Scale-Space Singularities

Allan D. Jepson and David J. Fleet

Department of Computer Science, University of Toronto, Toronto, Canada M5S 1A4

Introduction

In order to compute image velocity or binocular disparity it is necessary (in some sense) to localize structure in an image sequence and track it across frames, or to match it between left and right stereo views. For example, differential-based velocity techniques measure the translation of level contours of either constant intensity [6], or constant filter response [5], while zero-crossing approaches focus on the motion of zero-crossings in the output of band-pass filters [11]. Recently, the use of contours of constant phase has been suggested for the measurement of binocular disparity [3, 7, 8] and image velocity [2]. In choosing what type of structure to track it is important to consider its stability under common image deformations such as contrast variations, dilations, shears, and rotations in addition to simple translation (cf. [10]). One main advantage of phase information is that, except near certain points referred to here as *singularities*, phase is generally stable with respect to affine image deformations.

In this paper we illustrate the stability of phase information as compared to the amplitude of filter output. In addition, we discuss the existence of phase singularities, the neighbourhoods about them where phase is unreliable, and we present a simple method for their detection. Given this detection scheme, highly accurate and robust approaches to the measurement of optic flow and binocular disparity are possible. For example, based on the spatiotemporal gradient of phase information, Fleet and Jepson [2] reported a technique for the measurement of component (normal) image velocity for which approximately 90% of the accepted estimates had less than 5% relative error in cases of significant dilation and shear.

Finally, the results presented here are of general interest for several reasons. First, they also apply to zero-crossings of the filter output in that zero-crossings can be viewed as lines of constant phase. Second, similar results apply to 2-d signals. Third, the problems caused by deviations from image translation do not exist solely for phase-based techniques (cf. [9, 10]). The fact that these issues can be addressed within a phase-based framework is a major advantage for such approaches.

Gabor Scale-Space

To demonstrate the robustness of phase and its singularities we first consider a *Gabor scale-space* expansion of a 1-d signal that expresses the filter output as a function of spatial position and the principal wavelength to which the filter is tuned. It is defined by

$$S(x, \lambda) \;=\; \mathit{Gabor}(x;\, \sigma(\lambda), k(\lambda)) \,*\, I(x)\,, \tag{1}$$

where $*$ denotes convolution, $I(x)$ is the input, $\mathit{Gabor}(x;\, \sigma, k) \equiv e^{ixk} G(x;\sigma)$ denotes the Gabor kernel [4] where $G(x;\sigma)$ is a Gaussian, and λ is the scale parameter (i.e. wavelength). The peak tuning frequency is given by $k(\lambda) = 2\pi/\lambda$, and the radius of support by $\sigma(\lambda) = (2^\beta + 1)/(2^\beta - 1)k(\lambda)$, where β denotes the octave bandwidth (usually taken to be near one). Because $\mathit{Gabor}(x;\, \sigma, k)$ is complex-valued, $S(x, \lambda)$ may be written as $\rho(x, \lambda)\, e^{i\phi(x,\lambda)}$ where the amplitude and phase components are given by

$$\rho(x, \lambda) \;=\; |S(x, \lambda)|\,, \qquad \phi(x, \lambda) \;=\; \arg(S(x, \lambda)) \;\equiv\; \mathrm{Im}\,[\log S(x, \lambda)]\,. \tag{2}$$

