Segmentation in Dynamic Image Sequences by Isolation of Coherent Wave Profiles

David Vernon

Department of Computer Science
Maynooth College
Ireland

Abstract. A segmentation and velocity estimation technique is presented which treats each object (either moving or stationary) as a distinct intensity wave profile. The Fourier components of wave profiles — and equally of objects — which move with constant velocity exhibit a regular frequency-dependent phase change. Using a Hough transform which embodies the relationship between velocity and phase change, moving objects are isolated by identifying the subset of the Fourier components of the total image intensity wave profile which exhibit this phase relationship. Velocity is measured by locating local maxima in the Hough space and segmentation is effected by re-constituting the moving wave profile — the object — from the Fourier components which satisfy the velocity/phase-change relationship for the detected velocity.

1 Introduction

Traditional approaches to segmentation typically exploit one of two broad approaches. These are (a) boundary detection, which depends on the detection of spatial intensity discontinuities (using first or second order gradient techniques) and their aggregation into contour-based object descriptions, and (b) region-growing, which depends on the identification of local regions which satisfy some regional similarity predicate (see [1] and [2] for an overview).

Equally, the measurement of object velocity in images normally exploits one of two primary techniques. The first involves the computation of the spatio-temporal gradient, differentiating the (filtered or unfiltered) image sequence with respect to time and subsequently computing the optical flow field (*e.g.* [3]). 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 (*e.g.* [4]). Comparisons of the many variations of these approaches and the relationship between them can be found in [5, 6, 7]. A third, lesser-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 which 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 of the image velocity which can be computed if one knows the position of the δ-Dirac function and any spatial frequency [8]. 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 cater for situations involving noisy backgrounds [9], several objects [10, 11], and non-uniform cluttered backgrounds [12].

In this paper, we present a segmentation and velocity estimation technique which is based on an alternative formulation of the above spatio-temporal approach. This alternative uses the normal spatial Fourier Transform together with a Hough Transform, rather than the spatiotemporal Fourier Transform. Here, object velocity is computed and segmentation effected by treating each object as a distinct intensity wave profile, with Fourier components, and by identifying the Fourier components which exhibit the magnitude and phase changes which are consistent with detected velocity. This velocity detection is accomplished using an appropriate Hough transform[13] which operates on the phase component of the Fourier Transform of each image in the sequence. Segmentation is effected by re-constituting the moving wave profile — the object — from its identified Fourier components. 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 to the wave equation. The Fourier components of wave profiles — and equally of objects — which 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 which exhibit this phase relationship using an appropriately-designed 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 is not significant and, secondly, the formulation presented in this paper lends itself to direct extension for more complex motion. Consequently, objects which are visually or spatially complex and which would be difficult to analyse using either of the traditional spatiotemporal differentiation, feature-based, or region-based approaches can be effectively treated. The objective of this paper is to introduce the alternative formulation and to demonstrate its effectiveness. We will also discuss briefly on-going work in the direct extension of the technique to address situations exhibiting more complex object motion.

2 Overview of the Approach

Consider an image $g(x,y,t)$: a 2-D spatio-temporal 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. For example, consider the motion of an object such as a motor-car travelling toward you on a road. In this paper, however, we will be restricting 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 waveform in image-time space.

Thus:

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

The solution of this wave equation $\psi^i(x,y,t)$ is, in effect, a description of the object as a grey-level wave profile propagating with constant velocity v_i, *i.e.* $\psi^i(x,y,t) = f^i(x - v_{x_i}t, y - v_{y_i}t)$, 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 $\psi = g(x,y,t)$.

Let us assume that the total optical field comprises m objects which we characterize as waves:

$$\psi = g(x,y,t) = \sum_{i=1}^{m} \psi^i(x,y,t) \tag{1}$$

By the principle of linear superposition, the total optical field can also be be decomposed into constituent components:

$$\psi(x,y,t) = \sum_{j=1}^{n} c_j \psi_j(x,y,t)$$

Equally, the wave corresponding to a given object, $\psi^i(x,y,t) = f^i(x - v_{x_i}t, y - v_{y_i}t)$, can be so decomposed:

$$\psi^i(x,y,t) = \sum_{k=1}^{l_i} c_k^i \psi_k^i(x,y,t)$$

where $c_k^i \in \{c_j\}, \psi_k^i \in \{\psi_j\}$ and where $\psi_k^i(x,y,t)$ is also a solution to the wave equation.

