On Transparent Motion Computation

K. Langley

Dept. of Psychology, University College London,
London, U.K. WC1E 6BT

D. J. Fleet

Dept. of Computer Science, Queen's University,
Kingston, Canada K7L 3N6

T. J. Atherton,

Dept. Computer Science, University of Warwick,
Coventry, U.K CV4 7AL.

Abstract

We extend the principle of phase-based techniques for measuring optical flow and binocular disparity to multiple motion estimation. We analyse multiple optical flows by estimating phase gradients (instantaneous frequencies) from a set of independent bandpass quadrature filter pairs. Our approach is similar to that of Shizawa and Mase [22], in which nth-order differential operators are required to compute n simultaneous velocity estimates. The approach presented here only requires a set of band-pass filters and their first derivatives.

1 Introduction

Within motion analysis, there are two distinct points of view. There are techniques that detect and track edge or corner features over time (e.g [4]), and those that compute explicit measurements from the image intensity pattern. Edge-based approaches must typically deal with the aperture problem, while corner based approaches generally provide sparse measurements. For intensity-based approaches this only becomes a problem where the image intensity function has a one-dimensional structure.

Intensity-based approaches can themselves be subdivided into three different groups. Spatio-temporal energy models[1] compute image velocity from the relative amplitudes of the outputs of different band-pass filters. This approach, however, does not perform well when all of the power of the signal lies in the passband of a single filter. There are differential techniques that measure velocity from spatiotemporal derivatives of intensity of band-pass filter outputs (e.g. [23]). However, these techniques can be sensitive to noise, geometric deformations between frames, and photometric variations as they assume conservation of image intensity, or its filtered representation [6].

The approach taken here is phase-based [13, 16]. Measurements are based on a representation of the signal structure provided by a family of quadrature-pair bandpass filters. The convolution of an image with a linear bandpass

operator is given by

$$R(\mathbf{x}, t) = I(\mathbf{x}, t) * K(\mathbf{x}, t) = \rho(\mathbf{x}, t)\exp[i\phi(\mathbf{x},t)] \tag{1}$$

where $K(\mathbf{x}, t)$ is a complex-values bandpass kernel, and $\rho(\mathbf{x},t)$ and $\phi(\mathbf{x}, t)$ are the amplitude and phase components of the bandpass response. We are assuming that the filters are effectively quadrature pairs and can be expressed as the product of a lowpass envelope and a complex exponential, for example, Gabor filters.

Interestingly, differentiation of the above equation provides two independent equations that can in principle be applied to solve the aperture problem in a local frequency domain model. It should be noted that similar envelope/phase properties are also obtained from differential operators, but the traditional representation of these filter kernels does not pay attention to the bandpass signal. We first consider the motion constraint equation[11] applied to the bandpass signal, assuming $\frac{dR(\mathbf{x},t)}{dt} = 0$:

$$\exp[i\phi(\mathbf{x},t)]\left[(\rho_x + i\phi_x\rho)v_1 + (\rho_y + i\phi_y\rho)v_2 + (\rho_t + i\phi_t\rho)\right] = 0 \tag{2}$$

where subscripts refer to the direction of partial differentiation with respect to the coordinate frame. This gives two equations:

$$\mathcal{I}m\frac{d\ln R(\mathbf{x},t)}{dt} = \phi_x v_1 + \phi_y v_2 + \phi_t = 0 \tag{3}$$

$$\mathcal{R}e\frac{d\ln R(\mathbf{x},t)}{dt} = \frac{\rho_x}{\rho}v_1 + \frac{\rho_y}{\rho}v_2 + \frac{\rho_t}{\rho} = 0 \tag{4}$$

resulting in the following velocity measurement for single motion flow:

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \phi_x & \phi_y \\ \rho_x & \rho_y \end{bmatrix}^{-1} \begin{bmatrix} \phi_t \\ \rho_t \end{bmatrix} \tag{5}$$

providing that the phase and energy derivatives are independent.

