

Disparity Using Feature Points in Multi Scale*

Ilkay Ulusoy¹, Edwin R. Hancock², and Ugur Halici¹

¹ Computer Vision and Artificial Neural Networks Lab.
Middle East Technical University, Ankara, Turkey

{ilkay}@metu.edu.tr

<http://vision1.eee.metu.edu.tr/~halici/>

² Department of Computer Science, University of York
York, YO1 5DD, UK

Abstract. In this paper we describe a statistical framework for binocular disparity estimation. We use a bank of Gabor filters to compute multiscale phase signatures at detected feature points. Using a von Mises distribution, we calculate correspondence probabilities for the feature points in different images using the phase differences at different scales. The disparity map is computed using the set of maximum likelihood correspondences.

1 Introduction and Motivation

For many species with frontally located eyes including humans, binocular disparity provides a powerful and highly quantitative cue to depth. For primates, it has been shown that different neurons in a number of visual cortical areas signal distinct ranges of binocular disparities [1,2,3,4]. This observation has led to the use of Gabor filters to model the phase differences for the receptive fields and to act as disparity decoders. However, although promising this Gabor model of complex cell responses has a number of shortcomings. First, a phase selective complex cell model can not uniquely signal a given retinal disparity. Second, they can not signal disparities beyond the quarter cycle limit of the input. Qian [12,13,14] has improved the complex cell model so that it can uniquely signal definite disparities. Furthermore, the experimental data of Anzai et. al. suggest that there may be a possibility of positional differences in disparity encoding [1]. Complex Gabor filters have also been used for finding disparity from the region-based phase differences between the left and right images [15]. Potential problems with the use of phase as a disparity encoder have been identified by Jenkin and Jepson [6,7,8]. If the stereo images are subjected to affine image deformations such as scaling or shifting with respect to one-another, at certain locations phase may not be stable through scale. Since there is extensive physiological and psychophysical evidence which indicates the frequency selectivity of cortical receptive fields, many algorithms incorporate spatial filters of multiple scale or size to model the shift in peak spatial frequency. For instance, Pollard

* This study is partially supported by TUBITAK BDP and METU AFP 2000.07.04.01

et. al. refine stereo correspondences by checking their behaviour through scale [11]. Sanger combines disparities at different scales using a weighting method [15]. Fleet simply sums the energy responses at different scales [5], Qian has a simple method which averages over different scales [12]. Marr et. al argue for a coarse to fine search procedure [10].

The observation underpinning this paper is that there is considerable scope for combining multiscale phase information to improve the estimation of disparity. Our approach is as follows: We commence from feature points detected using the method of Ludtke, Wilson and Hancock [9]. Next, a phase vector is calculated for each feature point. Correspondences are estimated using the similarity of orientation and phase at multiple scales. In this way we avoid the singular points encountered in the method of Jenkin and Jepson [6]. After calculating disparity from the positional difference between corresponding points, fine-tuning is performed using the phase difference information. This is done using a probabilistic model based on a von Mises distribution for the phase difference. The outline of the paper is follows. Extraction of features and their usage is explained in Section 2. In Section 3 we discuss the use of multiple scales for correspondence. The probabilistic phase difference model is explained in Section 4. In section 5 the results are discussed.

2 Extraction of Features Used in the Correspondence Algorithm

Gabor filters are well known models of simple cells:

$$g_{cos}(x, y, \omega_0, \theta) = \exp[-(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2})] \cos[2\pi\omega_0(x\cos(\theta) + y\sin(\theta))] \quad (1)$$

$$g_{sin}(x, y, \omega_0, \theta) = \exp[-(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2})] \sin[2\pi\omega_0(x\cos(\theta) + y\sin(\theta))] \quad (2)$$

where σ_x, σ_y express width of 2D Gaussian envelope along x and y direction, ω_0 is the spatial frequency and θ gives the orientation in space. Experiments show that adjacent simple cells have the same orientation and spatial frequency, but are in quadrature pairs (i.e. they differ in spatial phase by 90°) [4]. Thus a simple cell pair can be expressed by a complex Gabor filter:

$$g(x, y, \omega_0, \theta) = \exp\left\{-[\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}] + i2\pi\omega_0[x\cos\theta + y\sin\theta]\right\} \quad (3)$$

