A reprint from

OPTICAL ENGINEERING

September 1996

ISSN 0091-3286

PHASE-BASED MEASUREMENT OF OBJECT VELOCITY IN IMAGE SEQUENCES USING THE HOUGH TRANSFORM

David Vernon

Phase-based measurement of object velocity in image sequences using the Hough transform

David Vernon, MEMBER SPIE
Maynooth College
Department of Computer Science
Maynooth, County Kildare, Ireland
E-mail: david.vernon@may.ie

Abstract. An alternative to traditional spatiotemporal gradient and feature-based approaches to measuring object velocity in images is introduced. Here, the velocity is computed by treating each object as a distinct intensity wave profile, with Fourier components, and by identifying the Fourier components that exhibit the magnitude and phase changes that are consistent with anticipated velocity wave motion. This detection is accomplished using an appropriate Hough transform. The two major advantages of this technique are that because the analysis takes place in the Fourier domain, the spatial organization and the visual appearance of the moving object are not significant and, second, the formulation presented in this paper lends itself to direct extension for more complex motion. Consequently, objects that are visually or spatially complex and that would be difficult to analyze using either of the traditional spatiotemporal differentiation or feature-based approaches can be effectively treated. The proposed approach is demonstrated for scenes of varying complexity.

Subject terms: velocity measurements; object motion; Fourier transforms; frequency phases; wave models; Hough transforms.

Paper 31095 received Sep. 25, 1995; revised manuscript received Feb. 17, 1996; accepted for publication Feb. 17, 1996.

1 Introduction

Traditional approaches to measuring object velocity in images normally exploit one of two primary techniques. The first involves the computation of the spatiotemporal gradient, differentiating the (filtered or unfiltered) image sequence with respect to time and subsequently computing the optical flow field.[1] The second involves the segmentation of the object or feature in question using either region-based gradient (first or second order) filtering and analysis followed either by the computation of the optical flow field or by identification object correspondence, typically by matching contour or region primitives.[2] Comparisons of the many variations of these approaches and the relationship among them can be found in Refs. 3, 4, and 5.

A less-used approach exploits the regularity in spatiotemporal frequency representations of the image, such as the spatiotemporal Fourier transform domain, resulting from certain types of image motion. Briefly, it can be shown that the spatiotemporal Fourier transform of an image sequence in which the image content is moving with constant velocity results in a spatiotemporal frequency representation that is equal to the spatial Fourier transform of the first image multiplied by a δ-Dirac function in the temporal frequency domain. This δ-Dirac function is dependent on the image velocity, which can be computed if one knows the position of the δ-Dirac function and any spatial frequency.[6] Because this approach is based on image motion, rather than object motion, it normally assumes uniform (zero) background when evaluating object motion. Extensions of the technique have been developed to allow it to adjust for situations involving noisy backgrounds,[7] several objects,[8,9] and nonuniform cluttered backgrounds.[10]

In this paper, we present an alternative formulation of the above approach. This alternative uses the normal spatial Fourier transform together with a Hough transform, rather than the spatiotemporal Fourier transform. This approach lends itself to straightforward generalization to types of motion other than the uniform translation in a plane parallel to the image plane, as is normally required. Specifically, the velocity of objects is measured by treating each object (either moving or stationary) as a distinct intensity wave profile, each of which is an additive component of the total image intensity profile, and hence each of which is a solution of the wave equation. The Fourier components of wave profiles—and equally of objects—that move with constant velocity exhibit a regular phase change. The velocity of a moving object is measured by identifying the Fourier components of the total image intensity wave profile that exhibit this phase relationship using an appropriate Hough transform. This Hough transform embodies the relationship between velocity and phase change, and velocity is measured by locating local maxima in the Hough space.

The two major advantages of this technique are that, because the analysis takes place in the Fourier domain, the spatial organization and the visual appearance of the moving object are not significant and, second, the formulation presented in this paper lends itself to direct extension for more complex motion. Consequently, objects that are visually or spatially complex and that would be difficult to analyze using either of the traditional spatiotemporal differentiation or feature-based approaches can be effectively treated. In the following sections, the proposed approach is demonstrated for scenes of varying complexity.

