Phase-Based Disparity Measurement

David J. Fleet

Department of Computing and Information Science
Queen's University

Allan D. Jepson

Department of Computer Science
University of Toronto

Michael R.M. Jenkin

Department of Computer Science
York University

Index Terms: Stereopsis, Frequency Analysis, Gabor Filters, Phase, Scale-Space, Zero-Crossings

Running head: Phase-Based Disparity Measurement

Published in CVGIP: Image Understanding, 53(2): 198–210, 1991

Phase-Based Disparity Measurement

David J. Fleet
Department of Computing and Information Science
Queen's University

Allan D. Jepson
Department of Computer Science
University of Toronto

Michael R.M. Jenkin
Department of Computer Science
York University

Abstract

The measurement of image disparity is a fundamental precursor to binocular depth estimation. Recently, Jenkin and Jepson (1988) and Sanger (1988) described promising methods based on the output phase behaviour of band-pass Gabor filters. Here we discuss further justification for such techniques based on the stability of band-pass phase behaviour as a function of typical distortions that exist between left and right views. In addition, despite this general stability, we show that phase signals are occasionally very sensitive to spatial position and variations in scale, in which cases incorrect measurements occur. We find that the primary cause for this instability is the existence of singularities in phase signals. With the aid of the local frequency of the filter output (provided by the phase derivative) and the local amplitude information, the regions of phase instability near the singularities are detected so that potentially incorrect measurements can be identified. In addition, we show how the local frequency can be used away from the singularity neighbourhoods to improve the accuracy of the disparity estimates. Some experimental results are reported.

1 Introduction

The use of local, band-pass, linear filters (channels) has become the focus of considerable research on early biological and computational visual processing. It is readily accepted that the amplitude and zero-crossings of such filter output provide a useful source of information for texture analysis and the measurement of binocular disparity, image orientation, and optical flow (e.g. [Adelson and Bergen, 1985; Beck, et.al. 1987; Fogel and Sagi, 1989; Heeger 1988; Marr and Poggio 1979, Mayhew and Frisby 1981, Nishihara 1984]). Recently, considerable promise has been exhibited by a new technique for binocular (stereo) disparity measurement in which disparity is expressed in terms of phase differences in the output of local, band-pass filters applied to the left and right views [Jenkin and Jepson, 1988; Sanger, 1988; Langley, et.al., 1990]. (Burt et.al. (1989) have suggested a similar approach with respect to the measurement of image velocity.) Although Gabor filters have been a natural choice for the band-pass prefiltering of the input, any quadrature pair of band-pass, constant-phase filters will suffice.

The main advantage of such local, phase-based approaches is that disparity estimates are obtained with sub-pixel accuracy, without requiring explicit sub-pixel signal reconstruction or sub-pixel feature detection and localization. The measurements may be used directly, or iteratively as predictions for further, more accurate, estimates. Because there are no restrictions to specific values of phase (e.g. zeros) that must first be detected and localized, the density of measurements is also expected to be high. Finally, the computations may be implemented efficiently in parallel.

In this paper we introduce another major advantage of phase-based approaches, namely, the stability of band-pass phase behaviour with respect to image deformations that typically exist between left and right stereo views, so that similar structure is generally available for matching. In particular, we argue that phase information is stable under scale perturbations and smooth contrast variations. We then describe two sources of measurement error with phase-based approaches. The first is caused by the occasional sensitivity of phase to small changes in scale between left and right views. The second is the result of using an implicit local model of phase, which becomes inappropriate where the phase behaviour is too sensitive to changes in spatial position. Both of these problems are linked to the existence of singularities in the phase signal. We describe the properties of singularity neighbourhoods for 1-d signals, and show that they can be detected easily, so that incorrect measurements may be removed. It is also shown that the accuracy of the disparity measurements can also be improved, at little extra computational cost, using the same information that is already required to detect singularity neighbourhoods.

These results are of general interest for several reasons. First, they also apply to zero-crossings of the filter output in that zero-crossings can be thought of as lines of constant phase. For example, zero phase in the response of a complex Gabor filter corresponds to a zero in the response of an odd-symmetric sine-Gabor kernel. Similarly, zero-crossings of the Laplacian of a Gaussian amount to zeros in the phase response of its corresponding analytic filter, the imaginary part of which is provided by its Hilbert transform. Second, the issues and results are relevant to phase-based approaches to the measurement of image velocity (e.g. from the spatiotemporal phase gradient [Fleet and Jepson, 1989]), and to phase-correlation techniques (e.g. see [Kuglin and Hines, 1975; Yeshurun and Schwartz, 1989; Olson and Potter, 1989]). Third, problems that cause incorrect measurements, such as scale variations between left and right views are not unique to phase-based approaches. They exist for any technique that tries to match features between two views. Therefore, the general stability of phase information and the fact that we can easily detect regions of instability should be viewed as unique advantages of phase-based approaches.

2 Phase-Difference Techniques

If the geometric configuration of two cameras is known, then the *scene-point disparity* (i.e. the relative distance between the projections of a single scene point onto the two imaging surfaces) encodes the 3-d depth of the point. By convention, assume that crossed disparities, which correspond to scene features nearer than the fixation point, are negative. Uncrossed disparities, which correspond to scene features farther away than the fixation point, are positive. The computation of *image-point disparity* (or simply, image disparity) refers to the matching of image features in one view with those in the other in order to derive an approximation to scene-point disparity. One major constraint on candidate image properties for the measurement of image disparity is that they must be *stable* (nearly invariant) under the typical deformations that we expect between the left and right views.

Most approaches to disparity measurement assume an initial stage of linear, band-pass smoothing in order to reduce the effects of noise, and to isolate (i.e. separate) image properties of interest, such as multiple scales. Here we assume a stage of pre-filtering based on Gabor functions [Gabor, 1946]. Let $R_l(x,\, k_0)$ and $R_r(x,\, k_0)$ be the left and right views after convolution with a Gabor kernel given by

$$Gabor(x;\ \sigma, k_0) \quad \equiv \quad e^{ixk_0}\, G(x;\sigma)\ , \tag{1}$$

where $G(x;\sigma)$ is a Gaussian window with standard deviation σ, and k_0 is referred to as the peak tuning frequency (in radians) to which the filter responds maximally. The amplitude spectrum of (1) is Gaussian-shaped and centred at k_0. In vision applications it is common to define the extent of the amplitude spectra to be one standard deviation $\sigma_k = \sigma^{-1}$, with bandwidths close to one octave. Because the filters are complex, so are the responses $R_l(x,\, k_0)$ and $R_r(x,\, k_0)$. Therefore, they may be written as

$$R_l(x;\, k_0) \;=\; \rho_l(x)\, e^{i[k_0 x + \psi_l(x)]}\ , \qquad R_r(x;\, k_0) \;=\; \rho_r(x)\, e^{i[k_0 x + \psi_r(x)]}\ , \tag{2}$$

where $\rho(x)$ and $\phi(x) \equiv [k_0 x + \psi(x)]$ denote the amplitude and phase components of response. The phase component has been split into two terms, a linear term $k_0 x$ that is determined by the peak frequency of the Gabor filter, and $\psi(x)$. We do this to emphasize that the response $R(x;\, k_0)$ of a band-pass filter can be viewed as a slowly varying modulation of the base signal $e^{ik_0 x}$ to which the filter is tuned. More precisely, we can write $R(x,\, k_0)$ as

$$R(x;\, k_0) \quad = \quad M(x;\, k_0)\; e^{ik_0 x} \;, \tag{3}$$

where $M(x;\, k_0) \equiv \rho(x) e^{i\psi(x)}$ is a low-pass signal. (This follows from the modulation property of Fourier transforms [Bracewell, 1978].) Following Whitham (1974) or Papoulis (1965), the local (instantaneous) frequency can be defined as the spatial derivative of the phase signal, $k(x) \equiv \phi'(x)$. If the phase of $M(x,\, k_0)$ is linear, as in $\psi(x) = k_1 x$, then $R(x;\, k_0)$ is just an amplitude-modulated sinusoid with constant frequency $k_0 + k_1$. Otherwise, the phase derivative $\phi'(x) = k_0 + \psi'(x)$ provides a local, constant-frequency approximation to $R(x;\, k_0)$. The local frequency plays a key role in what follows.

