Biol. Cybern. 59, 405–418 (1988)

Biological Cybernetics

Stereo Disparity Computation Using Gabor Filters

T. D. Sanger
Massachusetts Institute of Technology, NE 43-743, Cambridge, MA 02139, USA

Abstract. A solution to the correspondence problem for stereopsis is proposed using the differences in the complex phase of local spatial frequency components. One-dimensional spatial Gabor filters (Gabor 1946; Marcelja 1980), at different positions and spatial frequencies are convolved with each member of a stereo pair. The difference between the complex phase at corresponding points in the two images is used to find the stereo disparity. Disparity values are combined across spatial frequencies for each image location. Three-dimensional depth maps have been computed from real images under standard lighting conditions, as well as from random-dot stereograms (Julesz 1971). The algorithm can discriminate disparities significantly smaller than the width of a pixel. It is possible that a similar mechanism might be used in the human visual system.

Introduction

Successful computer methods for solving the stereo correspondence problem often involve the use of specific "features" as matching primitives (Barnard and Thompson 1980). These features are assumed to correspond to real structural properties of objects in the scene. Marr and Poggio (1979) have proposed the use of zero-crossings of the filtered images as the matching primitive, since the zero-crossings may correspond to real edges. Grimson (1985) modified this algorithm to impose continuity along contours. Ohta and Kanade (1985) use derivatives to find edges on each scanline, and then link edges between different scanlines to provide global constraints. And Kass (1983) performs matching based upon "feature vectors" of the first and second partial derivatives at different spatial scales.

A problem with matching based upon specific features is that it may be possible to find images for which the features are either misleading, sparse, or absent (Mayhew and Frisby 1981). J.G. Daugman (personal communication) has shown that it is possible to transform any image into an equally recognizable one which contains no zero-crossings. One could interpret any such image by using additional features, such as those suggested by Kass (1983). Mayhew and Frisby (1981) have suggested the use of peaks of the filtered image in addition to zero-crossings, and Poggio and Poggio (1984) mention that zero-crossings are probably not the only matching primitive in the human visual system. But feature-based algorithms require the existence of a sufficient number of features in the image, and give only sparse data if the features occur infrequently. Often such algorithms rely on a scarcity of features in order to make unambiguous matches. Unfortunately, this means that disparity can only be computed at relatively few points in an image. As a consequence, the output of such algorithms may not be useful for reconstructing smoothly curved surfaces or complex objects.

Stereoacuity is often limited by the spacing of the feature detectors. Detectors are usually spaced one pixel apart, and the pixel width therefore determines the disparity resolution. Without hyperacuity, large camera spacings are required to obtain adequately large disparities (Verri and Torre 1986), and this leads to a greater possibility of ambiguous or missing matches due to occlusions and viewpoint differences. It is possible to interpolate between points in the original image or convolution product. This technique can be used to find the location of edges to subpixel accuracy (Marr et al. 1979; MacVicar-Whelan and Binford 1981).

Algorithms which are not based on matching features are often called "correspondenceless" algorithms. An example is the use of cross-correlation to match small areas of each image (Hannah 1974;

Panton 1978; Moravec 1980). Theoretically, depth estimates may be generated for every point in a scene, allowing dense depth maps and the interpretation of complicated surfaces. Correlation requires that the image have sufficient local texture at each point to be matched, but that the texture not be repetitive or discontinuous (Lim and Binford 1987). Such techniques are very sensitive to noise, as well as to contrast or lighting changes between the two images which may be caused by the differing locations of the cameras. Other "correspondenceless" algorithms involve subtracting the left and right images and using the difference as a measure of disparity (Sperling 1970; Krotkov 1986).

To avoid the problems of lighting or noise differences between the two images one can apply bandpass filters to the images before attempting either a correspondenceless or a feature-based analysis. The left and right images may match quite well within one or more spatial-frequency bands, and a combination of information from different bands may be able to further reduce inaccuracies due to bandlimited noise. Biological inspiration for such methods comes from experiments showing that the visual cortex encodes information using bandpass spatial-frequency filters (Campbell and Robson 1968; Julesz 1971; Pollen and Taylor 1971; Maffei and Fiorentini 1973; Julesz and Miller 1975; Levinson and Blake 1979; Shapley and Lennie 1985).

A useful bandpass filter is the spatial Gabor filter (Gabor 1946; Marcelja 1980) which has both limited spatial width and finite bandwidth, and whose shape is similar to the receptive field profile of simple cells in primate visual cortex (Marcelja 1980). Although many similar filters can also be used, the Gabor filters allow easy separation of the modulating component (which determines the spatial frequency) and the envelope (which determines the bandwidth). These filters have been used for motion sensing by several authors (Watson and Ahumada 1983; Adelson and Bergen 1984; Sperling and van Santen 1984; Watson and Ahumada 1985; van Santen and Sperling 1985; Adelson and Bergen 1985; Heeger 1986), often by extracting the local spectral energy and analyzing it over a finite time interval. Such an analysis is not applicable to a stereo vision system, however, since only two frames of image data are available and algorithms therefore do not have access to the continuous information for which energy analyses are particularly suited.

The method proposed here is to convolve the left and right images with a complex Gabor filter and to use the difference in complex phase at each point to indicate a local shift between the two images. The disparity at a point is linearly proportional to the local phase difference in approximately the same way that a uniform shift of the entire image is proportional to the change in complex phase of the Fourier coefficients. The algorithm ignores the amplitude of the local spectral components and so does not attempt to recognize any features of the scene. It senses disparity directly from a locally computed parameter of the pair of images, and it therefore requires no formal matching process. After disparity is calculated at each spatial scale, the different values are combined independently to produce a single depth estimate.

Oppenheim and Lim (1981), Julesz and Schumer (1981), and Howard and Richardson (1986) have suggested that phase information is critically important for human interpretation of visual scenes. (Note that these authors refer to the phase of the global Fourier transform, and their comments might not apply to the phase of local Gabor filters.) Further support for the use of phase information in primate visual cortex comes from the discovery that simple cells occur in pairs with quadrature relative phase (Pollen and Ronner 1981), and could therefore represent the phase of a complex filter (Robson 1975). Burgess and Ghandeharian (1984) found that for detection tasks, humans perform better than the best "phase-insensitive" detector. Their results imply that matching between images makes use of the complex phase information in local regions. For stereopsis, the data of Mayhew and Frisby (1981) shows that this phase information is recombined from different spatial scales to produce a unified perception of depth. Since humans cannot compare spatial phase between frequencies more than two octaves apart (Lawden 1983), it seems likely that the disparity is computed entirely independently at each spatial scale, and that the resulting values are compared subsequently.

The algorithm presented here allows many opportunities to eliminate noise and produce reliable depth estimates. Since disparity is computed independently at each scale, a comparison of the different disparity values can help to reduce the effects of bandlimited noise. At every step of computation, confidence values are generated which are used to affect computations in subsequent steps. The first-level confidence value is based upon the assumption that the amplitude of the Gabor response should match in the left and right images (as it would for objects near the fixation plane in the absence of noise). The second-level confidence is derived from the agreement between different bandpass filters on the disparity estimate. And the third-level confidence is based upon the assumption that most surfaces will be smooth, so that nearby disparity values should be similar. If the output of this algorithm is used for surface interpolation or object recognition, the confidence values may be quite useful. Their significance will depend upon the requirements of the system using stereopsis for input.

Since the change in phase is a continuous variable independent of the spatial quantization of the image, it can indicate disparity down to the physical limits on the system. This means that disparities significantly smaller than the width of a pixel can be accurately discriminated, thereby allowing smaller camera spacings and fewer errors from occlusions. A similar capability is important for human vision. Humans can discriminate disparities as small as 5 s of arc, although foveal photoreceptor spacing is no less than 30 s of the arc (Morgan and Watt 1982).

The method proposed here is almost equivalent to performing a cross-correlation analysis of a pair of bandlimited images. However, in cross-correlation algorithms, accuracy depends upon the ability to recognize and localize the cross-correlation peak, whereas in the algorithm proposed here, disparity is measured directly as a function of image properties. Since this is a "correspondenceless" algorithm, no particular feature set is selected, so the method will work on arbitrary real images. The many opportunities for elimination of noise allow the algorithm to perform well on both synthetic and real images, under a variety of natural and non-optimal lighting conditions.

