行政院國家科學委員會專題研究計畫 成果報告

智慧型視差監控攝影機之嵌入系統
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 95-3113-P-002-001-
執 行 期 間 ：95 年 02 月 01 日至 96 年 01 月 31 日
執 行 單 位 ：國立臺灣大學電機工程學系暨研究所

計 畫 主 持 人：林巍聳

計畫參與人員：博士班研究生-兼任助理：方俊雄
碩士班研究生-兼任助理：施宇駿、劉培德
碩士畢業生：林建同

處 理 方 式 ：本計畫涉及專利或其他智慧財產權，1 年後可公開查詢

中 華 民 國 96 年 02 月 13 日

智慧型視差監控攝影機之嵌入系統
Embedded System of Intelligent Disparity Estimation Camera

計畫編號：NSC-95-3113-P-002-001
執行期間：2006年2月1日 至 2007年1月31日
執行單位：國立台灣大學電機研究所
計畫主持人：林巍聳* 教授
計畫參與人員：林建同 施宇駿 劉培德 方俊雄
* Corresponding author; E-mail: weisong@cc.ee.ntu.edu.tw; No. 1, Sec. 4, Roosevelt Rd. Taipei 106, Taiwan

Abstract- Dense disparity map extracted from stereo images is three-dimensional information of the scene that can be used to locate objects and measure their motion parameters. Potential applications are obstacle detection and navigation of mobile robots. Correlation-based correspondence is a common approach of disparity measurement. But computation of the correlation measure and the subsequent maximization of this value turn out to be a computationally expensive process. Disparity map extraction by measuring the local phase shift is attractive mainly because of its efficiency and applicability to theories of stereopsis in the human visual cortex. But the working range of the phase shift approach is limited by the wavelength of the interested filter. Wherever the working range is smaller than real disparity, the result is distorted by the aliasing effects. In this paper, hierarchical phase-shift coherence measurement (HPCM) is proposed to obtain large working range so as to avoid the aliasing effects. Basically, HPCM measures the disparity in multi-resolution image scales. The disparity measurement in a fine resolution scale is on the base of the disparity maps obtained in less fine resolution scales. The results obtained in every scale are piled up to obtain the ultimate disparities. Experimental results show that HPCM outperforms other phase-based approaches.

Keywords- intelligent camera, computational stereo, local image phase, stereo vision, disparity map

1. Introduction

Intelligent stereo camera capable of extracting three-dimensional structure of a scene from stereo images has found applications in obstacle detection and autonomous navigation of mobile robots [1], [2]. Its computational effort pays mainly for searching for the image correspondences and reconstructing the three-dimensional pre-image. The solution to this computational stereo has been studied by the computer vision community for decades. Currently, there is correlation-, feature-, and phase-based approaches available [3]. Correlation-based correspondence minimizes the deviation or maximizes the correlation between patches of the left and right images [4]. The computation of the correlation measure and the subsequent maximization of this value turn out to be a computationally expensive process. In general, correlation-based approach encounters problems with perspective projection, inter-ocular illumination differences and images that contain too much or too little structure [5]. Once the corresponding points are determined, as shown in figure 1, the pre-image can be found at the intersection of the rays passing through these points and the associated pinholes. Then the depth of the pre-image can be calculated with the triangulation. Disparity which is the difference between the coordinates of the corresponding points is a sense of depth. While calculating the depth needs to know the baseline and focal length of the stereo camera, calculating the disparity needs no knowledge of the camera parameters. Thus applications of the depth map where the camera parameters are not available may invoke the disparity map as an alternative.

Feature-based approach converts tremendous intensity data into a set of features which are assumed to be a more stable image property than the raw intensities. The matching stage operates only on these extracted image features. Since only a few specific feature classes are utilized, most areas of the stereo images that fall into the no-feature-present class are ignored further in the matching process. This makes it impossible to obtain the dense disparity map [6].