2 Overview

Consider an image $g(x,y,t)$: a 2-D spatiotemporal representation of the reflectance function of a scene. This image is normally regarded and viewed as a time-varying two-dimensional representation of intensity values. However, the image $g(x,y,t)$ can also be regarded as a time-varying surface. Consider an object O_i to be moving in the image. If we view $g(x,y,t)$ as a time-varying surface, the height of each point on the surface defining the reflectance value at that point, then this object may be viewed as a wave, with a characteristic shape, propagating through the image space with a velocity $v_i(t)$. The velocity function v can, in general, be a function of image coordinates and time: $v(x,y,t)$; that is, it can vary with position and time. In this paper, however, we restrict our attention primarily to the situation where the velocity is constant and parallel to the image plane. This restriction means that the shape of the wave profile does not vary with position and that it propagates with constant velocity: $v_i(t)=v$, a constant. The task then becomes one of isolating the wave and computing this velocity.

Let us use the general form of the 2-D differential wave equation to model the object O_i or, equivalently, its wave form in image-time space. Thus:

$$\frac{\partial^2 \psi^i(x,y,t)}{\partial x^2}+\frac{\partial^2 \psi^i(x,y,t)}{\partial y^2}=\frac{1}{v_i^2}\frac{\partial^2 \psi^i(x,y,t)}{\partial t^2}.$$

The solution of this wave equation $\psi^i(x,y,t)$ is, in effect, a description of the object as a gray-level wave profile propagating with constant velocity v_i, that is,

$$\psi^i(x,y,t)=f^i(x-v_{x_i}t,y-v_{y_i}t),$$

where $f^i(x,y)$ is a solution to the wave equation at time $t=0$ and this solution describes the shape of the wave at time $t=0$. Our task is to solve this equation for all distinct v_i using only our knowledge of the total optical field $\psi=g(x,y,t)$. That is, we seek a decomposition of the total optical field $\psi=g(x,y,t)$ into components $c_jf_j(x-v_{x_j}t,y-v_{y_j}t)$, each of which is a solution of the wave equation:

$$\psi(x,y,t)=g(x,y,t)=\sum_{j=1}^{n} c_jf_j(x-v_{x_j}t,y-v_{y_j}t),$$

and to group them into m sets of the form

$$\psi^i(x,y,t)=\sum_{k=1}^{l_i} c_k^if_k^i(x-v_{x_k}^it,y-v_{y_k}^it)$$

such that v_{x_k}=constant and v_{y_k}=constant: that is, the components of ψ^i all have a constant propagation velocity. We note in passing that $\psi=g(x,y,t)=\Sigma_{i=1}^m\psi^i(x,y,t)$, where m is the number of objects comprising the total optical field. We use the discrete Fourier transform to accomplish the decomposition and the Hough transform to accomplish the grouping.

3 2-D Fourier Transform

The discrete Fourier transform $\mathscr{F}[f(x,y)]$ of a 2-D function $f(x,y)$ is given by:

$$\mathscr{F}[f(x,y)]=\mathsf{F}(k_x,k_y)=\sum_x\sum_y f(x,y)\exp[i(k_xx+k_yy)],$$

and the inverse discrete Fourier transform is:

$$\begin{aligned}f(x,y)&=\mathscr{F}^{-1}[\mathsf{F}(k_x,k_y)]\\&=\frac{1}{(2\pi)^2}\sum_{k_x}\sum_{k_y}|\mathsf{F}(k_x,k_y)|\exp[i\phi(k_x,k_y)]\\&\quad\times\exp[-i(k_xx+k_yy)],\end{aligned}$$

where $|\mathsf{F}(k_x,k_y)|$ is the real-valued amplitude spectrum and $\phi(k_x,k_y)$ is the real-valued phase spectrum.

In effect, $f(x,y)$ can be constructed from a linear combination of elementary functions having the form $\exp[-i(k_xx+k_yy)]$, each appropriately weighted in amplitude and phase by a complex factor $\mathsf{F}(k_x,k_y)$. For a function or waveform translating with constant velocity (v_x,v_y), $f(x,y)$ becomes $f(x-v_x\delta t,y-v_y\delta t)$. By the shift property[11] its Fourier transform is given by:

$$\begin{aligned}\mathscr{F}[f(x-v_x\delta t,y-v_y\delta t)]&=|\mathsf{F}(k_x,k_y)|\exp[i\phi(k_x,k_y)]\\&\quad\times\exp[-i(k_xv_x\delta t+k_yv_y\delta t)].\end{aligned}$$

