An Optical Device for Computation of Binocular Stereo Disparity with a Single Static Camera

David Vernon

CAPTEC Ltd., Malahide, Co. Dublin, Ireland

ABSTRACT

This paper presents a new device and technique for computing the stereo disparity of two binocular optical images using the data from a single sensor. The device, comprising a mirror and beamsplitter, superimposes the two views onto a single sensor to produce a single additive composite image. Local (*i.e.* windowed) Fourier analysis of this composite image yields the phase difference between the two component images and, thereby, the stereo disparity. The primary advantages of this approach is that it allows existing monocular cameras (digital or analogue, interlaced or non-interlaced) to be converted to stereo at relatively little cost and effort. Results are presented for both simulated images and images acquired with a prototype single-sensor stereo camera. As currectly conceived, the approach would probably not be appropriate as a general-purpose technique for the computation of detailed structure of the environment — and it certainly won't supplant existing multi-camera stereo systems for complex problems — but it is suitable for simple stereo-based applications, such as obstacle avoidance and segmentation.

Keywords: stereoscopic vision, sensors, optics, Fourier transform.

1. INTRODUCTION

In computer vision, the measurement of stereo disparity is one of the most common and important techniques for recovering the 3-D structure of the imaged scene and for object segmentation. Almost all stereo systems deploy two calibrated cameras in a binocular configuration although trinocular[1] and multi-view[2,3] stereo systems have been proposed to overcome some of the short-comings of binocular systems. Indeed, such systems are commercially available.[4]

There are essentially two stages in the processing of all stereo images: the search for matching points in the left and right stereo images, yielding conjugate pairs of points, and the computation of distance or range using (intrinsic) camera calibration data (*e.g.* the focal length of the lenses) and extrinsic data about the camera configuration (*e.g.* the base-line distance between the cameras). The first stage — frequently referred to as the correspondence problem — is by far the more difficult of the two stages. Typically, one of three different general approaches is used to identify the correspondence between conjugate pairs of points:

1. Correlation of local windows in the two images (three in the case of trinocular stereo);[5,6]
2. Feature extraction (*e.g.* edges or corners) followed by feature matching;[5,7–9]
3. Fourier frequency domain techniques in which the disparity, or shift, between two windows/sub-images is given directly by their phase difference.[10,11]

In all cases, the computational complexity of the correspondence problem can be significantly reduced by exploiting knowledge of the relative orientations of the cameras, *i.e.* their vergence, to compute the epi-polar geometry of the configuration. This geometry provides constraints on the possible set of matches between conjugate pairs: in effect, all conjugate pairs are constrained to lie on a unique line — the epi-polar line — given by the epi-polar geometry and the image coordinates of one of the points comprising a conjugate pair, reducing the search for the second point in the conjugate pair to a process with linear complexity. In many

Author e-mail: vernon@ieee.org

cases, the so-called fronto-parallel binocular camera configuration is used, *i.e.* the optical axes of both cameras are parallel and the relative positions and orientations of each camera are related by a simple translation of the base-line distance along one of the axes of the image plane. This configuration yields a special epi-polar geometry with all epi-polar lines being aligned in one direction parallel to the axis of translation. Typically, this means that all conjugate pairs lie on corresponding scan-lines in the image.

Despite the importance of stereo vision, stereo cameras have not yet become commonplace in either the research environment or in industrial and commercial applications, except, of course, in situations where stereo is actually the subject being studied. One possible reason for this is the additional financial cost associated with purchasing the second camera and interfaces and the set-up costs in configuring, calibrating, and synchronizing them. Without doubt, the new breed of dedicated fronto-parallel stereo heads from companies such as PointGrey[4] and Videre Design[12] which use CMOS and CCD sensors and the IEEE-1394 digital interface will help to alter this state of affairs but, although these cameras represent good value for money in relative terms, their cost is still significant in absolute terms, if only because they, by definition, require twice or three times the hardware (lenses, sensors, housing, interface circuitry).