The recognition that the human visual system is capable of separately analysing several independent motions at the same point in the image domain, has prompted some authors[20, 3] to investigate the computational rationale behind multiple optical flow analysis. It is hoped that algorithms suitable for measuring multiple image velocities in a single image neighbourhood will be helpful in a wide variety of circumstances, including the superposition of signals, nonlinear transparent phenomena, and several forms of occlusion. Towards this end, several methods have been proposed based on the superposition of two or more translating signals (e.g. [20, 22, 3, 12]). This paper discusses a variation on this theme, using the previous work of Shizawa and Mase[20, 22] as a starting point. It is shown that multiple motions may be computed without the need to estimate second or higher-order derivatives. The problem of computing multiple motion is posed instead in terms of constraints on local measures of instantaneous frequency from different band-pass filter outputs. With a preliminary implementation we find that this method produces reliable estimates of two simultaneous motions of superimposed signals. It also appears to be robust with respect to multiplicative combinations of translating signals, and differences in signal power with respect to transparent surfaces.

The occurrence of more than one legitimate image velocity in a single image neighbourhood may be caused by one of several common phenomena:

- specularities or mirror-like surface reflections like those off a polished floor;

- shadows under diffuse lighting conditions that are seen to move across a stationary surface;

- occlusion such as a single occluding boundary or the fragmented occlusion caused by natural vegetation or certain fences;

- translucency, in which light reflected from one surface is passed through another to the camera, such as stained (or dirty) glass;

- and atmospheric phenomena such as smoke, rain or snow.

In just the past few years several methods that address the problem of multiple image velocities have emerged. For example, some methods compute velocity histograms in relatively local regions of the image [9, 12]. Similarly, Fleet and Jepson[7] showed that local phase information can be used to compute multiple estimates of the normal component of 2-d velocity. However, these techniques do not address the segmentation of the different local measurements to compute separate 2-d velocity estimates. Langley and Fleet[17] have argued that the independence of phase and energy velocity does provide a basis to explain transparent motion to simple signals in human vision. They also noted that the group (energy) velocity of the image signal is not constrained to pass through the origin of the frequency domain, which is one of the properties of multiplicative motion transparency. Bergen et al[3], have derived an iterative method that initially locks onto the one of the motions, allowing it to be cancelled by substrating a deformed version of one frame from another. The same operations can then be applied to the resulting sequence to detect other motions that might exist. However, only Shizawa and Mase[20, 22] have attempted to obtain explicit constraint equations for the analysis of multiple flows from image sequences. Their approach requires that second or higher order derivatives be extracted in space and time, and averaged throughout local apertures in order to estimate the parameters of motion. In order to compute n image velocities simultaneously requires the application of nth-order differential operators.

By contrast, we recast the multiple-flow motion constraint equation in terms of a constraint on instantaneous frequencies of the signals. This as an extension to phase-based methods for measuring image velocity and binocular disparity from the output of band-pass filters [13, 16]. We use the phase gradient, to give a measure of the instantaneous frequency of the filter response as a function of space and time. Instantaneous frequency may be computed from the filter outputs directly, without explicitly representing the phase signal. In addition, we note that the specific form of band-pass filters is not crucial to the approach. Moreover, there exist recent results concerning the general stability of phase information, as well as its potential instabilities, that we may exploit to use instantaneous frequency in a reliable manner [13, 6].

2 Background Theory

With respect to motion transparency, much of the groundwork has already been covered by Shizawa and Mase[20, 22] in terms of understanding the necessary

constraints that are required to derive several motions from an image intensity function. Langley and Atherton[14, 15] use a related model to detect corners in images.