Following Jenkin and Jepson (1988) and Sanger (1988), the problem of computing binocular disparity can be formulated in terms of phase matching, that is, determining the shift required so that the phases of the left and right signals become equal.[1] The local image disparity, at a specific position x, for an initial guess d_i, is defined to be the shift $d(x)$ such that

$$\phi_l \left(x - \frac{d(x)}{2} \right) \quad = \quad \phi_r \left(x + \frac{d(x)}{2} \right) \;, \tag{4}$$

and $|d(x) - d_i|$ is as small as possible. Because of phase periodicity, this approach can only be expected to deal with shifts of less than half a wavelength to either side of the initial guess. If the disparity is too large (i.e. if the local wavelength is too small with respect to the shift) then the computed phase difference can be wrong by a multiple of 2π. This yields an incorrect disparity measurements. In order to use filters tuned to higher frequencies with relatively large disparities, a control strategy, such as coarse-to-fine propagation, is necessary [Nishihara, 1984; Jenkin and Jepson, 1988; Sanger, 1988; Langley, et.al., 1990]. In what follows we concentrate on the basic disparity measurements while assuming that the initial guess is sufficiently close to the true disparity; the control strategy is beyond the scope of this paper.

If the initial guess is sufficiently good and the filter outputs are sinusoidal with

[1] Some authors (e.g., [Sanger, 1988; Burt, et.al., 1989]) motivate the use of phase information from the Fourier shift theorem, which states that the translation of a signal amounts to a phase shift of each Fourier component by a proportion of its wavelength [Bracewell, 1978]. Here, we argue that the more important justification for the use of phase information is that it is a *stable* image *property* for matching (see Section 3).

constant frequency k_0, then the shift necessary to match the phases of the left and right filter outputs is given by (cf. Jenkin and Jepson (1988) and Sanger (1988))

$$\tilde{d}_0(x) \;\equiv\; \frac{[\phi_l(x) \,-\, \phi_r(x)]_{2\pi}}{k_0}\,, \tag{5}$$

where $[\theta]_{2\pi}$ denotes the principal part of θ, that is, $[\theta]_{2\pi} \in (-\pi, \pi]$. However, if the outputs are *not* sinusoidal with constant frequency k_0, then the disparity estimates provided by $\tilde{d}_0$ will not be exact. For example, if the left and right filter outputs are shifted versions of one another with disparity δ, then the phase difference in the numerator of (5) becomes

$$\Delta\phi(x) \;=\; \phi\left(x + \frac{\delta}{2}\right) \,-\, \phi\left(x - \frac{\delta}{2}\right) \,. \tag{6}$$

If $\phi(x)$ is smooth (singularities are discussed below), we can rewrite (6) with $\phi(x)$ expressed as a Taylor series about x as $\Delta\phi(x) = \delta\,\phi'(x) + O(\delta^3\,\phi'''(x))$. The disparity error, $\epsilon(x) = d(x) - \tilde{d}_0(x)$, with $\tilde{d}_0(x) = \frac{\Delta\phi(x)}{k_0}$ as in (5), can now be written as

$$\begin{aligned}\epsilon(x) \;&=\; \delta \,-\, \tilde{d}_0(x)\\ &=\; \delta\left(1 - \frac{\phi'(x)}{k_0}\right) \,+\, O\left(\frac{\delta^3\,\phi'''(x)}{k_0}\right) \,.\end{aligned} \tag{7}$$

Notice that the order-δ term in (7) arises from the discrepancy between the peak tuning frequency k_0 and the local frequency of the filter output $\phi'(x)$.

We can improve this technique by adopting a more general model, for which frequency should be locally constant, but not necessarily equal to the peak frequency k_0. More precisely, replace k_0 in (5) by the average local frequency between the left and right views. This yields

$$\tilde{d}_1(x) \;=\; \frac{[\phi_l(x) \,-\, \phi_r(x)]_{2\pi}}{\bar{k}(x)}\,, \qquad \text{where} \quad \bar{k}(x) \;=\; \frac{1}{2}(\phi_l'(x) + \phi_r'(x))\,. \tag{8}$$

Then, $\tilde{d}_1(x)$ yields the exact disparity when the left and right views both have locally constant frequencies, but not necessarily equal to the filter tuning k_0. When the left and right outputs are shifted versions of one another, but do not have constant frequency, the disparity error $\epsilon(x) = \delta - \tilde{d}_1(x)$ for (8) reduces to

$$\epsilon(x) \;=\; O\left(\frac{\delta^2\,\phi'''(x)}{\phi'(x)}\right) \,. \tag{9}$$

In comparison to (7), notice that (9) contains no order-δ term.

When the disparity estimates are not exact, then shifting the signals by $\tilde{d}_1(x)$ will not precisely match the phases in the left and right filter outputs as required by (4). Because the error depends on the initial disparity (i.e. δ in (7) and (9)), one way to improve the accuracy is to iterate the basic measurements. On one cycle of the iteration the images are first shifted according to the current disparity approximation, the local image disparity between the shifted left and right signals is computed using either $\tilde{d}_0$ or $\tilde{d}_1$, and the disparity approximation is updated with the result. That is, with an initial guess d_i, the disparity approximation at iteration $t+1$ is given by

$$\tilde{d}^{t+1}(x) \;=\; \tilde{d}^t(x) \;+\; \Delta\tilde{d}^t(x)\ , \tag{10}$$

where

$$\begin{aligned} \tilde{d}^0(x) \;&=\; d_i\ , \\ \Delta\tilde{d}^t(x) \;&=\; \frac{[\phi_l(x-\frac{\tilde{d}^t(x)}{2}) \;-\; \phi_r(x+\frac{\tilde{d}^t(x)}{2})]_{2\pi}}{k(x)}\ , \end{aligned}$$

and $k(x)$ is given by $\bar{k}(x)$ in (8) for the new predictor, or by k_0 for the old predictor (5). The estimate $\tilde{d}^t(x)$ converges to the exact disparity, as defined by (4). Note that with bandwidths of one octave the local frequency may be up to $k_0/2$ away from k_0, in which case the first-order errors in (7) can be as large as $\delta/2$. Therefore, an upper bound on the convergence rate for the old disparity predictor $\tilde{d}_0(x)$ is 1 bit per iteration. The new predictor $\tilde{d}_1(x)$ converges quadratically.

3 Phase Stability and Scale-Space Singularities

The transformation between the left and right stereo views is often assumed to be reasonably well approximated by image translation (i.e. $I_l(x) = I_r(x+d)$). We argue that an affine transform,

$$I_l(a(x)) \;=\; I_r(x)\ , \qquad \text{where} \quad a(x) \;=\; a_0 \;+\; a_1 x\ . \tag{11}$$

is a more realistic first approximation. For, even in the case of a nearly planar surface, scale variations of up to 20% between the two views are not uncommon [Ogle, 1950].

If we assume an affine transformation as in (11), then, because the filters have constant (octave) bandwidth, the two output signals R_l and R_r satisfy

$$R_l(a(x),\, k_1) \;=\; R_r(x,\, k_2)\ , \qquad \text{where} \quad k_2 \;=\; k_1\, a_1\ . \tag{12}$$

That is, a rescaling of the input can be matched by changing the peak tuning frequency so that the same image structure is extracted by the filters. In this case, the two outputs are related by the same affine transformation provided that differently tuned filters were applied to the left and right inputs. We might expect to recover $a(x)$ by matching $R_l(x, k_1)$ with $R_r(x, k_2)$. Unfortunately, k_2 depends on the unknown scale factor a_1. Instead, we attempt to recover $a(x)$ from the left and right responses of the same filter, that is, from $R_l(x, k_1)$ and $R_r(x, k_1)$. For this to succeed, with a useful result, the structure of $R_r(x, k_2)$ used for matching must be well approximated by the structure of $R_r(x, k_1)$. In other words, the image property used for binocular matching must be *stable* under small scale perturbations.

3.1 Gabor Scale-Space

To demonstrate the dependence of Gabor output on spatial position and scale we use a *Gabor scale-space expansion* that expresses the filter output as a function of spatial position and the principal wavelength to which the filter is tuned. Similar expansions have been considered by Witkin (1983), Koenderink (1984), Mallat (1989), Lindeberg (1990), and others based on Gaussians (low-pass filtering) and derivatives of Gaussians (band-pass filtering). Let the scale-space expansion of $I(x)$ be given by

$$S(x, \lambda) \;=\; Gabor(x;\ \sigma(\lambda), k(\lambda)) \;*\; I(x)\ , \tag{13}$$

where $*$ denotes convolution, $Gabor(x;\ \sigma, k)$ is defined in (1), and λ is the scale parameter. The peak tuning frequency of the filter is given by

$$k(\lambda) \;=\; \frac{2\pi}{\lambda}\ . \tag{14}$$

With a bandwidth of β octaves (assumed to be near 1), and with one standard deviation as a measure of the extent of the Gaussian envelope (1), the standard deviation of the amplitude spectra σ_k should satisfy $\beta = \log_2[(k(\lambda) + \sigma_k)/(k(\lambda) - \sigma_k)]$. From this, it is easy to show that the approximate radius of spatial support is given by

$$\sigma(\lambda) \;=\; \frac{1}{k(\lambda)} \left(\frac{2^\beta + 1}{2^\beta - 1} \right)\ . \tag{15}$$

The amplitude and phase components of $S(x, \lambda)$ are given by

$$\rho(x, \lambda) \;=\; |S(x, \lambda)| \;\equiv\; \sqrt{(\mathrm{Re}[S(x, \lambda)])^2 \,+\, (\mathrm{Im}[S(x, \lambda)])^2}\ , \tag{16a}$$

$$\phi(x, \lambda) \;=\; \arg[S(x, \lambda)] \;\equiv\; \mathrm{Im}\,[\log_e S(x, \lambda)]\ . \tag{16b}$$

Figure 1(top) shows an signal composed of white Gaussian noise concatenated with a scanline from a real image. The two middle images show $\rho(x, \lambda)$ and $\phi(x, \lambda)$ over an interval of two octaves. Level contours of constant amplitude and phase are shown below. In the context of the scale-space expansion, an image property is said to be *stable* where its level contours are vertical.