In this paper we use a bank of 8 complex Gabor filters of different orientation. From the output of the filter-bank, we compute a population vector [9]:

$$\mathbf{p}(x, y) = \begin{bmatrix} p_x(x, y) \\ p_y(x, y) \end{bmatrix} = \sum_{i=1}^n G(x, y, \omega_0, \theta_i) \mathbf{e}_i \quad (4)$$

where (x,y) is the position of the pixel in the image, n is the number of different orientation states, $G(x, y, \omega_0, \theta_i)$ is the response (energy) of a quadrature pair of Gabor filters with orientation θ_i and $\mathbf{e}_i = (\cos \theta_i, \sin \theta_i)^T$ is the unit vector in the direction θ_i . Here, the population vector is the vector sum of the $n=8$ filter response vectors and the resultant orientation is given by $\theta_{pop}(x, y) = \arctan[p_y(x, y)/p_x(x, y)]$. When compared to the tuning width of a single Gabor filter, the orientation estimate returned by the population vector is very accurate even though a relatively limited number of filters is used.

In our study, the feature points used for correspondence analysis are the locations where the length of population vector is locally maximum (see [9] for details). These points are located on object boundaries. In Figure 1a,b we show stereo images with numbered feature points on right image. Figure 1c,d shows the feature points from the images with the estimated orientation encoded as a grey-level.

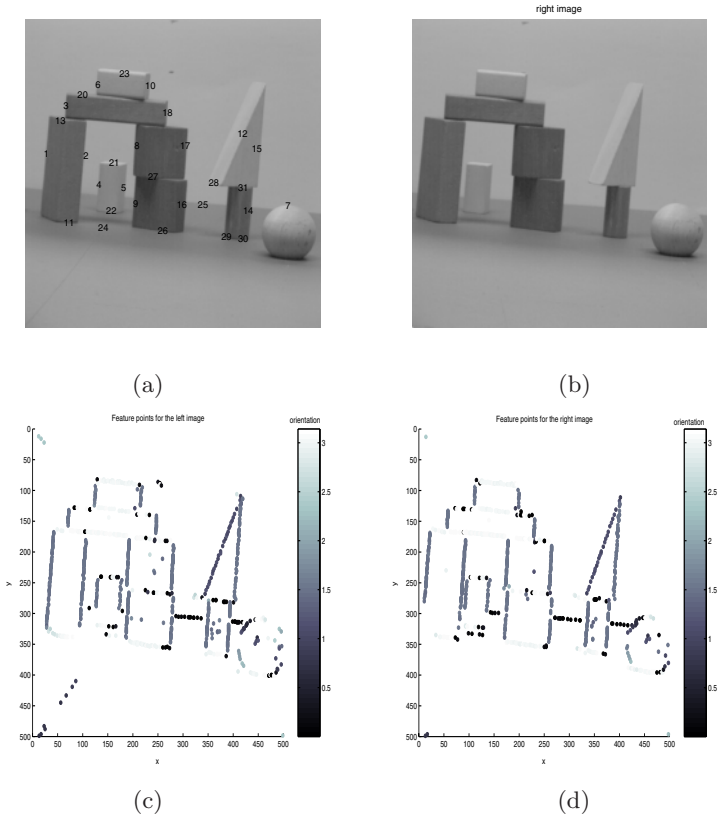


Fig. 1. (a) Right image of the stereo pair. (b) Left image of the stereo pair. (c) Feature points for right image. (d) Feature points for left image

3 Finding Corresponding Pairs and Disparity Using Multi-phase

The attributes used for the correspondence matching of feature points are orientation and phase. It is well known that phase based methods for disparity estimation are successful except in the neighbourhood of singularities [6]. In particular phase is stable with respect to geometric deformations and contrast variations between the left and right stereo views. In this paper, disparity is estimated from the region-based phase differences between the left and right images. Our estimate is obtained by first filtering the raw image data with a complex Gabor filter and computing the quantity

$$\phi_w(x, y) = \arctan \left[\frac{G_{sin}(x, y, w_0, \theta)}{G_{cos}(x, y, w_0, \theta)} \right] \quad (5)$$

where $G_{cos}(x, y, w_0, \theta)$ and $G_{sin}(x, y, w_0, \theta)$ are the cosine-phase and sine-phase filter responses of the image. We use the phase measurements for Gabor filters of different width, i.e. different scales, to locate correspondences. We use three filters each separated by one octave. The width of the narrowest filter is 6 pixels. For each feature point at the right image, we search over a window for feature points of similar orientation and phase in the left image. Let $\Phi_i = (\phi_1, \phi_2, \phi_3)^T$ be a vector of phase estimates obtained using the three filters. We measure the similarity of the phase-vectors by weighting the different components using the method described by Sanger [15]. Let C be the weighting matrix. The candidate j which has the closest weighted phase to the feature point i is the one that satisfies the condition $j = \arg \{ \min \{ \Phi_i C^{-1} \Phi_j^T \} \}$. The disparity is the distance between corresponding feature points. In performing this, position shift between the receptive fields of binocular disparity selective cells are mimicked [4].