Thus, a spatial shift of $(v_x\delta t,v_y\delta t)$ of a waveform in the spatial domain, i.e. $f(x,y)$ shifted to $f(v_x\delta t,v_y\delta t)$, only produces a change in the phase of the Fourier components in the frequency domain. This phase change is $\exp[-i(k_xv_x\delta t+k_yv_y\delta t)]$. Thus, in order to segment the image into its component wave forms, each of which corresponds to an object moving with constant velocity in the image, we simply need to identify the set of frequency components k_x and k_y which have all been modified by the same phase shift, i.e. $\exp[-i(k_xv_x\delta t+k_yv_y\delta t)]$. To accomplish this, we note that the phase spectrum for the shifted wave at time $t+\delta t$ is equal to the phase spectrum of the wave at time t multiplied by the phase change given before:

$$\begin{aligned}&\exp[i\phi_{t+\delta t}(k_x,k_y)]\\&\quad=\exp[-i(k_xv_x\delta t+k_yv_y\delta t)]\exp[i\phi_t(k_x,k_y)]\\&\quad=\exp\{i[\phi_t(k_x,k_y)-(k_xv_x\delta t+k_yv_y\delta t)]\}.\end{aligned}$$

Hence,

$$\phi_{t+\delta t}(k_x,k_y)=\phi_t(k_x,k_y)-(k_xv_x\delta t+k_yv_y\delta t).$$

That is, the phase at time $t+\delta t$ is equal to the initial phase at time t minus $(k_xv_x\delta t+k_yv_y\delta t)$. Since we require v_x and v_y, we rearrange as follows:

$$v_y=\frac{1}{k_y\delta t}[\phi_t(k_x,k_y)-\phi_{t+\delta t}(k_x,k_y)-k_xv_x\delta t]. \tag{1}$$

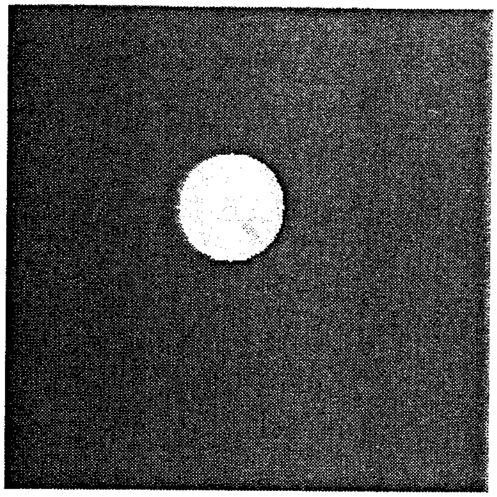

Fig. 1 Image sequence 1; image 1 (synthetic data).

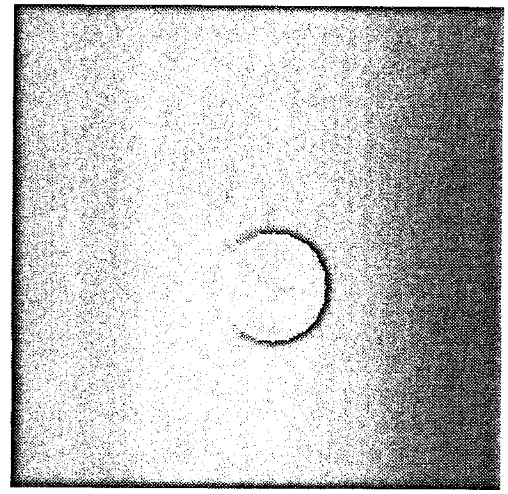

Fig. 2 Image sequence 1; image 8 (synthetic data).

This equation becomes degenerate if $k_y=0$ in which case we use an alternative rearrangement as follows:

$$v_x=\frac{[\phi_t(k_x,k_y)-\phi_{t+\delta t}(k_x,k_y)]}{k_x\delta t}. \tag{2}$$