It is clear from Figure 1 that $\rho(x,\,\lambda)$ depends significantly on scale because its level contours are not generally vertical. As a consequence, the amplitude signal and the raw filter output (which has a significant amplitude component) are not good candidate properties for matching. By contrast, the phase structure is generally stable with respect to scale perturbations. It is also clear, however, that in some regions phase *does* depend significantly on scale. As explained below, a major source of instability is the occurrence of singularities in the phase signal.

3.2 Singularity Neighbourhoods

For a general image $I(x)$, the scale-space defined by (13) is analytic, and contains a number of isolated zeros, where $S(x,\,\lambda) = 0$. In the amplitude signal shown in Figure 1, zeros appear as black spots. The phase signal $\phi(x,\,\lambda)$ is also analytic, except at the zeros of $S(x,\,\lambda)$ where $\phi(x,\,\lambda)$ is undefined. The expected density of these singularities is proportional to the peak tuning frequency [Papoulis, 1965; Fleet, 1990]. They can be generated easily by superimposing two complex exponentials with the same amplitudes and nearby frequencies. At a phase singularity, the band-pass signal passes through the origin of the complex plane. From the continuity of $S(x,\,\lambda)$ it can be argued that at scales just above and below the singular point the band-pass signal passes very close to the origin. As a consequence, phase and/or amplitude change quickly as a function of spatial position. Below we describe this characteristic behaviour of $\rho(x,\,\lambda)$ and $\phi(x,\,\lambda)$ near singular points; a more technical discussion and details can be found in [Fleet, 1990]. In what follows, let $(x_0,\,\lambda_0)$ denote the location of a singularity.

The neighbourhoods just above singular points (i.e. for $\lambda > \lambda_0$) are characterized by local frequencies that are significantly below the corresponding peak frequencies $k(\lambda)$ (see Figure 2, top panel). Within these neighbourhoods there exist *retrograde regions* within which $\phi_x(x,\,\lambda) < 0$. Along the boundaries of retrograde regions (which begin and terminate at singular points) the local frequency is zero; i.e. $\phi_x(x,\,\lambda) = 0$. This is significant because where $\phi_x(x,\,\lambda) = 0$ the level phase contours are horizontal, and not vertical as desired. Near these boundaries, both inside and outside the retrograde regions, the level contours are still nearly horizontal, which means that phase matching will be very sensitive to small changes in scale.

Below singular points (i.e. for $\lambda < \lambda_0$) the neighbourhoods are characterized by local frequencies that are significantly higher than corresponding peak tuning frequencies. In addition, the local frequency of response changes rapidly as a function of spatial location, with local maxima that roughly coincide with the singular locations. (This is evident in the bottom panel of Figure 2.) Although this rapid variation in phase behaviour is not a source of instability with respect to scale, it will cause problems for disparity computation that is based on a constant-frequency model of the local filter output as in (8). The numerical approximation of $\phi'(x)$ for use in (8) will be poor, the local domains of convergence will be extremely small, and the convergence will be slow.

Figure 3 shows a detailed example of level phase contours near a phase singularity. The singular point lies near the center at the point where the phase contours intersect, which is at the bottom of the small ellipsoidal contour that marks the retrograde boundary (i.e. the contour $\phi_x(x, \lambda) = 0$). The higher density of phase contours below the singularity reflects the higher local frequencies. The existence of nearly horizontal level contours above the singularity reflects the instability.

Finally, the neighbourhoods to the left and right (i.e. spatially adjacent) of singular points can be characterized in terms of amplitude variation. As we approach a singular point, $\rho(x, \lambda_0)$ goes to zero. Based on a simple linear model of amplitude, at a specific location x_1 near the singularity, the distance to the singular point is approximately $|\Delta x| = \rho(x_1, \lambda_0)/|\rho_x(x_1, \lambda_0)|$. Equivalently, as we approach the singularity $|\rho_x(x_1, \lambda_0)|/\rho(x_1, \lambda_0)$ increases. As a consequence, the neighbourhoods to the left and right of phase singularities can be characterized by large values of $|\rho_x(x, \lambda)|/\rho(x, \lambda)$.

3.3 Detection of Singularity Neighbourhoods

As explained above, it is necessary that singularity neighbourhoods be detected so that incorrect disparity estimates can be avoided. Here we describe constraints that may be used to identify singularity neighbourhoods without requiring the explicit localization of the singular points.

To detect the neighbourhoods above and below the singular points we suggest a constraint on the local frequency in relation to the peak tuning frequency. In particular, we constrain the distance between local frequency of response $\phi_x(x, \lambda)$ and the peak frequency $k(\lambda)$, which can be expressed as a function of the standard deviation of the amplitude spectrum $\sigma_k(\lambda)$. This amounts to a condition that

$$\frac{|\phi_x(x, \lambda) - k(\lambda)|}{\sigma_k(\lambda)} < \tau_k , \qquad \sigma_k(\lambda) = k(\lambda) \left(\frac{2^\beta - 1}{2^\beta + 1} \right) . \tag{17}$$

Level contours of (17) for different values of τ_k are shown in Figure 4 (top-left) for the same scale-space as that shown in Figure 1. They form 8-shaped regions with the singular points at their centres.

The neighbourhoods to the left and right of singular points can be detected with a constraint on local amplitude variation:

$$\sigma(\lambda) \frac{|\rho_x(x, \lambda)|}{\rho(x, \lambda)} < \tau_\rho , \tag{18}$$

where $\sigma(\lambda)$ defines the radius of the filter support (15). If $\rho(x_1, \lambda_0)/|\rho_x(x_1, \lambda_0)|$ is taken to be an approximation to the distance from a singularity, then (18) detects points that lie within a distance of $\sigma(\lambda)/\tau_\rho$ from a singularity. Levels contours of (18) for different values of τ_ρ are shown in Figure 4 (top-right). They form ∞-shaped regions with the singularities at their centres.

As the thresholds τ_k and τ_ρ decrease, the constraints become tighter and larger neighbourhoods are detected. Figure 4 (middle-left) shows the combined behaviour of the two thresholds: *1)* with $\tau_k = 1.2$, so that local frequencies are accepted up to 20% outside the nominal tuning range of the filters; and *2)* with $\tau_\rho = 1.0$, so that points within $\sigma(\lambda)$ of a singularity are discarded (in the experiments below, the filter outputs are sub-sampled with one (complex) sample every $\sigma(\lambda)$, so that $\tau_\rho = 1.0$ marks all points within one pixel of a singularity). Finally, the last three panels of Figure 4 show the level phase contours before the removal of singularity neighbourhoods (middle-right), the contours that survive the constraints (bottom-left), and the contours in those regions detected by the constraints (bottom-right). In total, these two constraints typically remove between 15% and 20% of the scale-space area.

Finally, it is important to note that a constraint on absolute amplitude is also desirable. If the input has no significant power at frequencies to which a filter is tuned, then noise may dominate the response, or there may be no variation in amplitude or local frequency. The constraints outlined above are designed to detect the pathological behaviour of phase signals where there is significant local energy. They do not detect a fundamental lack of signal. Therefore, whenever the local amplitude is sufficiently small, the filter output should be ignored. (In the experiments reported below, we discard all responses whose amplitude was less than 5% the maximum response of the filter over that image.)

4 Experimental Results

To demonstrate the improved disparity estimates that result from the removal of singularity neighbourhoods and the predictor $\tilde{d}_1(x)$ in (8), we have done experiments with affine deformations between the left and right views. The data for one set of experiments was collected from scanlines from real images (results from white noise input were similar in most respects). We also apply the new technique to a real pair of images in comparison to the technique of Jepson and Jenkin (1989).