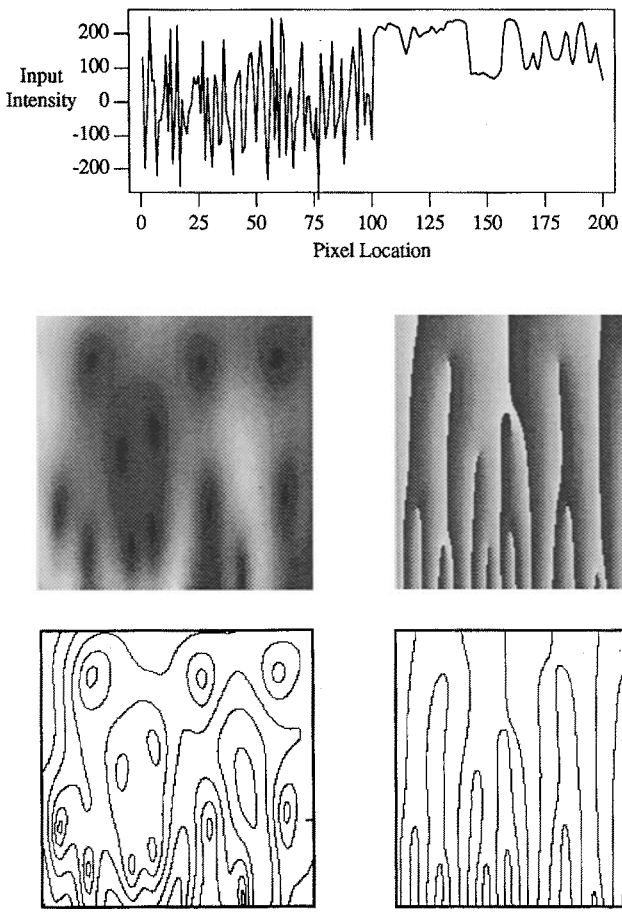

Figure 1. **Scale-Space Expansion:** *(top) 1-d input signal – a sample of white Gaussian noise concatenated with a scanline from a real image. (middle) $\rho(x, \lambda)$ and $\phi(x, \lambda)$ are shown with scale on the vertical axis spanning two octaves with $12 \leq \lambda \leq 48$ pixels. (bottom) Level contours of $\rho(x, \lambda)$ and $\phi(x, \lambda)$.*

The local frequency of $S(x, \lambda)$ can be defined as the spatial derivative of phase [12]: $\phi_x(x, \lambda)$. Although, $S(x, \lambda)$ is not expected to have constant frequency (linear phase), a first-order approximation to the spatial phase variation is generally accurate because of the band-pass filter tuning (cf. Fig. 2). This yields an amplitude-modulated, constant-frequency approximation to the local structure of $S(x, \lambda)$. Figure 1 shows a 1-d signal with the amplitude and phase components of its scale-space expansion.

As mentioned above, it is important that the signal property to be tracked is *stable* with respect to scale perturbations. This suggests that its level contours should be vertical in scale-space. To see this, consider two 1-d signals (e.g. left and right views) where one is a near-identity affine transformation of the other; i.e. let

$$I_r(a(x)) \;=\; I_l(x)\,, \qquad \text{where} \quad a(x) \;=\; a_0 + a_1 x\,, \tag{3}$$

with a_1 near 1. Because the filters have constant bandwidth the two output signals will satisfy

$$S_r(a(x), \lambda_1) \;=\; S_l(x, \lambda_2)\,, \qquad \text{where} \quad \lambda_1 \;=\; a_1 \lambda_2\,. \tag{4}$$

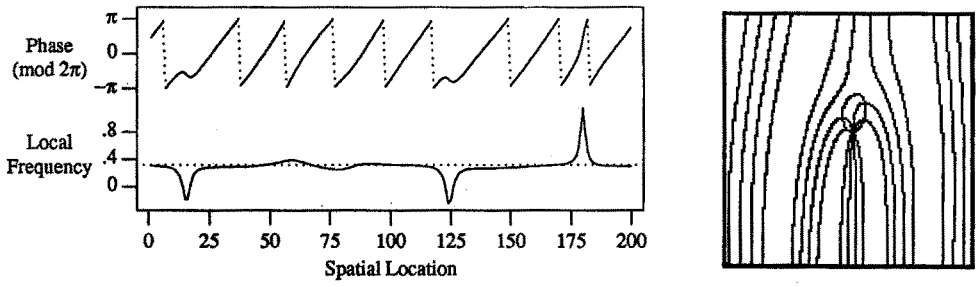

Figure 2. **Phase and Local Frequency Near Singularities:** *(left) $\phi(x, \lambda)$ and $\phi_x(x, \lambda)$ are shown for a slice of the scale-space in Figure 1 (with $\lambda = 20$). Vertical dotted lines denote phase wrapping (not discontinuities), and the horizontal dotted line marks the filter's peak tuning frequency $k(\lambda) = 0.314$. (right) Typical behaviour of level phase contours near a singularity. The singularity is the point in the centre through which several phase contours pass. The small ellipsoidal contour marks the retrograde boundary.*

That is, the two outputs would have similar structure if filters tuned to λ_1 and λ_2 had been applied to $I_r(x)$ and $I_l(x)$ respectively. However, in measuring disparity (or velocity) it is common to apply the same filters to $I_r(x)$ and $I_l(x)$ because the scale factor a_1 is unknown. In other words, we attempt to recover $a(x)$ by matching structure (features) of $S_r(a(x), \lambda_1)$ and $S_l(x, \lambda_1)$. To be successful, the structure of $S_l(x, \lambda_2)$ that is used for matching must be well represented by the structure of $S_l(x, \lambda_1)$. Equivalently, its level contours should be nearly vertical in scale-space.

