Phase-Based Computation of Stereo Disparity

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

Abstract

We present an alternative to traditional feature-matching approaches to the measurement of stereo disparity. Here, the disparity is computed by first registering the two images using an algorithm which combines both intensity and gradient information and then by estimating the relative local image shift using the phase of the Fourier components of local 64×64 pixel windows in each stereo image.

The window disparity is measured by exploiting the shift property of the Fourier transform. The Fourier components which exhibit phase changes which are consistent with a given shift are detected using a custom Hough transform. Results are presented for benchmark image data produced by the computer vision group in Carnegie Mellon University.

1 Introduction

Traditional approaches to the measurement of stereo disparity in images normally exploit one of two primary techniques (i) those which attempt to identify the left-to-right stereo image correspondence by matching regions and sub-regions in the image using, *e.g.* correlation techniques, and (ii) those which exploit feature-based matching using, *e.g.* edge-correspondence. [1, 2]. Typically, epipolar and smoothness constraints are used to reduce the complexity of the matching process and, normally, both matching procedures are carried out in the space domain.

In this paper, we present an alternative region-based approach which operates in the spatial frequency domain and not the spatial domain, exploiting the shift property of the Fourier transform to detect and measure image translation [3, 4, 5, 6, 7, 8, 9]. This approach comprises two distinct steps. First, the two stereo images are registered and, second, the image shift in each 64×64 pixel window centred at position (i, j) in the left and right images are computed by identifying the spatial-frequency dependent phase shifts which are consistent with a unique shift. The latter step is effected using a custom Hough transform to be described below.

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 generalization so that more complex circumstances (*e.g.* occlusion) can be addressed. Consequently, objects which are visually or spatially complex and which would be difficult to analyse using either of the traditional approaches can be effectively treated.

2 Theoretical Basis of the Approach

2.1 Image Registration andn Computation of Global Disparity

No two stereo image pairs are exactly alike. Nonetheless, they are approximately similar and this similarity makes it possible to register the two images, effectively computing the gross, or global, disparity between the two images. Once this is done, the point-by-point deviations from this global disparity can be identified. The two stereo images are registered by translating and amplitude shifting one with respect to the other, and evaluating a similarity measure d at each shift where d is defined in terms of both the image data and the gradient of the left image as follows:

$$d = \sqrt{\Sigma(L(i,j) - (R(i,j) + \delta)^2 * w)}$$

where w is a weighting factor which apportions relative emphasis of points as a function of their gradient value:

$$w = 1 - k + \frac{k * \nabla L(i,j)}{\nabla_{max} L(i,j)}$$

k is a user-specified percentage which specifies the relative weighting. The offset δ is used to normalise the amplitude of the right image $R(i,j)$ with respect to the left image $L(i,j)$ based on the difference of their average values. The registered image is the right image translated and amplitude-shifted by the values which minimise the difference measure d.

2.2 Computation of Local Disparity

Once the left and right images are registered, the displacements (or disparities) of all 64x64 pixel windows in the left image (with respect to the right image) are then computed. The window disparity is measured by exploiting the shift property of the Fourier transform, *i.e.* a shift in a signal only introduces a phase change in the Fourier components. Consequently, the Fourier transforms of the corresponding windows in the left and right images are computed and then the Fourier components which exhibit phase changes which are consistent with a given shift are detected. This detection is accomplished using a custom Hough transform which defines the possible x and y displacements as a function of the phase change of each spatial frequency component. The displacement with the greatest amount of evidence in the Hough space is taken to be the stereo disparity of that window and is computed to sub-pixel accuracy.

Specifically, the Fourier transform of an image $f(x,y)$ shifted by $(\delta x, \delta y)$ is given by [10]:

$$\mathcal{F}\left(f(x - \delta x, y - \delta y)\right) = |\mathsf{F}(k_x, k_y)|\, e^{i\phi(k_x,k_y)} e^{-i(k_x\delta x + k_y \delta y)}$$

Thus, a spatial shift of $(\delta x, \delta y)$ of an image in the spatial domain, *i.e.* $f(x,y)$ shifted to $f(x - \delta x, y - \delta y)$, only produces a change in the phase of the Fourier components in the frequency domain. This phase change is $e^{-i(k_x\delta x + k_y\delta y)}$. Thus, in order to identify the spatial shift of a given 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.* $e^{-i(k_x\delta x + k_y\delta y)}$. To accomplish this, we note that the phase spectrum for the shifted image (*i.e.* the left image) is equal to the phase spectrum of the right image multiplied by the phase change given above:

$$\begin{aligned} e^{i\phi_{t+\delta t}(k_x,k_y)} &= e^{-i(k_x\delta x + k_y\delta y)} e^{i\phi_r(k_x,k_y)} \\ &= e^{i(\phi_r(k_x,k_y) - (k_x\delta x + k_y\delta y))} \end{aligned}$$

Hence:

$$\phi_l(k_x, k_y) \;=\; \phi_r(k_x, k_y) - (k_x\delta x + k_y\delta y)$$

That is, the phase of the left image is equal to the phase of the right image minus $(k_x\delta x + k_y\delta y)$. Since we require δx and δy, we rearrange as follows:

$$\delta y = \frac{1}{k_y}\left(\phi_r(k_x, k_y) - \phi_l(k_x, k_y) - k_x\delta x\right) \tag{1}$$

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

$$\delta x = \frac{(\phi_r(k_x, k_y) - \phi_l(k_x, k_y))}{k_x} \tag{2}$$

Treating the equation above as a Hough transform, with a 2-D Hough transform space defined on $\delta x, \delta y$, then we can compute δy for all possible values of δx, and for all (known) values of

$k_x, k_y, \phi_r(k_x, k_y), \phi_l(k_x, k_y)$. Local maxima in this $\delta x, \delta y$ Hough transform space occur at the maximally-likely image shift, *i.e.* the local stereo disparity can be computed by identifying the position of the maximum in the Hough space.