4.1 Experiments with Affine Deformation

In what follows, Gabor filters were used with bandwidths of 0.8 octaves and principal wavelengths between 4 and 100 pixels. For illustrative purposes some outputs were over-sampled (as in Figures 1, 3, and 6), but all computations used an effective sub-sampling rate of one (complex) sample every $\sigma(\lambda)$, where $\sigma(\lambda)$ denotes the standard deviation of the Gaussian window. The computation of local frequency as well as the amplitude derivative is described in detail in the Appendix.

Figure 6 shows the errors that result near singularity neighbourhoods. The left and right phase signals were shifted versions of a slice of the scale-space (with $\lambda = 20$) shown in Figure 1. Figure 5 (left) shows this slice marked on the scale-space phase contours, and Figure 5 (right) shows the corresponding phase and local frequency as a function of spatial position. Notice from Figure 5 that this slice crosses three singularity neighbourhoods, two just above singular points and one just below. For the experiment, the error in the initial guess was uniformly 25% of the tuning wavelength (i.e. $\delta = 5$ pixels). Figure 6 (top) shows the results of $\tilde{d}_0(x)$ without the removal of singularity neighbourhoods. The middle result shows $\tilde{d}_0(x)$ with disparity measurements removed whenever either of the left or the right filter responses did not satisfy the two constraints (17) and (18) with $\tau_\rho = 1.0$ and $\tau_k = 1.2$. The removal of large errors caused by the poor local model are evident. Finally, Figure 6 (bottom) shows the more accurate results of $\tilde{d}_1(x)$, with the same singularity neighbourhoods removed.

Sanger (1988) avoids some of the errors that occur in singularity neighbourhoods with the use of smoothing applied to raw disparity estimates and a constraint on the difference in amplitude between the left and right signals. The amplitude difference will catch some of the incorrect estimates. In particular, when the amplitude derivative is large, there will often (but not always) be a large difference in amplitude between the left and right views. The smoothing reduces many of the remaining errors, but in doing so it sacrifices resolution and the accuracy of other estimates. Also note that Sanger did not iterate the approach (as in (10)), and therefore the use of the peak tuning frequency in (5) limits the magnitude of the errors produced. Iteration would push these errors further from the true disparity (this is shown below in Figure 11). Jepson and Jenkin (1989) employed a constraint on disparity gradients in terms of spatial phase differences in the individual views. This constraint is similar to that in (17) and removes some of the errors caused by phase behaviour in singularity neighbourhoods. But it was found to be unreliable.

Figure 7 shows mean disparity error and standard deviation bars for the old and new predictors ($\tilde{d}_0$ and $\tilde{d}_1$) as a function of filter scale (for $4 \le \lambda \le 64$). There was no scale change between the left and right views, and the error in the initial guess was again 25% of the tuned wavelengths (i.e. $\delta(\lambda) = 0.25\lambda$). Relative prediction error was computed as $e(x) = 100.0\,(\delta(\lambda) - \tilde{d}(x))/\delta(\lambda)$. Singularity neighbourhoods were detected using (17) and (18) as in Figure 6 with $\tau_\rho = 1.0$ and $\tau_k = 1.2$. The improved accuracy of $\tilde{d}_1(x)$ is evident. Also note the significant positive bias in $\tilde{d}_0(x)$. Because real images tend to have more power at lower frequencies, the local frequency $\phi'(x)$

will often be less than the peak tuning frequency. Therefore the order-δ term in (7) has a positive average value, which causes the observed bias. This bias is absent from the results of $\tilde{d}_1(x)$ since the local frequency is computed.

Figure 8 shows the dependence of $\tilde{d}_1(x)$ on the error in the initial guess $\delta(\lambda)$ with no scale change between the two views. Disparities were between 2% and 44% of the wavelengths to which the filters were tuned (i.e. $\delta(\lambda) = \alpha\lambda$, for $\alpha = 0.02, ..., 0.44$). The wavelengths used were between 25 and 100 pixels. Prediction error was computed as $e(x) = 100.0\,(\delta(\lambda) - \tilde{d}_1(x))/\lambda$. Although the results are extremely good, the errors increase dramatically for $\delta(\lambda) > 0.38\lambda$. To explain this, notice that with $\beta = 0.8$ octaves and $\tau_k = 1.2$, (17) removes all points with $\phi'(x) > 1.32\,k(\lambda)$. This means that local wavelengths may be as small as $0.76\,\lambda$, in which case the domain of convergence may be as small as 0.38λ. As the initial disparities $\delta(\lambda)$ increased beyond 0.38λ, a larger proportion of initial guesses fell within the wrong domain of convergence. When the boundary between the domains of convergence is crossed, the disparity estimates are wrong by about one wavelength (because the phase difference wraps around modulo 2π). For Figure 8 only positive disparities were used. Therefore, a histogram of disparity estimates should reveal two peaks, one at $\delta(\lambda)$, and the other near $\delta(\lambda) - \lambda$. This explains the positive error bias for $\delta(\lambda) > 0.38\lambda$ and the rapid increase in the standard deviation. For negative disparities a corresponding negative bias appears.

Figure 9 shows the dependence of $\tilde{d}_1(x)$ on scale differences (up to 35%) between the left and right views. The filter wavelengths used were between 12 and 48 pixels. Scale differences between the two views can be simulated by applying filters with different peak frequencies to a single input signal. The top panel shows prediction error as a function of scale variation between the left and right views when the initial guess was exact; i.e. there was scale change between the left and right views locally, but *no* spatial shift. In terms of scale-space expansions (e.g. Figure 1), the prediction error here is caused solely by the horizontal drift of level phase contours as scale changes. It is encouraging to note that the prediction errors remain tolerable even for scale changes of 20% or more. The next three plots, from top to bottom, show the prediction error as we introduce spatial shifts (errors in the initial guess) of 10%, 20%, and 30% of the tuned wavelength. Again, it is encouraging that the prediction errors increase only slightly. In all cases, note that up to 20% scale variation may occur between the left and right views while the disparity errors remain mainly below 10% of the filter wavelength.

Finally, it is important to note that all disparity estimates obtained in singularity neighbourhoods (as defined by (17) and (18) with $\tau_\rho = 1.0$ and $\tau_k = 1.2$) were not included in the statistics reported in Figures 6 – 8. Whereas the inaccuracies shown here are generally limited to a small fraction of a wavelength, the errors that occur in singularity neighbourhoods are often as large as plus or minus one wavelength. As a consequence, when they are included the error variances grow dramatically and the error behaviour shown and discussed above, especially with respect to Figures 7 and 8, is not readily discernable.

4.2 Application to Real Images

The previous examples have shown the effect of this new technique in 1-d. In order to demonstrate the results further, an existing algorithm [Jepson and Jenkin, 1989] has been modified to incorporate the improved disparity predictor and the removal of singularity neighbourhoods. A coarse-to-fine control strategy based on that of Nishihara (1984) is used. The Gabor filters had bandwidths of one octave, and the different wavelengths to which the filters were tuned were 4, 8, 16, and 32 pixels. The coarsest scale begins computation with initial guesses of $d_i = 0$. Computed disparities at one level are then passed down as initial guesses for the next finer level. The results reported below, for both the new and the old methods, show only those disparity estimates computed from the finest channel. The technique is demonstrated on a stereogram of a car part,[2] the left view of which is shown in Figure 10. Although not known accurately, the actual disparities range between -1 and 20 pixels.

As mentioned in Section 4.1 the techniques described in Sanger (1988) and Jepson and Jenkin (1989) used several conditions and smoothing to remove some of these estimates. The goal of this experiment, and Figures 11 and 12, is illustrate how important and common the measurement errors due to phase singularities can be. Towards this end, Figure 11 shows the output of the old predictor with 1 and then 11 iterations, without the removal of singularity neighbourhoods or any of the *ad hoc* methods of removing incorrect estimates. The occurrences of incorrect measurements are clear. In both plots disparity is encoded as height. The incorrect estimates are more obvious in the second case as the disparity estimates are allowed to converge. With only one iteration, the disparity estimates, $\tilde{d}_0$, are bounded by the model of constant frequency (taken to be the peak tuning frequency). With further iterations the sizes of estimated disparity are not so constrained. Thus, with more iterations the adverse effects of singularity neighbourhoods become clearer.

