SINGLE-EYE RANGE ESTIMATION BY USING DISPLACED APERTURES WITH COLOR FILTERS

Yasufumi Amari
Email: amari@secom.co.jp
SECOM Intelligent Systems Laboratory*

Edward H. Adelson
Email: adelson@media.mit.edu
The Media Laboratory, Massachusetts Institute of Technology**

Abstract

A single-eye range sensing system is useful because of its compactness. Generally, a single-eye range sensing system can be realized by using the sub-regions of the main lens as plural lenses. Depth estimation analyzing a defocused image[1], which uses this method implicitly, is an application of this single-eye range sensing.

A simple single-eye range sensing system can be easily constructed by introducing two displaced apertures with color filters on the main lens. The color image obtained can be separated into two images, corresponding to views from two viewpoints, with an RGB decoder. Range data can then be estimated from the disparity between the two images.

1 Introduction

In a single lens imaging system, when the geometry is such that the imaging equation:

$$\frac{1}{a}+\frac{1}{b}=\frac{1}{f} \tag{1}$$

is not satisfied, then the object will not be in focus. (a is the distances between the subject and the lens, b is distance between the lens and the screen, and f is the focal length of the lens.)

Focal error contains range information of subject to the camera. This range information can be extracted by analyzing the defocused image[1]. Generally, focal error can be determined by varying the aperture parameters, such as size and location of the system. Single lens range sensing systems have been implemented by these methods. The Pentland Depth by Defocus Range Camera[2] used two apertures of different size to extract range information. The Plenoptic Camera[3] described by Adelson and Wang uses a lenticular lens to form effective apertures at different locations for the same purpose.

In this paper, we describe a simple single lens passive range camera that uses displaced apertures with color filters. The color filters provide a convenient way to separate the image obtained into two images corresponding to images taken with the aperture at different locations. Range information can then be extracted by analyzing these two images.

2 Workings of the Sensor

To simplify the explanation of this sensor, we initially assume a white light point subject and a two-dimensional single lens imaging system with two displaced apertures with red and green filters as shown in Figure 1.

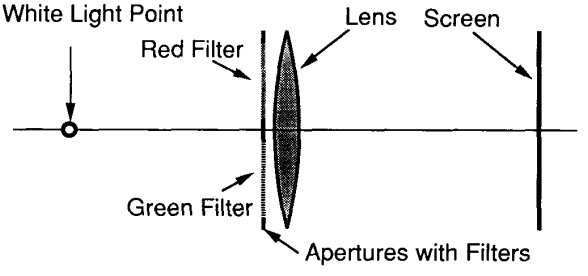

Figure 1: Schematic Diagram of the Sensor

When the system is in focus as Figure 2 (a), i.e. the subject exists at the optical conjugate point of the screen, red and green images of the subject by the apertures perfectly overlap on the screen. So, if we observe the red and the green images at the same point on the screen, we can use Equation 1 to estimate the distance between the lens and the subject as below.

$$Distance = a = \frac{b \cdot f}{b - f} \tag{2}$$

Meanwhile, in the case where the subject exists nearer than the optical conjugate point of the screen, so-called far focus, rays from the red and the green apertures make two independent red and green images at different points on the screen as shown in Figure 2 (b). In this case, the imaging equation is can be expressed as

$$\frac{1}{a-d}+\frac{1}{b+l}=\frac{1}{f} \tag{3}$$

where d is the displacement of the subject from the conjugate point of the screen and l is the distance between the screen and the conjugate point of the actual subject location. From similar triangles, we can derive

$$\frac{l}{s}=\frac{b+l}{D} \tag{4}$$

where s and D are the distances between the both images and between the centers of the apertures, respectively. From Equation 3 and 4, the distance between the lens and the subject can be estimated as

$$Distance = a - d = \frac{b \cdot f \cdot D}{b \cdot D - f \cdot D + f \cdot s}. \tag{5}$$

Conversely, when the subject exists further than the optical conjugate point of the screen, so-called near focus, as Figure

*6-1-1 Sakaemachi Tachikawa, Tokyo 190 Japan
**20 Ames Street Cambridge, MA 02139

0-7803-0582-5 /92$3.00©1992 IEEE

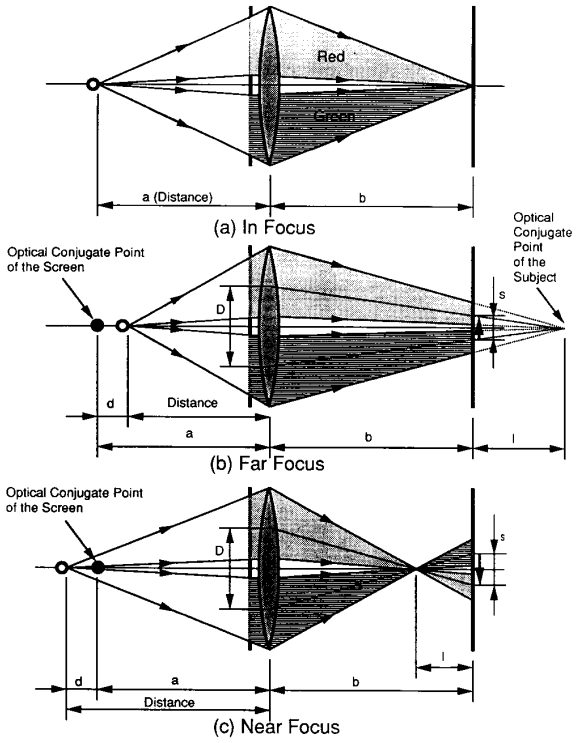

Figure 2: Principles of the Range Sensor

2 (c), the locations of the red and green images on the screen switch. In this case, the distance between the lens and the subject can be expressed as