If we have several images taken at time $t=t_0,t_1,t_2,t_3,\ldots,$ we can compute ϕ_{t_0} in particular and $\phi_{t_0+n\delta t}$, in general. Treating the equation above as a Hough transform, with a 2-D Hough transform space defined on v_x, v_y, then we can compute v_y for all possible values of v_x, and for all (known) values of n, k_x, k_y, $\phi_{t_0+n\delta t}(k_x,k_y)$. Local maxima in this v_x,v_y Hough transform space signify Fourier components that comprise wave forms—objects—in the spatial domain that are moving with constant velocity v_x,v_y. Note that in both Eqs. (1) and (2), $\phi(k_x,k_y)$ represents the absolute phase of frequency (k_x,k_y). However, in the Fourier domain, phase is bounded by $\pm 2\pi$ and phase values will "wrap" as they cross this threshold. In effect, phase values are represented modulo 2π. In this implementation, we have allowed for this by solving Eqs. (1) and (2) for the given phase values $\phi(k_x,k_y)+2n\pi,\phi>0$; $\phi(k_x,k_y)-2n\pi,\phi<0$, for all n such that $2n\pi<|k_xv_{x_{\max}}\delta t|+|k_yv_{y_{\max}}\delta t|$.

We note in passing that the computational complexity of the algorithm is $\mathcal{O}(nm^2)$ where m is the dimension of the Fourier domain and n is the dimension of the Hough transform space (and is proportional to the measurable accuracy) since Eq. (1) must be computed for all x and y spatial frequencies and for all possible values of v_x.

4 Results

Figures 1 through 12 demonstrate the results of applying the technique to three scenarios of increasing complexity. Image sequence 1 (Figs. 1 to 3) illustrates the technique when applied to synthetic images: Figs. 1 and 2 show the first and last images in an eight-image sequence; Fig. 3 shows the resultant Hough transform, which displays a well-localized maximum. Figures 4 and 5 depict a translat-

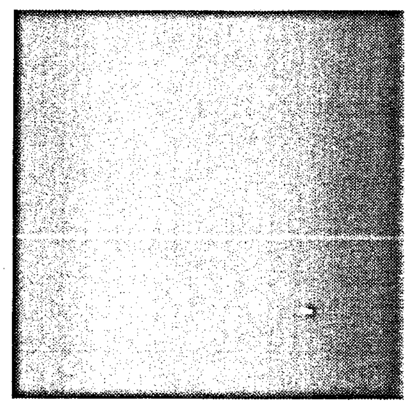

Fig. 3 Image sequence 1; Hough transform (v_x,v_y) space derived from images 1 to 8, inclusive.

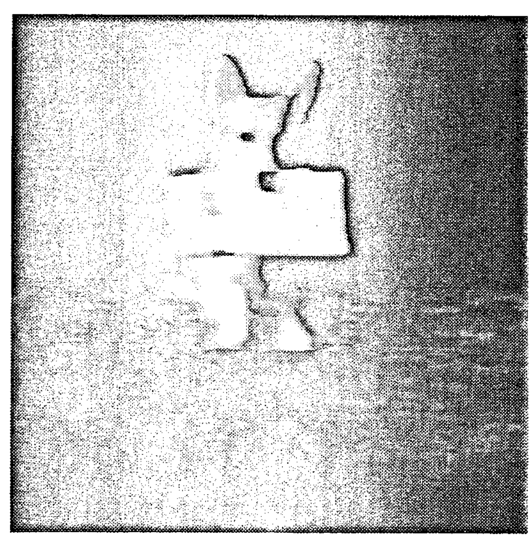

Fig. 4 Image sequence 2; image 1 (constant background).

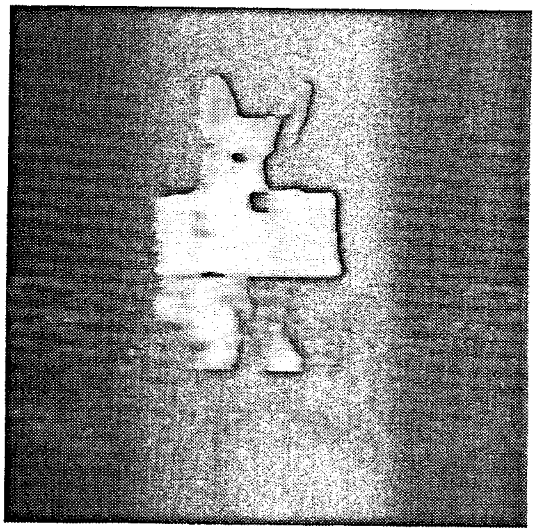

Fig. 5 Image sequence 2; image 8 (constant background).