It is clear from Figure 1 that amplitude structure depends significantly on scale in that its the level contours are *not* vertical. As a consequence, the filter response (which depends significantly on amplitude) is also unstable. By contrast, note that except for several isolated regions, phase is generally stable with respect to scale perturbations. As explained below, the major source of this instability is the occurrence of singularities in the phase signal $\phi(x, \lambda)$.

Singularity Neighbourhoods

For a general image $I(x)$, the scale-space defined by (1) is analytic, and contains a number of isolated zeros, where $S(x, \lambda) = 0$. In $\rho(x, \lambda)$ shown in Figure 1, zeros appear as black spots. The phase signal in (2) is also analytic, except at the zeros of $S(x, \lambda)$. The expected density of these *phase singularities* is proportional to the peak tuning frequency. Here, we describe the characteristic behaviour of $S(x, \lambda)$ in neighbourhoods about singular points. In what follows, let (x_0, λ_0) denote the location of a typical singularity.

The neighbourhoods just above and below singular points can be characterized in terms of the behaviour of phase $\phi(x, \lambda)$, and local frequency $\phi_x(x, \lambda)$. Above singular points (for $\lambda > \lambda_0$) they are characterized by local frequencies that are significantly below the corresponding peak tuning frequencies $k(\lambda)$. Within these neighbourhoods there exist *retrograde regions* where local frequencies are negative, i.e. $\phi_x(x, \lambda) < 0$. Along the boundaries of retrograde regions (which begin and terminate at singular points) the local frequency is zero; i.e. $\phi_x(x, \lambda) = 0$. The significance of this is that, where $\phi_x(x, \lambda) = 0$, the level phase contours are horizontal, and not vertical as desired. Nearby this boundary, both inside and outside the retrograde regions, the level contours are generally far from vertical, which, as discussed above, implies considerable phase instability. Below singular points (for $\lambda < \lambda_0$) the neighbourhoods are characterized by local frequencies of response that are significantly higher than the peak tuning frequencies. In addition, the local frequency changes rapidly as a function of spatial location.

To illustrate this behaviour Figure 2 (*left*) shows a 1-d slice of $\phi(x, \lambda)$ and $\phi_x(x, \lambda)$ from the scale-space in Figure 1 at a single scale ($\lambda = 20$). This slice (marked on the amplitude contour plot in Fig. 1) passes

through three singularity neighbourhoods, two just above singularities (near locations 17 and 124), and one just below a singularity (near location 180). Notice the low (sometimes negative) and high local frequencies near the singularities. Figure 2 (*right*) shows the typical behaviour of level phase contours near a singularity. The phase singularity is the point in the middle through which several of the phase contours pass. The small elliptical contour marks the retrograde boundary where $\phi_x(x, \lambda) = 0$. The instability above the singular point is clear from the nearly horizontal level phase contours. Directly below the singular point, the high local frequencies are evident from the high density of phase contours.

Finally, the neighbourhoods spatially adjacent to 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 $\rho(x, \lambda_0)$ at x_1 near x_0, the distance to the singularity $|x_0 - x_1|$ is approximately $\rho(x_1, \lambda_0)/|\rho_x(x_1, \lambda_0)|$. Therefore, as we approach the singularity $|\rho_x(x_1, \lambda_0)|/\rho(x_1, \lambda_0)$ increases.

Detection of Singularity Neighbourhoods

In order to use phase information reliably toward the measurement of image velocity or binocular disparity, singularity neighbourhoods must be detected so that measurements in them may be discarded. Here we introduce constraints on local frequency and amplitude that can be used to identity locations within singularity neighbourhoods, while avoiding the explicit localization of the singular points.

To detect the neighbourhoods above and below the singular points we constrain the distance between the local frequency of response and the peak tuning frequency. This can be expressed as a function of the extent of the amplitude spectrum (measured at one standard deviation $\sigma_k(\lambda)$) as follows:

$$\frac{|\phi_x(x, \lambda) - k(\lambda)|}{\sigma_k(\lambda)} < \tau_k , \qquad \sigma_k(\lambda) = k(\lambda)\frac{(2^\beta - 1)}{(2^\beta + 1)} . \tag{5}$$

The neighbourhoods adjacent to singular points can be detected with a local amplitude constraint:

$$\sigma(\lambda)\frac{|\rho_x(x, \lambda)|}{\rho(x, \lambda)} < \tau_\rho , \tag{6}$$

where $\sigma(\lambda)$ defines the radius of filter support. Level contours of (5) for different values of τ_k form 8-shaped regions with the singular points are their centres, while level contours of (6) form ∞-shaped regions. (see Figure 3 (*top row*)). As τ_k and τ_ρ decrease, the constraints become tighter and larger neighbourhoods are detected. Figure 3 (*top-right*) shows the combined behaviour of (5) and (6) as applied to the scale-space in Figure 1, with $\tau_k = 1.2$ (i.e. local frequencies are accepted up to 20% outside the nominal tuning range of the filters) and $\tau_\rho = 1.0$ (i.e. points within $\sigma(\lambda)$ of a singularity are discarded). These constraints typically remove about 15% the scale-space area. Finally, Figure 3 (*bottom row*) also shows the original level phase contours of Figure 1, the contours that survive the constraints, and the contours in those regions removed. Notice the stability of the contours outside the singularity neighbourhoods.

