Estimation of Optical Flow using the Fourier Transform

D. Vernon
Department of Computer Science
National University of Ireland, Maynooth
Ireland
dvernon@cs.may.ie

Abstract

A technique for computing the optical flow of a sequence of four or more images is presented. The velocity at each point in the sequence can be computed by treating a local region as a distinct sub-image which is translating with some velocity and by identifying the Fourier components which exhibit the magnitude and phase changes which are consistent with this velocity. The approach presented is more general than previously developed Fourier approaches to optical flow estimation because it does not assume that the entire sub-image is translating with a unique velocity. Instead, it allows that there may be (and certainly will be at occluding boundaries) two objects moving in the image, each having a distinct velocity. Consequently, it computes both possible velocities and chooses as the velocity which possesses the greater number of Fourier components.

The proposed approach is evaluated using Otte and Nagel's benchmark image sequence, [6], for which ground-truth data is available, and both maximum and RMS errors of velocity magnitude and direction are computed.

1 Introduction

The measurement of optical flow has received a great deal of attention by the computer vision community and the literature is replete with a very large number of publications on the topic. Surveys and comparisons of the different approaches can be found in comprehensive works by Barron *et al.* [5] and by Otte and Nagel [6]. Most approaches to the measurement of optical flow in images normally exploit one of two primary techniques. The first involves the computation of the spatio-temporal derivatives (first-order or second-order), differentiating the (filtered or unfiltered) image sequence with respect to time and thus computing the optical flow field (*e.g.* [3]). The second involves either feature- or region-based matching (such as normalized cross-correlation) of local iconic information such as raw image data or segmented features or objects. Comparisons of the many variations of these approaches and the relationship between them can be found in [5, 6, 7, 8].

A 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 [9, 10, 11, 12, 13].

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. The velocity at each point in the sequence is computed by treating a local region as a distinct sub-image which is translating with some velocity and by identifying the Fourier components which exhibit the magnitude and phase changes which are consistent with this velocity. The approach presented is more general than previously developed Fourier approaches to optical flow estimation because it does not assume that the entire sub-image is translating with a unique velocity. Instead, it allows that there may be (and certainly will be at occluding boundaries) two objects moving in the image, each having a distinct velocity.

Consequently, it computes both possible velocities and chooses as the velocity which possesses the greater number of Fourier components. The approach is based on the construction and solution of a set of four simultaneous quadratic equations in four complex unknowns. These unknowns are the Fourier components of two objects translating in the image and the phase change associated with their displacement. A Hough transform is used to group each of these solutions into two distinct sets each of which exhibits a common phase change and, hence, a common velocity. Because it bases the velocity computation on the Fourier phase information, the approach is able to estimate the velocity on any signal function except a uniform flat field (which has an ambiguous flow field in any case) and, consequently, the approach lends itself to the production of arbitrarily dense optical flow fields. In addition, the technique facilitates the computation of the velocity vectors with sub-pixel accuracy.

The proposed approach is evaluated using two images from Otte and Nagel's benchmark image sequence [6], for which ground-truth data is available, and both maximum and RMS errors of velocity magnitude and direction are reported.

2 Theoretical Background

It is shown elsewhere [18] that an image sequence comprising two additive images can be decomposed into the constituent images, provided each constituent image exhibits a distinct velocity. This decomposition is effected by solving a a set of four simultaneous quadratic equations in four complex unknowns. These unknowns are the Fourier components of two objects translating in the image and the phase change $\delta\phi(k_x, k_y)$ associated with the displacement of each of the two constituent images. This phase change is a spatial frequency-dependent function of the velocity of the image and is given by:

$$\delta\phi(k_x, k_y) = -(v_x^i k_x \delta t + v_y^i k_y \delta t) \tag{1}$$

This phase change will differ for each object i. It only remains, then, to identify the two velocities (v_x^1, v_y^1) and (v_x^2, v_y^2). We do this using a Hough transform. From equation 1 we have:

$$v_y = \frac{1}{k_y \delta t}\left(\delta\phi(k_x, k_y) - k_x v_x \delta t\right) \tag{2}$$