1 Description of the Algorithm

If we are presented with two images shifted relative to each other by an amount Δx, then the Fourier shift theorem states that the difference in the phase of a (one-dimensional) horizontal Fourier component of spatial frequency ω between the two images is $\omega \Delta x$. We can therefore derive the relative position of the two images from the Fourier phase information. If the two images constitute a stereo pair, then the value Δx is the disparity and will be proportional to distance from the fixation plane. The algorithm presented here is motivated by the expectation that a similar relation between the disparity and the change in complex phase will hold when local filters are used instead of a global Fourier transform. We use local filters to allow for images which do not have the same disparity at all points.

The well-known uncertainty theorem for spatial width and spatial bandwidth (first proved by Gabor (1946) in the time/frequency domain) dictates that a local filter must have a nonzero bandwidth. We use a set of different filters with differing ranges of frequency sensitivity to span the entire spatial frequency spectrum. We convolve the image with each of these filters, and use the information gained from the entire set to compute disparity. While any single filter can provide a phase value, the use of multiple filters guarantees that at least one will have sufficient response to give an accurate disparity estimate.

1.1 Gabor Filters

As we will see in Sect. 2.8, it is important that both the spatial width of the filters and the spatial frequency bandwidth be small. One-dimensional filters which minimize the product of spatial width and bandwidth were first described in the time/frequency domain by Gabor (1946), and take the form:

$$G(x-x_0)=e^{-(x-x_0)^2/2\sigma^2}\cdot e^{i\omega_0(x-x_0)}. \tag{1}$$

These "spatial Gabor filters" are thus seen to be the product of a complex harmonic function and a Gaussian envelope. x_0 is the spatial location of the filter, and ω_0 is the frequency of the harmonic component which will be the central spatial frequency of the power spectrum. The Fourier transform has the same functional form:

$$g(\omega-\omega_0)=e^{-(\omega-\omega_0)^2/2\tau^2}\cdot e^{-ix_0(\omega-\omega_0)}, \tag{2}$$

where σ and τ are the spatial half-width and spatial-frequency half-bandwidth of the filter, and the product $\sigma\tau$ is 1. This is the theoretical minimum for all complex-valued linear filters (Gabor 1946). The real (even) and imaginary (odd) components of the filter are given by:

$$G_{\text{even}}(x-x_0)=e^{-(x-x_0)^2/2\sigma^2}\cdot\cos[\omega_0(x-x_0)], \tag{3}$$

$$G_{\text{odd}}(x-x_0)=e^{-(x-x_0)^2/2\sigma^2}\cdot\sin[\omega_0(x-x_0)], \tag{4}$$

and are shown in Fig. 1.

For our purposes here, we will maintain the shape of the filters constant at all scales. This means that the spatial width σ will be proportional to the wavelength $2\pi/\omega_0$, and the number of cycles of the harmonic function within the Gaussian envelope will be the same for all filters. Therefore, τ is proportional to ω_0, and we refer to this as a "constant relative bandwidth" relation. The relative full bandwidth in octaves is obtained from:

$$\lambda=\log_2\left(\frac{\omega_0+\tau}{\omega_0-\tau}\right)=\log_2\left(\frac{1+t}{1-t}\right), \tag{5}$$

where t is the proportionality constant such that $\tau=t\omega_0$, and $t<1$. If $t=0.33$, for example, then we obtain a constant relative full bandwidth of one octave.

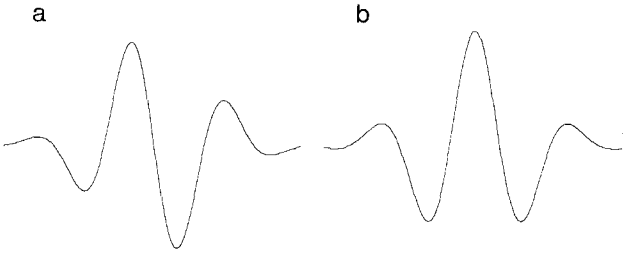

Fig. 1. a Odd, and **b** Even-symmetric Gabor filter impulse response. Relative bandwidth is one octave

Note that σ and τ represent the standard deviations of the Gaussian envelopes, and are not the same as the widths at half-height often used by other authors. The definitions of σ and τ are related to the half widths at half-height by $\sigma_{\text{half}} = \sigma\sqrt[4]{2}$, and $\tau_{\text{half}} = \tau\sqrt[4]{2}$. The relative full bandwidth at half-height in octaves is therefore given by:

$$\lambda_{\text{half}} = \log_2\left(\frac{\omega_0 + \tau\sqrt[4]{2}}{\omega_0 - \tau\sqrt[4]{2}}\right) = \log_2\left(\frac{1 + t_{\text{half}}}{1 - t_{\text{half}}}\right), \tag{6}$$

where $t_{\text{half}} = \tau_{\text{half}}/\omega_0 = t\sqrt[4]{2}$.

1.2 The Complex Phase Difference

If there is a constant disparity across some part of the scene (a flat surface in the frontoparallel plane) then in that region we have $L(x, y) = R(x + \Delta x, y)$ where $L(x, y)$ and $R(x, y)$ are the left and right images, respectively. For any point $\langle x_0, y_0 \rangle$ and particular Gabor filter G with central frequency ω_0, we can form the one-dimensional products:

$$c_R(x_0, y_0) = \int R(x, y_0) G(x_0 - x)\mathrm{d}x, \tag{7}$$

$$c_L(x_0, y_0) = \int L(x, y_0) G(x_0 - x)\mathrm{d}x = \int R(x + \Delta x, y_0) G(x_0 - x)\mathrm{d}x \tag{8}$$

and we compute the complex phase difference $\arg[c_L] - \arg[c_R] = \Delta\Phi$.

We now approximate Δx by $\Delta\Phi/\omega_0$. This is strictly valid only for filters of infinitesimal bandwidth (a Fourier decomposition), arising directly from the Fourier shift theorem. It is approximately true for Gabor filters under certain conditions (see Sect. 2).

1.3 Combination of Disparities from Different Spatial Scales

The combination of disparities across spatial scales is performed by a simple weighted average, with the weighting being provided by the first-level confidence values. Therefore, if bandlimited noise is present, disparity estimates computed from that spectral region will be ignored. We compute

$$d = \frac{\sum_{\omega} \Delta x_\omega r_1(\omega)}{\sum_{\omega} r_1(\omega)}, \tag{9}$$

where d is the average disparity, $r_1(\omega)$ is the first-level confidence (between 0 and 1) at each spatial frequency ω, and Δx_ω is the computed disparity for each spatial frequency.

1.4 Computation of Confidence Values

Since the amplitudes of the Gabor responses $|c_R|$ and $|c_L|$ are not used for disparity determination, they can be used as an independent measure of confidence. At the lowest level, if there is insufficient amplitude for a particular filter in one or both images, then the phase information can be considered meaningless, and a confidence of zero assigned.

If there is sufficient amplitude in both images, then a comparison of the value of $|c_R|$ and $|c_L|$ can be made. By analogy with the Fourier shift theorem, we would expect that if one image were a uniform shift of the other (a flat surface at constant disparity), then the Gabor amplitudes would be similar. The amplitudes are exactly the same only in the Fourier case of infinitesimal bandwidth, but for Gabor filters they are approximately the same if the disparity is not too great. If the Gabor amplitudes are very different, this may indicate the presence of noise in the image, or a disparity which is beyond the detection limit of a particular filter. Here, we use the ratio of the smaller to the larger so that a value of 1 indicates a high confidence:

$$r_1(\omega) = \min\left[\frac{|c_R|}{|c_L|}, \frac{|c_L|}{|c_R|}\right]. \tag{10}$$

At the next level, confidence is derived from a comparison of the disparities computed at each spatial scale. The weighted average is computed as in (9), and the extent to which the disparity estimates deviate from this gives the confidence:

$$r_2 = \frac{\sum_{\omega} r_1(\omega)\,|\Delta x_\omega - d|}{\sum_{\omega} r_1(\omega)}. \tag{11}$$

The deviations $|\Delta x_\omega - d|$ are weighted by the first-level confidence $r_1(\omega)$, as in (9), so that the second-level confidence reflects the contributions of different spatial frequencies to the computed disparity d. Note that high confidence will be indicated by values of r_2 close to zero. To convert to a more usable form, we set

$$r_2' = \frac{1}{1 + r_2} \tag{12}$$

which gives a value between 0 and 1, with 1 indicating high confidence.

