Stereoscopic Phase-Differencing: Multiscale Synthesis

James Monaco and Alan C. Bovik
The University of Texas at Austin
Department of Electrical and Computer Engineering
Austin, TX 78712-1084 USA
{monaco,bovik}@ece.utexas.edu

Lawrence K. Cormack
The University of Texas at Austin
Department of Psychology
Austin, TX 78712-0187 USA

Abstract

Exploiting the quasi-linear relationship between local phase and disparity, phase-differencing registration algorithms provide a fast, powerful means for disparity estimation. Unfortunately, phase-differencing techniques suffer from a significant impediment: the neglect of multiscale information. In this work, we introduce a novel registration algorithm that combines strategies of both phase-differencing and local correlation. This hybrid approach retains the advantageous properties of phase-differencing while incorporating the multiscale aspects of local correlation.

1 Introduction

As an alternative to typical feature based algorithms, phase-based techniques, such as phase-differencing, have proven to be a powerful means for stereo image correspondence [1, 6, 8–10, 12]. These techniques can produce disparity maps to sub-pixel accuracy, without requiring explicit sub-pixel reconstruction. Since relevant phase information is available at any point in the image, not only at specific tokens as in feature based methods [3], phase-based methods yield dense disparity maps. Additionally, phase is relatively insensitive to typical inter-image differences such as lighting, shadows, and noise [2]. Unfortunately, phase-differencing techniques suffer from a significant impediment: the neglect of multiscale information.

Phase-differencing algorithms require coarse to fine processing schemes to allow larger scales to bring the images into near registration before proceeding to higher scales. At each scale disparity information is procured from a small range of frequencies. Images with relatively little energy in these frequency bands tend to produce erroneous estimates that further contaminate higher scales as the iteration proceeds. To mitigate this problem, alternative means for using phase information have been suggested. A typical approach is to apply a local Fourier Transformation [4, 10] or wavelet decomposition [12] and then globally search between images for locations that minimize the difference in their coefficients. This difference is predominately a function of the local phase. These methods are time consuming, disregarding the linearity of phase that allows phase-differencing to rapidly convergence to high resolution solutions. Furthermore, their resolution is inherently limited by the sampling of the transformations, and consequently require ancillary methods for sub-pixel estimation. Other approaches use phase-differencing to realize disparity estimates at multiple frequencies then heuristically combine them at the end [7, 11].

The intent of this paper is to introduce a novel algorithm for combining multiscale disparity information in a theoretically sound approach that retains the fast, high resolution convergence properties of phase-differencing. Developing this approach, we demonstrate the intimate relation between phase-differencing and local correlation.

2 Review: Phase Differencing and Local Correlation

In this section we first review strategies for implementing a phase-differencing registration algorithm. We then provide an overview of local correlation and its value in disparity estimation. Finally, we briefly compare and contrast both methodologies, highlighting their advantages and shortcomings.

2.1 Phase-Differencing

Phase-differencing exploits the quasi-linear relationship between phase and translation to estimate disparity in a single step using the expression $\delta = \Delta\theta/\tilde{\omega}$, where δ is the local disparity between the stereo images, $\Delta\theta$ is the phase difference, and $\tilde{\omega}$ is some approximation of the instantaneous frequency $\frac{\partial}{\partial x}\theta$. This estimate can be refined in a Newton

iteration

$$\delta^{(n+1)} = \frac{\theta_2\left(x + \delta^{(n)}\right) - \theta_1\left(x\right)}{\tilde{\omega}\left(x, \delta^{(n)}\right)}, \qquad (1)$$

where $\tilde{\omega}\left(x, \delta\right)$ is judiciously chosen as the average instantaneous frequency over both images (though other formulations may suffice)

$$\tilde{\omega}\left(x, \delta\right) = \frac{1}{2}\left[\frac{\partial\theta_2}{\partial x}\left(x + \delta\right) + \frac{\partial\theta_1}{\partial x}\left(x\right)\right]. \qquad (2)$$

Since phase wrapping forces $|\Delta\theta| < \pi$, in order to detect absolute disparities of size D we must have $\tilde{\omega} < \pi/D$. That is, high disparities must be detected at low frequencies. This suggests a coarse to fine process that allows large scales to bring the images into near registration before continuing to smaller scales.