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

$$\psi^i(x,y,t) = \sum_{k=1}^{l_i} c_k^i f_k^i(x - v_{x_k}^i t, y - v_{y_k}^i t) \tag{2}$$

such that v_{x_k} = constant, v_{y_k} = constant, $\forall k$: that is, the components of ψ^i all have a constant propagation velocity. Substituting (2) into (1), we have:

$$\begin{aligned}\psi &= g(x,y,t)\\ &= \sum_{i=1}^{m}\sum_{k=1}^{l_i} c_k^i f_k^i(x - v_{x_k}^i t, y - v_{y_k}^i t)\end{aligned}$$

That is, the image $\psi = g(x,y,t)$ is the sum of all the individual wave profiles ψ_i, each wave profile ψ_i comprising components with constant velocity $(v_{x_k}^i, v_{y_k}^i)$.

Similarly,

$$\begin{aligned}\psi(x,y,t) &= g(x,y,t)\\ &= \sum_{j=1}^{n} c_j \psi_j(x,y,t)\\ &= \sum_{j=1}^{n} c_j f_j(x - v_{x_j} t, y - v_{y_j} t)\end{aligned}$$

The task, then, is to decompose the image into constituent components $c_j f_j(x - v_{x_j}t, y - v_{y_j}t)$, each of which satisfies the differential wave equation, and to group them into m sets $c_k^i f_k^i(x - v_{x_k}^i t, y - v_{y_k}^i t)$, $1 \le i \le m$ such that $(v_{x_k}^i, v_{y_k}^i)$ is constant.

We will use the discrete Fourier transform to accomplish the decomposition and the Hough transform to accomplish the grouping.

3 The 2-D Fourier Transform

The Fourier transform $\mathcal{F}(f(x,y)) = \mathsf{F}(k_x,k_y)$ of a 2-D function $f(x,y)$ is given by:

$$\begin{aligned}\mathcal{F}(f(x,y)) &= \mathsf{F}(k_x,k_y)\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{i(k_x x + k_y y)}\mathrm{d}x\mathrm{d}y\end{aligned}$$

where k_x and k_y are the spatial frequencies in the x and y directions. Note that $\mathsf{F}(k_x, k_y)$ is defined on a complex domain with $\mathsf{F}(k_x, k_y) = \mathsf{A}(k_x, k_y) + i\mathsf{B}(k_x, k_y)$ where

$$\mathsf{A}(k_x, k_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x', y') \cos(k_x x' + k_y y') \mathrm{d}x' \mathrm{d}y'$$
$$\mathsf{B}(k_x, k_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x', y') \sin(k_x x' + k_y y') \mathrm{d}x' \mathrm{d}y'$$

$\mathsf{F}(k_x, k_y)$ may also be expressed in terms of its magnitude and phase:

$$\mathsf{F}(k_x, k_y) = |\mathsf{F}(k_x, k_y)|\, e^{i\phi(k_x, k_y)}$$

where:

$$|\mathsf{F}(k_x, k_y)| = \sqrt{\mathsf{A}^2(k_x, k_y) + \mathsf{B}^2(k_x, k_y)}$$
$$\phi(k_x, k_y) = \arctan \frac{\mathsf{B}^2(k_x, k_y)}{\mathsf{A}^2(k_x, k_y)}$$

$|\mathsf{F}(k_x, k_y)|$ is the real-valued *amplitude spectrum* and $\phi(k_x, k_y)$ is the real-valued *phase spectrum*. The inverse Fourier transform is given by:

$$\begin{aligned} f(x, y) &= \mathcal{F}^{-1}\left(\mathsf{F}(k_x, k_y)\right) \\ &= \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathsf{F}(k_x, k_y) e^{-i(k_x x + k_y y)} \mathrm{d}k_x \mathrm{d}k_y \\ &= \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\mathsf{F}(k_x, k_y)|\, e^{i\phi(k_x, k_y)} e^{-i(k_x x + k_y y)} \mathrm{d}k_x \mathrm{d}k_y \end{aligned}$$

In the discrete case, the Fourier transform becomes:

$$\begin{aligned} \mathcal{F}\left(f(x, y)\right) &= \mathsf{F}(k_x, k_y) \\ &= \sum_x \sum_y f(x, y) e^{i(k_x x + k_y y)} \end{aligned}$$

and the inverse discrete Fourier transform is:

$$\begin{aligned} f(x, y) &= \mathcal{F}^{-1}\left(\mathsf{F}(k_x, k_y)\right) \\ &= \frac{1}{(2\pi)^2} \sum_{k_x} \sum_{k_y} |\mathsf{F}(k_x, k_y)|\, e^{i\phi(k_x, k_y)} e^{-i(k_x x + k_y y)} \end{aligned}$$

In effect, $f(x, y)$ can be constructed from a linear combination of elementary functions having the form $e^{-i(k_x x + k_y y)}$, each appropriately weighted in amplitude and phase by a complex factor $\mathsf{F}(k_x, k_y)$. However, this construction is valid only for a given time $t = t_0$, say, and, since we are dealing with a wave function $\psi(x, y, t)$ rather than a 2-D spatial image $\psi(x, y)$, we need to develop this formulation of the Fourier and inverse Fourier transforms.