1.5 Implementation Details

Gabor filters are typically chosen with a bandwidth between one-half and one octave, and with overlapping regions of spectral sensitivity. The choice of center spatial frequencies is arbitrary, but should be low enough so that the spatial width of the filters is larger than the largest expected disparity. Better results can be obtained if something is known about the image in advance, so that spectral regions of bandlimited noise

can be avoided. Using a greater number of spatial frequencies to compute the average disparity will lead to increased accuracy, although only three or four different filters were needed to generate the examples given in Sect. 3.

Since the output of the algorithm was not being used as input to some other system, the final confidence values were used to generate a smooth surface. Under the assumption that nearby points should have similar disparities, each point of the final disparity map was assigned a weighted average of its own and neighboring disparities, with the weighting determined by the confidence at each point. The smoothing radius could be adjusted as needed. Note that this is not the same as a simple smoothing of the final disparity surface, since each point is weighted by its confidence so that high confidence points will tend to determine the actual location of the surface.

2 Analysis of Performance

2.1 Formulation

Let $R(x)$ and $L(x)$ be two one-dimensional images, and let $m(x)$ be a mapping such that $L(x)=R(m(x))$ (in general, $m(x)$ may not be surjective or injective). We also define $\delta(x)=m(x)-x$ to be the disparity at each point x, where x is measured from the fixation point (defined to be the origin, for which $\delta(0)=m(0)=0$). $G(x)$ is taken to be the Gabor filter defined in (1), and $g(\omega-\omega_0)$ is its Fourier transform. (Without loss of generality, we take $x_0=0$ in the definition of the filter so that $g(\omega-\omega_0)$ will be real.) $r(\omega)$ and $l(\omega)$ are the Fourier transforms of $R(x)$ and $L(x)$, respectively, and are given by:

$$r(\omega)=\int R(x)\,e^{-i\omega x}dx,$$
$$l(\omega)=\int L(x)\,e^{-i\omega x}dx, \tag{13}$$

$$R(x)=\frac{1}{2\pi}\int r(\omega)\,e^{ix\omega}d\omega,$$
$$L(x)=\frac{1}{2\pi}\int l(\omega)\,e^{ix\omega}d\omega, \tag{14}$$

(Throughout the following, all spatial frequencies ω are in radians/pixel.) Therefore,

$$L(x)=R(m(x))=\frac{1}{2\pi}\int r(\omega)\,e^{i\omega m(x)}d\omega. \tag{15}$$

We now form the convolution products $c_R=G*R$ and $c_L=G*L$, as

$$G*R(x_0)=\int G(x)R(x_0-x)dx, \tag{16}$$

$$G*L(x_0)=\int G(x)L(x_0-x)dx$$
$$=\int G(x)\left[\frac{1}{2\pi}\int r(\omega)\,e^{i\omega m(x_0-x)}d\omega\right]dx. \tag{17}$$

Changing variables gives:

$$G*L(x_0)=\frac{-1}{2\pi}\int r(\omega)\left[\int G(x_0-m^{-1}(x))\times e^{i\omega x}\left(\frac{dm^{-1}(x)}{dx}\right)dx\right]d\omega, \tag{18}$$

where $m^{-1}(x)$ is the inverse of the mapping (if it exists).

In general, $m^{-1}(x)$ may not exist. To circumvent this problem we assume that the spatial width σ of the filters is small enough that $m(x)$ is monotonic and differentiable within this range for almost all the filters. Further, to obtain an easily understandable solution we assume that $m(x)$ can be approximated locally by a linear function $m(x)\approx Ax+B$. The values of A and B will be different in different parts of the image, but will be assumed to be constant over the width of any given filter. A represents the expansion of one image relative to the other (slant), and B represents the average disparity (absolute distance from the fixation plane) near x_0 (the center of the filter). From Eq. 17 we then obtain the result:

$$G*L(x_0)=\frac{1}{2\pi}\int g(A\omega-\omega_0)r(\omega)\,e^{i\omega Ax_0}\,e^{i\omega B}d\omega. \tag{19}$$

Note that $r(\omega)$ is complex, but that $g(A\omega-\omega_0)$ is real.

2.2 Accuracy of the Absolute Disparity Estimate

We first take the simplest case for which $A=1$, and the disparity is constant in the region to be examined. Computing the disparity then involves comparing the complex phases of the two values:

$$c_L=\frac{1}{2\pi}\int g(\omega-\omega_0)r(\omega)\,e^{i\omega x_0}\,e^{i\omega B}d\omega \tag{20}$$

and

$$c_R=\frac{1}{2\pi}\int g(\omega-\omega_0)r(\omega)\,e^{i\omega x_0}d\omega. \tag{21}$$

We set $\Delta\Phi=\arg[c_L]-\arg[c_R]$, restricted to the range $(-\pi,\pi]$. Disparity is computed by setting $\delta(x_0)=\Delta\Phi/\omega_0$. Therefore, the error in the disparity estimate is $B-\Delta\Phi/\omega_0$, and we thus are interested in the extent to which $\Delta\Phi$ deviates from the value $\omega_0 B$. Since $g(\omega-\omega_0)$ is a Gaussian centered at ω_0, we assume that most of the energy of the product $g(\omega-\omega_0)r(\omega)$ is located within the width of g. We therefore restrict the range of integration to $[\omega_0-\tau,\ \omega_0+\tau]$. The value of $\Delta\Phi$ will thus be within the range $[(\omega_0-\tau)B, (\omega_0+\tau)B]$, and we expect the maximum deviation from $\omega_0 B$ to be on the order of $B\tau$. When the disparity $B=\delta(x_0)$ is calculated by the relation $\delta(x_0)=\Delta\Phi/\omega_0$, the error will be proportional to $B\tau/\omega_0$. If we maintain a constant relative bandwidth of $\tau=t\omega_0$, then the disparity error is on the

order of Bt, or $(100)t\%$ of the calculated disparity. For a one-octave bandwidth filter, for example, $t=0.33$ and we expect the maximum error to be about 33%. Note that the error is independent of the size of the filter, if we maintain a constant relative bandwidth. By similar reasoning, when $A \neq 1$, we expect a maximum error on the order of $(Ax_0+B)\tau/\omega_0=(Ax_0+B)t$.

This approximation depends on $r(w)$ being well-behaved in the spatial-frequency band being examined. The true maximum error is proportional to $B\tau'/\omega_0$ where τ' is the half-bandwidth of the product $g(\omega-\omega_0)r(\omega)$. τ' may be significantly larger than τ if $r(\omega)$ contains large spikes outside the $\omega_0-\tau$ or $\omega_0+\tau$ boundaries. Such an occurrence is not frequent for strongly textured images or random-dot stereograms, which may have nearly flat spectra near most values of ω_0.

2.3 Accuracy of Relative Disparities

Although the absolute disparity measurement may be inaccurate by a significant percentage of the true disparity, the change in disparity between neighboring points is small. If $\arg[c_R]$ (or $\arg[c_L]$) is computed at a point x_0, then moving to a point x_0+d will change its value by $\omega_0 \mathrm{d} \pm \tau \mathrm{d}$, as described in Sect. 2.2. Therefore, the phase difference $\Delta\Phi$ computed at x_0+d will change by at most $2\tau \mathrm{d}$. $\Delta\Phi/\omega_0$ will change by at most $2t\mathrm{d}$. Note that comparisons should only be made between nearby points, since if d is larger than $(\pi/2)\sigma$, the error in $\Delta\Phi$ $(\pm 2\tau \mathrm{d})$ will be larger than π, and the phase difference will be meaningless. Since the relative disparity error becomes arbitrarily small for points close together, we see that the algorithm will generate a smooth surface and the value of the absolute error will not change much over short distances.

There is evidence that humans are considerably better at measuring the relative disparity between objects than at calculating the absolute disparity (the true distance from the fixation plane). Westheimer (1979) mentions that while we can discriminate relative disparity to a few seconds of arc, we can only measure absolute disparity to within about a minute of arc. And Anstis et al. (1978) have demonstrated that a depth edge can determine the perceived disparity of regions around it, thereby implying that relative disparity computations are only made over short distances.

2.4 Disparity Limits

The effective disparity limit is set by the filter width σ. Since σ is proportional to $1/\tau$, and τ is proportional to the spatial frequency ω in a constant relative bandwidth system, we see that σ is proportional to $1/\omega$. Therefore, the maximum disparity which can be fused should vary inversely with spatial frequency for band-limited images.