Phase-based approach takes advantage of the fact that the disparity from band-pass signals is equivalent to the local phase difference between them [7], [8]. This approach is attractive mainly because of its parallel efficiency and applicability to theories of stereopsis in the human visual cortex [9]. Sanger [10] pioneered measuring disparities with phase differences. This simple algorithm is called the phase-shift measurement (PM) in the following context. The working range of PM is limited by the wavelength of the interested filter, usually a Gabor filter. Surely using larger wavelength may enlarge the working range of PM; unfortunately this may result in inaccurate measurement. Phase-shift coherence measurement (PCM) [6] was shown an effective method to improve the accuracy of PM. The change is that the phase shift is measured with a stack of filters and only the coherent filters are taken into account in determining the disparity. Nevertheless PCM has working range limited by the wavelength of the filter stack and still suffers from the aliasing effects.

In this paper, the hierarchical phase-shift coherence measurement (HPCM) measures disparities in multi-resolution image scales instead of single resolution one. Beginning with the lowest resolution scale, the disparity map of a resolution scale is used to shift the left image of the next finer resolution scale so that the modified finer image pair has all disparities fall within the working range of the filter stack. Since disparity measurement in a fine resolution scale is based on the results of the coarse resolution scales, the aliasing effects are avoided. The disparities obtained in every scale are piled up to obtain the ultimate disparity. This hierarchical approach can extend the working range of PCM or PM to meet the needs in real application environments.

The rest of this paper is organized as follows: Section 2 formulates the hierarchical phase-shift coherence measurement; Section 3 shows two experiments to compare the performance of HPCM with that of PM, PCM, and hierarchical phase-shift measurement (HPM); Section 4 is the conclusion.

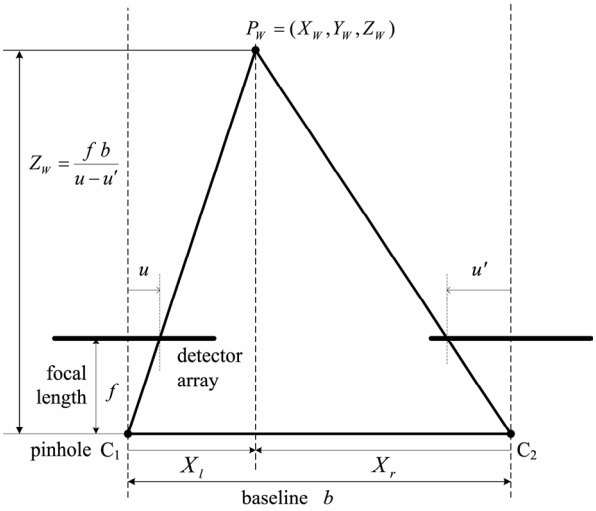


Fig. 1 Image correspondence and depth reconstruction

2. Hierarchical phase-shift coherence measurement

HPCM aims at using phase-based approach to extract dense disparity map from stereo image pair. The technical bases are firstly the same pre-image appearing in low resolution scale image has phase shift smaller than that of in high resolution scale image; and secondly, the disparity from band-pass signals is equivalent to the local phase shift between them.

2.1 Phase-shift measurement

Phase, as a function of space, is defined as the complex argument of a complex-valued band-pass signal. The band-pass signal is usually generated by the spatial Gabor filter [11]. The spatial Gabor filter has both limited spatial width and finite bandwidth, and whose shape is similar to the receptive field profile of simple cells in primate visual cortex. Although many other filters can be used, the Gabor filters allow easy separation of the modulating component which determines the spatial frequency, and the envelope which determines the bandwidth.

The spatial Gabor filters take the form

$$G(x-x_0) = e^{-\frac{(x-x_0)^2}{2\sigma^2}} e^{i\omega_0(x-x_0)} \tag{1}$$

where x_0 is the spatial location of the filter, σ is the spatial half-width of the filter, and ω_0 is the frequency of the harmonic component which will be the center spatial frequency of the power spectrum. These spatial Gabor filters are seen to be the product of a complex harmonic function and a Gaussian envelope. The Fourier transform of (1) has the same functional form

$$g(\omega-\omega_0) = e^{-\frac{(\omega-\omega_0)^2}{2\tau^2}} e^{-ix_0(\omega-\omega_0)} \tag{2}$$

where τ is the spatial-frequency half-bandwidth of the filter, and $\sigma\tau = 1$.