Consider a waveform $f(x, y)$ and an identical waveform shifted to coordinates (x_δ, y_δ), *i.e.*, $f(x - x_\delta, y - y_\delta)$. If the waveform is travelling with constant velocity, then $f(x - x_\delta, y - y_\delta) = f(x - v_x\delta t, y - v_y\delta t)$. The Fourier transform of $f(x - x_\delta, y - y_\delta)$, equivalently $f(x - v_x\delta t, y - v_y\delta t)$, is given by:

$$
\begin{aligned}
&\mathcal{F}\left(f(x - v_x\delta t, y - v_y\delta t)\right)\\
&= \mathsf{F}(k_x, k_y)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x - v_x\delta t, y - v_y\delta t) e^{i(k_x(x - v_x\delta t) + k_y(y - v_y\delta t))} \mathrm{d}x\mathrm{d}y\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x - v_x\delta t, y - v_y\delta t) e^{i(k_x x + k_y y) - i(k_x v_x\delta t + k_y v_y\delta t)} \mathrm{d}x\mathrm{d}y\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x - v_x\delta t, y - v_y\delta t) e^{i(k_x x + k_y y)} e^{-i(k_x v_x\delta t + k_y v_y\delta t)} \mathrm{d}x\mathrm{d}y\\
&= e^{-i(k_x v_x\delta t + k_y v_y\delta t)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x - v_x\delta t, y - v_y\delta t) e^{i(k_x x + k_y y)} \mathrm{d}x\mathrm{d}y\\
&= e^{-i(k_x v_x\delta t + k_y v_y\delta t)} \mathsf{F}(k_x, k_y)\\
&= |\mathsf{F}(k_x, k_y)|\, e^{i\phi(k_x, k_y)} e^{-i(k_x v_x\delta t + k_y v_y\delta t)}
\end{aligned}
$$

Thus, a spatial shift of $(x_\delta, y_\delta) = (v_x\delta t, v_y\delta t)$ of a waveform in the spatial domain, *i.e.* $f(x, y)$ shifted to $f(x_\delta, y_\delta) = f(v_x\delta t, v_y\delta t)$, only produces a change in the phase of the Fourier components in the frequency domain. This phase change is given by:

$$e^{-i(k_x v_x\delta t + k_y v_y\delta t)}$$

Thus, in order to segment the image into its component waveforms, 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.* $e^{-i(k_x v_x\delta t + k_y v_y\delta t)}$. 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 above:

$$
\begin{aligned}
e^{i\phi_{t+\delta t}(k_x, k_y)} &= e^{-i(k_x v_x\delta t + k_y v_y\delta t)} e^{i\phi_t(k_x, k_y)}\\
&= e^{i(\phi_t(k_x, k_y) - (k_x v_x\delta t + k_y v_y\delta t))}
\end{aligned}
$$

Hence:

$$\phi_{t+\delta t}(k_x, k_y) = \phi_t(k_x, k_y) - (k_x v_x\delta t + k_y v_y\delta t)$$

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

$$v_y = \frac{1}{k_y\delta t}\left(\phi_t(k_x, k_y) - \phi_{t+\delta t}(k_x, k_y) - k_x v_x\delta t\right)$$

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

$$v_x = \frac{(\phi_t(k_x, k_y) - \phi_{t+\delta t}(k_x, k_y))}{k_x \delta t}$$

If we have several images taken at time $t = t_0, t_1, t_2, t_3, ...$, we can compute ϕ_{t_0}, in particular, and $\phi_{t_0+n\delta t}$, in general. Treating the equation above an 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 which comprise waveforms – objects – in the spatial domain which are moving with constant velocity v_x, v_y.

4 Results

Figures 1 through 12 demonstrate the results of applying the technique. Figures 1 and 2 show the first and last images in an eight image sequence featuring a moving dog; figure 3 shows the velocity Hough Transform derived from the phase values of the Fourier Transform of each image in the sequence. The velocity computations are summarized in Table 1. Figures 4 to 6 depict the re-constructed, segmented, image where only frequency components whose phase satisfies the velocity/phase-change relationship embodied in the Hough Transform for the detected velocity are used in the reconstruction; figures 4, 5, and 6 show the results when a minimum phase change threshold of 0.4, 0.3, and 0.2 radians, respectively, is imposed.