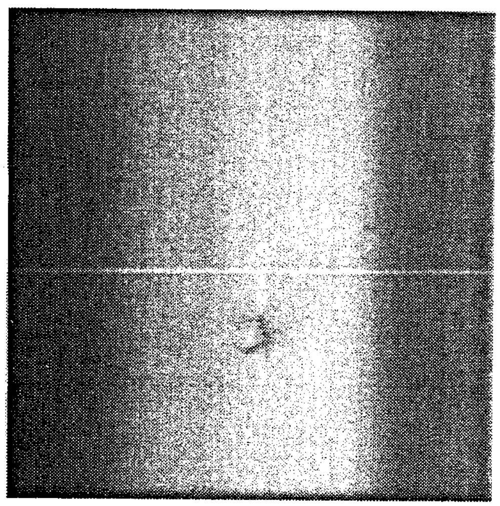

Fig. 6 Image sequence 2; Hough transform (v_x, v_y) space derived from images 1 to 8, inclusive.

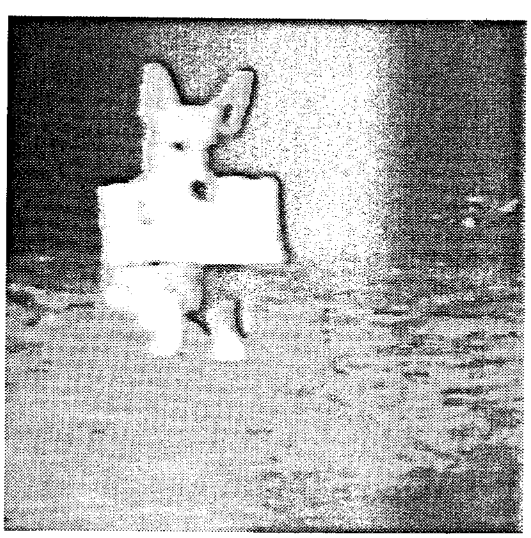

Fig. 7 Image sequence 3; image 1 (slight variation in background).

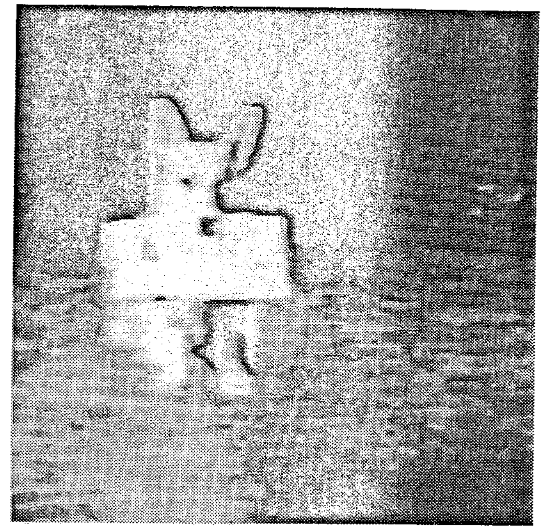

Fig. 8 Image sequence 3; image 8 (slight variation in background).

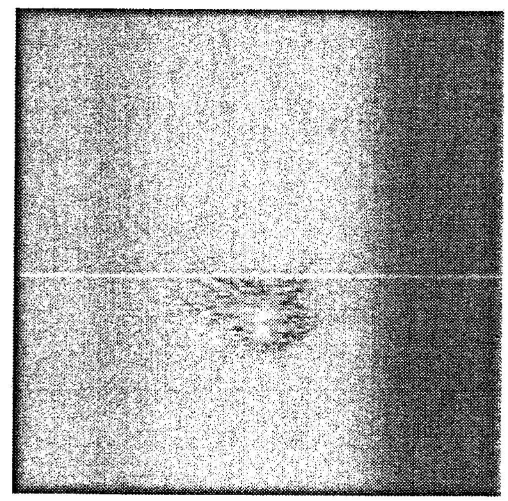

Fig. 9 Image sequence 3; Hough transform (v_x, v_y) space derived from images 1 to 8, inclusive.

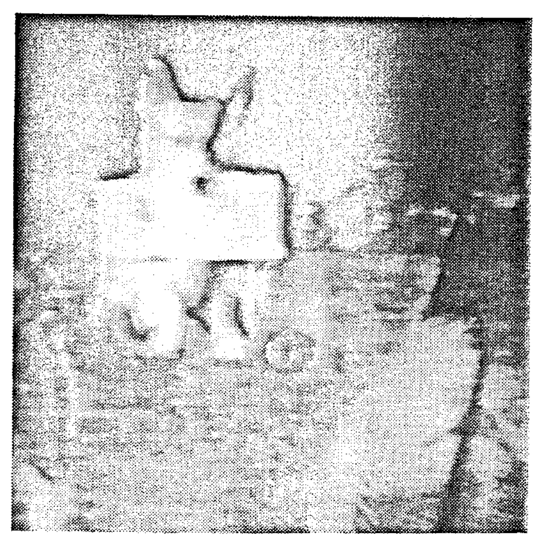

Fig. 10 Image sequence 4; image 1 (reflected image superimposed on scene).