Figure 12 shows the response of the same algorithm taking advantage of the improved phase-based disparity predictor $\tilde{d}_1(x)$, with only one iteration and the removal of singularity neighbourhoods. Singularity neighbourhoods were detected using constraints on local frequency and amplitude. The results do not change significantly with

[2]The Renault Car part stereogram is a widely available stereo pair that is often used in the literature. The images are 256x256 8-bit images. The car part is roughly 'T' shaped, with the 'T' being slightly inclined in depth.

this method if more iterations are allowed. Figure 12 (top-left) shows the estimated disparity encoded as intensity. In this image it is clear, at least, that the results are well isolated to the car part (areas of black indicate discarded measurements). For illustrative purposes, these estimates were crudely blurred (square-wave averaging over active neighbours with radius of 2 pixels). The result, shown in Figure 12 (top-right), better illustrates the depth variation. The same result is also displayed below as a perspective height plot so that comparisons can be made with Figure 11. The structure found throughout the car part is somewhat difficult to compare to the old method at first glance. However, the noise throughout the background region around the part has been clearly removed. Moreover, upon closer inspection, the accuracy and robustness of the new technique can be seen. It is especially clear at the boundaries of the part.

5 Summary

Phase information in the output of band-pass linear filters is useful for the measurement of binocular disparity. Although often motivated with the Fourier shift theorem, we argue that one of the major advantages of phase information is its stability with respect to affine deformations and contrast differences between the left and right views. In particular, phase is considerably more stable than amplitude. This is important if the image-point disparity is to be used as an approximation to scene-point disparity.

It was shown that, despite the general stability of phase information, there exist regions of pathological behaviour near phase singularities where disparity estimates should not be trusted. It was shown further that simple constraints can be used to detect these regions and discard incorrect measurements. Finally, the local frequency of the filter response was shown to play a key role in both the detection of singularity neighbourhoods and in the construction of a more accurate disparity predictor. Experiments in the final sections of the paper showed that affine deformations between the left and right views are handled well provided that singularity neighbourhoods are removed. In fact, scale changes of up to 20% were handled very well. These results provide a contribution to the existing techniques of Jenkin and Jepson (1988) and Sanger (1988) with respect to measurement accuracy, robustness, and our theoretical understanding.

Acknowledgements: This research was financially supported by NSERC and ITRC.

References

Adelson, E.H. and Bergen, J.R. (1985) Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* 2, pp. 284-299

Beck, J., Sutter, A., and Ivry, R. (1987) Spatial frequency channels and perceptual grouping in texture segregation. *Comp. Vision Gr. Image Proc.* 37, pp. 299-325

Bracewell, R.N. (1978) The Fourier Transform and its Applications. McGraw-Hill, New York

Burt, P.J., et.al. (1989) Object tracking with a moving camera. *Proc. IEEE Workshop on Motion*, Irvine, California, pp. 2-12.

Fleet, D.J. (1990) Measurement of Image Velocity. PhD Dissertation, Department of Computer Science, University of Toronto.

Fleet, D. and Jepson, A. (1989) Computation of normal velocity from local phase information. *Proc. IEEE CVPR*, San Diego, pp. 379-386

Fogel, I. and Sagi, D. (1989) Gabor filters for texture discrimination. *Biol. Cybern.* 61, pp. 103-113

Gabor, D. (1946) Theory of communication. *J. IEE* 93, pp. 429-457

Heeger, D.J. (1987) A model for the extraction of image flow. *J. Opt. Soc. Am. A* 4, pp. 1455-1471

Jenkin, M. and Jepson, A.D. (1988) The measurement of binocular disparity. *in* **Computational Processes in Human Vision**, (ed.) Z. Pylyshyn, Ablex Publ. New Jersey

Jepson, A.D. and Jenkin, M. (1989) Fast computation of disparity from phase differences. *Proc. IEEE CVPR*, San Diego, pp. 398-403

Koenderink, J.J. (1984) The structure of images. *Biol. Cybern.* 50, pp. 363-370

Kuglin, C. and Hines, D. (1975) The phase correlation image alignment method. *Proc. IEEE Int'l Conf. Cybern. Society*, pp. 163-165

Langley, K., Atherton, T.J., Wilson, R.G., and Larcombe, M.H.E. (1990) Vertical and horizontal disparities from phase. Proc. 1st ECCV, Antibes, Springer-Verlag Lecture Series in Computer Science, vol. 427, pp. 315-325

Lindeberg, T. (1990) Scale-space for discrete signals. *IEEE Trans. PAMI* 12, pp. 234-254

Mallat, S. (1989) A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. PAMI* 11, pp. 674-693

Marr, D. and Poggio, T. (1979) A computational theory of human stereo vision. *Proc. R. Soc. Lond.* B204, pp. 301-328

Mayhew, J. and Frisby, J. (1981) Computational studies toward a theory of human stereopsis. *Artificial Intelligence* 17, pp 349-385

Nishihara, K. (1984) Prism: A practical real-time imaging stereo matcher. A.I. Memo 780, AI Lab, MIT

Ogle, K.N. (1950) Research in Binocular Vision, Saunders, Philadelphia

Olson, T. and Potter, R. (1989) Real-time vergence control. *Proc. IEEE CVPR*, San Diego, pp. 404-409

Papoulis, A. (1965) Probability, Random Variables, and Stochastic Process. McGraw-Hill, Singapore

Sanger, T. (1988) Stereo disparity computation using Gabor filters. *Biol. Cybern.* 59, pp. 405-418

Whitham, G.B. (1974) Linear and Nonlinear Waves. John Wiley and Sons, New York

Witkin, A. (1983) Scale-space filtering. *Proc. IJCAI*, Karlsruhe, pp. 1019-1022

Yeshurun, Y. and Schwartz, E. (1989) Cepstral filtering on a columnar image architecture: A fast algorithm for binocular stereo segmentation. *IEEE Trans. PAMI* 11, pp. 759-767

Appendix: Computation of Local Frequency and Amplitude Derivative

Here we describe a straight-forward method that may be used to compute the phase derivative (local frequency) $\phi'(x)$, and the relative amplitude derivative $\rho'(x)/\rho(x)$. The computation is based on the following identities:

$$\phi'(x) \;=\; \frac{\mathrm{Im}[R^*(x)\ R'(x)\ ]}{|R(x)|^2}\ , \tag{19a}$$

$$\frac{\rho'(x)}{\rho(x)} \;=\; \frac{\mathrm{Re}[R^*(x)\ R'(x)\ ]}{|R(x)|^2}\ , \tag{19b}$$

where $R^*(x)$ denotes complex conjugate of $R(x)$, and $\mathrm{Im}[z]$ denotes the imaginary part of z. In terms of the real and imaginary parts of $R(x)$ and $R'(x)$, (19) reduces to

$$\phi'(x) \;=\; \frac{\mathrm{Im}[R'(x)]\,\mathrm{Re}[R(x)] \;-\; \mathrm{Re}[R'(x)]\,\mathrm{Im}[R(x)]}{\mathrm{Re}[R(x)]^2 + \mathrm{Im}[R(x)]^2}\ , \tag{20a}$$

$$\frac{\rho'(x)}{\rho(x)} \;=\; \frac{\mathrm{Re}[R'(x)]\,\mathrm{Re}[R(x)] \;+\; \mathrm{Im}[R'(x)]\,\mathrm{Im}[R(x)]}{\mathrm{Re}[R(x)]^2 + \mathrm{Im}[R(x)]^2}\ . \tag{20b}$$

Thus from (20), it is clear that we may concentrate on the differentiation of $R(x)$, and hence avoid the issues of phase unwrapping and phase singularities that would exist if we were to compute $\phi'(x)$ directly from a sub-sampled representation of $\phi(x)$.

Because $R(x)$ is a band-pass filtered signal, we expect its power to be concentrated about the peak frequency k_0. For efficiency, we should also expect the response to be sub-sampled at a rate determined by the bandwidth (the extent of the amplitude spectrum). To see that $R(x)$ may be sub-sampled and reconstructed from the subsampled representation, note that $M(x) \equiv R(x)e^{-ik_0 x}$ amounts to shifting the power of $R(x)$ to the origin so that $M(x)$ is low-pass (cf. Section 2). Therefore, a sub-sampled representation of $M(x)$ can be interpolated and differentiated using standard methods.

Alternatively, it is easy to show that we may sub-sample $R(x)$ directly, and then interpolate (or differentiate) with only a slightly more complicated kernel. To see this, let

$$R(x) \;=\; M(x)\,c(x)\ , \qquad \text{where} \quad c(x) \;=\; e^{ik_0 x}\ . \tag{21}$$

The derivative of $R(x)$ has the form

$$R'(x) \;=\; M'(x)\,c(x) \;+\; M(x)\,c'(x)$$

$$= \; c(x) \sum_n \left(M(n) h(x-n) \right) \; + \; ik_0 R(x) \; , \tag{22}$$

where $M(n)$ denotes a discrete, sub-sampled version of $M(x)$, $h(x)$ is a standard kernel for numerical differentiation (expressed as convolution), and $c'(x) = ik_0 c(x)$. If we replace $M(n)$ by $R(n)\, c(-n)$, then (22) becomes

$$\begin{aligned} R'(x) \; &= \; c(x) \sum_n \left(R(n) c(-n) h(x-n) \right) \; + \; ik_0 R(x) \\ &= \; c(x) c(-x) \sum_n \left(R(n) c(x-n) h(x-n) \right) \; + \; ik_0 R(x) \\ &= \; \sum_n \left(R(n)\, H(x-n) \right) \; + \; ik_0 R(x) \; , \end{aligned} \tag{23}$$

where $H(x) = h(x)\, c(x)$ is the new kernel for differentiating $R(x)$ according to (23). In other words, the new kernel is a modulated form of whatever kernel was chosen for numerical differentiation of the low-pass signal.