The results of Shizawa and Mase[20, 22] are based on the superposition of two translating signals. In this case, not only does the motion constraint equation apply to each of the component signals, but there is a combined constraint that applies to their superposition. For example, let $f(\mathbf{x}, t)$ be the sum of two translating signals, $f_1(\mathbf{x}, t)$ and $f_2(\mathbf{x}, t)$, with velocities $\mathbf{v}_1 = (u_1, v_1)$ and $\mathbf{v}_2 = (u_2, v_2)$. Individually, the signals satisfy the motion constraint equations:

$$(\mathbf{v}_j, 1) \cdot \nabla f_j(\mathbf{x}, t) = 0 , \qquad j = 1, 2 , \tag{6}$$

where $\nabla = [\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial t}]$. Their superposition $f(\mathbf{x}, t)$ then satisfies:[1]

$$((\mathbf{v}_1, 1) \cdot \nabla) ((\mathbf{v}_2, 1) \cdot \nabla) f(\mathbf{x}, t) = 0 , \tag{7}$$

where $(\mathbf{v}, 1) \cdot \nabla = [u\frac{\partial}{\partial x}, v\frac{\partial}{\partial y}, \frac{\partial}{\partial t}]$. When (7) is expanded, the individual terms are found to be :

$$u_1 u_2 f_{xx} + v_1 v_2 f_{yy} + (u_1 v_2 + u_2 v_1) f_{xy} + (u_1 + u_2) f_{xt} +$$
$$(v_1 + v_2) f_{yt} + f_{tt} = 0. \tag{8}$$

Using differential measurements of $f(\mathbf{x}, t)$ at five points Shizawa and Mase [21] describe how to compute the individual 2-d velocities. The fitting of three velocity planes through the origin of the frequency domain is a direct extension of this formalism to include higher-order differential operators.

3   Constraints on Instantaneous Frequency

It is well-known that the translation of a 2-d pattern has all its power concentrated on a plane in the frequency domain [5]; that is, the Fourier transform of (6) satisfies:

$$\hat{f}_j(\mathbf{k}, \omega) = \hat{h}(\mathbf{k}) \delta(\mathbf{v}_j \cdot \mathbf{k} + \omega) , \tag{9}$$

where $\mathbf{k}$ and ω are spatial and temporal frequency variables, $\delta(\cdot)$ is a Dirac delta function, and $\hat{h}(\mathbf{k})$ represents the 2-d Fourier transform of the 2-d pattern that is translating. The velocity constraint in (9) is:

$$\mathbf{v}_j \cdot \mathbf{k} + \omega = 0 , \tag{10}$$

which also follows from the Fourier transform of (6).

In these terms, finding a solution to (8) for the two velocities amounts to simultaneously fitting two planes to the power of $f(\mathbf{x}, t)$ in the frequency domain. Towards this end, note that the Fourier transform of (8) is given by:

$$i u_1 u_2 \, k_1^2 \hat{f}(\mathbf{k}, \omega) + i v_1 v_2 \, k_2^2 \hat{f}(\mathbf{k}, \omega) + i(u_1 v_2 + u_2 v_1) \, k_1 k_2 \hat{f}(\mathbf{k}, \omega) + \tag{11}$$
$$i(u_1 + u_2) \, k_1 \omega \hat{f}(\mathbf{k}, \omega) + i(v_1 + v_2) \, k_2 \omega \hat{f}(\mathbf{k}, \omega) + i \omega^2 \hat{f}(\mathbf{k}, \omega) = 0 .$$

---

[1]This derivation assumes more than the conservation of f_1 and f_2, as would be required by (6) alone. In (7), because of the cascaded differentiation, it is important that the two velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ be constant as a functions of space and time.

If we factor out $i\hat{f}(\mathbf{k}, \omega)$ from (12), we are left with the constraint:

$$u_1 u_2, k_1^2 + v_1 v_2 k_2^2 + (u_1 v_2 + u_2 v_1)k_1 k_2 + (u_1 + u_2)k_1 \omega +$$
$$(v_1 + v_2)k_2 \omega + \omega^2 = 0 . \tag{12}$$

In effect, (12) constrains the locations of nonzero power in the frequency domain $(\mathbf{k}, \omega)$ to lie on one of two planes.

