar. Note that the lower spatial-frequency tuned channel has less noise responses. This also is expected as the noise tends to be of a high spatial frequency, and is attenuated by the use of band-pass channels.

As a final example, consider the case of noise added to a bar tilted in depth (such as the bar in Fig. 2). Previous figures show the action of the correlation- (Fig. 3), correspondence- (Fig. 5), and phase-based methods (Fig. 13). As before, we add Gaussian white noise to the left and right signals (Fig. 24) and present the output of the DOG operator to the stimuli (Fig. 25). As with the frontoparallel bar there are many zero-crossings in the signals which do not correspond to the edges of the bar. The correlation method also produces a result similar to that given in Figure 22, with noise adding spurious texture to the correlation surface with no relationship to the underlying structure (Fig. 26). Finally, the phase matching technique

left   right

FIG. 20.   Left and right intensity profiles of corrupted bar target.
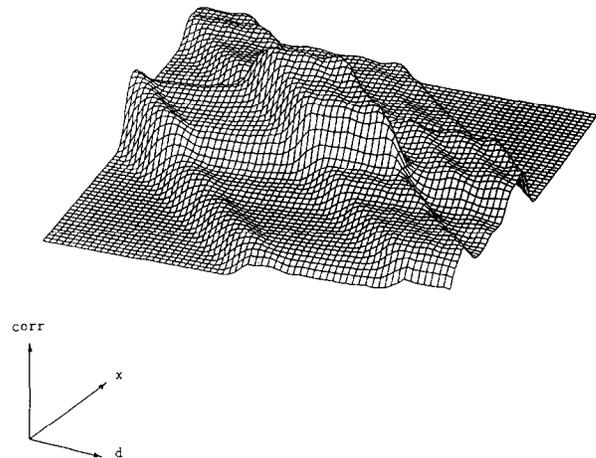
corr

FIG. 22.   Correlation surface for corrupted bar target.
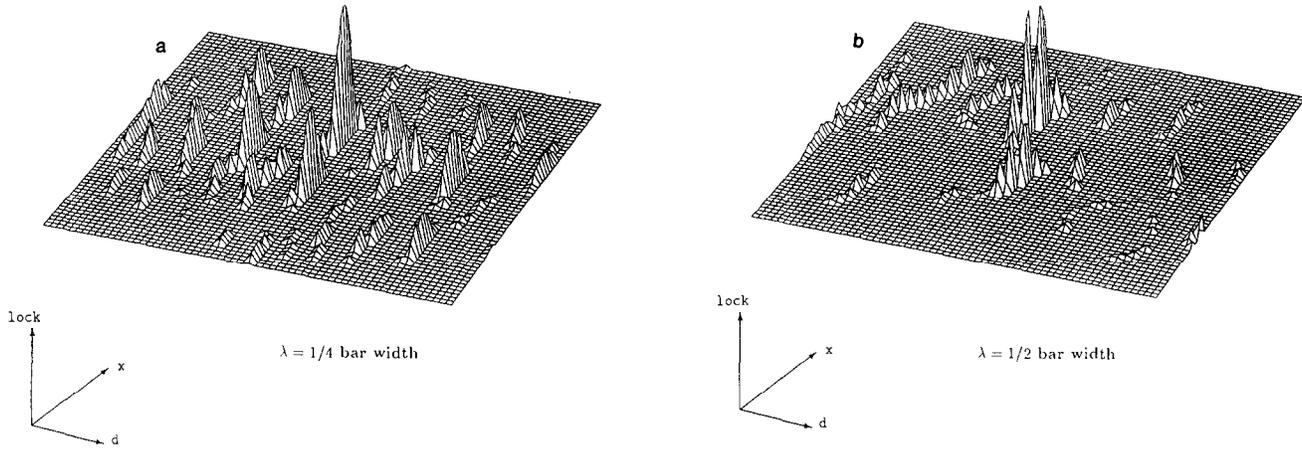
FIG. 23.   Phase difference response for corrupted bar target.

responses are shown in Fig. 27a, b. The results are similar to those shown in Fig. 23 with the exception that the tilted nature of the surface is more pronounced in Fig. 27a than in 27b. This is consistent with the results obtained with the saw tooth surface. At lower spatial frequencies, the finer details of the disparity surface are lost.

The three techniques compute the actual position of measured points in disparity space in very different ways. The actual position determined by the algorithms is dependent on the process used to find the peak or zero-crossing required. For example, to find the actual position of a zero-crossing, the surface could be locally interpolated and the exact position of the zero-crossing found from an explicit solution of the interpolant for its zeros.

The correspondence-based technique should find two matches (corresponding to the edges of the bar), and this is found in practice. When noise is added, a number (48 for both the bar and tilted bar experiments) of false matches occur. 46 of these matches would have to be removed by later levels of processing. For the correlation-based technique, the theoretical peak in the correlation space (65025) is almost exactly matched by (65024

and 65025) by the implementation. When noise is added the correlation space peaks in a number of places and the maximum peak is now much larger than the peak found by correlating the uncorrupted signals (72562 and 73455). An examination of Figs. 22 and 26 show that the correlation surface peaks due to the accidental correlation of noise in the left image with noise in the right.