If there is a constant disparity across some part of the scene then in that region we have $l(x-\Delta x) = r(x)$, where l and r represent left and right images, respectively, and Δx is the disparity. Because the Gabor filter is complex, so is the response to a given image. For any point (x_0, y_0) and particular Gabor filter with center frequency ω_0, we obtain the convolution products

$$\begin{aligned} H_R(x_0,y_0) &= \int r(x,y_0)G(x_0-x)dx \\ &= \int l(x-\Delta x,y_0)G(x_0-x)dx \\ &= P_R + iQ_R \end{aligned} \tag{3}$$

$$\begin{aligned} H_L(x_0,y_0) &= \int l(x,y_0)G(x_0-x)dx \\ &= P_L + iQ_L \end{aligned} \tag{4}$$

where P and Q denote the real and imaginary parts, respectively. Then the complex phase difference is [12]

$$\begin{aligned} \Delta\phi(x_0,y_0) &= \angle H_L(x_0,y_0) - \angle H_R(x_0,y_0) \\ &= \tan^{-1}\left(\frac{Q_L P_R - P_L Q_R}{P_L P_R + Q_L Q_R}\right) \end{aligned} \tag{5}$$

Since the disparity from band-pass signals is equivalent to the local phase shift between them [10], the disparity $\Delta x(x_0, y_0)$ is calculated as

$$\Delta x(x_0,y_0) = \Delta\phi(x_0,y_0)/\omega_0 \tag{6}$$

Considering the inherent uncertainties in the instrumentation and data, (6) is usually evaluated for a stack of several Gabor filters with slightly different center frequencies and all responding to a common view direction. The combination of disparities across spatial scales is performed by a simple weighted average, with the weighting being provided by the confidence values. The weighted-average disparity d is calculated as

$$d = \frac{\sum_{\omega} \Delta x_\omega r_1(\omega)}{\sum_{\omega} r_1(\omega)} \tag{7}$$

where $r_1(\omega)$ is the confidence value at each center spatial frequency ω, and Δx_ω is the computed disparity for each center frequency. When the focal length f and baseline b of the stereo camera are known, the depth Z_W is calculated as

$$Z_W = \frac{f\,b}{u-u'} = \frac{f\,b}{d} \tag{8}$$

where u and u' are defined in figure 1.

Since the amplitudes of the Gabor responses $|H_R|$ and $|H_L|$ are not used for disparity calculation, they can be used as an independent measure of confidence. If there is insufficient amplitude for a particular filter in one or both images, then the phase information can be considered meaningless, and a confidence of zero is assigned. On the other hand, if there is sufficient amplitude in both images of the pair, then a comparison of the values of $|H_R|$ and $|H_L|$ can be made. We would expect that if one image were a uniform shift of the other, then the Gabor amplitudes would be similar. In

contrary, if the Gabor amplitudes are very different, this may indicate the presence of noise in the image, or a disparity, which is beyond the measurement limit of a particular filter. Hence, in (7), the confidence value at each center frequency is calculated as [10]

$$r_1(\omega) = \min\left[\frac{|H_R|}{|H_L|}, \frac{|H_L|}{|H_R|}\right] \tag{9}$$

2.2 Phase-shift coherence measurement

PM can only account for disparities smaller than a half of the wavelength of the filter being used. That means its working range is

$$d < \frac{1}{2}\lambda = \frac{\pi}{\omega_0} \tag{10}$$

Out of this working range, random aliasing effects may occur depending in a complicated way on image content and working range of the Gabor filter in question. Therefore, direct application of (7) suffers from the random aliasing effects and the results may be blurring or even distorted. PCM employs a coherence mechanism to exclude the aliasing units as

$$d = \left(\frac{\sum_{\omega} \Delta x_{\omega} r_1(\omega)}{\sum_{\omega} r_1(\omega)}\right)_{\omega \in C} \tag{11}$$