Following from equation (3) in the signal domain and noting that the expectation of phase derivatives (instantaneous frequency) under certain circumstances relate to Fourier frequencies[18], we are assuming that a number of independent measurements can be obtained from bandpass filters:

$$\mathcal{I}m[\frac{d\ln R}{dt}] = (\phi_x u_i + \phi_y v_i + \phi_t) = 0 \text{ for } i = 1, 2. \tag{13}$$

when combined for two flows gives:

$$(\phi_x u_1 + \phi_y v_1 + \phi_t)(\phi_x u_2 + \phi_y v_2 + \phi_t) = 0 \tag{14}$$

such that each bandpass filter selectively responds to an individual component of the multiple flow field. Independent measurements are possible by using filters tuned to different scales and orientation. The assumption that individual filters are individually selective to components of the image velocity is not without problems. For example, when transparent image sequences are defined within the passband of an individual filter kernel, reliable discrimination cannot be expected. Further, marked differences in the signal power of transparent sequences may also restrict velocity estimates to the dominant signal using the approach presented here.

Our approach to solving for the two velocities involves finding a solution to the coefficients in (12) that contain the components of $\mathbf{v}_1$ and $\mathbf{v}_2$. We then compute the individual velocities from these terms. For convenience, we rewrite (12) directly in vector form as:

$$\mathbf{a}^T \mathbf{m} = 0 , \tag{15}$$

where $\mathbf{m} = (\phi_x^2, \phi_y^2, \phi_x \phi_y, \phi_x \phi_t, \phi_y \phi_t, \phi_t^2)^T$, and $\mathbf{a} = (u_1 u_2, v_1 v_2, u_1 v_2 + u_2 v_1, u_1 + u_2, v_1 + v_2, 1)^T$. At least 5 independent measurements of instantaneous frequency (at which there is significant power) are required to solve for the five unknown elements of $\mathbf{a}$. Given six or more measurements of instantaneous frequency we have an overconstrained system, and can solve for the elements of $\mathbf{a}$ more robustly. We do this by minimizing the squared error between the model and the instantaneous frequencies; that is, we minimize

$$\sum_j (\mathbf{a}^T \mathbf{m}_j)^2 \tag{16}$$

with respect to $\mathbf{a}$. Differentiating (16) with respect to $\mathbf{a}$, and setting the result to zero produces the linear normal equations:

$$M \mathbf{a} = 0 , \quad \text{where } M \equiv \sum_j \mathbf{m}_j \mathbf{m}_j^T . \tag{17}$$

Equation (17) constrains **a** to lie in the null space of M, the matrix of outer products. Ideally, in the case of two motions, M has a rank of 5, with a 5-dimensional column space and a 1-dimensional null space. Following Barman et al[2] the goodness of fit of the model may be determined by ordering the eigenvalues of M ($\lambda_1 > \lambda_2... > \lambda_6$) and comparing the ratio of $\frac{\lambda_5 - \lambda_6}{\lambda_1}$ to unity. Note if M has a rank of 3 or 4, the flow fields are ambiguous and cannot be individually determined. The null space is spanned by the eigenvector corresponding to the zero eigenvalue. Therefore, to solve for the elements of **a**, we compute the smallest eigenvalue of M (which should be zero for two motions), and its corresponding eigenvector. We then scale the eigenvector so that its last element is unity, which produces our least squares estimate of **a**.

From the elements of **a**, as described in [21], the individual velocities are found as follows: For convenience, let the computed elements of **a** be denoted a_j, $j = 1, ..., 5$. Then, the two components of velocity are given by solutions to:

$$u_j = \frac{1}{2}a_4 \pm \sqrt{\frac{1}{4}a_4^2 - a_1}, \quad v_j = \frac{1}{2}a_5 \pm \sqrt{\frac{1}{4}a_5^2 - a_2}. \tag{18}$$

The correct combinations of these roots to one another to obtain estimates of $\mathbf{v}_1$ and $\mathbf{v}_2$ are then determined by a_3. In particular, note that their are only two ways to combine the different estimates of velocity in the x and y directions. Only one of these two will equal the third component of **a**.

4   Computing Instantaneous Frequency

In order to compute various measurements of instantaneous frequency, we assume that a family of band-pass filters, such as those used by Heeger, or Fleet and Jepson [10, 5], are applied to the image sequence. If the tuning of the filters is sufficiently different, we can assume that the frequency measurements represent independent degrees of freedom of instantaneous frequency measurements.