The phase-difference-based technique has the advantage of using spatial-frequency tuned channels that reduce much of the effect of the noise. A number of false locks are found, but strong locks are still identified at the edges of the bar. Although more formal analysis is required, the phase difference method appears to perform quite well in the presence of noise.

5. SUMMARY

We have presented current techniques for the measurement of structure in space from disparate views. Current techniques can be classified according to their underlying computational task. We have identified two such tasks: correspondence or correlation. Using a few simple
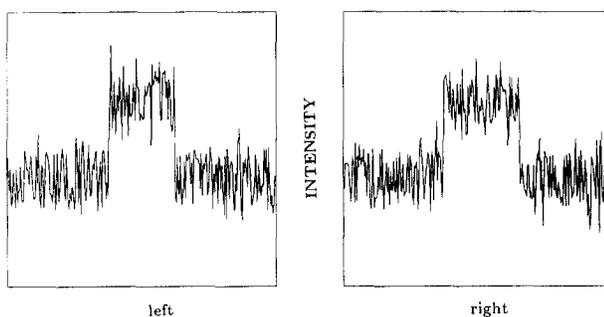
FIG. 24.   Left and right intensity profiles of corrupted, tilted bar target.
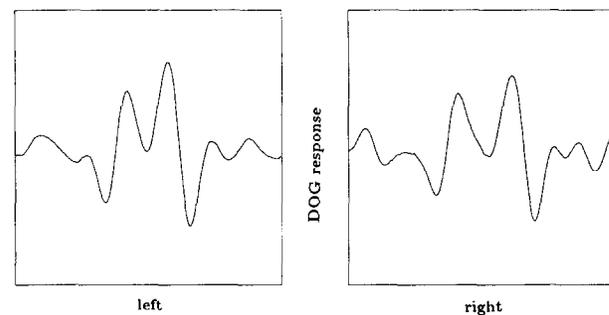
FIG. 25.   DOG response for corrupted, tilted bar target.
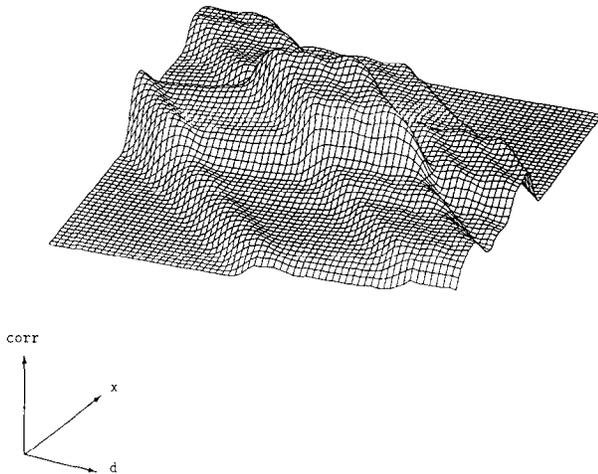
FIG. 26. Correlation surface for corrupted, tilted bar target.

dence-based algorithms, on the other hand, undersample the true disparity of the surface, and have difficulties when presented with surfaces in which intensity changes are tied to changes in disparity.

By developing a notion of disparity in a bandpass spatial-frequency tuned channel, we have constructed a disparity demon which can be tuned for particular disparities. \These disparity detectors can be applied independently and are capable of producing dense and robust responses to surfaces that can only be marginally processed using current techniques. In addition these detectors can be tuned to deal with interocular differences such as orientation, contrast, and nonzero disparity gradients that cannot be easily measured with current disparity measurement techniques.

ACKNOWLEDGMENTS

examples we have shown that neither of these techniques is particularly robust. Interocular differences in lighting and the effects of the perspective imaging of binocular systems pose difficult problems to correlation algorithms that fundamentally assume that the left and right eyes see simply shifted versions of the same structure. Correspon-

FIG. 27. Phase difference response for corrupted, tilted bar target.

REFERENCES

1. D. Arnold, Local context in matching edges for stereo vision, *Proceedings of the Image Understanding Workshop, May, 1978*, pp. 65–72.
2. S. T. Barnard and W. B. Thompson, Disparity analysis of images, *IEEE Trans. Pattern Anal. Machine Intelligence* PAMI-2 (4), 1980, 333–340.
3. C. Blakemore, A. Fiorentini, and L. Maffei, A second neural mechanism of binocular depth discrimination, *J. Physiol.* 226, 1972, 725–749.
4. P. J. Burt, Fast filter transforms for image processing, in *Multiresolution Image Processing and Analysis* (A. Rosenfeld, Ed.), Springer-Verlag, Berlin/New York, 1983.
5. J. Clark and P. Lawrence, *Scale Space Analysis of Stereo Disparity Errors*, Technical Report, Joint Robotics and Teleoperator Laboratory, No. 3, February, 1985.
6. M. A. Crombie, Coordination of stereo image registration and pixel classification, *Photogramm. Eng. Remote Sens.* 49, (4), April, 1983.
7. J. L. Crowley and R. M. Stern, Fast computation of the difference of low-pass transform, in *IEEE Trans. Pattern Anal. Machine Intelligence*, 1984.
8. D. Fleet and A. Jepson, The extraction of orientation and 2-d velocity through hierarchical processing, *Proc. SPIE—Int. Soc. Opt. Eng.* 594, 1985, 10–20.
9. W. Förstner, Quality assessment of object location and point transfer using digital image correlation techniques, *Proceedings, 15th ISPRS Congress, Rio de Janeiro, 1984*, pp. 169–191.
10. W. Förstner, A feature-based correspondence algorithm for image matching, in *ISP Comm. III, Rovaniemi, 1986*.
11. W. Grimson, *From Images to Surfaces: A Computational Study of the Human Early Visual System*, MIT Press, Cambridge, MA, 1981.
12. B. K. P. Horn, Non-Correlation Methods for Stereo Matching, *Photogramm. Eng. Remote Sens.* 49 (5), 535–536, 1983.