where C denotes the largest cluster of coherent filters. A cluster of coherent filters satisfies $\left|\Delta x_{\omega_i} - \Delta x_{\omega_j}\right| \le \varepsilon \quad \forall i, j$, where ε is a nonnegative constant. Figure 2 compares the clusters used in PCM calculation (11) with that of PM calculation (7). Figure 3 shows the computational paradigm of PCM. L and R represent the left and right scan-lines and d is the measured disparity. Labels gab1~gab6 represent the Gabor filters with different center frequencies. In this example, a filter stack which is embraced by the rectangle consists of six Gabor filters to calculate the disparity of a single image point. The coherence detection presented in [6] is similar to PCM. Except the filter stack was constructed by leading a Gabor filter with random pre-filtering to its input data.

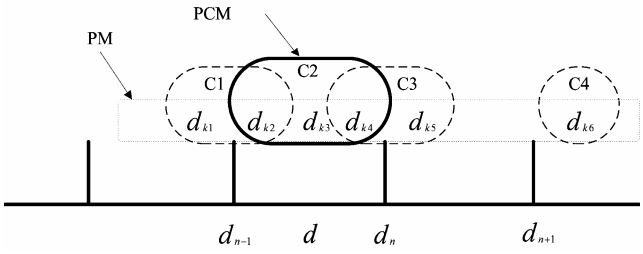


Fig. 2 Clusters considered in PM and PCM

2.3 HPCM and HPM

Depending on whether (11) or (7) are used, the hierarchical disparity measurement is called HPCM or HPM. Figure 4 shows the computational paradigm of HPCM and HPM. It includes mainly four procedures:

1) *Reduce*: Reduce the stereo image pair into multi-resolution scales; then beginning with the lowest resolution scale do Step 2 ~ Step 4.
2) *Measure*: Measure the disparity with PCM or PM to construct the disparity map
3) *Expand*: Expand the disparity map to the next finer resolution scale
4) *Shift*: Shift the left image of the next finer scale by according to the expanded disparity map.

The above procedures are repeated from Step 2 to Step 4 until the finest resolution scale, i.e. the ground truth image scale, is reached at Step 2. The disparity maps obtained in every resolution scale are piled up to obtain the ultimate disparity map.