Measurement of Binocular Disparity

To illustrate the problems caused by phase instability and the rapid variation of local frequency that occur in singularity neighbourhoods, we compare the results of a technique for disparity measurement with and without their detection. Following Jenkin and Jepson [7] and Sanger [8], estimates of binocular disparity are computed as

$$d(x) = \frac{[\phi_l(x) - \phi_r(x)]_{2\pi}}{k_0} , \tag{7}$$

where $\phi_l(x)$ and $\phi_r(x)$ denote the phase responses of the left and right views, k_0 denotes the peak tuning

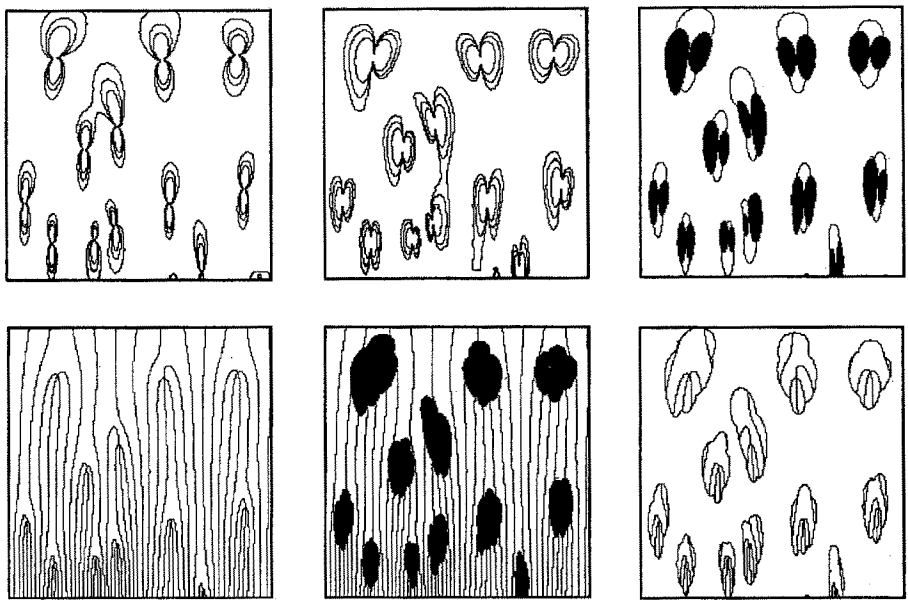

Figure 3. **Detection of Singularity Neighbourhoods**: *(top-right) Level contours of (5) for $\tau_k = 1$, 1.5 and 2 for the scale-space in Fig. 1. (top-middle) Level contours of (6) for $\tau_\rho = 0.75$, 1, and 1.5. (top-right) Neighbourhoods removed by (6) with $\tau_\rho = 1$ are shown in black, while the contours show the remaining regions marked by (5) with $\tau_k = 1.2$. (bottom) Level phase contours for scale-space (cf. Fig. 1), contours that survive the constraints, and the phase contours in those regions removed with $\tau_\rho = 1$ and $\tau_k = 1.2$.*

frequency of the filter, and $[\theta]_{2\pi} \in (-\pi, \pi]$ denotes the principal part of θ. This computation presumes a model of local phase given by $\phi(x) = k_0 x + \phi_0$; when the left and right signals are shifted versions of one another, and the filter outputs have constant frequency k_0, then (7) yields the exact result. Toward a more general model we can replace k_0 in (7) by the average local frequency in the left and right outputs $(\phi_l'(x) + \phi_r'(x))/2$. This allows frequencies other than k_0 with accurate results [3]. Making the local model explicit is important because the measurement accuracy and reliability depend on the appropriateness of the local model. For example, in neighbourhoods above and below singular points, which are characterized by a high variation in local frequency, the linear phase model is inappropriate and the numerical approximation of $\phi'(x)$ will be poor. The removal of these regions is therefore important.