For example, for numerical differentiation of a low-pass signal $M(n)$ at the nodes of the sampling lattice, with a sub-sampling distance of s, we may convolve $M(n)$ with a discrete kernel $h(n) = \frac{1}{12s}(-1,\, 8,\, 0,\, -8,\, 1)$ based on a 4-pt central-difference formula. To apply a corresponding difference scheme to $R(x)$ we use a discrete kernel $H(n) = \frac{1}{12s}(-e^{-i2sk_0},\, 8e^{-isk_0},\, 0,\, -8e^{isk_0},\, e^{i2sk_0})$. Discrete convolution between $R(n)$ and $H(n)$ yields the first term in (23) . In summary, the computation of the amplitude derivative and the local frequency in (18) amounts to three steps:

1. convolution of the sub-sampled version of $R(x)$ with the complex kernel $H(x)$;

2. addition of the result to $ik_0 R(x)$ as in (23);

3. projection of $R'(x)$ onto $R^*(x)$ as in (20).

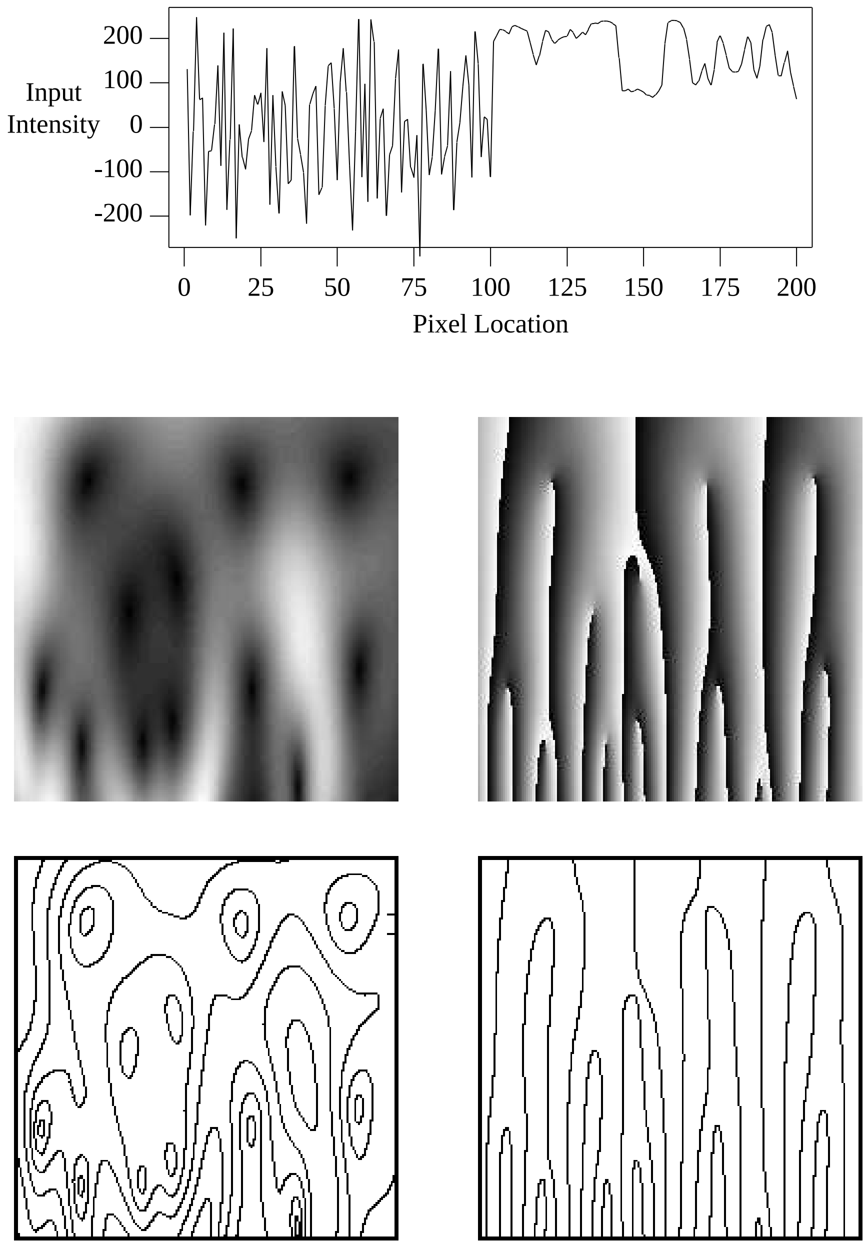

Figure 1. **Gabor Scale-Space Expansion:** *The input signal (top) consists of white Gaussian noise (left), and a scanline from a real image (right). The middle images show the amplitude and phase components of* $S(x, \lambda)$ *for* $12 \le \lambda \le 48$ *(in pixels). The vertical and horizontal axes represent scale and spatial position. The bottom two figures show level contours of constant* $\rho(x, \lambda)$ *and* $\phi(x, \lambda)$*. Scales for* $\lambda = 38$ *and* $\lambda = 41$ *are marked by ticks on the upper right side of (bottom-left) for reference in Figure 2.*

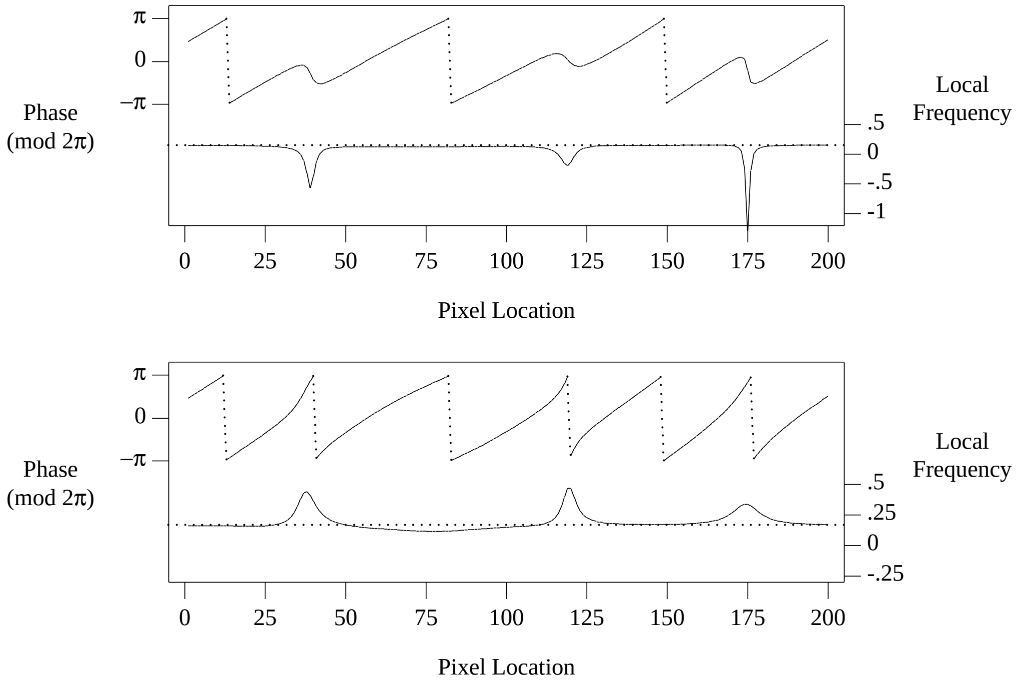

Figure 2. **Phase and Local Frequency Behaviour Near Singularities:** *Phase and local frequency are shown as a function of spatial position for two nearby scales from the scale-space in Fig. 1 (marked on Figure 1(bottom-left)). The first scale ($\lambda = 41$) lies just above three zeros of $\rho(x, \lambda)$. The second ($\lambda = 38$) lies just below the same zeros. Here, the vertical dotted lines denote phase wrapping between $-\pi$ and π (not discontinuities). The horizontal dotted lines mark the peak tuning frequencies of the two scales.*