Another cause of disparity limits is "phase wrap-around". For an image consisting of a single component ω_i, if the magnitude of the disparity at x_0 exceeds π/ω_i then the phase difference $\Delta\Phi$ will "wrap around", and the computed value of $\Delta\Phi$ will be $\omega_i\Delta x-2\pi$. For any given value of B there will be frequencies which wrap around. In general, for a particular filter with central frequency ω_0 and half bandwidth τ, the maximum allowable disparity before wrap-around effects will cause inaccuracy is $\pi/(\omega_0+\tau)$. The effective disparity limit for the system is determined by the highest filter frequency chosen. Note that, if wrap-around occurs, disparity values computed from different filters with center frequencies ω_i will be dependent on ω_i by the relation $\Delta x_i=\Delta\Phi_i/\omega_i=\Delta x-2\pi/\omega_i$. The system can detect the error when these disparity values do not correspond.

A decrease in maximum fuseable disparity with increasing spatial frequency was found psychophysically by Schor et al. (Schor et al. 1984). Unfortunately, there is other evidence that disparities well beyond the predicted limits can be measured (Mayhew and Frisby 1979; Mowforth et al. 1981), so the significance of this result is unclear.

2.5 Disparity Gradient Limits

If $A \neq 1$, we must consider the effects of frequency domain scaling (due to the term $g(A\omega-\omega_0)$) on the frequencies to which the filter is sensitive. For now, let $B=0$ (which we can accomplish by shifting the origin of coordinates). Then the region in the spatial frequency domain to which $g(A\omega-\omega_0)$ is sensitive is $[(1/A)(\omega_0-\tau), (1/A)(\omega_0+\tau)]$. For $A>1$, $g(A\omega-\omega_0)$ and $g(\omega-\omega_0)$ will have no part of the spectrum in common if $(1/A)(\omega_0+\tau)<\omega_0-\tau$. We may assume that phase values computed from different parts of the spectrum will be unrelated, so for the phase difference to be meaningful we require $(1/A)(\omega_0+\tau)>\omega_0-\tau$, or

$$A<\frac{\omega_0+\tau}{\omega_0-\tau} \quad (A>1). \tag{22}$$

Similarly, for $A<1$, we require

$$A>\frac{\omega_0-\tau}{\omega_0+\tau} \quad (A<1). \tag{23}$$

If we maintain a constant relative bandwidth $\tau=t\omega_0$, we can rewrite this expression as

$$\frac{1-t}{1+t}<A<\frac{1+t}{1-t}. \tag{24}$$

If the relative bandwidth is one octave, for instance, then $t=0.33$ and we must have $1/2<A<2$. If a scale

change greater than $(1+t)/(1-t)$ is presented, the system can detect this fact since it will be signalled by a difference in the Gabor amplitude values which now reflect completely different ranges of the spatial-frequency spectrum.

Tyler (1973), Tyler and Sutter (1979), and Burt and Julesz (1980) have found psychophysically that there is a maximal rate of disparity change that can be fused. Tyler and Sutter (1979) found that when two different sinewave gratings are presented stereoscopically, the maximum ratio of spatial frequencies (or the maximum scale change) which can be fused is a factor of 1.5, although a change of up to 2 is possible if the images are moving. The psychophysical value found by Burt and Julesz (1980) for fusion of dots is closer to 3. (Note that they define the disparity gradient differently: their gradient is equal to $2(A-1)/(A+1)$.) They also found that a disparity gradient limit exists in all directions. The theory presented here predicts such a limit only in the horizontal direction. In the macaque monkey, the mean spatial frequency bandwidth in the fovea is 1.4 octaves (Foster et al. 1985). For this bandwidth, $t=0.45$, and the discussion of Sect. 2.5 shows that we must have $A<2.63$. This value is in reasonable agreement with the psychophysical data mentioned above.

2.6 Other Causes of Error

Sharp depth boundaries in real-world scenes lead to occlusions. There will be areas of each image which do not occur in the other, and matching of any sort will be impossible in these regions. Near boundaries, Gabor filters will give unpredictable phase information, and disparity values computed from these filters will be meaningless. The occurrence of this problem will be signalled by the amplitudes of the components $|c_R|$ and $|c_L|$ being different. When significant amplitude differences are found across a range of spatial frequencies, the image location should be considered ambiguous. If a continuous contour of such ambiguous points exists, it is strong evidence for the existence of a depth boundary in the image.

2.7 Noise Tolerance

If additive noise is present, we have $R(x)+N_R(x)$ and $L(x)+N_L(x)$. The convolution products can be written

$$G*(R+N_R)=c_R+n_R \tag{25}$$

and

$$G*(L+N_L)=c_L+n_L. \tag{26}$$

The effect of the noise on the complex phase is limited by

$$\alpha_{\max}=\arcsin(|n|/|c|), \tag{27}$$

for each image (if $|n|<|c|$). Therefore, the change in the computed value of $\Delta\Phi$ is limited to $2\alpha_{\max}$, and the disparity will be accurate to within $2\alpha_{\max}/\omega_0$. If $|c|\gg|n|$, then we can approximate $\alpha_{\max}\approx|n|/|c|$ which is the reciprocal of the signal to noise ratio. Therefore, the loss of accuracy occurs smoothly with increasing image noise, and is most severe at low spatial frequencies. If the noise is bandlimited, then the inaccuracy will be limited to those filters which respond to the noise frequencies.

2.8 The Importance of Gabor Filters

From Sect. 2.2 we see that the disparity error is on the order of Bt. Therefore, a smaller relative bandwidth will allow a more accurately computed depth. However, a larger relative bandwidth will increase the disparity gradient limit $(1+t)/(1-t)$ and decrease the spatial width of the filters, so the approximation of $m(x)$ by a linear function will be more likely to be valid. Smaller spatial width also means that there is less chance of any given filter falling across a depth boundary. Gabor (1946) showed that (in the time domain) Gabor filters achieve the theoretical minimum product of spatial width and bandwidth for complex-valued linear filters. It is for this reason that Gabor filters are a good choice for this algorithm. Note that many similar filter shapes achieve near-minimality, and that the algorithm would not suffer severely if a different set of quadrature-phase filters were substituted.

2.9 Accuracy of Depth Estimates

Verri and Torre (1986) calculate that the error in computation of depth can be described by

$$\Delta z=\left(\frac{Z^2}{fd}\right)\Delta x_{\min}, \tag{28}$$

where Δz is the depth error, Z is the absolute depth of the point being measured (near the fixation plane), f is the focal length of the lens used, d is the inter-camera distance, and $\Delta x_{\min}$ is the minimum measurable disparity. For the image of a mask given in Sect. 3, the values of these parameters are $Z=2m$, $f=50\,\text{mm}$, $d=6.4\,\text{cm}$, and the pixel spacing is $24\,\mu$. Since there are 64 greylevels in the Autovision system, the theoretical limit of disparity measurement is $24\,\mu/64$ (a shift smaller than this will have no effect on the cameras). This means that the minimum depth discrimination is 0.47 mm. If we assume that, in noisy real images, the algorithm can only measure down to $1/20^{\text{th}}$ of a pixel width, then the theoretical depth discrimination is 1.5 mm.

However, absolute depth estimates may be considerably worse than discrimination ability. From

Sect. 2.2 we note that the absolute disparity estimate is accurate to within only 33% of the total disparity. This means that measurement of absolute disparities far from the fixation plane will be considerably less accurate than for those closer to the fixation plane. Nevertheless, the *relative* disparities between adjacent points will be discriminable down to the limits determined by noise in the imaging system. So although the depth of a point 2 cm from the fixation plane (which generates a disparity of about 1 pixel) can only be known to within 7 mm (assuming 33% accuracy), the relative depths between this point and neighboring points can be computed with up to 0.5 mm precision (depending on the local rate of slant).

3 Examples

The following sections show the performance of the algorithm on a variety of test images. Computation time depended on the number of different scales of filter used, and the complexity of the image. Typical times were one minute for six spatial scales. The output is presented in the form of a three-dimensional depth map. This map shows the relative depths at each point (the overall height of the map is adjusted for ease of interpretation). The true depths are not computed.