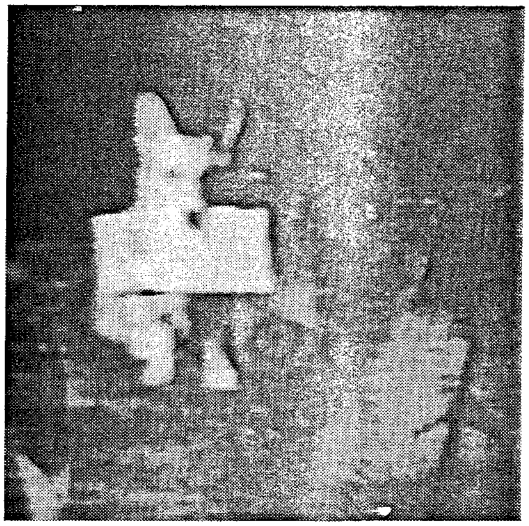

Fig. 11 Image sequence 4; image 8 (reflected image superimposed on scene).

ing object superimposed on a constant background. As with Fig. 3, the maximum in the Hough transform space is well localized (see Fig. 6). Figures 7 and 8 depict an object translating over a background that varies slightly in each image with the Hough transform shown in Fig. 9. In this instance, the maximum in the Hough transform space is less well localized (see Fig. 9). Finally, the sequence in Figs. 10 and 11 demonstrates the results in a more complex scenario where there is a foreground image of a cat superimposed on the (moving) background scene depicted in Figs. 7 and 8. Such a situation arises when, for example, an observer views a scene through a window and sees both the external scene and its own reflection. Again, one can see a degradation in the localization of the maximum in the Hough transform space (Fig. 12). Table 1 summarizes the actual and computed velocities in all cases. Finally, in order to provide an indication of the accuracy of the technique, the graph in Fig. 13 shows the percentage error in the computed velocity of the object in Fig. 4 when translated at 1, 2,...9 pixels per frame over a constant background.

Table 1 Summary of measured velocities.

Image Sequence	Actual Velocity (pixels/frame)		Computed Velocity (pixels/frame)	
	v_x	v_y	v_x	v_y
1	5.72	5.72	5.7	5.6
2	0	3.0	0	2.9
3	0	3.07	0.56	2.8
4	0	3.07	0.56	3.0

A number of things should be noted about these results. Most important, the implementation of the FFT that is being used produces a frequency domain representation that is scaled to fit the size of the destination image. Thus, the interpixel distance in the Fourier magnitude image is not a constant spatial frequency interval, but varies with image content. As a consequence, interpixel distances in the Hough transform space also vary with image content and do not correspond to absolute velocity values. The result of this is that velocity values cannot be read directly from the Hough transform space. To overcome this, we have calibrated the Hough Transform velocity space for each image sequence by manually identifying a fiducial mark in the moving object and identifying its position in the first and last images. This yields the actual interframe velocity values in units of pixels/frame. Hough transform velocity interpixel values are then computed by identifying the maximum in the Hough transform generated using a similar image sequence. While this situation is not ideal, in most practical situations a calibration phase is not uncommon. In any event, an implementation of an FFT that will yield absolute spectral values is planned.

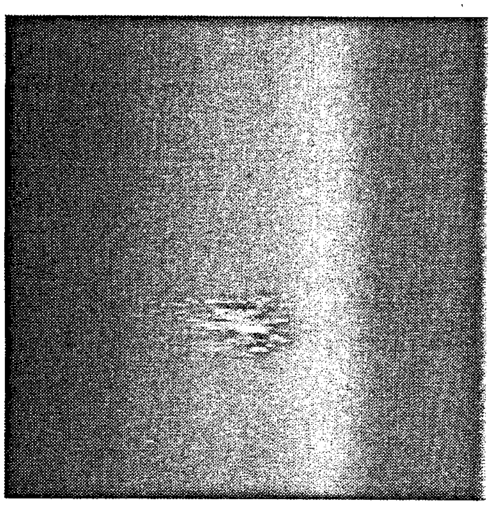

Fig. 12 Image sequence 4; Hough transform (v_x, v_y) space derived from images 1 to 8, inclusive.