To illustrate this, assume a simple situation in which the left and right views are shifted versions of the 1-d signal shown in Figure 1 (*top*). Let the disparity be 5 pixels, and let the Gabor filters be tuned to a wavelength of 20 pixels. Thus the left and right phase signals are shifted versions of the scale-space slice shown in Figure 2, which crosses three singularity neighbourhoods. Figure 4(*top*) shows the results of (7) with the crude linear model [7, 8], and without the removal of singularity neighbourhoods. Figure 4 (*middle*) shows the consequence of removing any disparity measurement for which the left or the right filter responses did not satisfy (5) or (6) with $\tau_\rho = 1.0$ and $\tau_k = 1.2$ (as in Fig. 3). In [8] a heuristic constraint on amplitude differences between left and right signals and subsequent smoothing were used to lessen the effects of such errors. Unfortunately, this smoothing will sacrifice the resolution and accuracy of nearby estimates. Finally, Figure 4 (*bottom*) shows the improvements obtained with the more general linear model.

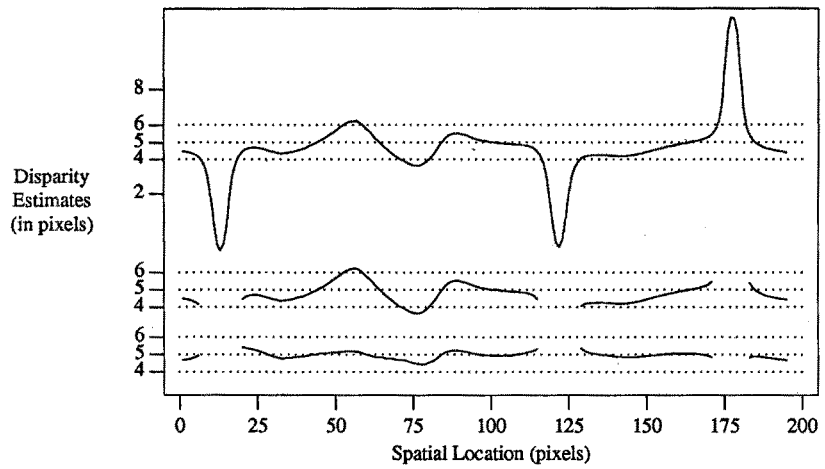

Figure 4. **Disparity Measurement:** *The top two plots show the disparity estimates based on (7) without, and then with, the removal of singularity neighbourhoods. Notice the substantial errors in the first case versus the second. The bottom plot shows the improved technique in which the local frequency is used instead of the peak frequency in (7). The same neighbourhoods have been removed.*

Summary

Phase-based techniques for the measurement of binocular disparity and image velocity are encouraging, especially because of the stability of band-pass phase information with respect to deviations from image translation that are typical in projections of 3-d scenes. Despite this stability, phase is unreliable in the neighbourhoods of phase singularities. This instability was described, and it was shown that singularity neighbourhoods may be detected using simple constraints on the local frequency and the amplitude of the filter output. Finally, these results were discussed briefly in the context of binocular disparity measurement.

References

1. Burt, P.J., et.al. (1989) Object tracking with a moving camera. *IEEE Motion Workshop*, Irvine, p 2-12
2. Fleet, D. and Jepson, A. (1989) Computation of normal velocity from local phase information. *Proc. IEEE CVPR*, San Diego, pp 379-386
3. Fleet, D. Jepson, A. and Jenkin M. (1990) Phase-based disparity measurement. *submitted*
4. Gabor, D. (1946) Theory of communication. *J. IEE* 93, pp. 429-457
5. Glazer, F. (1987) Hierarchical gradient-based motion detection. *Proc. DARPA IUW*, LA., pp 733-748
6. Horn, B.K.P. and Schunck, B.G. (1981) Determining optic flow. *Artif. Intel.* 17, pp. 185-204
7. Jenkin, M. and Jepson, A.D. (1988) The measurement of binocular disparity. *in* **Computational Proceses in Human Vision**, (ed.) Z. Pylyshyn, Ablex Press, New Jersey
8. Sanger, T. (1988) Stereo disparity computation using Gabor filters. *Biol. Cybern.* 59, pp. 405-418
9. Schunck, B.G. (1985) Image flow: fundamentals and future research. *Proc. IEEE CVPR*, San Francisco, pp 560-571
10. Verri, A. and Poggio, T. (1987) Against quantitative optic flow. *Proc. IEEE ICCV*, London, p 171-179
11. Waxman, A.M., Wu, J., and Bergholm, F. (1988) Convected activation profiles: Receptive fields for real-time measurement of short-range visual motion. *Proc. IEEE CVPR*, Ann Arbor, pp 717-723
12. Whitham, G.B. (1974) **Linear and Nonlinear Waves.** John Wiley and Sons, New York