3.1 Random-Dot Square

Figure 2A and B is the right and left images of a random-dot stereo pair which represents a central square raised above a flat background (viewed by crossing the eyes). The disparity of the central square is 5 pixels. To compute the depth map of Fig. 2C, five different Gabor filters with 1/2-octave bandwidths were used. The resulting map was smoothed based on confidence values in a 3×3 neighborhood of each point, and the depth was linearly interpolated over regions of excessively low confidence.

3.2 Random-Dot Ramp

Figure 3A and B gives the right and left images of a random dot pair representing a smooth ramp in the center of the image. The maximum disparity of the ramp is 10 pixels. The depth map of Fig. 3C was not locally smoothed, but linear interpolation was performed over low-confidence regions. Note that the vertical magnification in Fig. 3C is half that of Fig. 2C, so that the height of the two figures appears the same although the maximum disparity in 3C is twice as great.

3.3 Random-Dot Hyperacuity Test

Figure 4A and B gives the right and left images of a square which is $1/64^{th}$ of a pixel behind the background. The background has a uniform $1/32^{nd}$ disparity relative to the fixation plane. Subpixel shifts were performed using the following formula:

$$L_n = (\tfrac{63}{64})R_n + (\tfrac{1}{64})R_{n+1},$$

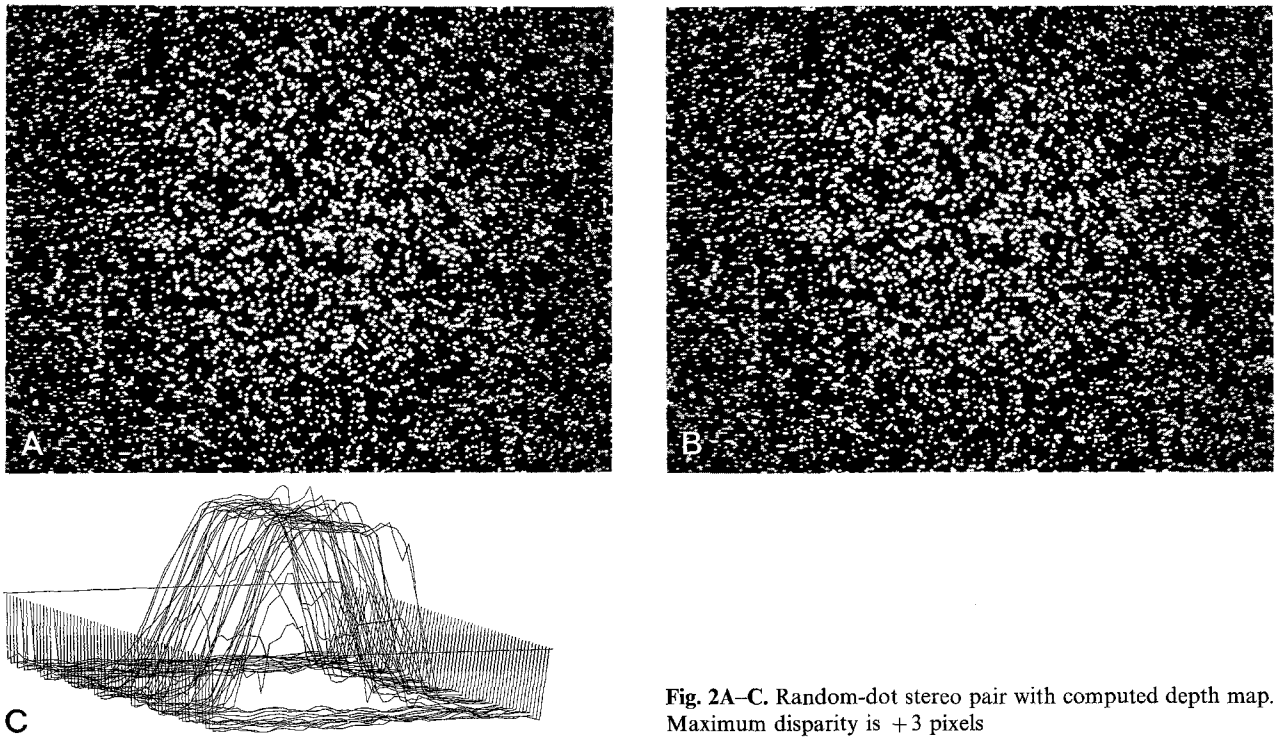

Fig. 2A–C. Random-dot stereo pair with computed depth map. Maximum disparity is +3 pixels

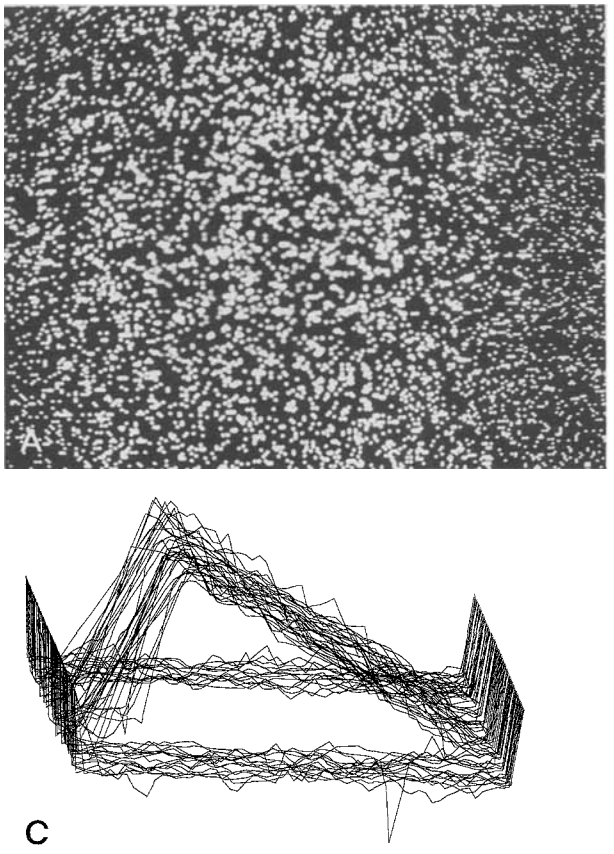

Fig. 3A–C. Random-dot stereo pair with computed depth map. Maximum disparity is ± 5 pixels