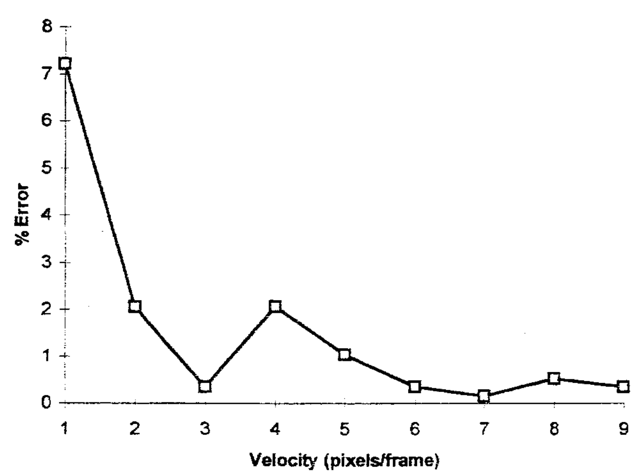

Fig. 13 Error in velocity measurement for constant background.

5 Discussion

5.1 *Implications of Using the Wave Model*

The velocity of objects moving parallel to the image plane is measured by treating each object as a distinct intensity wave profile. It is assumed that each wave profile is an additive component of the total image intensity profile and hence that each is a solution to the wave equation.

Strictly speaking, this is not a valid assumption since the intensity profile of an object does not add to the intensity profile of the background visual environment; rather, it occludes it. As a result, the occluded part of the background changes as the object moves and this distorts the phase components of the background profile. As it happens, this problem is in fact useful in circumstances where one is attempting to estimate the motion of transparent or translucent objects where the object and background intensity profile are, to an approximation, additive (such as the situation shown in Figs. 10 and 11).

5.2 *Implications of Using the Fourier Transform*

The Fourier transform represents a signal comprising two or more distinct components of identical frequency but different phase by a single component in the frequency domain, with the magnitude determined by the relative phase difference, and the phase equal to the sum of the phases of the two components. This causes problems for the phase-based velocity estimation procedure in that the model assumes independent distinct (sinusoidal) components, possibly of the same frequency but with different and varying phases, for each object or wave profile. The Fourier domain does not allow such a model and conflates the two components of identical frequency into one frequency component with additive phase and magnitude which is a function of the relative phase of the components. This means that we cannot be sure that all the components of a profile will be present in the resultant image (due to destructive interference) or that the phase change of a component will be exclusively due to the translation of that profile: it could be that the phase changes are also a result of the shift in the components of another translating profile or object. Note that this phase problem does not arise when there is only one translating object since all phase changes are a result of its translation even though the absolute phase value may well contain an additive component due to the stationary profile or the image background.

The consequence of this component conflation is that when two translating profiles share a common spectrum (i.e., both comprise common spatial frequencies), the changes in their additive phase will be a function of their joint velocities, rather than the velocities of the individual profiles. This problem can be surmounted, albeit at significant additional computational cost, by extending the Hough transform to deal explicitly with two distinct velocities $[(v_{x_1},v_{y_1})$ and $(v_{x_2},v_{y_2})]$ rather than one (v_x,v_y). This results in a 4-D accumulator space rather than a 2-D one; identification in this space of local maxima of all possible composite velocities allows the isolation of the individual object velocities as well as their composite velocities.

5.3 *Implications of Using the Hough Transform*

As noted in the introduction, spatiotemporal frequency and phase-based approaches to the estimation of object velocity, while not as popular as other approaches, have been successfully used for situations where the objects are translating on a plane parallel to the image plane. The advantage of using the Hough transform to group the Fourier components rather than a temporal Fourier transform is that the grouping criterion can be arbitrarily complex (although with a consequent increase in computational cost). In this paper we have restricted ourselves to the normal translational motion, but the technique can be extended in a very straightforward manner to deal with more complex circumstances as follows.

Object scaling. If the object translates either toward or away from the camera, then the perspective lens distortion will result in a scaled image of the object. Such a scaling results in an inverse scaling of the spatial frequencies and this is easily incorporated into the equation defining the Hough transform. This results in a 3-D Hough transform defined in terms of v_x, v_y, s: the x and y velocity components and the scaling factor, respectively.