Instantaneous frequency is defined as the spatiotemporal phase gradient [8, 19] It gives a local approximation to the structure of the filter response in terms of an amplitude-modulated, sinusoidal signal. We measure the instantaneous frequency of the filter output $R(\mathbf{x}, t)$ using the identity:

$$\phi_x(\mathbf{x}, t) = \frac{\mathcal{I}m[R^*(\mathbf{x}, t) R_x(\mathbf{x}, t)]}{|R(\mathbf{x}, t)|^2}, \tag{19}$$

where $R^*(\mathbf{x}, t)$ is the complex conjugate of $R(\mathbf{x}, t)$.

But not all phase gradients are useful in constraining the multiple motions that may exist in the image. First, it is important that more weight be given to those frequencies that correspond to greater amounts of local energy, given by the amplitude of the filter output $|R|$. Second, it is important that the measurements of instantaneous frequency be ignored in regions where the phase of the filter output is overly sensitive to small variations in spatial position of the scale of input. These are detected using the theory of phase singularities described by Jepson and Fleet [13, 5]. They occur because of interference between energy maxima in the power spectrum. Finally, we only expect to be able to isolate transparent motion when there is some parameter (scale,

orientation or speed) that can be used to distinguish the different motions. We require at least 5 independent measurements of instantaneous frequencies from filters that respond primarily to only one of the two motions.

5 Implementation and Results

We have completed an implementation of this approach that works for 1-d and 2-d signals. One-dimensional signals allow phase gradients to be displayed as images, an aid to an intuitive grasp of the approach. That is we are using the properties of the bandpass filter to discriminate the phase velocity of transparent motion fields. The details are as follows: Figure 1 shows a 1-d space-time image of superimposed random dots moving at $\pm 18.26°$; velocity is conveniently viewed here as orientation. The energy and phase responses of the bandpass filters are then shown. In the first case (top-right and bottom-left) the filter responds mainly to the rightward moving stimulus. Then we show the phase response of a filter whose tuning bisects the two stimuli. This shows some of the distortion that occurs when applying filters that respond in part to both patterns. For the signal in Figure 1, the mean velocity (orientation) error was $-0.327°$ for leftward motion, and $-0.046°$ for rightward motion, with standard deviations $1.63°$ and $1.58°$.

The experiment was repeated (see Figure 2) with the same stimuli multiplicatively combined rather than superimposed. In this case, we find mean errors of $0.398°$ for left and $0.053°$ for rightward motion with standard deviations of $3.98°$ and $5.1°$. At present, we lack an explanation for the stability of phase information in cases of multiplicative transparency, except that the local support of the filters appears to play a major role in separating the two signals, whether combined linearly or not. We also find that this method is stable with respect to 66% differences in contrast between the two signals. The final results to 1-d sequences (Figure 2) are presented to a translating and dilating noise pattern. In particular, the phase contours from this sequence provide some indication of the problems that arise when two independent flow fields are defined within the neighborhood of support of individual bandpass filter kernels.

The final example in figure 3 shows the velocity and error field in the case of 2-d motion, with two an added random noise pattern superimposed upon the translating tree sequence. Mean errors to both flow fields were found to be $1.2°$ with a standard deviation of $7.0°$. A total of 16 independent Gabor filter kernels and their derivatives were used, defined over a 20x20x20 neighborhood of support and tuned to a frequency magnitude of 0.2 cycles per pixel.