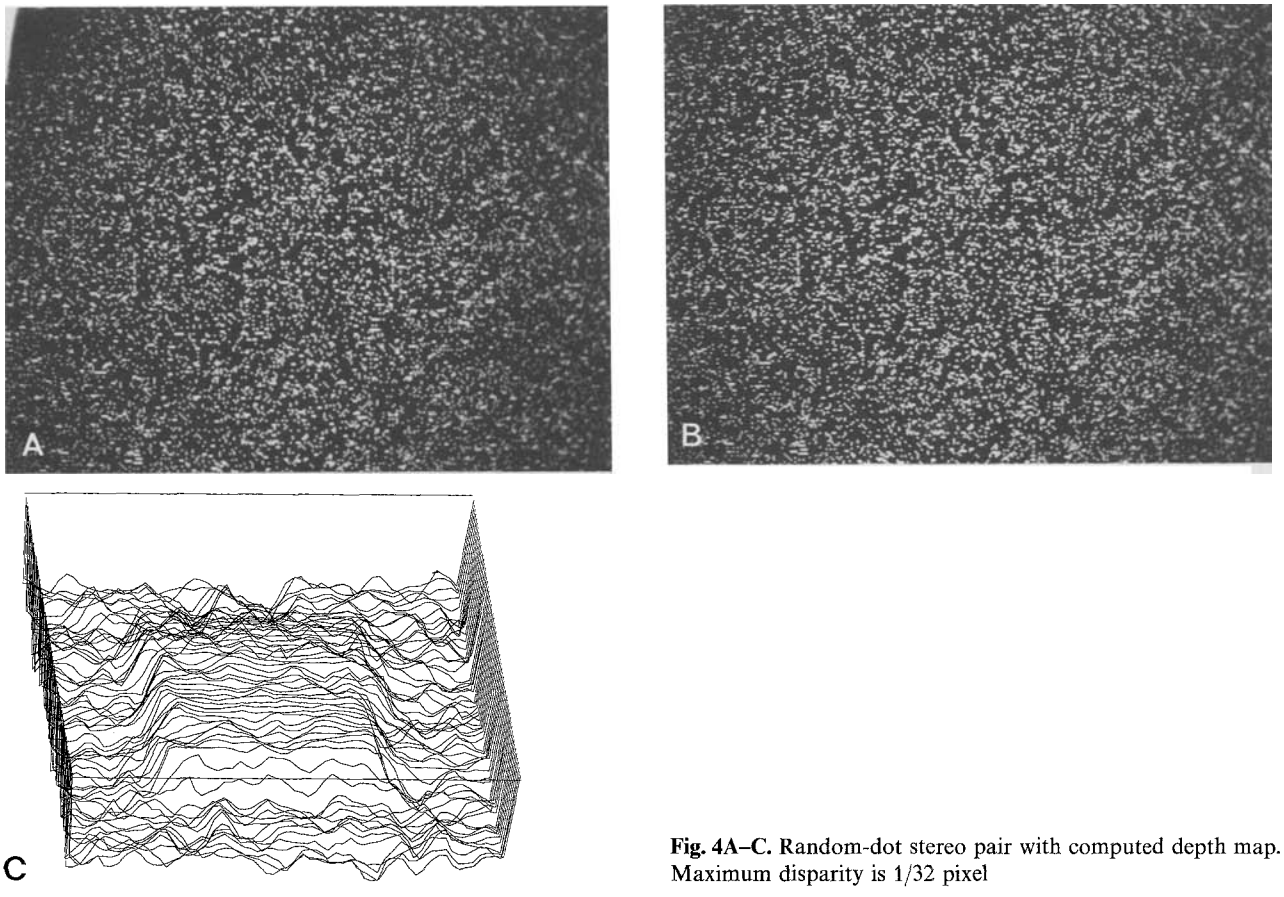

Fig. 4A–C. Random-dot stereo pair with computed depth map. Maximum disparity is 1/32 pixel

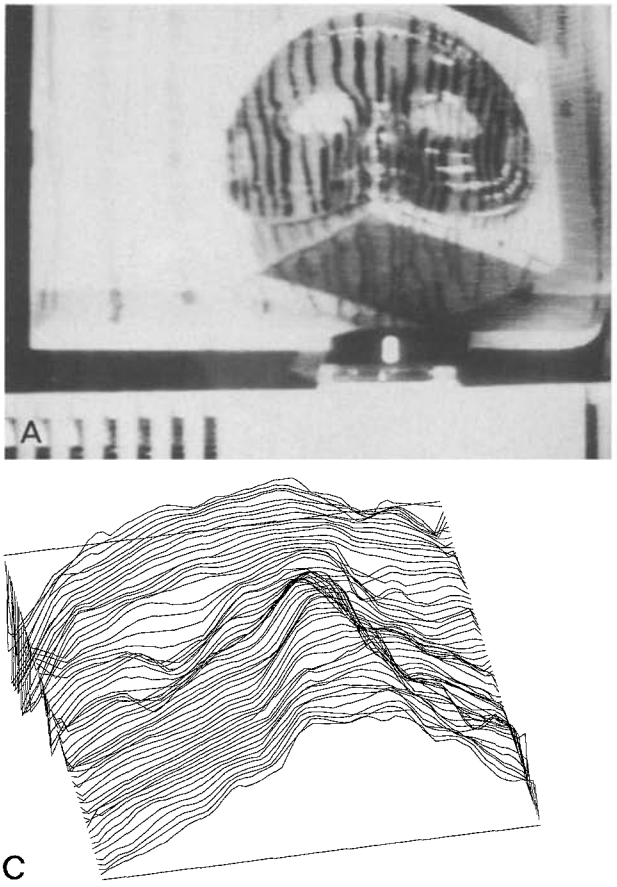

Fig. 5A–C. Test image of mask, and map computed from a small region around the mask. Maximum disparity is approximately 1.5 pixels

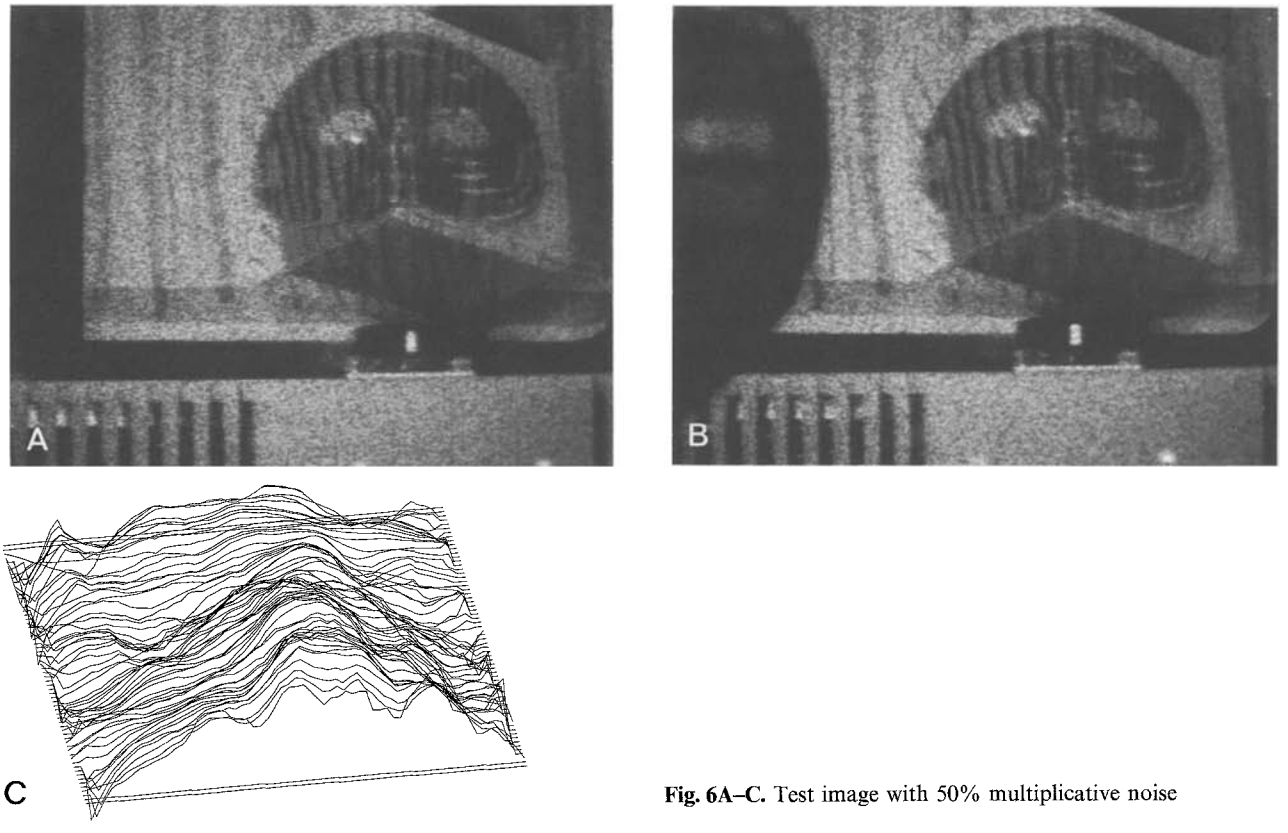

Fig. 6A–C. Test image with 50% multiplicative noise

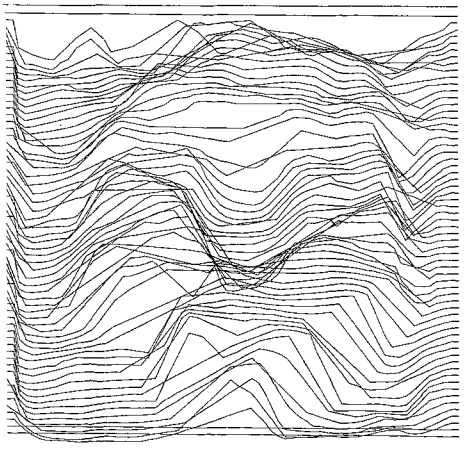

Fig. 7A–C. Aerial view of the Pentagon. Maximum disparity is approximately 7 pixels

where R_n and R_{n+1} are the intensities of two neighboring pixels in the right image, and L_n is the assigned intensity of a pixel in the left image. (The difference between the two images may not be visible due to nonlinearities of the display and reproduction process.) Figure 4C gives the computed disparity map. The vertical scale on this map is 100 times the scale of Fig. 2C. Smoothing was performed in a 3×3 region, and low-confidence regions were interpolated.