Curvilinear motion. Objects that do not describe translation in a straight line, i.e. $v_x=v_x(x,y)$ and $v_y=v_y(x,y)$, can also be estimated if their velocity profiles are continuous. In this case, the object will generate a velocity curve (or crest) in the Hough transform space rather than a single peak.

Nonuniform velocity. It is often the case that objects do not move with constant velocity and the velocity is time dependent, i.e. $v_x=v_x(t)$ and $v_y=v_y(t)$. In this case, the Hough transform can be extended by two additional dimensions $v_x(t)$ and $v_y(t)$ to cater for expected velocity profiles. For example, let $v_x=at$ and $v_y=bt$, in which case we have a 4-D transform space defined on v_x, v_y, a, and b.

Rotation about an axis. The situation where an object rotates about a vertical or horizontal axis parallel to the image plane as it translates can be adjusted for by allowing independent scaling of the object in the horizontal and vertical directions, respectively.

Segmentation. Since the Hough transform effects a grouping process, explicitly identifying the component of the moving object or wave profile, this object can be emphasized—segmented—by noting the spatial frequencies that correspond to these local maxima and by blocking all other frequencies prior to the application of an inverse Fourier transform.

6 Conclusions

A flexible and extendible technique for estimating the velocity of objects moving in image sequences has been presented and its efficacy has been demonstrated. It now remains to validate and evaluate the proposed extensions.

References

1. J. H. Chou and T.-C. Chou, "On the detection and the computation of optical flow," *IEEE Trans. Pattern Anal. Machine Intelligence* **14**(3), 346–352 (1992).

2. H. Shariat and K. E. Price, "Motion estimation with more than two frames," *IEEE Trans. Pattern Anal. Machine Intelligence* **12**(5), 417–434 (1990).
3. M. Otte and H.-H. Nagel, "Optical flow estimation: advances and comparisons," *Lecture Notes in Computer Science*, J. O. Eklundh, Ed., *Computer Vision-ECCV '94*, pp. 51–60, Springer-Verlag, Berlin (1994).
4. M. Tistarelli, "Multiple constraints for optical flow," *Lecture Notes in Computer Science*, J. O. Eklundh, Ed., *Computer Vision-ECCV '94*, pp. 61–70, Springer-Verlag, Berlin (1994).
5. L. Jacobson and H. Wechsler, "Derivation of optical flow using a spatiotemporal-frequency approach," *Comp. Vision, Graphics, Image Proc.* **38**, 29–65 (1987).
6. M. P. Cagigal, L. Vega, and P. Prieto, "Object movement characterization from low-light-level images," *Opt. Eng.* **33**(8), 2810–2812 (1994).
7. M. P. Cagigal, L. Vega, and P. Prieto, "Movement characterization with the spatiotemporal Fourier transform of low-light-level images," *Appl. Opt.* **34**(11), 1769–1774 (1995).
8. S. A. Mahmoud, M. S. Afifi, and R. J. Green, "Recognition and velocity computation of large moving objects in images," *IEEE Trans. Acoustics, Speech, Signal Proc.* **36**(11), 1790–1791 (1988).
9. S. A. Mahmoud, "A new technique for velocity estimation of large moving objects," *IEEE Trans. Signal Proc.* **39**(3), 741–743 (1991).
10. S. A. Rajala, A. N. Riddle, and W. E. Snyder, "Application of one-dimensional Fourier transform for tracking moving objects in noisy environments," *Comp. Vision, Graphics, Image Proc.* **21**, 280–293 (1983).
11. A. Rosenfeld and A. Kak, *Digital Picture Processing*, Vol. 1, p. 12 Academic Press, New York (1982).

David Vernon began his professional career in 1979 as a software engineer with Westinghouse Electric, Inc, having graduated in engineering from Trinity College, Dublin. In 1983, he was appointed a lecturer in the Department of Computer Science in Trinity College and completed his PhD in robot vision in 1985. From 1991 to 1993, he worked as a scientific officer in the European Commission and from 1994 to 1995 he was responsible for the development of a policy on information systems for Trinity College. He was appointed to the Chair of Computer Science at Maynooth College, Ireland, in 1995. His research activities are in computer vision and autonomous systems theory. He is the author and editor of four books on computer vision.