This equation represents a Hough transform in the two variables (v_x, v_y) which we solve for all $\delta\phi(k_x, k_y), k_x, k_y$, and v_x. Note that $\delta\phi(k_x, k_y) = \arctan(\Im\Delta\Phi(k_x, k_y), \Re\Delta\Phi(k_x, k_y))$. Local maxima in this Hough transform space represent the velocities which are exhibited by the frequency components. In this case, there are two velocity maxima, one corresponding to wave 1 and the other to wave 2. The location of these maxima give us (v_x^1, v_y^1) and (v_x^2, v_y^2) and, thus, we can proceed to sort the components.

Note that the Hough transform equation 2 becomes degenerate if $k_y = 0$ in which case we use an alternative re-arrangement as follows:

$$v_x = \frac{(\delta\phi(k_x, k_y))}{k_x \delta t} \tag{3}$$

3 Results

In the estimation of velocity at a point, we base the estimate on the translation which is apparent in a local region (or window). Unfortunately, the image data in such a region will, in fact, exhibit a change due not only to the signal shift but also the translation of objects into the window and out of the the window. Consequently, there is a change in the spectral content of the window and not just a phase change as is assumed in the model. In order to reduce the impact of this 'edge effect', image data in a region is weighted as a function of its distance from the region centre. In this paper, a Gaussian weighting function is used and the Gaussian's standard deviation σ chosen such that the weighting at a some distance from the region centre is 50% of that at the region centre, where w is the length of the side of the 2-D region. Results are presented for Gaussian weighting functions of three standard deviations, each representing increased attenuation of image data toward the edge of the image (the three functions provide 50% weighting at $\frac{w}{8}$, $\frac{2w}{8}$, and $\frac{3w}{8}$ from the region centre). In the following, we will denote the three Gaussian functions as $\sigma_{\frac{w}{8}}$, $\sigma_{\frac{2w}{8}}$, and $\sigma_{\frac{3w}{8}}$.

Figures 1 through 2 demonstrate the results of applying the technique to two images in Otte and Nagel's ground-truth test sequence [6].

Table 1 summarizes the mean magnitude and the mean direction of ground-truth data and the measured velocities; table 2 provides a summary of the RMS and mean errors of the measured velocities.

Note that all of the results presented in this paper are the unprocessed output of the algorithm (apart from interpolation); each velocity vector has been estimated independently and the vector field has not be subjected to median or mean filtering.

4 Conclusions

The technique presented estimates the optical flow of an image sequence with at least four images by decoupling the Fourier transform of a local Gaussian-weighted window centred at every point at which the flow field is to be computed. The results compare very favourably with ground truth optical flow data for the benchmark image sequence used to test the approach.

Problems remain at occluding contours. However, since the technique computes the velocities of both objects in the local window around the occluding boundary (*i.e.* the occluding object and the occluded background) it should be possible, in principle at least, to compute the correct flow field on either side of the velocity discontinuity (*i.e.* the occluding boundary).

References

[1] D. Vernon, *Machine Vision* Prentice-Hall International, London (1991).

[2] D. Vernon and G. Sandini, *Parallel Computer Vision — The* VIS $\vec{a}$ 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] J.L. Barron, D.J. Fleet, and S. Beauchemin, "Performance of optical flow techniques", *International Journal of Computer Vision*, **12(1)**, 43-77 (1994).

[6] 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).

[7] 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).

[8] L. Jacobson and H. Wechsler, "Derivation of optical flow using a spatiotemporal-frequency approach", *Computer Vision, Graphics, and Image Processing*, **38**, 29-65 (1987).

[9] M.P. Cagigal, L. Vega, P. Prieto, "Object movement characterization from low-light-level images", *Optical Engineering*, **33(8)**, 2810-2812 (1994).

[10] 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).

[11] 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).

[12] S. A. Mahmoud, "A new technique for velocity estimation of large moving objects", *IEEE Transactions on Signal Processing*, **39(3)**, 741-743 (1991).

[13] 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).