6 Conclusion

This paper outlines a new method for computing multiple optical flows using quadrature-pair filters and their first-order derivatives. The approach is extendable to several independent velocities by increasing the number of filters and modifying our constraint equation. The basic approach is a variation on the theme discussed in detail by Shizawa and Mase. But it offers a substantially different perspective, since it requires only first-order filters, and a mechanism
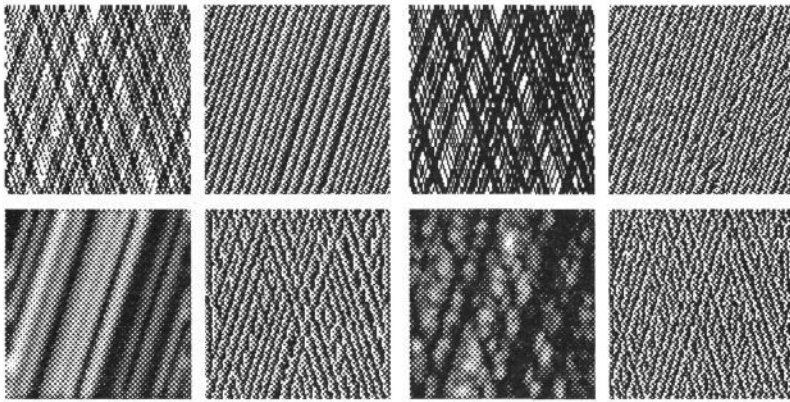
Figure 1: *Phase and energy contours for 1-d motion sequences (a)* **Left** *Additive transparency (b)* **Right** *Multiplicative transparency. For each image;* **Top left** *Image intensity sequence* **Top** right *phase and* **Bottom left** *amplitude response for a filter tuned to the motion sequence,* **Bottom right** *Phase contours for a filter equally sensitive to both components of motion sequence.*

to estimate instantaneous frequency.

There are a number of advantages in the approach that we have chosen. In particular, our processing paradigm allows higher order (deformation, dilation, rotation) properties of the optic flow field to be derived from further differentiation of the bandpass signal representation ($\nabla \frac{dR(\mathbf{x}, t)}{dt} = 0$).

However, the approach presented here still retains a number of difficulties. The foremost problem is the ability to determine precisely when reliable measurements can be obtained from bandpass filters under transparent motion. Our preliminary results suggest that the energy derivative may play an important role in supporting phase measurements from similar bandpass filters when both provide similar velocity measurements.

References

[1] Adelson.E.H and Bergen.J.R. Spatiotemporal energy models for the perception of motion. *J.Opt. Soc. Am. A*, 2,2:284–299, 1985.

[2] Barman.H, Haglund.L, Knutsson.H, and Granlund.G.H. Estimation of Velocity, Acceleration and Disparity in Time sequences. *IEEE motion workshop*, 44–51, 1991.

[3] Bergen.J.R, Burt.P, Hingorani.R, and Peleg.S. Computing two motions from three frames. *Proc. 3rd ICCV, Osaka,Japan*, 27–32, 1990.

[4] Deriche.R and Faugeras.O. Tracking line elements. In Faugeras.O, editor, *Computer Vision-ECCV 90*, pages 341–345, Antibes,France, 1989. Springer-Verlag.

[5] Fleet.D.J. PhD thesis, University of Toronto, 1990.

[6] Fleet.D.J and Jepson.A. Stability of Phase information. *IEEE workshop on Visual Motion, Princeton.*, 1991.

[7] Fleet.D.J and Jepson.A.D. The computation of normal velocity from local phase information. *IJCV*, 5:77–104, 1990.

[8] Franks.L. *Signal Analysis*. Prentice-Hall,Inc,N.J., 1969.

[9] Girod.B and Kuo.D. Direct Estimation of Displacement Histograms. *Proc. Image Understanding and Machine Vision.*, pages 73–76, 1989.

[10] Heeger.D.J. A model for the extraction of image flow. *J.Opt.Soc.Am*, 4:1455–1471, 1987.

[11] Horn.B.K.P and Schunk.B.G. Determining optic flow. *Artificial Intelligence*, 17:185–204, 1981.

[12] Jasinschi.R and Rosenfield.A. Sumi.K. The perception of visual motion coherence and transparency: a statististical model. Technical report, Univ. Maryland, 1990.

[13] Jepson.A.D. and Fleet.D.J. *Scale-Space Singularities*. (Ed.) Faugeras.O, Proc. ECCV,Antibes , Springer-Verlag, 1990.