On the other hand, Gluckman and Nayer have pointed out several radiometric and geometric advantages of a single camera stereo configuration.[13–15] For example, the optical and sampling characteristics of the system (lens distortion, blurring, focal length, spectral response, gain, offset, pixel size) are identical for both stereo images, thereby facilitating stereo matching. The calibration procedure is also simpler since there is only one set of extrinsic calibration parameters and camera synchronization is no longer an issue. Finally, the acquisition and storage of stereo images is easier than with conventional two-camera stereo.

In this paper, we will show how one can effect stereo imaging with a single sensor by using a simple arrangement of a mirror and a beam-splitter, thereby significantly reducing the capital and set-up costs. This arrangement can be used with any existing (monocular) camera, digital or analogue, interlaced or non-interlaced.

2. PREVIOUS APPROACHES

Adelson and Wang[16] describe a single lens stereo system that uses a plenoptic camera whereby light distribution in a local region (typically 5×5 pixels) is analysed to yield an estimate of the scene depth at the centre of that region. The system exploits a conventional finite-aperture imaging lens and a lenticular array placed at the sensor plane. Each lenticular element acts as an individual pin-hole camera and images the local light distribution onto a 5×5 array of sensor photosites. This results in a lenticular array of images which effectively capture the scene from a continuum of viewpoints, thereby facilitating depth estimation by measuring relative image displacement. The depth resolution is limited by the ratio of the lens aperture to object distance so that the plenoptic camera is best suited to applications involving small objects placed close to the camera (*e.g.* parts inspection) rather than, say, a robot navigation application.

A number of researchers have used mirrors to effect stereoscopic viewing using a single camera although many of these necessitate camera rotation and so are of use only with static scenes.[17,18]

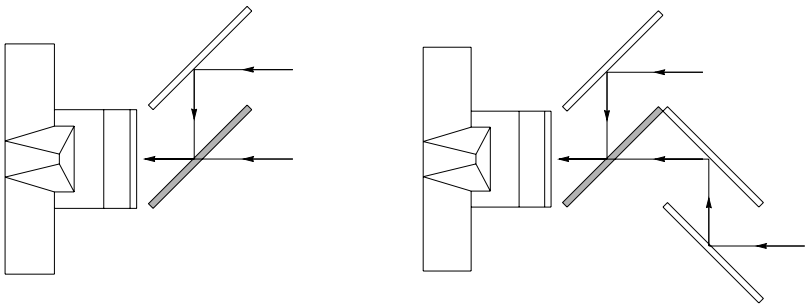
More conventional binocular stereopsis can also be be achieved using mirror-based catadioptric configurations. For example, Gluckman and Nayer[13] present a calibrated single camera stereo system which captures stereo image pairs on a single sensor using light reflected from one or more planar mirrors. However, the left and right stereo images are projected on the left and right halves of the sensor, respectively, thereby reducing by half the horizontal resolution and field of view of the stereo system. They subsequently report a similarly configured system which guarantees rectified stereo images in which the epi-polar lines are parallel to the image scan-lines, just as in a fronto-parallel configuration.[14,15]

Similar systems which use mirrors to reflect images from two viewpoints onto two distinct halves of the sensor have also been reported.[19–23]

Lee *et al.* have reported a single camera stereo system based on the use of a biprism.[24] In this case, the two stereo views are formed by refraction through the oblique faces of the prism rather than by reflection in a set of mirrors, each (rectified) view being imaged on separate halves of the camera sensor in much the same way as the catadioptric systems.

(a) (b) (c)

Figure 1. (a) left view; (b) right view; (c) left and right superimposed and imaged by a single sensor



(a) (b)

Figure 2. Optical configuration (a) different distances to each point in scene for left and right views; (b) identical distances to each point in the scene for left and right views. In both cases, the beam-splitter is the shaded optical element; all others are first-surface mirrors.