Figures 7 through 12 demonstrate the results in a more complex scenario where there is a moving foreground image of a cat superimposed on the background scene depicted in figures 1 and 2. Such a situation arises when, for example, an observer (the cat) views a scene through a window and sees both the external scene and its own reflection.

Image Sequence	Actual Velocity (pixels/frame)		Computed Velocity (pixels/frame)	
	v_x	v_y	v_x	v_y
1	0	3.07	0	3.2
2	0	3.07	0.4	2.8

Table 1. Summary of Measured Velocities

Fig. 1. Image 1 of sequence 1

Fig. 2. Image 8 of sequence 1

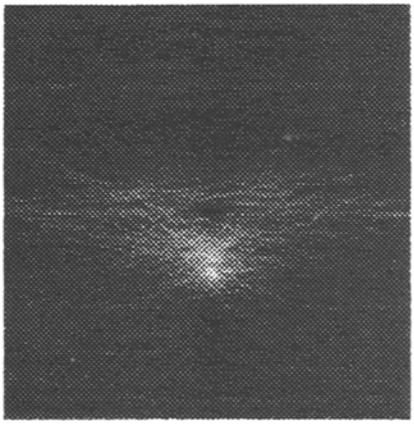

Fig. 3. Hough transform (v_x, v_y) space derived from images 1-8

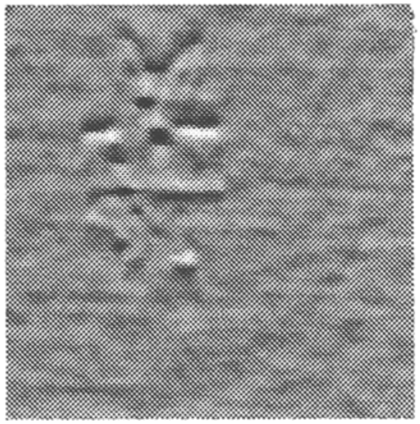

Fig. 4. Re-constructed image (min. phase change of 0.4 radians)

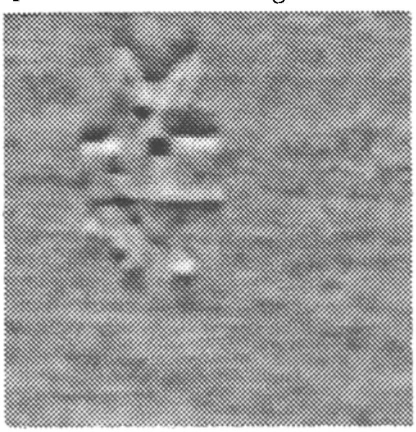

Fig. 5. Re-constructed image (min. phase change of 0.3 radians)

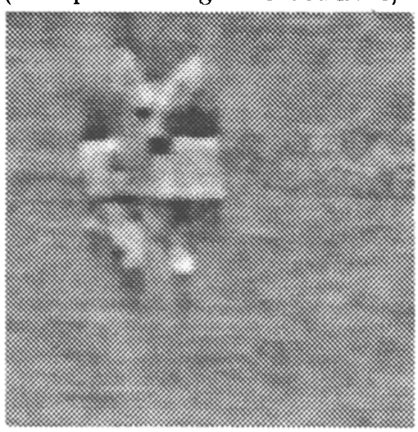

Fig. 6. Re-constructed image (min. phase change of 0.2 radians)