13. D. H. Hubel and T. N. Wiesel, Cells sensitive to binocular depth in area 18 of the macaque monkey cortex, *Nature* **255**, 1970, 41–42.

14. M. R. M. Jenkin, *Visual Stereoscopic Computation*, PhD Thesis, Department of Computer Science, University of Toronto, 1988.

15. M. R. M. Jenkin and P. A. Kolers, Some problems with correspondence, in *Motion Understanding: Robot and Human Vison* (W. N. Martin and J. K. Aggarwal, Eds.), pp. 269–295, Kluwer, Norwell, MA, 1988.

16. A. D. Jepson and M. R. M. Jenkin, The fast computation of disparity from phase differences, in *Proceedings, IEEE CVPR, San Diego, 1989*, pp. 398–403.

17. P. E. Kelly, P. R. H. McConnell, and S. J. Mildenberger, The gestalt photomapping system, *Photogramm. Eng. Remote Sens.* **43**, (11), 1977, 1407–1417.

18. C. D. Kuglin and D. C. Hines, The phase correlation image alignment method, in *Proceedings, IEEE International Conference on Cybernetics and Society, 1975*.

19. A. L. Lowrie, *Multiple Resolution Row Matching for Stereo Vision*, Masters Project Report, Carnegie–Mellon University, 1984.

20. D. Marr, *Vison*, Freeman, San Francisco, 1982.

21. D. Marr and T. Poggio, Cooperative computation of stereo disparity, *Science* **194**, 1976, 283–287.

22. D. Marr and T. Poggio, A theory of human stereo vision, AI Memo No. 451, MIT, 1977.

23. J. Mayhew and J. Frisby, The computation of binocular edges, *Perception* **9**, 1980, 69–86.

24. J. Mayhew and J. Frisby, Psychophysical and computation studies towards a theory of human stereopsis, *Artif. Intelligence* **17**, 1981, 349–385.

25. H. P. Moravec, Towards automatic visual obstacle avoidance, *Proceedings of the 5th IJCAI, Massachusetts, 1977*, p. 54.

26. H.-H. Nagel and W. Enkelman, Investigation of second order grey-value variations to estimate corner point displacements, in *Proceedings of the 5th IJCAI, Massachusetts, 1977*.

27. Y. Ohta and T. Kanade, Stereo by intra- and inter-scanline search using dynamic programming, *IEEE Trans. Pattern Anal. Machine Intelligence* **PAMI-7** (2), 1985, 139–154.

28. G. Poggio and B. Fischer, Binocular interaction and depth sensitivity in striate and prestriate cortex in behaving rhesus monkey, *J. Neurophysiol.* **40**, (6), 1977, 1392–1405.

29. S. Pollard, J. Mayhew, and J. Frisby, PMF: A stereo correspondence algorithm using a disparity gradient limit, *Perception* **14**, 1985, 449–470.

30. K. Prazdny, Detection of binocular disparities, *Biol. Cybern.* **52**, 1985, 93–99.

31. W. Richards, H. K. Nishihara, and B. Dawson, *Cartoon: A Biologically Motivated Edge Detection Algorithm*, AI Memo No. 668, June, 1982.

32. T. D. Sanger, Stereo disparity computation using Gabor filters, *Biol. Cybern.* **59**, 1988, 405–418.

33. J. Stokes, Image matching with phase shift methods, in *ISPRS Comm. III. Symp., Rovaniemi, 1986*.

34. D. Terzopoulos, *Multi-level Reconstruction of Visual Surfaces; Variational Principles and Finite Element Representation*, AI Memo No. 671, MIT, 1982.

35. C. E. Thorpe, An analysis of interest operators for FIDO, in *Proceedings of the Workshop on Computer Vision: Representation and Control, Annapolis, 1984*.

36. A. P. Witkin, D. Terzopoulis, and M. Kass, Signal matching through scale space, in *Proceedings 5th National Conference on AI, Philadelphia, 1986*, pp. 714–719.

37. G. A. Wood, Realities of automatic correlation problem, *Photogramm. Eng. Remote Sens.* **49**, 537–538, 1983.