2.2 Local Correlation

Local correlation is another disparity estimation technique that relies heavily on phase information; though, unlike phase-differencing, no explicit computation of phase is performed [4, 8]. To better elucidate this methodology consider the regular cross-correlation of the two images $i_1\left(x\right)$ and $i_2\left(x\right)$ and its representation in the frequency domain:

$$\begin{aligned} p\left(s\right) &= \int i_1\left(x\right)i_2\left(x + s\right)dx \\ &= \frac{1}{2\pi}\int I_1\left(\omega\right)I_2^*\left(\omega\right)e^{-j\omega s}dw, \qquad (3) \end{aligned}$$

where $I_1\left(\omega\right)$ and $I_2\left(\omega\right)$ are the Fourier Transforms of $i_1\left(x\right)$ and $i_2\left(x\right)$, respectively. If $i_1\left(x\right)$ and $i_2\left(x\right)$ are related by a simple translation, $i_2\left(x\right) = i_2\left(x - \delta\right)$, then (3) reduces to

$$p\left(s\right) = \frac{1}{2\pi}\int |I_1\left(\omega\right)|^2 e^{-j\omega(s-\delta)}dw, \qquad (4)$$

which obviously reaches its maximum when $s = \delta$. Removing the magnitudes from the previous expression (leaving only phase information) produces the phase-correlation

$$\tilde{p}\left(s\right) = \frac{1}{2\pi}\int e^{-j\omega(s-\delta)}dw = \delta\left(s - \delta\right), \qquad (5)$$

where $\delta\left(\cdot\right)$ is the Dirac Delta function.

These results assume a global shift between images and must be spatially localized. Localization is achieved by first convolving each image with the ith quadrature kernel $h_i\left(x\right)$

$$O_{1i}\left(x\right) = h_i\left(x\right) * i_1\left(x\right) = A_{1i}\left(x\right)e^{j\theta_{1i}(x)}, \qquad (6)$$

$$O_{2i}\left(x\right) = h_i\left(x\right) * i_2\left(x\right) = A_{2i}\left(x\right)e^{j\theta_{2i}(x)}. \qquad (7)$$

Usually, each $h_i\left(x\right)$ is tuned to a unique center frequency ω_i; and $O_{\cdot i}$ represents one channel of a bandpass decomposition of the image. Using (6) and (7), the local cross correlation can be expressed as

$$\begin{aligned} p\left(x, s\right) &= \sum O_{1i}\left(x\right)O_{2i}^*\left(x + s\right) \\ &= \sum A_{1i}(x)A_{2i}(x + s)\,e^{j[\theta_{1i}(x) - \theta_{2i}(x+s)]}.(8) \end{aligned}$$

If the kernels are identically windowed complex exponentials with linearly spaced tuning frequencies, we can consider (6) and (7) as samples of the Short Time Fourier Transforms of $i_1\left(x\right)$ and $i_2\left(x\right)$. Recoving the disparity at x_o requires finding peaks in the real portion of $p\left(x_o, s\right)$ or zeros in the imaginary part. Since both are equivalent we can express the local correlation using only the real part as

$$p\left(x, s\right) = \sum A_{1i}(x)A_{2i}(x + s)\cos\left[\theta_{1i}\left(x\right) - \theta_{2i}\left(x + s\right)\right]. \qquad (9)$$

In general, the values $A_{1i}\left(x\right)$ and $A_{2i}\left(x\right)$ can be considered weighting coefficients for each cosine function. Local correlation bases these weights on the magnitude of the filtered responses for each frequency channel. Phase-correlation assumes all weights are equivalent. In our experiments we do not normalize the coefficients, as we view the contributions of the magnitude information as beneficial. However, it is important to note that our approach (outlined in a subsequent section) does not depend on any specific choice of weights. This is not to suggest that the choice of weights is unimportant, but only that the following descriptions are independent of their selection.