3. THEORETICAL BACKGROUND

The essential idea behind the single-sensor stereo developed in this paper is to additively superimpose the two optical images from physically-distinct binocular viewpoints on a single image sensor, yielding a single image of the sum of both binocular views (see Figure 1).* This approach, which has the advantage over catadioptric systems that it exploits the full horizontal spatial resolution of the sensor for both left and right images, can be accomplished in a cost-effective manner using a simple arrangement of mirrors and beamsplitters. Two configurations are proposed in this paper. The first uses one mirror to reflect the left image to the beamsplitter and a beamsplitter to reflect the left image to the camera and at the same time additively combine the right image (see Figure 2 (a)).

Since the camera is a projective imaging system and since the path of the light rays from sensor to imaged point is slightly longer for the left image than for the right image, by an amount equal to the base-line distance, this optical arrangment will cause the left view to be slightly smaller than the right image. In most cases this will not be a problem but if it is then an alternative configuration can be used which guarantees that both paths are identical in length (see Figure 2 (b)).

*This approach is *not* the same as the now fairly established technique used by some video camera devices in which an interlaced signal captures the left and right images in the even and odd fields respectively with the aid of an LCD shuttering system that is synchronized with the field pulses in the video signal.[25]

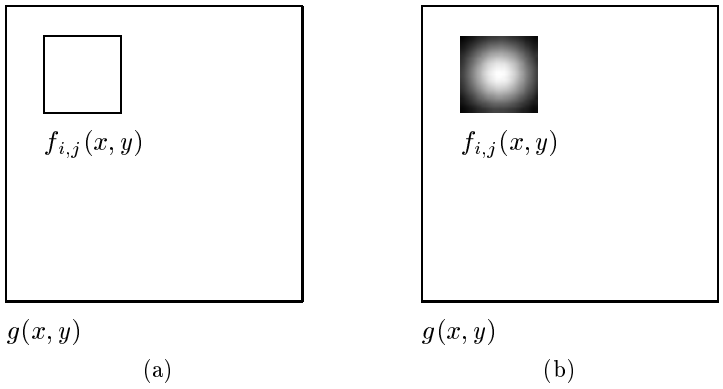


Figure 3. A window $f_{ij}(x,y)$, centred at coordinates (i,j), in a grey-level image $g(x,y)$; (a) simple box window, and (b) apodized window using a Gaussian weighted function.

The problem is now to compute the stereo disparity at each point in this single additive image, *i.e.* to identify the relative shift between the points comprising each conjugate pair. To solve this problem, we first observe that this configuration is fronto-parallel so that the epi-polar lines are all horizontal and hence all shifts are horizontal.

Next, we consider a local region or window $f_{ij}(x,y)$, centred at coordinates (i,j), in a grey-level image $g(x,y)$ — see Figure 3. This window contains the sum of a corresponding window in the left and right images $f^L(x,y)$ and $f^R(x,y)$:

$$f(x,y) = f^L(x,y) + f^R(x,y)$$

To a first approximation, the image in the right window is the same as that in the left window, except that it is shifted horizontally by some amount (the amount will vary inversely with the distance of the object from the camera). That is

$$f^R(x,y) = f^L(x+\delta x, y)$$

Note that this is not strictly accurate since image content will change at the periphery of the window because of the shift due to image content entering and exiting the window. However, for reasonably small baselines or relatively distant objects, this will not be significant. Even so, to lessen the potential effects of this problem, we use a Gaussian multiplicative apodizing windowing function. Thus, we extract as the windowed image function $f'_{i,j}(x,y)$:

$$f'_{i,j}(x,y) = G(x,y,\sigma) f_{i,j}(x,y)$$

where $G(x,y,\sigma)$ is a Gaussian function with standard deviation σ. Throughout this paper we use a Gaussian with a standard deviation σ chosen such that the weighting at a distance $\frac{nw}{8}$ pixels from the region centre is 50% of that at the region centre, where w is the length (in pixels) of the side of the 2-D region and $n = 1, 2, 3$.

In the following, we will drop the i,j subscripts and prime superscript, and simply use $f(x,y)$ to represent a window in an image, on the explicit understanding that the window is an apodized sub-region of the additive image $g(x,y)$, evaluated at coordinates (i,j).