[14] D. Vernon, "Phase-Based Measurement of Object Velocity in Image Sequences using the Hough Transform", *Optical Engineering* (1996).

[15] D. Vernon, "Segmentation in Dynamic Image Sequences by Isolation of Coherent Wave Profiles", Proceedings of the 4th European Conference on Computer Vision, Springer-Verlag, 293-303 (1996).

[16] P.V.C. Hough, 'Method and Means for Recognising Complex Patterns' U.S. Patent 3,069,654, (1962).

[17] L. Hahn, *Complex Numbers and Geometry*, The Mathmatical Association of America, Washington, D.C. (1994).

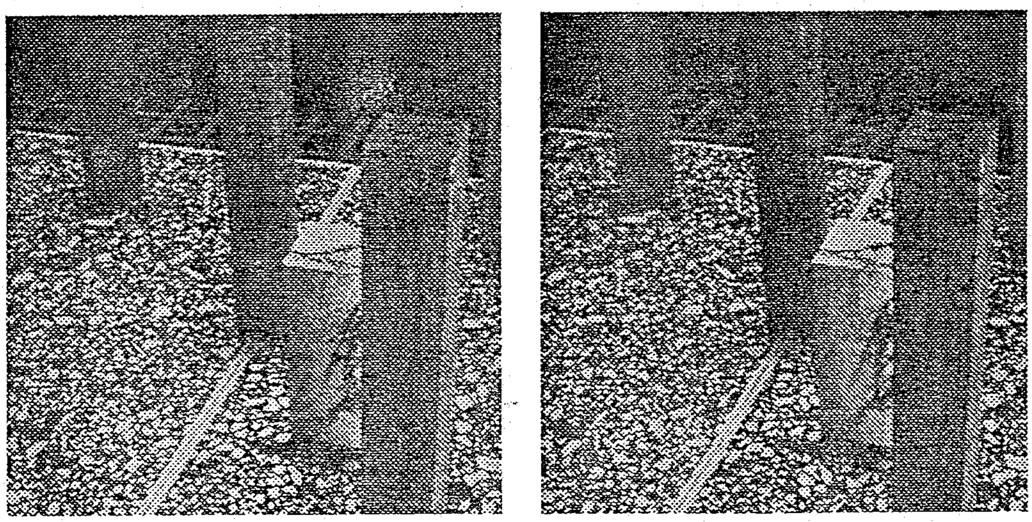
[18] D. Vernon, "Removal of Superimposed Reflections from Dynamic Image Sequences", Proceedings of the First Pan-Irish Conference on Machine Vision and Image Processing (1997).

Image	Magnitude		Direction	
	mean	standard deviation	mean	standard deviation
Ground Truth	1.353	0.807	0.272	0.231
$\frac{3w}{8}$	0.959	0.766	0.240	0.374
$\frac{2w}{8}$	0.991	0.793	0.249	0.289
$\frac{w}{8}$	0.975	0.772	0.246	0.289

Table 1: Summary of mean magnitude and the mean direction of ground-truth data and measured velocities (using Gaussian weighting functions with 50% weight at $\frac{3w}{8}$, $\frac{2w}{8}$, and $\frac{w}{8}$ pixels from window centre, respectively, where w = window size).