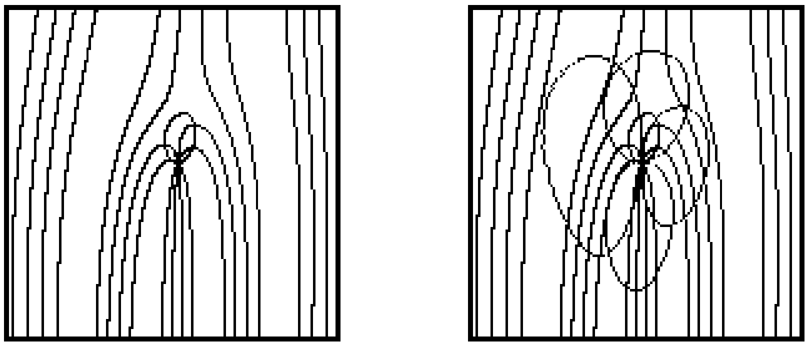

Figure 3. **Level Phase Contours in Singularity Neighbourhoods:** *Level phase contours are shown near a singularity. (left) The small ellipsoidal contour marks locations at which $\phi_x(x, \lambda) = 0$. The singularity lies at the bottom of this contour. (right) The added contours in the right panel show the regions delineated by the constraints used to detect the singularity neighbourhoods (cf. Section 3.3).*

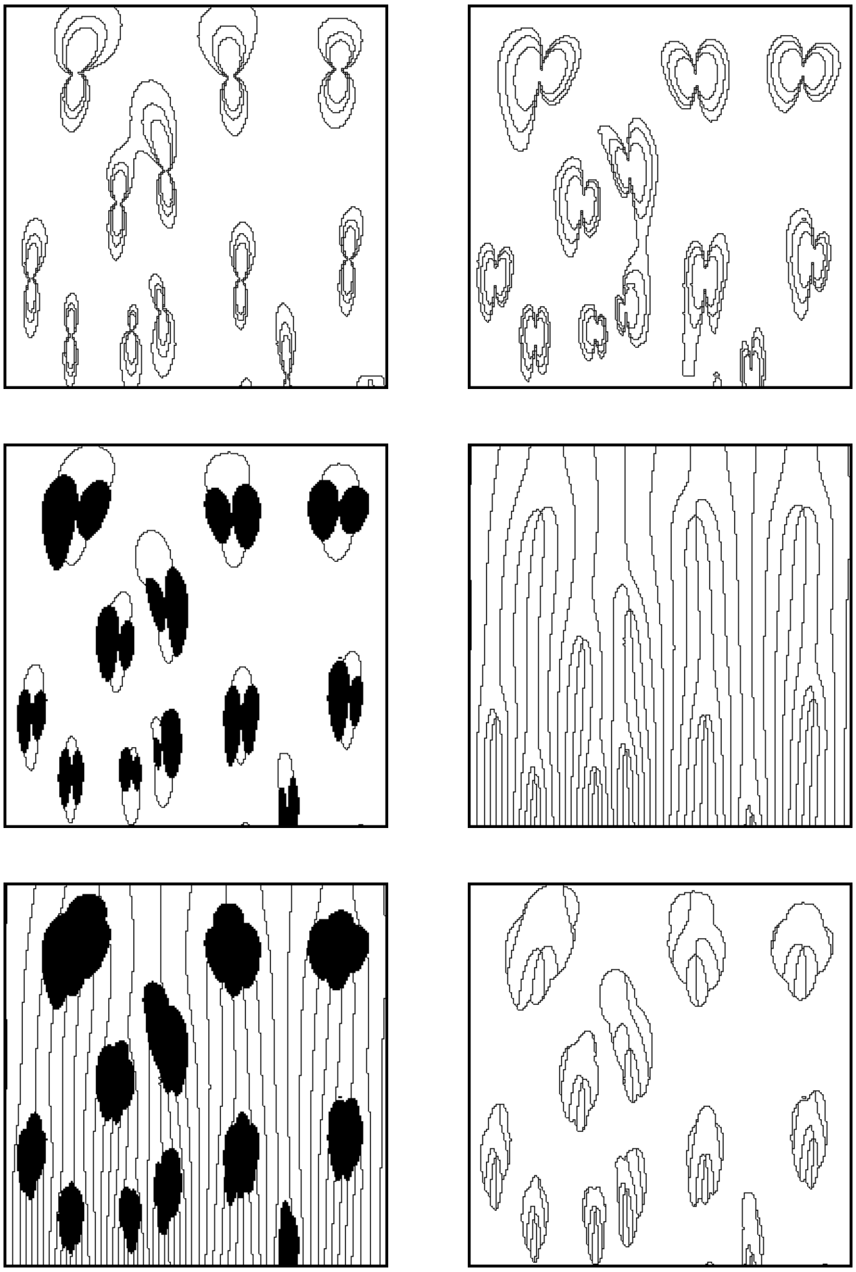

Figure 4. **Detection of Singularity Neighbourhoods:** *(top-left) Levels contours of the local frequency constraint (17) for* τ_k = 1.0, *1.5, and 2.0. (top-right) Level contours of the amplitude constraint (18) for* τ_ρ = 0.75, *1.0, and 1.5. (middle-left) Combined regions removed by both constraints with* $\tau_\rho = 1.0$ *and* $\tau_k = 1.2$. *(cf. Fig. 3.) Here, for convenience the region removed by (18) is shown in black, while the contours delineate the remaining regions detected by (17). (middle-right) Level phase contous corresponding to Figure 1. (bottom-left) Phase contours that survive the constraints. (bottom-right) Phase contours in regions removed by the constraints. Note that the constraints capture the unstable regions.*

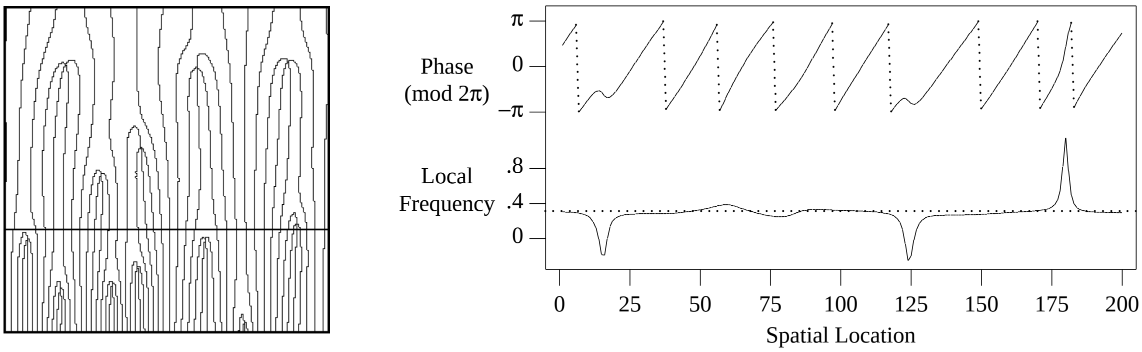

Figure 5. **Scale-Space Slice:** *The horizontal lines shows the slice of the scale-space at $\lambda = 20$ pixels. On the right we show phase and local frequency as a function of spatial position for this slice.*

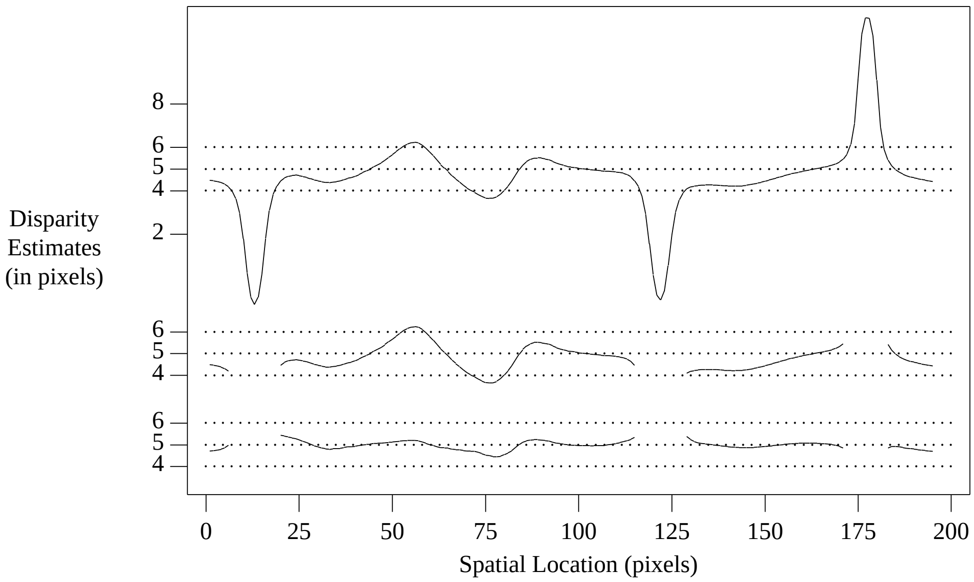

Figure 6. **Disparity Measurement:** *The top two plots show the disparity estimates of $\tilde{d}_0(x)$ without, and then with, the singularity neighbourhoods removed. The bottom plot shows the results of the improved predictor $\tilde{d}_1(x)$, with the singularity neighbourhoods removed. The tuned wavelength λ was 20 pixels and the disparity (error in the initial guess) was 5 pixels.*