Thus, our model of the extracted region is:

$$f(x,y) = f^L(x,y) + f^L(x+\delta x, y) \tag{1}$$

We next use the Fourier transform and the Fourier shift property to estimate δx. Taking the Fourier transform of $f(x,y)$, we have:

$$\begin{aligned} \mathcal{F}\left(f(x,y)\right) &= \mathcal{F}\left(f^L(x,y) + f^L(x+\delta x, y)\right) \\ &= \mathcal{F}\left(f^L(x,y)\right) + \mathcal{F}\left(f^L(x+\delta x, y)\right) \end{aligned}$$

By the Fourier shift property whereby:

$$\begin{aligned}\mathcal{F}\left(f(x-\delta x, y-\delta y)\right) &= \mathcal{F}\left(f(x,y)\right) e^{-i(\omega_x \delta x + \omega_y \delta y)} \\ &= \mathsf{F}(\omega_x, \omega_y) e^{-i(\omega_x \delta x + \omega_y \delta y)}\end{aligned}$$

we have:

$$\mathcal{F}\left(f(x+\delta x, y)\right) = \mathsf{F}(\omega_x, \omega_y) e^{i(\omega_x \delta x)} \tag{2}$$

And, thus:

$$\mathsf{F}(\omega_x, \omega_y) = \mathsf{F}^L(\omega_x, \omega_y) + \mathsf{F}^L(\omega_x, \omega_y) e^{i(\omega_x \delta x)} \tag{3}$$

The term $e^{i(\omega_x \delta x)}$ represents a spatial-frequency dependent phase-shift, *i.e.* a rotation of each spatial-frequency phasor (or Fourier component) by an angle $\omega_x \delta x$. Now, at some spatial frequencies, ω_{x_n} say, $\omega_{x_n} \delta x = \pi + 2\pi n, n = 0, 1, 2, \ldots$, *i.e.* the angle of rotation will be π radians modulo 2π. At this spatial frequency, $\mathsf{F}^L(\omega_x, \omega_y)$ and $\mathsf{F}^L(\omega_x, \omega_y) e^{i(\omega_x \delta x)}$ will be in anti-phase and will sum to zero. Thus, in order to estimate δx we only need to identify the spatial frequencies ω_{x_n} at which the magnitude of the Fourier components $|\mathsf{F}(\omega_x, \omega_y)| = \left|\mathsf{F}^L(\omega_x, \omega_y) + \mathsf{F}^L(\omega_x, \omega_y) e^{i(\omega_x \delta x)}\right|$ are zero and then compute δx from $\delta x = \frac{\pi}{\omega_{x_0}}$.

Since this rotation depends only on the ω_x spatial frequency these zero-components will occur at all ω_y frequencies. We use this fact to improve the robustness of the technique by summing all Fourier component along the ω_y axis to yield a 1-D signature F^s of Fourier component magnitudes as a function of ω_x:

$$\mathsf{F}^s(\omega_x) = \sum_{\omega_y} \mathsf{F}(\omega_x, \omega_y)$$

This signature periodically approaches zero at frequencies $\omega_{x_n} = \frac{\pi + 2\pi n}{\delta x}$ (see figures 5.e – 6.e and 5.f – 6.f).

4. RESULTS

The foregoing technique was tested using both simulated images and using images acquired using a prototype device based on the single mirror/beam-splitter configuration (see Figure 2.a).

Figure 5 shows the results of applying this technique to an example formed by simply adding the left and right images $f^L(x, y)$ and $f^R(x, y)$ of a conventional stereo pair (note that the base line is very small: approx. 1 cm). Figure 5.a, 5.b, and 5.c show the left, right, and composite views. Figure 5.d shows a sample apodized 64×64 pixel window $f(x, y)$ extracted from the centre of the image. Figure 5.e shows the corresponding unprocessed Fourier transform of $\mathsf{F}(\omega_x, \omega_y)$ together with $\mathsf{F}^s(\omega_x)$, the Fourier tranform of the composite image after it has been summed in along the ω_y axis; Figure 5.f shows a plot of $\mathsf{F}^s(\omega_x)$. Figure 5.g shows the disparity field computed every 10 pixels with the same field superimposed on the left image being shown in Figure 5.h, while Figure 5.i shows the an interpolated disparity field encoded as a grey-level image where brightness is proportional to the magnitude of the disparity (in this case, disparity was computed every 5 pixels).