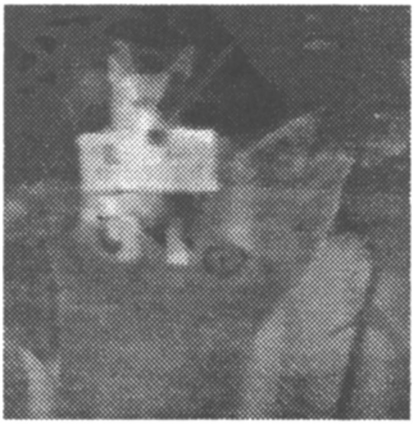

Fig. 7. Image 1 of sequence 2

Fig. 8. Image 8 of sequence 2

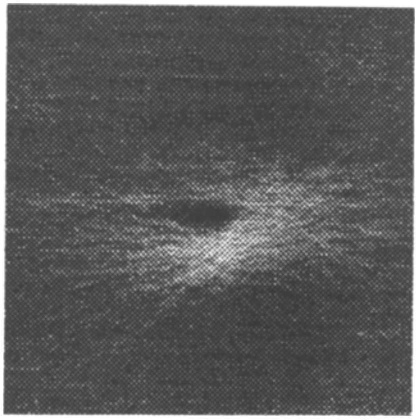

Fig. 9. Hough transform (v_x, v_y) space derived from images 1-8

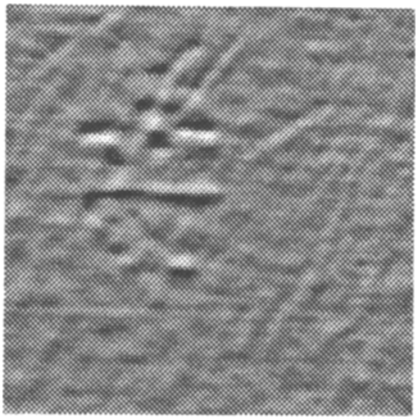

Fig. 10. Re-constructed image (min. phase change of 0.4 radians)

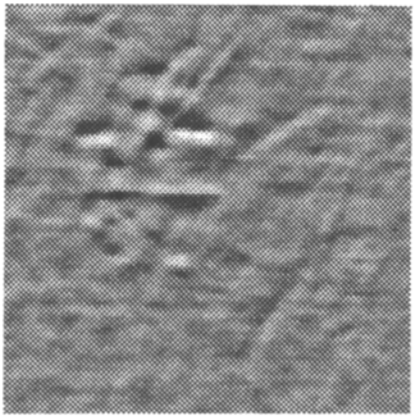

Fig. 11. Re-constructed image (min. phase change of 0.3 radians)

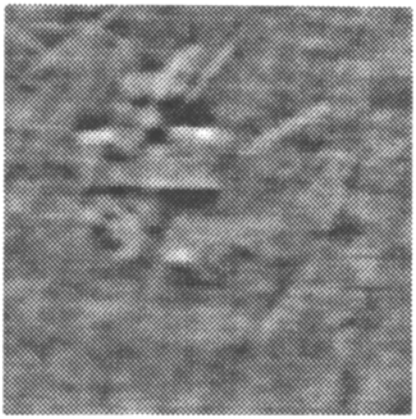

Fig. 12. Re-constructed image (min. phase change of 0.2 radians)

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 may in fact be 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 figures 7 and 8).

5.2 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 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 which 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.

Non-Uniform 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 catered for by allowing independent scaling of the object in the horizontal and vertical directions, respectively.

6 Conclusions

A flexible and extendable technique for segmenting and 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 in situations of varying complexity.

References

1. D. Vernon, *Machine Vision* Prentice-Hall International, London (1991).
2. D. Vernon and G. Sandini, *Parallel Computer Vision — The* VIS a VIS *System*, Ellis Horwood, London (1992).
3. J.H. Duncan and T.-C. Chou, "On the detection and the computation of optical flow", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14(3)**, 346-352 (1992).
4. H. Shariat and K.E. Price, "Motion estimation with more than two frames", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12(5)**, 417-434 (1990).
5. 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*, Springer-Verlag, Berlin, 51-60 (1994).
6. M. Tistarelli, "Multiple constraints for optical flow", *Lecture Notes in Computer Science*, J.O. Eklundh (Ed.), *Computer Vision – ECCV '94*, Springer-Verlag, Berlin, 61-70 (1994).
7. L. Jacobson and H. Wechsler, "Derivation of optical flow using a spatiotemporal-frequency approach", *Computer Vision, Graphics, and Image Processing*, **38**, 29-65 (1987).
8. M.P. Cagigal, L. Vega, P. Prieto, "Object movement characterization from low-light-level images", *Optical Engineering*, **33(8)**, 2810-2812 (1994).
9. M.P. Cagigal, L. Vega, P. Prieto, "Movement characterization with the spatiotemporal Fourier transform of low-light-level images", *Applied Optics*, **34(11)**, 1769-1774 (1995).
10. S. A. Mahmoud, M.S. Afifi, and R. J. Green, "Recognition and velocity computation of large moving objects in images", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **36(11)**, 1790-1791 (1988).
11. S. A. Mahmoud, "A new technique for velocity estimation of large moving objects", *IEEE Transactions on Signal Processing*, **39(3)**, 741-743 (1991).
12. S.A. Rajala, A. N. Riddle, and W.E. Snyder, "Application of one-dimensional Fourier transform for tracking moving objects in noisy environments", *Computer Vision, Graphics, and Image Processing*, **21**, 280-293 (1983).
13. P.V.C. Hough, 'Method and Means for Recognising Complex Patterms' U.S. Patent 3,069,654, (1962).