2.3 Discussion

Having completed our review, we now compare and contrast local correlation and phase-differencing. The most significant advantage of local correlation over phase-differencing is its ability to consider the entire frequency spectrum simultaneously. Since phase wrapping is no longer a problem (or at least it is mitigated), we can include high frequencies in the determination of large disparities. Unfortunately, this advantage come at the cost of dramatically increased computation time. Furthermore, a direct implementation would seem to require sampling $p\left(x, s\right)$, tying the resolution of the disparity estimates to the sampling density. Phase-differencing has no theoretical upper bound on its resolution.

3 Hybrid Approach

In this section we introduce a novel combination of phase-differencing and local correlation that retains the advantages of each, but without the normally attendant complications. We begin by incorporating the most important

aspect of phase-differencing, linear phase, into the local correlation model as follows:

$$p(x, s) = \sum A_{1i}(x) A_{2i}(x) \cos[\tilde{\omega}_i(x)(s - \delta_i)], \quad (10)$$

where $\delta_i = \frac{1}{\tilde{\omega}_i}[\theta_{2i}(x) - \theta_{1i}(x)]$ is the disparity found by phase-differencing. For a given point x_o, the function $p(x_o, s)$ becomes a simple sum of cosine waves. Equation (10) can be seen as a weighted sinusoidal voting for the best disparity, where the weights are the products of the magnitudes. This sinusoidal voting establishes a framework for reconciling disparities across multiple frequencies. If there is only one frequency channel then the maxima in s exactly correspond to the choice of disparities that would result from phase-differencing (in phase-differencing the maximum closest to zero is selected). As additional frequencies are added, the occurrence of global maxima becomes less and less frequent. Within the acceptable range of disparities, we hope to ultimately reduce this number to one. Furthermore, for a given $p(x_o, s_o)$, a confidence value (similar to that proposed in [4]) can be realized by normalizing the correlation by the sum of the weights over all frequency channels

$$c(x, s) = \frac{\sum A_{1i}(x) A_{2i}(x) \cos[\tilde{\omega}_i(x)(s - \delta_i)]}{\sum A_{1i}(x) A_{2i}(x)}. \quad (11)$$

When all the disparities δ_i coincide, the confidence reaches its maximum value of 1. As they begin to separate, the confidence degrades. Finally, it is important to note that (10) is only valid when phase is relatively linear with respect to translation. For frequencies channels outside regions of linearity, we would choose to exclude their respective cosine functions from (10) and (11). See [5] for a treatment of phase instability.

For an image point x_o, the goal is to find the value of s (within a predetermined range of disparities) that maximizes (10). Since $p(x_o, s)$ will generally have many local maxima, initially we will need to perform a global search. Once we are within the domain of convergence of the global maximum, we can proceed with a Newton iteration. First consider the iterative scheme implied by the following equations:

$$s^{(n+1)} = s^{(n)} + s_r^{(n)} \quad (12)$$

where, $s_r^{(n)}$ represents the residual disparity that maximizes

$$\begin{aligned} p\left(x_o, s^{(n)}\right) &= \sum A_{1i}(x_o) A_{2i}\left(x_o + s^{(n)}\right) \times \\ &\quad \cos\left(\tilde{\omega}_i^{(n)}\left[s_r^{(n)} - \delta_i^{(n)}\right]\right), \quad (13) \end{aligned}$$

and $\delta_i^{(n)}$ and $\tilde{\omega}_i^{(n)}$ represent each channel's disparity estimate and instantaneous frequency for the current iteration.