Figure 6 shows the results obtained for images acquired using the mirror/beam-splitter configuration shown in figure 4.

5. CONCLUSION

This paper has shown how one can compute stereo disparity using a single sensor and a static camera. The main benefits of the approach are the low financial and set-up cost: the device is simple and can be deployed in existing monocular systems. The result is a system which can measure relatively small disparities, relatively accurately: there are theoretical limitations restricting its ability to detect disparities less than half the width of the window being used and, in practice, the maximum measurable disparity is somewhat less again due to sampling noise and the smoothing inherent in computing $\mathsf{F}^s(\omega_x)$.

Figure 4. Prototype single sensor stereo device.

Although in this paper we demonstrated the technique for a fronto-parallel device, it is feasible to control the vergence of the system by altering the angle of the side mirror(s). In this case, the zero Fourier components will aligned in a direction normal to the disparity shift. Control of the vergence angle — either manual or automatic — would be necessary if the disparity of foreground objects became large in relation to the size of the windowing function. Current work is focussed on dealing with vergence and on developing a more robust method for estimating the period of the zero Fourier components, especially when they are not aligned with the vertical axis (*i.e.* when the epipolar lines cannot be assumed to be horizontal). This should improve the quality of the results obtainable from the device and allow it to be tuned for specific applications. Very recent preliminary results are promising (see Figure 7).

The approach presented is particularly suitable for small base-line tasks (the minimum base-line is equal to the diameter of the lens). The baseline can be extended almost arbitrarily, although it would be necessary to scale the size of the side mirror(s) accordingly and vergence control would also be required. Note, however, that since the technique is predicated on distortion-free image translation, such wide-baseline stereo would not work particularly well with close objects due to the variation in object appearance.

Ultimately, and mindful of these restrictions, this is a novel and cost-effective way of converting existing monocular cameras to yield coarse stereo images. As currectly conceived, the approach would probably not be appropriate as a general-purpose technique for the computation of detailed structure of the environment — and it certainly won't supplant existing multi-camera stereo systems for complex problems — but it is suitable for simple stereo-based applications, such as obstacle avoidance and segmentation.

REFERENCES

1. N. Ayache and F. Lustman, "Fast and reliable passive trinocular stereovision," in *Proc. Int. Conference on Computer Vision*, pp. 422–427, 1987.
2. K. Kutulakos, "Approximate n-view stereo," in *Computer Vision – ECCV 2000, Lecture Notes in Computer Science* **1842**, pp. 67–83, Springer-Verlag, (Berlin), 2000.
3. D. W. Murray, "Recovering range using virtual multicamera stereo," *Computer Vision and Image Understanding* **61**, pp. 285–291, Mar. 1995.
4. P. G. R. Inc., "Stereo head vendor.," **http://www.ptgrey.com**.
5. O. D. Faugeras, P. Fua, B. Hotz, R. Ma, L. Robert, M. Thonnat, and Z. Zhang, "Quantitative and qualitative comparison of some area and feature-based stereo algorithms," in *International Workshop on Robust Computer Vision: Quality of Vision Algorithms*, W. Forstner and S. Ruwiedel, eds., pp. 1–26, (Karlsruhe, Germany), March 1992.
6. D. Vernon and G. Sandini, *Parallel Computer Vision — The* VIS $\vec{a}$ VIS *System*, Ellis Horwood, London, 1992.
7. S. Alibhai and S. W. Zucker, "Contour-based correspondence for stereo," in *Computer Vision – ECCV 2000, Lecture Notes in Computer Science* **1842**, pp. 314–330, Springer-Verlag, (Berlin), 2000.