The matching algorithm explained above is cross checked for left-right correspondences and right-left correspondences. In this way we may discard occluded feature points. For the stereo shown in Figure 1a,b we find correspondences for 537 of the 980 feature points in the right-hand image (Figure 1b). The final disparity values are displayed as gray scale values in Figure 2a and height plot in Figure 2b. Also in Figure 3 three main depth layers are shown separately. Out of the 537 matched feature points only 62 are in error, hence the success rate is 90%. Most of the errors are for feature points having a population vector orientation in the disparity direction. In order to obtain subpixel accuracy, a phase shift model of binocular cell receptive fields can be used [4]. Here, the subpixel disparity is calculated from the interocular phase differences for between corresponding points using the quantity $\Delta d = \frac{\phi_{ij}\lambda}{2\pi}$, where Δd is the fine tuning in disparity, $\phi_{ij} = \phi_i - \phi_j$ is the measured phase difference, i and j are the left and the right feature point identities respectively. In this way, the rough disparity estimate found by using only the position shift model is tuned by the phase shift model. As an example, the rough disparities on the edge-segment numbered 12 in Figure 1a shows a stair shaped structure (see Figure 2c top plot (*)). After fine tuning, the disparity varies more smoothly (see Figure 2c top plot (line)).

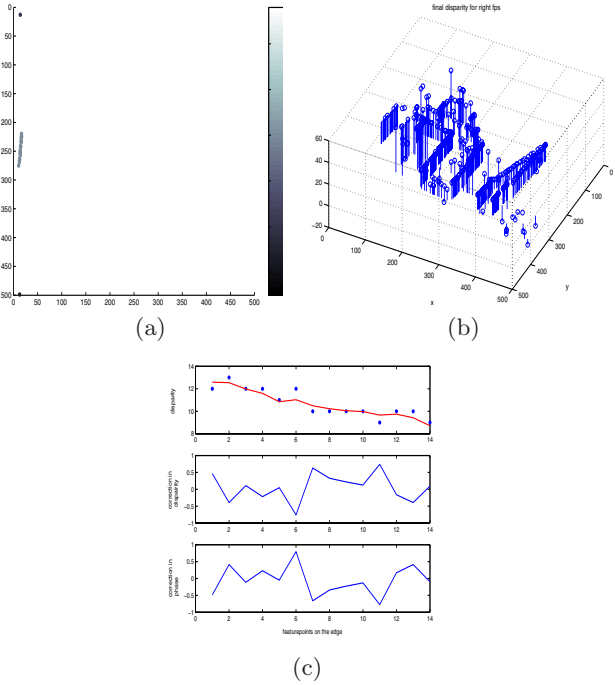


Fig. 2. (a),(b) Disparity. (c) Fine tuning result. Top: Coarse disparity (*), and fine disparity (line). Middle: Subpixel disparity. Bottom: Phase difference

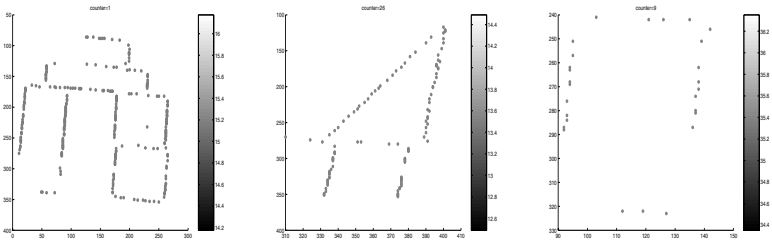


Fig. 3. Different depth layers

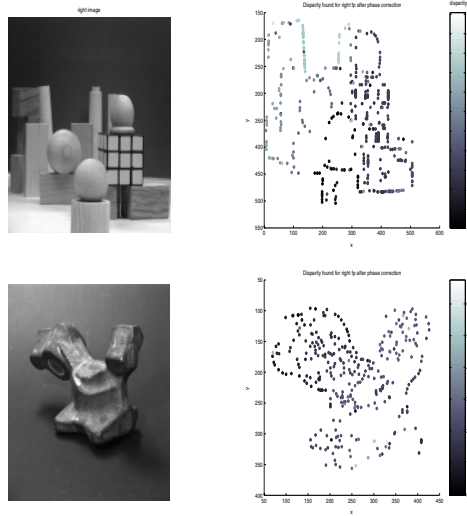


Fig. 4. Left: right images; Right: disparities

Disparity results for other image pairs are shown in Figure 4. Although the shapes in the images have very different characteristics, the results are still satisfactory.