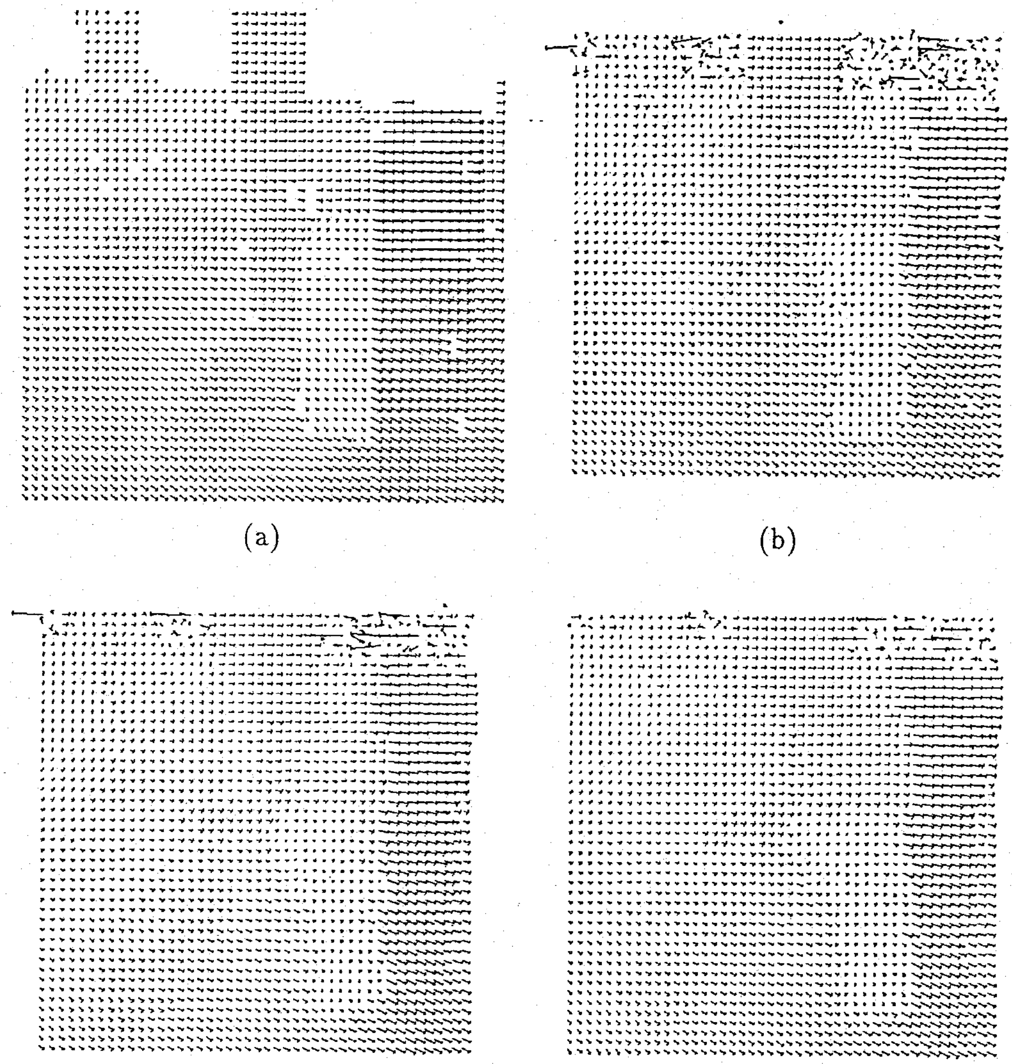
Gaussian Weighting	RMS Error		Mean Error	
	Magnitude (pixels)	Direction (radians)	Magnitude (pixels)	Direction (radians)
$\frac{3w}{8}$	0.522	0.202	0.280	0.070
$\frac{2w}{8}$	0.531	0.235	0.273	0.102
$\frac{w}{8}$	0.510	0.282	0.287	0.106

Table 2: Summary of Errors in measured velocities (using Gaussian weighting functions with 50% weight at $\frac{3w}{8}$, $\frac{2w}{8}$, and $\frac{w}{8}$ pixels from window centre, respectively, where w = window size).



(a) (b)

Figure 1: (a) and (b) Images number 40 and 43 of Otte and Nagel's ground-truth sequence.



(a) (b) (c) (d)

Figure 2: (a) True optical flow field computed from Otte and Nagel's ground-truth data (sampled every ten pixels); Optical flow field computed using phase information: Gaussian weighting function with 50% weight at (b) $\frac{w}{8}$ pixels (c) $\frac{2w}{8}$ pixels (d) $\frac{3w}{8}$ pixels from window centre, w = window size).

School of Information and Software Engineering
Faculty of Informatics

Irish Machine Vision and Image Processing Conference (IMVIP-97)

and

Eight Ireland Conference on Artificial Intelligence (AI-97)

Volume 1