$$Distance = a + d = \frac{b \cdot f \cdot D}{b \cdot D - f \cdot D - f \cdot s} \qquad (6)$$

in the same way as the far focus case.

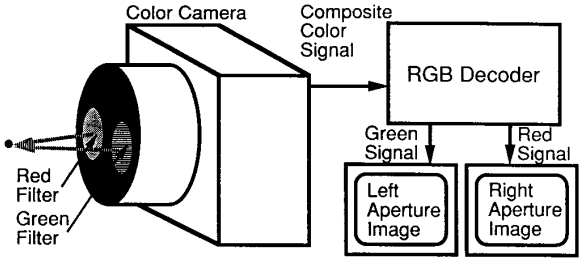

Figure 3: Schematic Diagram of the Range Sensor System

The schematic of the range camera is shown in Figure 3. The diagram shows a color camera with two red and green colored apertures at the main lens. Ideally, images corresponding to views from the locations of the red and the green apertures can be easily decoded from the output of the camera with an RGB decoder.

3 Processing

3.1 Estimation of the Right and the Left Aperture Images

In a real system, however, the color responses of the filters will not match that of the camera. The images obtained from the red and the green channels of the camera do not correspond to views seen by each of the apertures. This means that the red and green channels of the color camera will each have some contribution from both the apertures. This crosstalk has to be minimized to obtain a good range estimate.

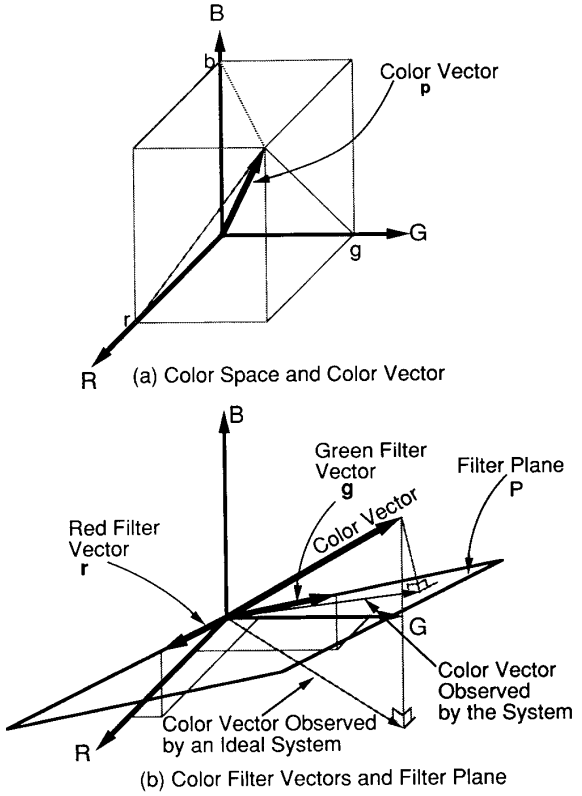

Figure 4: Color Space, Color Vector, and Filter Plane

The color system of a color camera can be regarded as a three dimensional vector space with red, green, and blue components as shown in Figure 4 (a). Any color from the camera can be expressed as a vector (**p**) spanned by the three RGB channels of the color camera in this RGB space. The system is ideal when the red filter vector is parallel to the red axis, and the green filter vector is also parallel to the green axis in this color space. Then any color observed through the red and the green filters will result in a color vector that lies on the RG plane. The color vector on the RG plane can be expressed as projection from the original color vector of the subject to the RG plane. The red and the green filters eliminate the blue component of the original color vector in this case. Images from the different view points are easily obtained by taking the red and the green outputs from the RGB decoder as shown in shown in Figure 4.

However, when the system is not ideal, the red and the green filter vectors will have red, green, and blue components as shown

in Fig. 4 (b). Any color vector observed by the system will lie in the subspace spanned by the red filter vector **r** and the green filter vector **g**. This subspace is a plane we call the filter plane. As in the ideal case, the color vector observed by the system can be expressed as a projection from the original color vector of the subject to this filter plane.