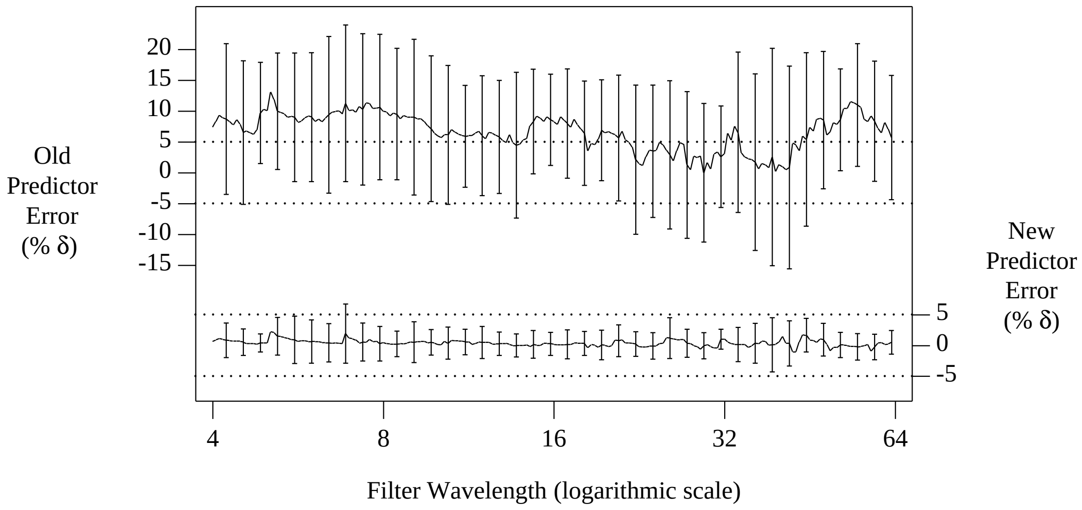

Figure 7. **Disparity Error for Old and New Predictors:** *Mean prediction error and standard deviation bars for* $\tilde{d}_0(x)$ *and* $\tilde{d}_1(x)$ *are shown as a function of filter wavelength. The initial guess was in error by* $\delta = 0.25\lambda$*. Disparity error was computed as* $e(x) = 100.0\,(\delta - \tilde{d}(x))/\delta$*. Singularity neighbourhoods were removed as in Figures 4 and 6.*

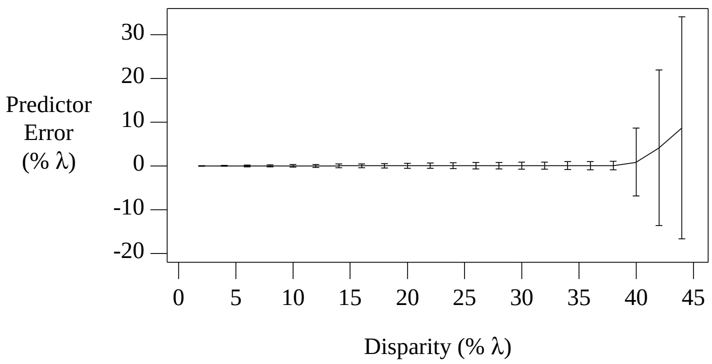

Figure 8. **Disparity Error versus True Disparity:** *Mean disparity error and standard deviation bars are shown as a function of the initial guess error* $\delta(\lambda) = \alpha\lambda$ *for* $\alpha = 0.02, ..., 0.44$*. Disparity error was computed as* $e(x) = 100.0\,(\delta(\lambda) - \tilde{d}_1(x))/\lambda$*.*

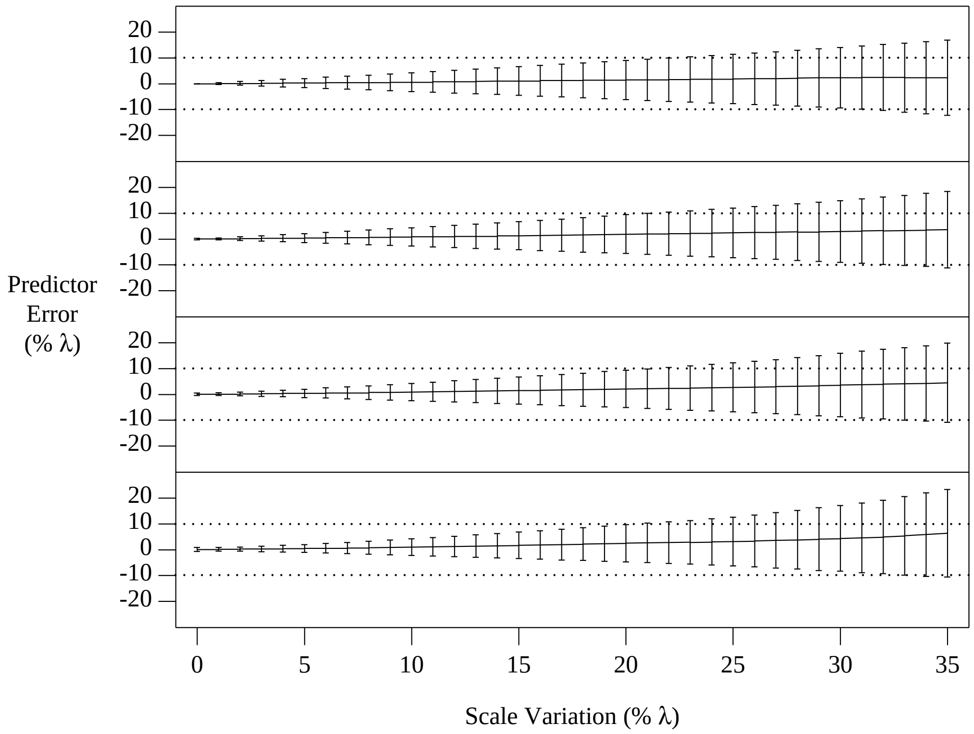

Figure 9. **Disparity Error versus Scale Variation:** *Mean disparity error and standard deviation bars are shown as a function of scale differences between left and right views. Disparity error was computed as* $e(x) = 100.0\,(\delta(\lambda) - d(x))/\lambda$. *From top to bottom, the four plots correspond to errors in the initial guess of* $\delta(\lambda) = 0.0, 0.1\lambda, 0.2\lambda$, *and* 0.3λ.

Figure 10. **Car Part Stereogram:** *Left image of Renault car part stereo pair.*

.

Figure 11. **Old Predictor Results:** *Results of $\tilde{d}_0(x)$ on the car part stereogram without removal of singularity neighbourhoods, after 1 iteration (top) and after 10 iterations (bottom). The results show the disparity estimates obtained at the finest scale of the coarse-to-fine algorithm.*

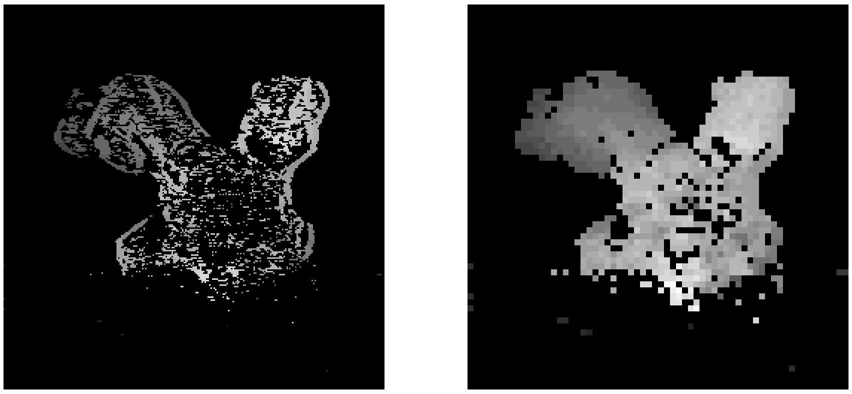

Figure 12. **New Predictor Results:** *Results of* $\tilde{d}_1(x)$ *at the finest level of the coarse-to-fine algorithm on car part stereogram with singularity neighbourhoods removed after 1 iteration (top-left). The results are also shown after simple blurring of the disparity estimates over* 4×4 *neighbourhoods (top-right). The blurred results are also shown as a perspective height plot for comparison with Figure 11.*