They are found as follows:

$$\delta_i^{(n)} = \frac{\theta_{2i}\left(x_o + s^{(n)}\right) - \theta_{1i}(x_o)}{\tilde{\omega}_i^{(n)}} \quad (14)$$

$$\tilde{\omega}_i^{(n)} = \frac{1}{2}\left[\frac{\partial}{\partial x}\theta_{2i}\left(x_o + s^{(n)}\right) - \frac{\partial}{\partial x}\theta_{1i}(x_o)\right] \quad (15)$$

The global search coarsely samples (13) and then locates the maximizing $s_r^{(n)}$. Once $s_r^{(n)} < \pi/\omega_N$, where ω_N is the highest center frequency, we can be reasonable sure that the algorithm has converged into a region in which even the highest frequency has only one maximum. We can now begin the high fidelity Newton iteration. (As a side note, if during the global search the instantaneous frequency $\tilde{\omega}_i^{(n)}$ in (13) and (14) is approximated by the center frequency ω_i of the quadrature filter, the cosines can be precomputed, greatly reducing the computational complexity. Our computer implementation uses this approximation.)

The Newton iteration requires the calculation of the derivative of (13). In order to derive an expression for this derivative we first rewrite (13) in the following more manageable form:

$$f(s) = \sum a_i \cos(\tilde{\omega}_i[s - \delta_i]). \quad (16)$$

We now determine its first derivative and then approximate it using a first order Taylor series in s

$$\begin{aligned} f'(s) &= \sum -a_i\tilde{\omega}_i \sin(\tilde{\omega}_i[s - \delta_i]) \\ &\approx \sum \left[a_i\tilde{\omega}_i \sin(\tilde{\omega}_i\delta_i) + \right.\\ &\qquad \left. a_i\tilde{\omega}_i^2 \cos(\tilde{\omega}_i\delta_i)\right]s. \quad (17) \end{aligned}$$

Without this Taylor expansion of $f'(s)$, the equation $f'(s) = 0$ is transcendental and must be solved iteratively. Using the first order approximation of $f'(s)$, we can easily solve for the s such that $f'(s) = 0$. This solution followed by further approximating the trigonometric functions by a first order Taylor series in δ_i we have

$$s = \frac{\sum a_i\tilde{\omega}_i \sin(\tilde{\omega}_i\delta_i)}{\sum a_i\tilde{\omega}_i^2 \cos(\tilde{\omega}_i\delta_i)} \approx \frac{\sum a_i\tilde{\omega}_i^2\delta_i}{\sum a_i\tilde{\omega}_i^2}. \quad (18)$$

We now have a new means for finding $s_r^{(n)}$

$$s_r^{(n)} = \frac{\sum A_{1i}(x_o) A_{2i}\left(x_o + s^{(n)}\right)\left[\tilde{\omega}_i^{(n)}\right]^2 \delta_i^{(n)}}{\sum A_{1i}(x_o) A_{2i}\left(x_o + s^{(n)}\right)\left[\tilde{\omega}_i^{(n)}\right]^2}. \quad (19)$$

Interestingly, (19) estimates the overall disparity by performing a weighted average of the disparities at each frequency channel. Since the individual disparity estimates are linear with respect to phase, so is their summation in (19).

Consequently, we can expect a convergence rate similar to that of typical phase-differencing.

Just as with phase-differencing, the hybrid approach uses a coarse to fine process. The coarseness is a function of the spatial extent of filters $h_i(x)$. Since filters with larger spatial extent are able to detect higher disparities, they first help bring the images into near registration before applying the next smaller filters.