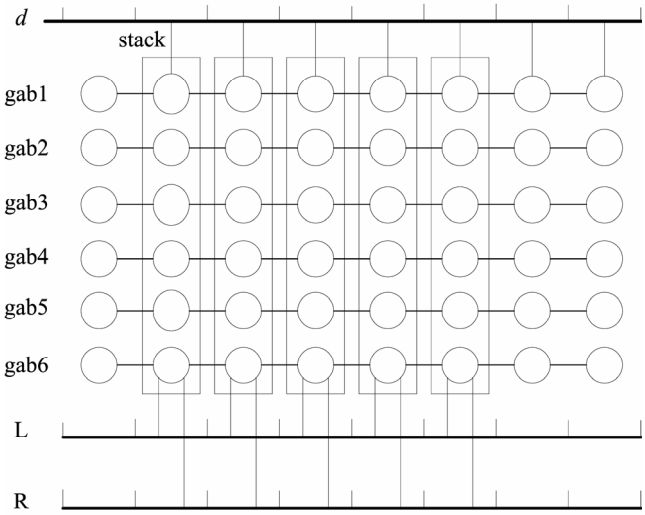


Fig. 3 The computational paradigm of PCM

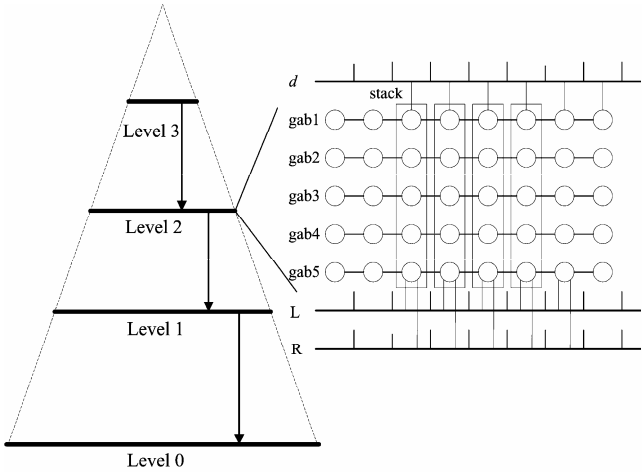


Fig. 4 The computational paradigm of HPCM and HPM

Reducing and expanding an image or disparity map are simple. Suppose the initial image is represented by the array V_0 which contains η_0 columns and γ_0 rows of pixels. This image becomes level 0 of the Gaussian pyramid. Level 1 of the pyramid is the image V_1. Each value in V_1 is computed as a weighted average of values in V_0 within an M by N window. Similarly, each value in V_2 is obtained from values in V_1 by applying the same pattern of weights and so on. The level-to-level averaging process is expressed as [13]

$$V_l(i,j)=\sum_{m=-\frac{M-1}{2}}^{\frac{M-1}{2}}\sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w(m,n)V_{l-1}(2i+m,2j+n) \quad (12)$$

where l represents the l-th level, w represents the weighted window, (i, j) is the point of the image in level l, $0 < i < \eta_l$, $0 < j < \gamma_l$, η_l and γ_l are the size of the image in level l. Similarly, an equation of expanding the image is expressed as

$$V_{l-1}(i,j)=4\sum_{m=-\frac{M-1}{2}}^{\frac{M-1}{2}}\sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w(m,n)V_l\left(\frac{i-m}{2},\frac{j-n}{2}\right) \quad (13)$$

Only terms for which $(i-m)/2$ and $(j-n)/2$ are integers are accounted in this equation.

The number of resolution scales needed in HPCM and HPM is dependent of the peak disparity appearing in the ground truth stereo images and the wavelength of the filter stack. Let the peak disparity appearing in the images be D. Then in the highest sampling level n, the peak disparity is

$$d_n=\frac{D}{2^{n-1}} \quad (14)$$

To avoid the aliasing effects, the criterion is that d_n must fulfill (10) for the majority of the Gabor filters of PCM, or for all the Gabor filters of PM.

3. Experimental results

The results of two experiments are shown in this section. Experiment 1 compares the disparity map extraction of HPCM with that of PM, PCM, and HPM. In experiment 2, HPCM, PM, PCM, and HPM are compared by measuring the depth. The filter stack is composed of six Gabor filters with 5, 6, 7, 8, 9, 10 pixels in the center wavelength, respectively.

Experiment 1: The ground truth image pair is shown in figure 4 [14]. It consists of a sculpture near the camera and a bookshelf far from the camera. The sculpture is so close to the camera that their visual disparities are out of the working range of the filter stack. Using PM and PCM to measure the disparity, the resulting disparity maps are shown in figures 5(a) and 5(b), respectively. In these figures, light grey and dark grey represent large and small disparities, respectively. Since the sculpture is very close to the camera, the corresponding regions in figures 5(a) and 5(b) should be relatively very bright. But it is found there is no significant difference against the region corresponding to the bookshelf. That means the measured disparities for the sculpture are wrong. The cause is that, according to (10), the aliasing effect makes PM and PCM treat all the disparities smaller than the working range of the filter stack. Using the resulting disparity maps to shift the original left image, the results are shown in figure 6(a) and 6(b). If the measured disparities are absolutely correct, the shifted left image should be identical to the ground truth right image as shown in figure 4(b). But in this case, the results are far from correctness.