[14] Langley.K. Ph.D thesis. *University of Warwick*, 1990.

[15] Langley.K and Atherton.T.J. Inferring the structure of images using multi-local filters. *BMVC91 University of Glasgow*, pages 111–118, 1991.

[16] Langley.K., Atherton.T.J., Wilson.R.G., and M.H.E.Larcombe. *Vertical and Horizontal Disparities from Phase.* (Ed.) Faugeras.O, Proc. ECCV,Antibes , Springer-Verlag, 1990.

[17] Langley.K. and D.J. Fleet. Using group and phase velocity to explain coherent and transparent motion. *AVA conference, manchester*, 1–2, 1992.

[18] Mandel.L. Interpretation of Instantaneous Frequency. *AJP*, 42:840–845, 1974.

[19] Papoulis.A. *Systems and Transforms with Applications in Optics.* McGraw-Hill, New York, 1968.

[20] Shizawa.M and Mase.K. Determining multiple optic flow using spatio-temporal filters. *ICPR90, atlantic City*, pages 274–278, 1990.

[21] Shizawa.M and Mase.K. Principle of Superposition: A common Computational Framework for Analysis of Multiple Motion. *IEEE workshop on visual motion,Princeton,N.J.*, 1991.

[22] Shizawa.M and Mase.K. A unified computational theory for motion transparency and motion boundaries based on eigenenergy analysis. *CVPR91, Hawaii*, pages 289–294, 1991.

[23] Uras.S, Girosi.F, Verri.A, and Torre.V. A Computational Approach to Motion Perception. *Biological Cybernetics*, 60:79–87, 1988.
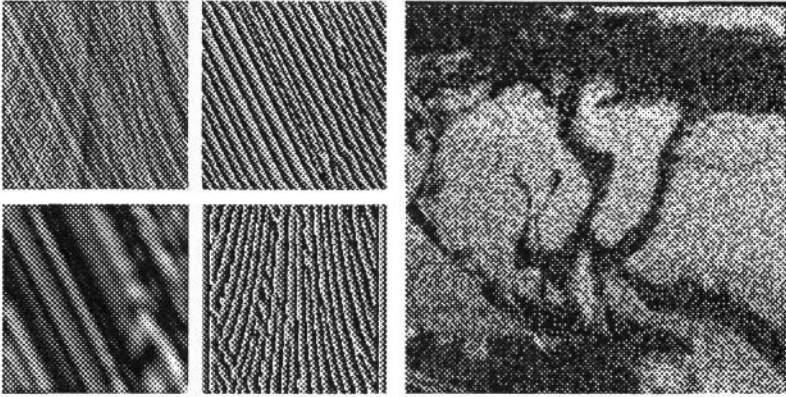
Figure 2: **Left** *A 1-d motion sequence consisting of a dilating and translating noise pattern additively combined. Phase contours are shown for a filter tuned to the translating component and a filter tuned to zero velocity.* **Right** *Frame 15 from a noise pattern moving vertically at one pixel per frame added to a translating tree sequence moving with a mean velocity of two pixels per frame to the right.*
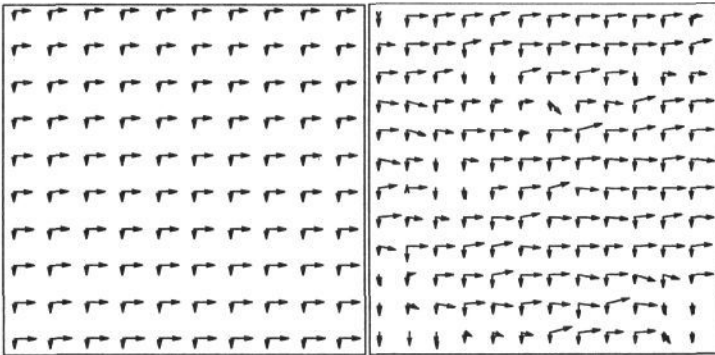

Figure 3: **Left** *True flow field from figure 2.* **Right** *Measured flow field.*