3.4 Natural Image

Figure 5A and B depicts a real scene containing a mask. The mask is 16 cm wide and 6 cm deep from the nose to the rear edges. Cameras were placed 6.4 cm apart at a distance of 2 m from the mask. Convergence was established near the back of the mask, and Gabor filters were chosen at 1/2-octave intervals. The computed depth map is shown in Fig. 5C. Smoothing was performed over 3×3 regions, and low confidence points were interpolated. The vertical magnification is twice that of Fig. 2C, and the maximum disparity is therefore about 1.5 pixels.

3.5 Natural Image with Multiplicative Noise

Figure 6A and B shows the same image as Fig. 5A and B, but with the addition of noise. The noise was generated by choosing, for each pixel, a pseudo-random number between 0 and 0.5 from a uniform distribution. The pixel was modified by multiplication with this number. The results of computation are shown in Fig. 6C. (The vertical magnification is slightly less here.)

3.6 Natural Image with Larger Disparity

Figure 8A and B shows an image with maximum disparity of around 7 pixels. The computed depth map is shown in Fig. 8C. All parameters were the same as for Fig. 5C, although the vertical magnification is less. Smoothing was performed over 5×5 regions.

4 Discussion

4.1 Relation to Other Algorithms

The algorithm proposed here may be considered to be a "correspondenceless" algorithm and should therefore be understood within this context. It performs essentially the same calculation as cross-correlation (Hannah 1974; Panton 1978; Moravec 1980), the primary difference being that the correlation is within bandlimited regions, and the disparity values from each region are computed independently and combined subsequently. Nishihara (1983, 1984) used a coarse-to-fine implementation of bandlimited cross-correlation to find disparities. He locates the correlation peak using an interpolation method which theoretically allows subpixel precision. Similarly, Krotkov (1986) locates the centroid of a small region in the left and right images and compares the positions with high precision. Unlike the complex phase, these methods do not give a disparity measurement directly, but the theoretical basis is the same.

Another very similar method involves subtracting the two images (Sperling 1970; Krotkov 1986). Krotkov proposed that the differences in the left and right images near zero-crossings could be used to discriminate crossed from uncrossed disparities. Although this method was not extended to actually determine the disparities, it is theoretically equivalent to finding the phase difference. The images were bandpass filtered with the Laplacian of a Gaussian, so as Krotkov notes, the method is "equivalent to finding the phase difference in the power spectra" (although he actually only computes the *sign* of the phase difference).

4.2 *Advantages of this Algorithm*

This paper presents a new method for solving the correspondence problem for stereopsis. The use of Fourier phase information allows disparities to be computed in an object-independent manner, without the necessity for explicitly matching local features. The algorithm applies equally well to random-dot stereograms and natural images, showing that the correspondence problem is solved both for the most abstract case with no monocular cues, and for the case of a real image with the normal range of spatial frequencies. The main advantages of the algorithm are:

1. No specific feature set needs to be identified, so we will be able to compute disparities for arbitrary images wherever the image is not perfectly unmodulated. There is no need for the presence of any particular type of feature such as a line ending or zero-crossing within the width of the filter. The resulting depth map will therefore be dense and can be used for identifying complex surfaces and depth boundaries.

2. The algorithm is not significantly affected by moderate amounts of noise. Disparity is computed independently at several different spatial scales, so band-limited noise can be easily eliminated. Combination of information from different scales increases the accuracy of the final disparity estimate. Locally computed confidence values are provided along with the depth map, and can be used either for "intelligent" smoothing or as input to subsequent processing systems. Because of its noise-reducing abilities, the algorithm will perform well for real image processing applications in which specular reflectance changes, lighting intensity changes, camera noise, and quantization errors may not be controllable.

3. Hyperacuity is possible down to the physical limits on the image acquisition system. If there are 256 greylevels, for instance, this means that disparities as small as $1/256^{th}$ of a pixel can be discriminated. Cameras can thus be placed very close together, thereby reducing occlusions due to viewpoint differences. In addition, each detector can sense a range of disparities from less than a pixel width to half the width of its receptive field, and therefore has a large range of possible depths to which it will respond.

4. All operations are completely local, allowing efficient parallel implementations. Both disparity and local confidence values can be computed based entirely on small regions of the image. Therefore, serial computers can save considerable amounts of computation by checking disparity at only a few points in a time-varying image. Certain points which are useful, "interesting", or which have given reliable disparity values in the past can be used to give continually updated information about the depth of objects in the scene without the necessity for performing a large-scale convolution and associated processing. This fact makes it possible to implement the algorithm on microprocessors without special hardware.

5. Operation is fast, with typical computation times on ordinary hardware being 1 min to compute a 64×64 array of disparities. On the connection machine at MIT, computation of depth for 256×256 points required less than two seconds.

4.3 *Biological Relevance*

In some sense, any system which recognizes disparity can be said to be sensitive to differential phase, since any change in the image will change the phase of local spatial frequency components. The main difference in the algorithm presented here is that the phase is measured *directly*, and that the change in the relative responses of odd and even symmetric filters is used to indicate disparity. It is important to realize that such an algorithm does not require that the phase be made explicit anywhere in the visual system. It seems unlikely that there are "arctangent cells" capable of computing the complex phase from the arctangent of the ratio of responses of the odd and even receptive field cells. What is important is that the phase information is implicit in the relative responses of these two filters. The implicit phase information could be propagated upward through the system in the form of four distinct response profiles (positive and negative even and odd filters). Groups of four responses could be compared to each other directly, without any need to compute the phase difference itself. What the algorithm predicts is that the complex phase is available to the system in a fairly direct way, and that the system might therefore make use of this information.

There is biological evidence supporting different aspects of this method. The use of bandpass spatial-frequency filters (Campbell and Robson 1968; Julesz 1971; Pollen and Taylor 1971; Maffei and Fiorentini 1973; Julesz and Miller 1975; Levinson and Blake 1979; Shapley and Lennie 1985) or quadrature-phase

Gabor filters in particular (Marcelja 1980; Pollen and Ronner 1981), the combination of information computed independently at different spatial scales (Mayhew and Frisby 1981; Lawden 1983), the importance of phase information (Oppenheim and Lim 1981; Julesz and Schumer 1981; Howard and Richardson 1986; Pollen and Ronner 1981; Robson 1975; Burgess and Ghandeharian 1984; Lawden 1983), and the ability to perform stereoacuity tasks (Morgan and Watt 1982) are in agreement with predictions of this algorithm. The algorithm also correctly predicts the maximum rate of change of disparity (Tyler 1973; Tyler and Sutter 1979; Burt and Julesz 1980) and the response to spatial frequency differences between the two eyes (Tyler and Sutter 1979). Its ability to determine relative disparities more accurately than absolute disparities agrees with human performance (Westheimer 1979; Anstis et al. 1978). Finally, a wavelength dependence of Panum's fusional area is predicted, in agreement with some (Schor et al. 1984) but not all (Mayhew and Frisby 1979; Mowforth et al. 1981) psychophysical evidence.

4.4 Limitations

While this algorithm performs well on smooth surfaces, it gives very poor results near depth boundaries. Occlusions lead to regions of one image which are not present in the other, and a filter which is across a depth boundary will therefore give meaningless phase information. This may be signalled by a poor first-level confidence value, but it will nevertheless be impossible to obtain an accurate depth near a depth edge.

Another problem is the limit on maximum disparities. Due to "wrap-around", the maximum disparity which a filter can accurately determine is one-half the wavelength of its central spatial frequency. This means that large filters are necessary to determine large disparities, since the filter width is usually two to three times the wavelength of the central spatial frequency. Unfortunately, these large filters will only give information about low frequencies in the image. This problem can be helped by using a coarse-to-fine strategy which allows larger filters to bring smaller ones into registration.

For images whose spectra are not flat, the absolute disparity estimates may be very inaccurate. While this may not be a problem for a system attempting to recognize objects based upon surface contours, it could adversely affect a system attempting to navigate a vehicle through the environment.

4.5 Improvements

It is possible to respond to larger disparities if the fixation point can be moved during processing. If the widest receptors in one region of an image signal very large disparities, then the fixation point can be moved in such a way as to bring more of the narrower high-frequency receptors into alignment. This trick was suggested by Marr and Poggio (1979) who used low-frequency information to guide vergence movements. A change in vergence is equivalent to comparing phase values from receptors at different positions in the image, and this need only be done locally in the regions where large disparities are expected.