Figures 5(c) ~ 5(h) compares the disparity maps obtained by HPM and HPCM for 2, 4, and 5 resolution scales, respectively. The notations for them are HPM-2 ~ HPM-5 and HPCM-2 ~ HPCM-5. The corresponding shifted left images are shown in figures 6(c) ~ 6(h). The peak disparities can be measured correctly by HPM-2, HPM-4, and HPM-5 are 5, 20, and 40 pixels (center wavelength 5 pixels); and 8, 32, and 64 pixels (center wavelength 8 pixels) by HPCM-2, HPCM-4, and HPCM-5, respectively. The real disparities of the sculpture fall around 40 pixels. Therefore, around the sculpture, both HPM-2 and HPCM-2 suffer from the aliasing effects. Figures 5(c) and 5(d) show the disparities are not correct and figures 6(c) and 6(d) show the left images around the sculpture distort significantly due to the incorrect shift.

The working range of HPM-4 and HPCM-4 almost cover the whole image except that around the nose. Therefore, it is seen in figures 5(e) and 5(f) the disparity map corresponding to the sculpture is bright relative to the background. Figures 6(e) and 6(f) show the left images are shifted nearly nice.

The disparities of the whole image fall within the working range of HPM-5 and HPCM-5. Hence, figures 5(g) and 6(h) show the sculpture emerges significantly from the background in the disparity maps. Figures 6(g) and 6(h) show the shifted left images match the ground truth right image nicely.

To discriminate the ability of HPM-4, HPCM-4, HPM-5, and HPCM-5, define the root-mean-square error (RMS) of the shifted left image against the ground truth right image as

$$RMS=\sqrt{\frac{\sum\left(l(x)-r(x)\right)^2}{N}} \quad (15)$$

where $l(x)$ and $r(x)$ denotes the intensities of pixel x, and N is the total number of pixels. The RMS associated with figures 6(e), 6(f), 6(g), and 6(h) are tabulated in table 1.

Table 1. RMS associated with figures 6(e), 6(f), 6(g), and 6(h)

RMS	HPM	HPCM
2 scales	N.A.	N.A.
4 scales	11.96	10.13
5 scales	10.78	9.54

It is found that using five resolution scales is better than using four scales, and furthermore HPCM outperforms HPM. But the RMS values show HPCM-5 is still not perfect. In addition to some data missing due to occlusions, the main cause is that some detailed information is disappeared in low resolution scale. This may result in the shift procedure simply takes care of the coarse and leaves the detail behind.

Experiment 2: PM, PCM, HPM-5 and HPCM-5 are compared in measuring depths. The stereo image pair for test is shown in figure 7. It contains a box and a cylinder at 140 cm and 100cm apart from the camera, respectively. The disparity maps obtained with HPM-5 and HPCM-5 are presented in figure 8. The disparity curves crossing the box and the cylinder (the 125th row) are shown in figure 9(a) and 9(b), respectively. Since the front surfaces of the box and the cylinder are flat and arc, respectively; the disparity curves should present these features. But it is seen the resulting disparity curves do not faithfully display this fact. One of the causes is that the correspondence between the left and right image pixels is inaccurate. Another is that the stereo images are not rectified exactly. Since that of HPCM-5 seems better by inspection, its disparity curve is converted into depths. The average depth of the box is 135.69 cm, and the nearest point of the cylinder is 97.17 cm. These numbers are close to

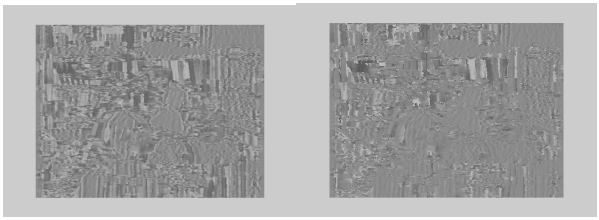
real depths but not so accurate. On the other hand, when PM or PCM is used, figures 9(c) and 9(d) show the results are nonsense. The cause is that their small working ranges result in the aliasing effects. In words, the experimental results verify that HPCM is an excellent solution in applications of computational stereo.