4 Results

We implemented the previously discussed hybrid technique using $N = 20$ Gabor kernels at each scale in the coarse to fine process. At the coarsest scale for the pentagon images (see Figure 1) the center frequencies ω_i of each of the N kernels were spaced linearly, covering the range $[\pi/16, \ 15\pi/16]$. All filters had identical bandwidths of $\sigma_\omega = \pi/48$. Three levels were used, with each step increasing the resolution by a factor of two, i.e. for the second level the center frequencies covered $[\pi/8, \ 7\pi/8]$ and $\sigma_\omega = \pi/24$. The left image and resultant disparity map are shown in Figure 1. Other than the addition of one coarser scale, the baseball images were processed identically. Figure 2 illustrates the left baseball image and its corresponding disparity map.
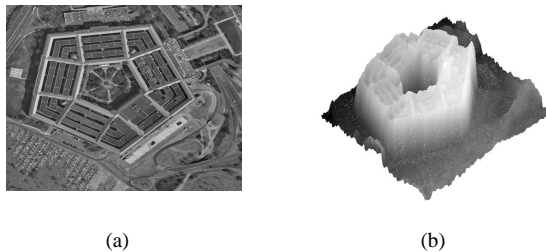

(a)   (b)

Figure 1. 1(a) Left image of pentagon. 1(b) Surface plot of disparity map.

5 Conclusion

We addressed the difficulty of multiscale fusion in phase-differencing by introducing a novel registration technique that incorporates strategies of both phase-differencing and local correlation. This hybrid approach allowed us to combine the rapidly calculated, high resolution results associated with phase-differencing with the multiscale information available to local correlation. Finally, we qualitatively demonstrated the performance on two pairs of stereo images.
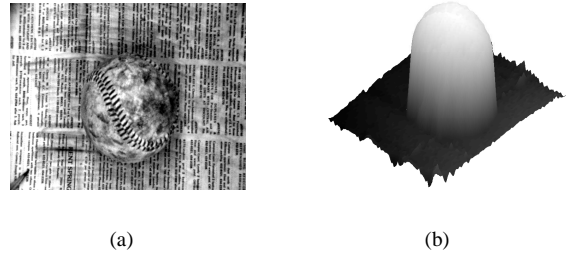

(a)   (b)

Figure 2. 2(a) Left image of baseball. 2(b) Surface plot of disparity map.

References

[1] T.-Y. Chen, A. C. Bovik, and L. K. Cormack. Stereoscopic ranging by matching image modulations. *IEEE Transactions on Image Processing*, 8(6):785–797, 1999.

[2] A. Cozzi, B. Crespi, F. Valentinotti, and F. Woergoetter. Performance of phase-based algorithms for disparity estimation. *Machine Vision and Applications*, 9(5-6):334–340, 1997.

[3] U. R. Dhond and J. Aggarwal. Structure from stereo - a review. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1489–1510, 1989.

[4] D. Fleet. Disparity from local weighted phase-correlation. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 1:48–54, 1994.

[5] D. Fleet and A. Jepson. Stability of phase information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(12):1253–1268, 1993.

[6] A. D. Jepson and M. Jenkin. Fast computation of disparity from phase differences. *Proceedings of the IEEE Computer Vision and Pattern Recognition*, pages 398–403, 1989.

[7] M. Ouali, D. Ziou, and C. Laurgeau. Cooperative multiscale phase-based disparity algorithm. *Proceedings of the IEEE International Conference on Image Processing*, 3:145–149, 1999.

[8] D. Papadimitriou and T. Dennis. Stereo disparity analysis using phase correlation. *Electronics Letters*, 30(18):1475–1477, 1994.

[9] T. Sanger. Stereo disparity computation using gabor filters. *Biological Cybernetics*, 59:405–418, 1988.

[10] J. Weng. Windowed fourier phase. completeness and signal reconstruction. *IEEE Transactions on Signal Processing*, 41(2):657–666, 1993.

[11] Y. Xu, L. Du, Z. Hou, and G. Jin. Scale-adaptive phase-based stereo matching method. *Tien Tzu Hsueh Pao/Acta Electronica Sinica*, 27(7):38–41, 1999.

[12] Y. Xu, J. Zhou, and G. Zhai. 2d phase-based matching in uncalibrated images. In *IEEE Workshop on Signal Processing Systems Design and Implementation*, pages 325–330, 2005.