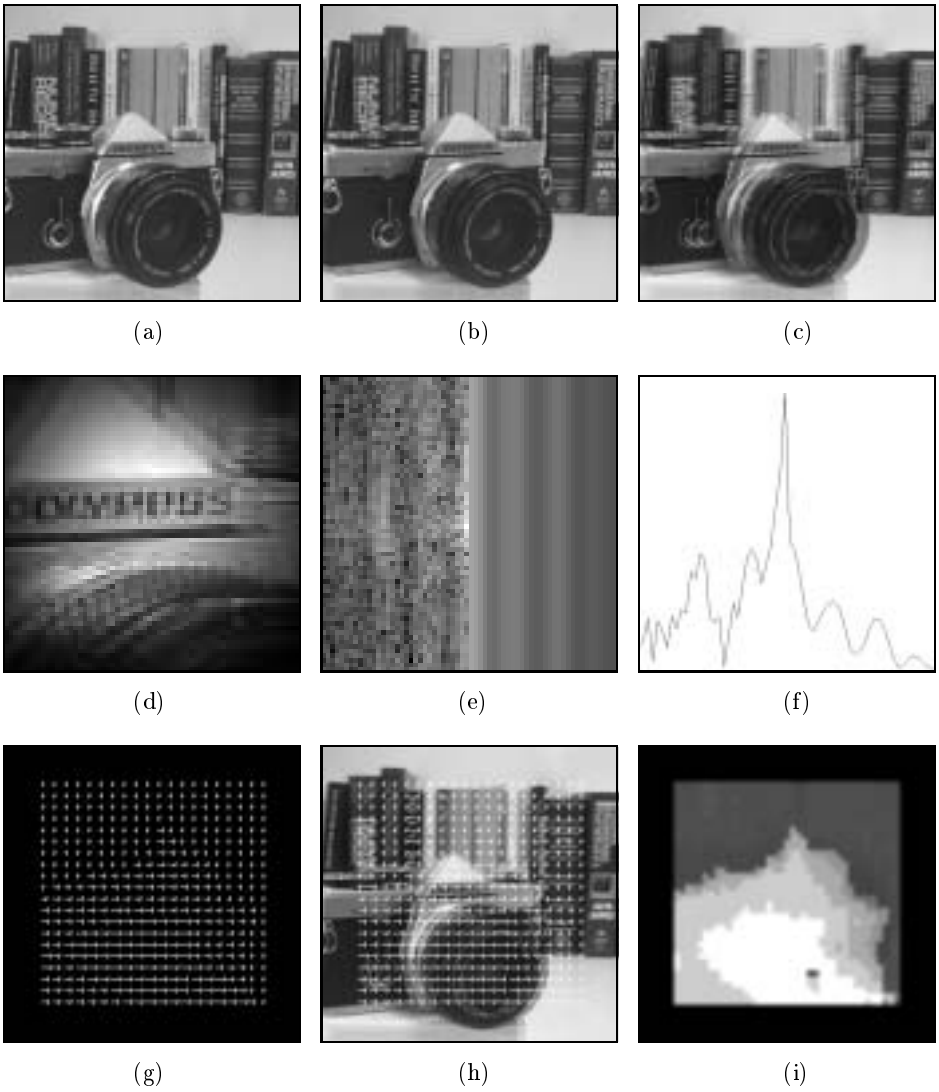


Figure 5. Simulated example (real stereo images added together) (a) left image $g^L(x,y)$; (b) right image $g^R(x,y)$; (c) composite image $g(x,y)$; (d) $f(x,y)$, apodized 64×64 sub-region of $g(x,y)$; (e) left-hand side:$\mathsf{F}(\omega_x, \omega_y)$, unprocessed Fourier transform of $f(x,y)$ – right-hand side:$\mathsf{F}^{\mathsf{s}}(\omega_x)$, Fourier tranform of the composite image after it has been summed in along the ω_y axis; (f) plot of $\mathsf{F}^{\mathsf{s}}(\omega_x)$; (g) the disparity field computed every 10 pixels; (h) disparity field superimposed on the left image (i) interpolated disparity field encoded as a grey-level image where brightness is proportional to the magnitude of the disparity (in this case, disparity was computed every 5 pixels).