4. Conclusion

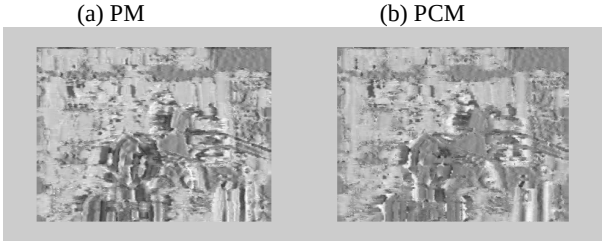
Other than feature- and correlation-based approaches, PM was shown able to determine disparity from the phase shift. PCM improved the accuracy of PM by introducing the coherence mechanism into the disparity calculation. Only the outputs of coherent filters were taken into account in determining the disparity. An apparent drawback of PM and PCM was that their working ranges were limited by the wavelength of the filter stack. In real application environments, this small working range might result in the aliasing effects so that the measured disparities were far from real values. HPCM and HPM obtained large working range and avoided the aliasing effects. Though HPCM was more accurate, both HPM and HPCM were useful in practical applications. Given the camera parameters, the dense disparity map could be transformed into depth map. Although the depth map was three-dimensional, it was found not so accurate for precise positioning. Practical applications should take this feature into consideration.

(a) Left image, (b) Right image

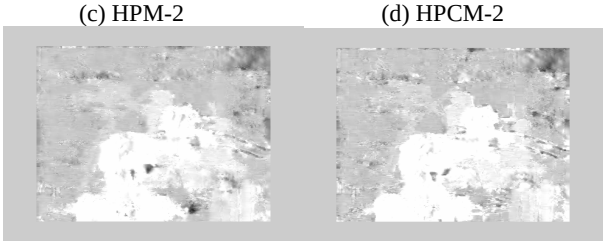
Fig. 4 The stereo image pair used in experiment 1



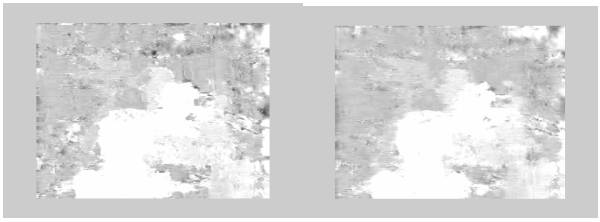
(a) PM (b) PCM



(c) HPM-2 (d) HPCM-2



(e) HPM-4 (f) HPCM-4

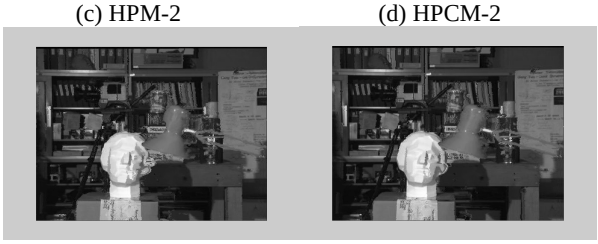


(g) HPM-5 (h) HPCM-5

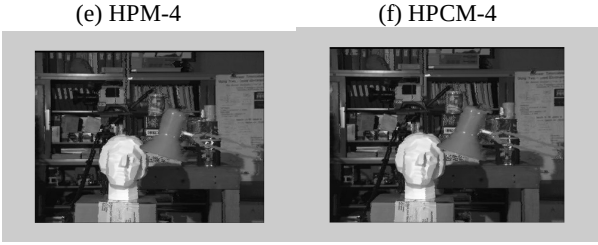
Fig. 5 Disparity maps of figure 4 obtained with PM, PCM, HPM, and HPCM