Note that, in both equations (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 (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 $2n\pi <\mid k_x v_{x_{max}} \delta t \mid + \mid k_y v_{y_{max}} \delta t \mid$.

3 Results

Figures 1(a)through 1(d) demonstrate the results of applying the technique to a stereo image taken from the Carnegie Mellon University stereo database [11]. Figures 1(a) and 1(b) show the left and right images, respectively. Figure 1(c) shows the computed local stereo disparity vector field computed every ten pixels superimposed on the left image and figure 1(d) shows the full stereo disparity image which was computed by bi-linear interpolation between each local stereo estimate.

4 Discussion

The model used in the approach presented in this paper assumes that the image in any window in the left image is simply a shifted version of the image in the corresponding window in the right image (and, hence, that there is only a phase difference between the Fourier components of the left and right images, respectively). Whilst this is usually a good approximation, is it invalid if there is a significant distortion in the left and right windows due *e.g.* to the presence of occluding boundaries in the image scene. Consequently, there may be significant changes in the spatial frequency magnitude in the window as well as the expected phase change. The solution of this problem necessitates a generalization of the spatial translation model upon which the approach is founded to embrace not just the phase-shift related rotation of each spatial frequency vector (or phasor) as it stands at present but also to embrace the variation in the magnitude of the phasor. Work is proceeding on this generalization and early theoretical and empirical results are encouraging.

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

[4] 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).

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

[6] 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).

[7] 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).

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

[9] 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).

[10] A. Rosenfeld and A. Kak, *Digital Picture Processing (Volume 1)*, Academic Press, New York, pp. 12 (1982).

[11] Carnegie Mellon University Calibrated Image Laboratory, CIL Stereo Dataset, dataset no. CIL-001, www.cs.cmu.edu/afs/cs/project/cil/ftp/html/cil-ster.html.

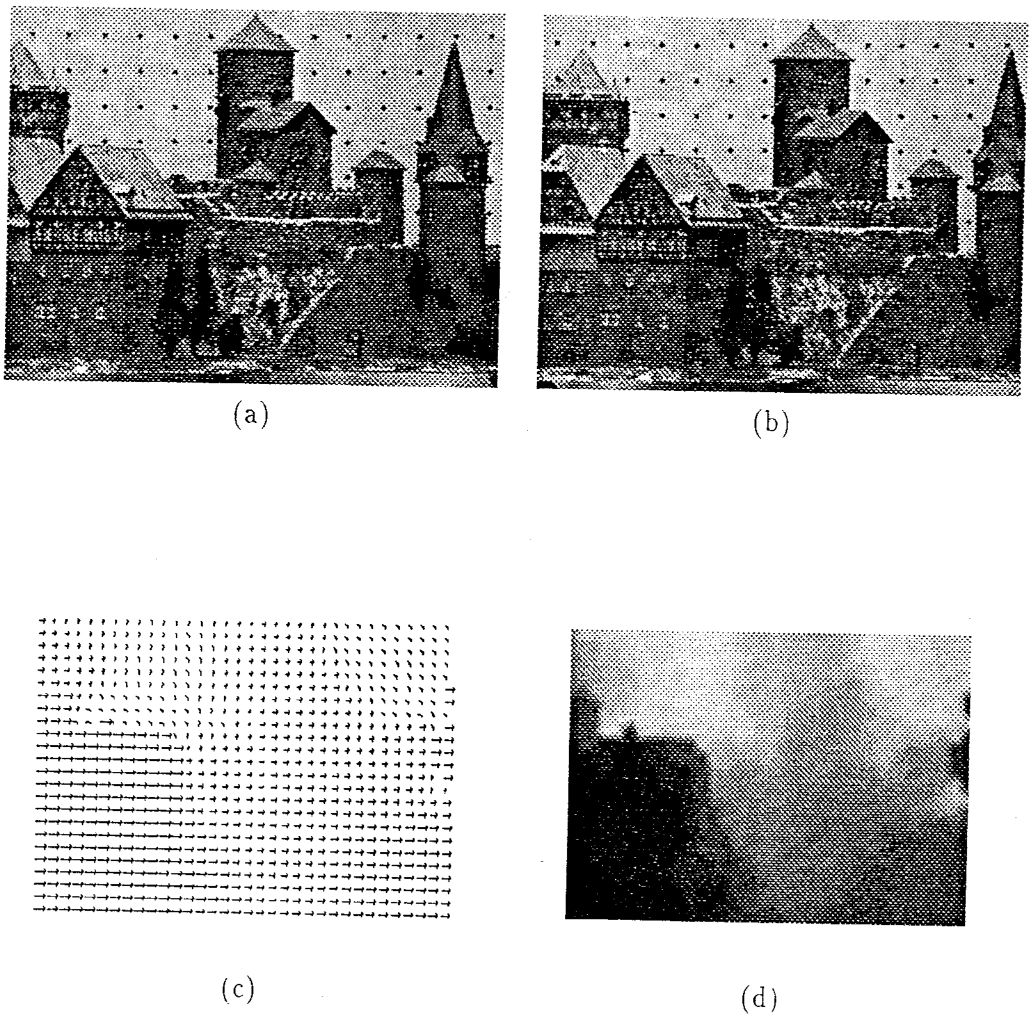

Figure 1: (a) Left (b) and right stereo images taken from the Carnegie Mellon University stereo database (c) Computed stereo disparity vector field (d) Stereo disparity map computed by bilinear interpolation of the magnitude of the stereo disparity vector field.

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