Although the examples given here use the raw image data as input, it is possible that performance might be enhanced for certain types of images by preprocessing. If it were known, for instance, that a particular feature in an image contained all the useful information, and that this feature was not too sparse, a map of the feature locations could be supplied to the Gabor phase algorithm for matching. Of course, acuity would be limited by the minimum feature spacing.

Any linear filtering operation could be used to reduce noise without affecting the acuity of the system, however. In addition, other types of information could be supplied as input. Color, texture, or motion information could be used for matching, and the disparity values computed from each type of data could be combined in the same way as for disparity values from different scales.

Acknowledgements. The author wishes to thank Roger Brockett, Dan Pollen, and Tomaso Poggio for their comments on the manuscript, and John Daugman, David Mumford, and Joel Wein for their comments and helpful discussions. This research was inspired by a course taught by Richard Kronauer. The work was performed during 1985 in the Harvard Robotics Laboratory under the direction of Roger Brockett. It was supported in part by the Office of Naval Research under contract number N 000-14-84-K-0504 with Harvard University. The author was supported during part of this research by a National Science Foundation Graduate Fellowship.

References

Adelson EH, Bergen JR (1984) Motion channels based on spatiotemporal energy. Invest Opthalmol Vis Sci [Suppl] A 25:14

Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. J Opt Soc Am A 2:284–299

Anstis SM, Howard IP, Rogers B (1978) A Craik-O'Brien-Cornsweet illusion for visual depth. Vision Res 18:213–217

Barnard ST, Thompson WB (1980) Disparity analysis of images. IEEE PAMI 2:333–340

Burgess A, Ghandeharian H (1984) Visual signal detection I. Ability to use phase information. J Opt Soc Am A 1:900–905

Burt P, Julesz B (1980) A disparity gradient limit for binocular fusion. Science 208:615–617

Campbell FW, Robson JG (1968) Application of Fourier analysis to the visibility of gratings. J Physiol 197:551–566

Foster KH, Gaska JP, Nagler M, Pollen DA (1985) Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. J Physiol 365:331–363

Gabor D (1946) Theory of communication. JIEE 93:429–459
Grimson WEL (1985) Computational experiments with a feature based stereo algorithm. IEEE PAMI 7:17–34
Hannah MJ (1974) Computer matching of areas in stereo imagery. Ph.D. thesis, Stanford University
Heeger DJ (1986) Depth and flow from motion energy. In: Proceedings of the Fifth National Conference on Artificial Intelligence (AAAI-86) Philadelphia, pp 657–661
Howard JH, Richardson KH (1986) Absolute phase uncertainty in sinusoidal grating detection. Catholic University of America, Human Performance Lab Tech Report ONR-86-27
Julesz B (1971) Foundations of cyclopean perception. University of Chicago Press, Chicago
Julesz B, Miller JR (1975) Independent spatial-frequency-tuned channels in binocular fusion and rivalry. Perception 4:125–143
Julesz B, Schumer RA (1981) Early visual perception. Ann Rev Psychol 32:575–627
Kass M (1983) A computational framework for the visual correspondence problem. In: Proceedings IJCAI-8, Karlsruhe, FRG, pp 1043–1045
Krotkov EP (1986) Visual hyperacuity: representation and computation of high precision position information. Comput Vis Graphics Image Proc 33:99–115
Lawden MC (1983) An investigation of the ability of the human visual system to encode spatial phase relationships. Vision Res 23:1451–1463
Levinson E, Blake R (1979) Stereopsis by harmonic analysis. Vision Res 19:73–78
Lim HS, Binford TO (1987) Stereo correspondence: a hierarchical approach. In: Proceedings of the Image Understanding Workshop, Washington, DC, pp 234–241
MacVicar-Whelan PJ, Binford TO (1981) Line finding with subpixel precision. In: Proceedings of the Image Understanding Workshop, Washington, DC, pp 211–216
Maffei L, Fiorentini A (1973) The visual cortex as a spatial frequency analyzer. Vision Res 13:1255–1267
Marcelja S (1980) Mathematical description of the responses of simple cortical cells. J Opt Soc Am 70:1297–1300
Marr D, Poggio T (1979) A computational theory of human stereo vision. Proc R Soc Lond B 204:301–328
Marr D, Poggio T, Hildreth E (1979) The smallest channel in early human vision. J Opt Soc Am 70:868–870
Mayhew JEW, Frisby JP (1979) Convergent disparity discriminations in narrow-band-filtered random-dot stereograms. Vision Res 19:63–71
Mayhew JE, Frisby JP (1981) Psychophysical and computational studies towards a theory of human stereopsis. Artif. Intell 17:349–385
Moravec HP (1980) Obstacle avoidance and navigation in the real world by a seeing robot rover. Stanford Computer Science Dept. Report STANCS-880-813
Morgan MJ, Watt RJ (1982) Mechanisms of interpolation in human spatial vision. Nature 299:553–555
Mowforth P, Mayhew JEW, Frisby JP (1981) Vergence eye movements made in response to spatial-frequency filtered random-dot stereograms. Perception 10:299–304
Nishihara HK (1983) PRISM: a practical realtime imaging stereo matcher. In Rooks B (ed) Proceedings of the 3rd International Conference on Robotic Vision and Sensory Controls, Cambridge, Mass
Nishihara HK (1984) Prism: a practical real-time imaging stereo matcher. MIT AI Memo 780
Ohta Y, Kanade T (1985) Stereo by intra- and inter-scanline search using dynamic programming. IEEE PAMI 7:139–154
Oppenheim AV, Lim JS (1981) The importance of phase in signals. Proc IEEE 69:529–541
Panton DJ (1978) A flexible approach to digital stereo mapping. Photogramm Eng Remote Sensing 44:1499–1512
Poggio GF, Poggio T (1984) The analysis of stereopsis. Ann Rev Neurosci 7:379–412
Pollen DA, Ronner SF (1981) Phase relationships between adjacent simple cells in the visual cortex. Science 212:1409–1411
Pollen DA, Taylor JH (1971) How does the striate cortex begin the reconstruction of the visual world? Science 173:74–77
Robson JG (1975) Receptive fields: neural representation of the spatial and intensive attributes of the visual image. In: Carterette EC, Friedman MP (eds) Handbook of perception, vol 5. Academic Press, New York, pp 81–116
Schor C, Wood I, Ogawa J (1984) Binocular sensory fusion is limited by spatial resolution. Vision Res 24:661–665
Santen JPH van, Sperling G (1985) Elaborated Reichardt detectors. J Opt Soc Am A 2:300–321
Shapley R, Lennie P (1985) Spatial frequency analysis in the visual system. Ann Rev Neurosci 8:547–583
Sperling G (1970) Binocular vision: a physical and a neural theory. Am J Psychol 83:461–534
Sperling G, van Santen JPH (1984) Temporal covariance model of human motion perception. J Opt Soc Am A 1:451–473
Tyler CW (1973) Stereoscopic vision: cortical limitations and a disparity scaling effect. Science 181:276–278
Tyler CW, Sutter EE (1979) Depth from spatial frequency difference: an old kind of stereopsis? Vision Res 19:859–865
Verri A, Torre V (1986) Absolute depth estimate in stereopsis. J Opt Soc Am A 3:297–299
Watson AB, Ahumada AJ (1983) A look at motion in the frequency domain. NASA tech report 84352
Watson AB, Ahumada AJ (1985) Model of human visual-motion sensing. J Opt Soc Am A 2:322–341
Westheimer G (1979) Cooperative neural processes involved in stereoscopic acuity. Exp Brain Res 36:585–597

Received: May 18, 1988

Terence D. Sanger
Massachusetts Institute of Technology
NE43-743
Cambridge, MA 02139
USA

Note Added in Proof. After this paper went to press, two reference errors were discovered in the Introduction: (1) "J. G. Daugman (Personal Communication)" appeared as: Daugman JG (1988) Pattern and motion vision without Laplacian zero crossings. J Opt Soc Am A 5(7):1142–1148. (2) "(Robson 1975)" should be "(Pollen and Ronner 1975)" referring to the article: Pollen DA, Ronner SF (1975) Periodic excitability changes across the receptive fields of complex cells in the striate and parastriate cortex of the cat. J Physiol 245:667–697. This is a common and unfortunate mistake considering the importance of Pollen and Ronner's work