(a) PM (b) PCM

(c) HPM-2 (d) HPCM-2



(e) HPM-4 (f) HPCM-4

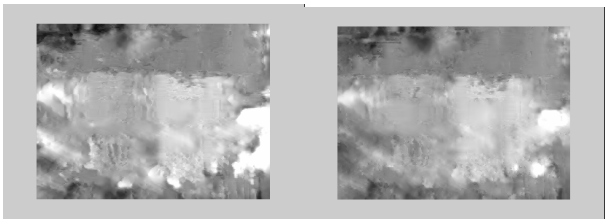


(g) HPM-5 (h) HPCM-5

Fig. 6 The shifted left images obtained with PM, PCM, HPM, and HPCM

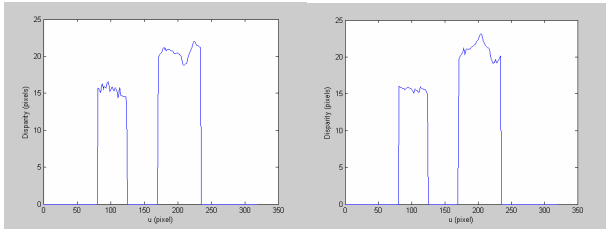
(a) The left image (b) The right image

Fig. 7 The stereo image pair used in experiment 2; the box and the cylinder are 140 cm and 100cm apart from the camera, respectively

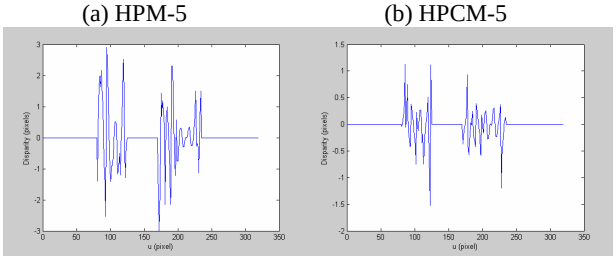


(a) HPM-5 (b) HPCM-5

Fig. 8 The disparity maps of figure 7 obtained with five resolution scales



(a) HPM-5 (b) HPCM-5



(c) PM (d) PCM

Fig. 9 The disparity curves crossing the box and the cylinder (along the 125th row)

5. Acknowledgements

The authors gratefully acknowledge National Science Council of Taiwan for grant under NSC95-3113-P-002-001.

6. References

[1] Wei-Song Lin, Ming-Kang Chuang and Glorious Tien, "Autonomous mobile robot navigation using stereovision," *Proceedings of the IEEE International Conference on Mechatronics*, Taipei, Taiwan, pp. 410-415, July 10-12, 2005.

[2] B. M. Collins and A. L. Kornhauser, "Stereo vision for obstacle detection in autonomous navigation," *Princeton University.* May 24, 2006.

[3] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *IEEE Trans. On Pattern Recognition and Machine Intelligence*, vol. 25, no. 8, pp. 993-1008, Aug. 2003.

[4] O. Faugeras, B. Hotz, H. Mathieu, T. Viéville and Z. Zhang, et al. "Real time correlation-based stereo: algorithm, implementations and applications," project report, Unité de recherché INTRIA Sophia-Antipolis, France, 1993.

[5] J. Wiklund, C-J. Westelius, H. Knutsson, "Hierarchical phase based disparity estimation," *Proceedings Second International Singapore Conference on Image Processing*, pp. 128-131, Sept. 1992.

[6] R. D. Henkel, "Fast stereovision by coherence detection," *Proc of the Int. Conf. on Computer Analysis of Images and Patterns*, (eds. G. Sommer, K. Daniilidis & J. Pauli), LNCS1296, Springer, Berlin, pp. 297-304, 1997.

[7] D. J. Fleet, A. D. Jepson, and M. Jenkin, "Phase based disparity measurement," *CVGIP: Image Understanding*, vol. 53, no. 2, pp. 198-210, 1991.

[8] D. J. Fleet, A. D. Jepson, "Stability of phase information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 14, pp.1253-1268, 1993.

[9] J. Hoey, "Stereo disparity from local image phase," 1999, available on website http://citeseer.ist.psu.edu

[10] T. D. Sanger, "Stereo disparity computation using Gabor filters," *Biological Cybernetics*, vol. 59, pp.405-418, 1988.

[11] D. Gabor, "Theory of communication," *JIEE*, vol. 93, pp. 429-459, 1946.

[12] F. Solari, S.P. Sabatini and G.N. Bisio, "Fast technique for phase-based disparity estimation with no explicit calculation of phase," *Electronics Letters*, vol. 37, no.23, Nov. 2001.

[13] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communication*, vol. COM-31, pp. 532-540, 1983.

[14] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, issue 1-3, pp. 7-42, 2002.