4 Probabilistic Model of the Disparity Algorithm

After finding correspondences and computing the associated disparities, we refine the correspondences using a probabilistic model for the distribution of phase differences. This model is based on the assumption that the measured phase follows the von Mises distribution:

$$p(\phi_{ij}|\kappa, \mu) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(\phi_{ij} - \mu)] \tag{6}$$

where the distribution width or standard deviation is κ , the mean is μ and I_0 is the zero order Bessel function. For each scale, we fit a mixture of von Mises distributions to the measured phase differences. We use the EM algorithm to estimate the parameters of the mixture components κ_w and μ_w . At iteration $n + 1$ of the algorithm the expected log likelihood functions for the estimation process is

$$Q = \sum_{i=1}^N \sum_{j=1}^N \sum_{w=1}^W P(w|\phi_{i,j}, \kappa_w^{(n)}, \mu_w^{(n)}) \ln p(\phi_{ij}|\kappa_w^{(n+1)}, \mu_w^{(n+1)}) P(w) \tag{7}$$

where N is the total number of phase difference measurements, and W is the total number of von Mises distributions in the mixture model. In the E or expectation

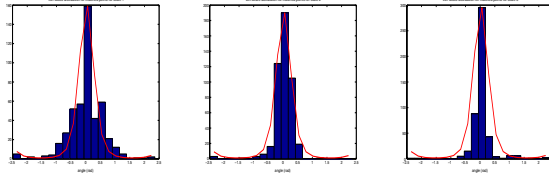


Fig. 5. Von Mises distributions fitted at three of the scales

step we compute the updated a posteriori probabilities

$$P(w|\phi_{i,j}, \kappa_w^{(n)}, \mu_w^{(n)}) = P_{i,j}^{(n)}(w) = 1/N \sum_{ij}^N p(\phi_{i,j}|\kappa_w^{(n)}, \mu_w^{(n)}) \tag{8}$$

In the M-step, the distribution means are given by

$$\mu_w^{(n+1)} = 1/2 \arctan \left[\frac{\sum_{i,j}^N P_{i,j}^{(n)}(w) \sin(2\phi_{ij})}{\sum_{i,j}^N P_{i,j}^{(n)}(w) \cos(2\phi_{ij})} \right] \tag{9}$$

The distribution widths are more difficult to obtain, and involve computing the quantity

$$R = \frac{I_1(\kappa_w^{(n+1)})}{I_0(\kappa_w^{(n+1)})} = \frac{\sum_{ij}^N p(\kappa_w^{(n)}, \mu_w^{(n)}|\phi_{ij}) \cos(2(\phi_{ij} - \mu_w^{(n)}))}{p(\kappa_w^{(n)}, \mu_w^{(n)}|\phi_{ij})} \tag{10}$$

For small values of R $\kappa_w^{(n+1)} \simeq (1/6)R(12 + 6R^2 + 5R^4)$ while when R is large $\kappa_w^{(n+1)} \simeq 1/(2(1 - R) - (1 - R^2) - (1 - R^3))$. The result of fitting the von Mises mixture at different scales is shown in Figure 5.

With the parameters of the mixture model to hand, we can estimate correspondence probabilities from the phase differences. The correspondence probabilities are taken to be a posteriori probability of the mixture with the smallest mean μ_{min} at convergence of the EM algorithm. Suppose that $S_{i,j}^s$ is the a posteriori correspondence probability for scale s . The overall correspondence probability is the product of correspondence probabilities computed at the different scales, i.e. $q_{i,j} = \prod_{s=1}^3 S_{i,j}^s$. The correspondences are taken so as to maximise $q_{i,j}$.

Applying the correspondences located in this way the computed disparities were very similar to those found using the method described in the previous section. The main differences are at horizontal edges as can be seen in the Figure 6.

5 Conclusion

We have presented a stereo correspondence method which is motivated by physiological and biological information. To do this we have modelled visual cortex