8. W. E. L. Grimson, "Computational experiments with a feature based stereo algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **7**(1), pp. 17–34, 1985.
9. D. Sherman and S. Peleg, "Stereo by incremental matching of contours," *IEEE Trans. Pattern Analysis and Machine Intelligence* **12**(11), pp. 1102–1106, 1990.
10. F. Valentinotti, G. Di Caro, and B. Crespi, "Real-time parallel computation of disparity and optical flow using phase difference.," *Machine Vision and Applications* **9**(3), pp. 87–96, 1996.
11. D. Vernon, *Fourier Vision — Segmentation and Velocity Measurement using the Fourier Transform*, Kluwer Academic Publishers, Boston, 2001.
12. V. D. Inc., "Stereo head vendor.," **http://www.videredesign.com**.

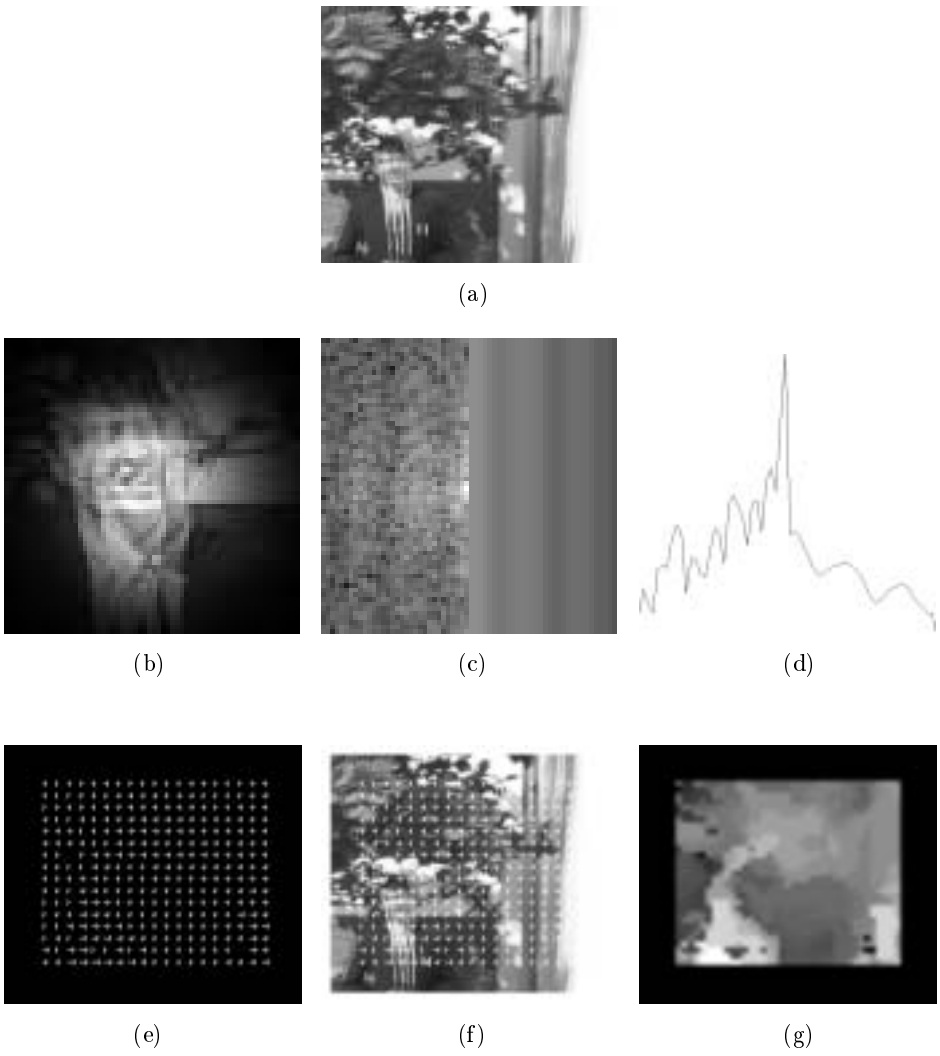
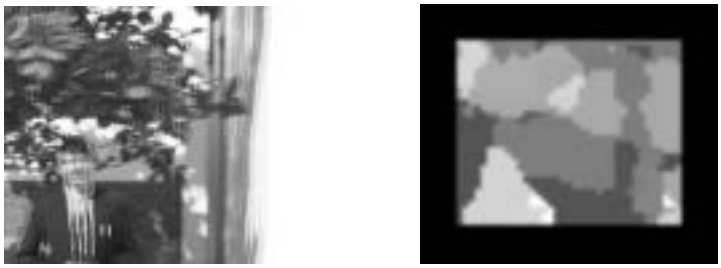