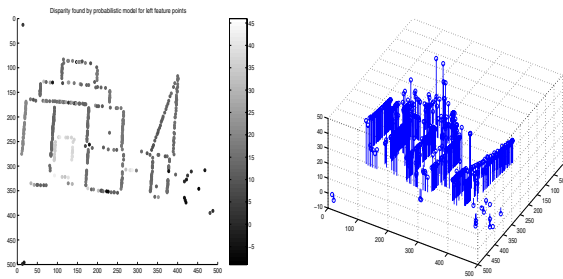


Fig. 6. Disparity found by probabilistic model

cell receptive fields using Gabor functions. Hypercolumns are encoded using population vectors. Thus, instead of calculating disparities using oriented Gabor filters and pooling the results over different orientations, a single orientation for each feature is obtained prior to disparity computation. The population vector estimate of stimulus orientation found using this method is very accurate given the small number of filters used. Although the feature points are sparse, since they are the points of high contrast edges that define the bounding contours of objects, they still prove to be informative. Correspondences between similarly oriented feature points are located using the phase information. This idea is also biologically grounded. The reason for this is that simple binocular cells occur in pairs that are in quadrature phase. Also, phase is sensitive to spatial differences, and hence it provides fine image detail which is helpful in discriminating neighbouring image regions. Phase is also robust to small scale differences. Unfortunately, there are image locations where phase is singular and can not be reliably used. In this study, by performing phase comparisons at multiple scales and by using confidence information we overcome these difficulties. We use the confidence weighting to augment phase information with information concerning the magnitude of the population vector to improve the correspondence method. Our use of multiple scales is also biologically plausible. The reason for this is that disparity encoding binocular cells are sensitive to different spatial wavelengths. We explore two routes to locating feature-point correspondences. Using the position shift model, rough disparity values are obtained and a large range of disparities can be calculated, but to a limited accuracy. Using the phase shift model, fine tuning is performed without encountering the quarter cycle limit. This tuning scheme also allows a continuum of disparity estimates to be obtained. The algorithm proves to be effective for textureless images, especially at depth boundaries. The next step is to use the computed disparity values for surface reconstruction.

References

1. Anzai, A., Ohzawa, I., Freeman, R. D.: Neural mechanisms for encoding binocular disparity: Receptive field position vs. phase. *Journal of Neurophysiology*, vol. 82, no. 2, pp. 874-890, 1999. [320](#)
2. Anzai, A., Ohzawa, I., Freeman, R. D.: Neural mechanisms for processing binocular information I. Simple cells. *Journal of Neurophysiology*, vol. 82, no. 2, pp. 891-908, 1999. [320](#)
3. Anzai, A., Ohzawa, I., Freeman, R. D.: Neural mechanisms for processing binocular information II. Complex cells. *Journal of Neurophysiology*, vol. 82, no. 2, pp. 909-924, 1999. [320](#)
4. DeAngelis, G.: Seeing in three dimension: the neurophysiology of stereopsis. *Trends in Cognitive Science*, vol. 4, no. 3, pp. 80-89, 2000. [320](#), [321](#), [323](#)
5. Fleet, D. J., Wagner, H., Heeger, D. J.: Neural encoding of binocular disparity: Energy models, position shifts and phase shifts. *Vision Research*, vol. 36, no. 12, pp. 1839-1857, 1996. [321](#)
6. Jenkin, M. R. M., Jepson, A. D.: Recovering local surface structure through local phase difference measurements. *CVGIP: Image Understanding*, vol. 59, no. 1, pp. 72-93, 1994. [320](#), [321](#), [323](#)
7. Jepson, A. D., Fleet, D. J.: Scale space singularities. *Lecture Notes in Computer Science*, vol. 427, pp. 50-55, 1990. [320](#)
8. Jepson, A. D., Fleet, D. J.: Phase singularities in scale space. *Image and Vision Computing*, vol. 9, no. 5, pp. 338-343, 1991. [320](#)
9. Ludtke, N., Wilson, R. C., Hancock, E. R.: Tangent files from population coding. *Lecture Notes in Computer Science*, vol. 1811, pp. 584-593, 2000. [321](#), [322](#)
10. Marr, D., Poggio, T.: A computational theory of human stereo vision. *Proceedings of the Royal Society of London*, B207, pp. 187-217, 1979. [321](#)
11. Pollard, S. B., Mayhew, J. E. W., Frisby, J. P.: PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, vol. 14, pp. 449-470, 1985. [321](#)
12. Qian, N.: Computing stereo disparity and motion with known binocular cell properties. *Neural Computation*, vol. 6, no. 3, pp. 390-404, 1994. [320](#), [321](#)
13. Qian, N., Zhu, Y.: Physiological computation of binocular disparity. *Vision Research*, vol. 37, no. 13, pp. 1811-1827, 1997. [320](#)
14. Qian, N.: Relationship between phase and Energy methods for disparity computation. *Neural Computation*, 12, pp. 279-292, 2000. [320](#)
15. Sanger, T. D.: Stereo disparity computation using Gabor filters. *Biol. Cybern.*, 59, pp. 405-418, 1988. [320](#), [321](#), [323](#)