Figure 6. Example acquired using prototype shown in figure 4. (a) composite image $g(x, y)$; (b) $f(x, y)$, apodized 64×64 sub-region of $g(x, y)$; (c) left-hand side:$\mathsf{F}(\omega_x, \omega_y)$, unprocessed Fourier transform of $f(x, y)$ – right-hand side:$\mathsf{F}^s(\omega_x)$, Fourier tranform of the composite image after it has been summed in along the ω_y axis; (d) plot of $\mathsf{F}^s(\omega_x)$; (e) the disparity field computed every 10 pixels; (f) disparity field superimposed on the left image (g) interpolated disparity field encoded as a grey-level image where brightness is proportional to the magnitude of the disparity (in this case, disparity was computed every 5 pixels).

13. J. Gluckman and S. K. Nayar, "Planar catadioptric stereo: Geometry and calibration," *Proc. IEEE Conference on Computer Vision and Pattern Recognition* , pp. 22–28, 1999.
14. J. Gluckman and S. K. Nayar, "Rectified catadioptric stereo sensors," *Proc. IEEE Conference on Computer Vision and Pattern Recognition* , pp. 380–387, 2000.
15. J. Gluckman and S. K. Nayar, "Rectified catadioptric stereo sensors," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**, pp. 224–236, February 2002.
16. E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(2), pp. 99–106, 1992.



(a) (b)

Figure 7. The estimation of the period of the zero Fourier components is the key to this approach and very recent preliminary results suggest that techniques for which don't assume exact fronto-parallel geometry will yield improved results; (a) composite image $g(x, y)$ aquired using the single-sensor stereo device; (b) interpolated disparity field.

17. S. Peleg and M. Ben-Ezra, "Stereo panorama with a single camera," *Proc. IEEE Conference on Computer Vision and Pattern Recognition* , pp. 395–401, 1999.
18. P. Peer and F. Solina, "Panoramic depth imaging: single standard camera approach," *International Journal of Computer Vision* **47**, pp. 149–160, 2002.
19. H. Mitsumoto, S. Tamura, K. Okazaki, N. Kajimi, and Y. Fukui, "3-d reconstruction using mirror images based on a plane symmetry recovering method," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(9), pp. 941–946, 1992.
20. A. Goshtasby and W. A. Gruver, "Design of a single-lens stereo camera system," *Pattern Recognition* **26**, pp. 923–936, 1993.
21. M. Inaba, T. Hara, and H. Inoue, "A stereo viewer based on a single camera with view control mechanism," in *Proc. International Conference on Robots and Systems*, 1993.
22. H. Mathieu and F. Devernay, "Systeme de miroirs pour la stereoscopie," *INRIA Technical Report 172* , 1995.
23. Z. Y. Zhang and H. T. Tsui, "3d construction from a single view of an object and its image in a plane mirror," in *Proc. International Conference on Pattern Recognition*, pp. 1174–1176, 1998.
24. D. H. Lee, I. S. Kweon, and C. R., "A biprism-stereo camera system," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'99)*, 1999.
25. I. 3-D Video, "Manufacturer of the nuview stereo camcorder adaptor.," **http://www.3-dvideo.com**.