Any color vector observed by the system can be expressed as a linear combination of the red filter vector and the green filter vector on the filter plane. This linear combination relationship can be written

$$\mathbf{p} = a\mathbf{r} + b\mathbf{g} = a\begin{bmatrix} r_r \\ g_r \\ b_r \end{bmatrix} + b\begin{bmatrix} r_g \\ g_g \\ b_g \end{bmatrix} = \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{7}$$

where $\mathbf{r} = [r_r\ g_r\ b_r]^T$ is the red filter vector, $\mathbf{g} = [r_g\ g_g\ b_g]^T$ is the green filter vector, and $\mathbf{p} = [R\ G\ B]^T$ is the observed vector by the system. In this equation, a and b represents the amount of light gathered by the red and the green colored apertures, respectively. Two grayscaled images can be obtained by determining these coefficients for every pixel in the color image.

Let $\mathbf{A} = \begin{bmatrix} r_r & r_g \\ g_r & g_g \\ b_r & b_g \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix}$, we see that

$$\mathbf{A}\,\mathbf{x} = \mathbf{p}. \tag{8}$$

Multiplying both side of this equation by the pseudo-inverse matrix $\mathbf{A}^{-1}$, we can estimate the red colored and the green colored aperture images as:

$$\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{p}. \tag{9}$$

The pseudo-inverse matrix $\mathbf{A}^{-1}$ of a matrix $\mathbf{A}$ is defined as:

$$\mathbf{A}^{-1} \equiv (\mathbf{A}^T\,\mathbf{A})^{-1}\,\mathbf{A}^T. \tag{10}$$

3.2 Disparity Extraction

In stereo matching algorithms, characteristics of corresponding parts of images are assumed to be almost same. The images from the different viewpoints are tagged by different colors in this single-eye method. Consequently, especially for objects that are colored, images obtained from the two colored apertures will have different scale and contrast. Performance of disparity extraction or stereo matching algorithms would improve by using edge information obtained from bandpassed images. We use a Laplacian-Pyramid method to reduce the density difference from pair of the images.

A one-dimensional least-squares gradient technique, similar to a technique used by Lucas and Kanade[4], is applied to the bandpassed images to extract the disparity. A gradient technique usually performs well for small disparities, so a coarse-fine tracking architecture described previously [5, 6] is used in our algorithm to improve estimates for larger disparities. The least-squares gradient method can be summarized as:

$$disparity(x,y) = \frac{\sum_p \frac{\partial I(x,y,\nu)}{\partial x}\frac{\partial I(x,y,\nu)}{\partial \nu}}{\sum_p [\frac{\partial I(x,y,\nu)}{\partial x}]^2}$$

$$= \frac{\sum_p [\frac{\frac{\partial I_1(x,y)}{\partial x} + \frac{\partial I_2(x,y)}{\partial x}}{2}][I_2(x,y) - I_1(x,y)]}{\sum_p [\frac{\frac{\partial I_1(x,y)}{\partial x} + \frac{\partial I_2(x,y)}{\partial x}}{2}]^2} \tag{11}$$

where $I(x,y,\nu)$ is a grayscaled image, ν is a viewpoint , I_1 and I_2 are images from different viewpoints and p is the summation patch.

After extracting the disparity between both of the images, we use the disparity-distance relationship obtained before to estimate range between the camera and the subject.

4 Implementation and Calibration

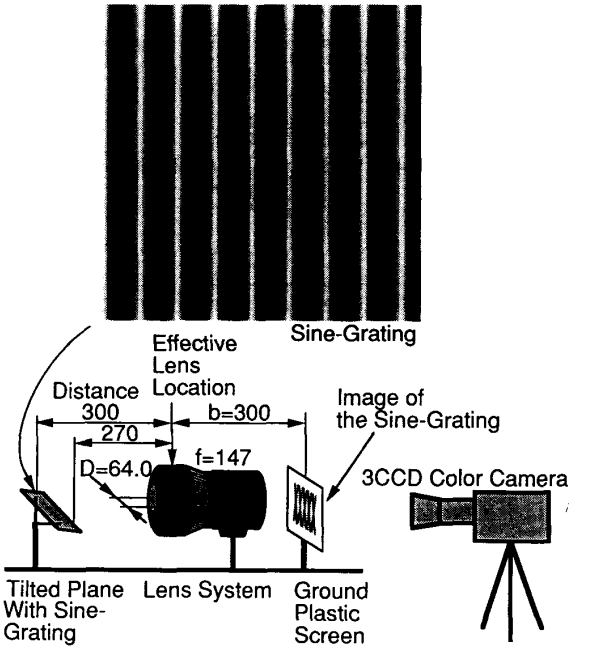

Figure 5: Experimental Apparatus

The described method is implemented by using a three-gun TV projector lens system. Apertures with red and green filters are located at the effective lens location of the lens system. The red and the green filters are carefully selected from a set of assorted plastic color filters for photographic effect purpose.

Figure 5 shows the experimental apparatus. Parameters of the experimental apparatus are also shown in this figure.

Calibration of the color response of the range camera requires measurement of the red filter vector **r** and green filter vector **g**. The red filter vector **r** is obtained from the means of the R, G, and B images of a white sheet of paper with the red colored aperture open and the green colored aperture closed. Similarly, we obtain the green filter vector **g** with the green colored aperture open and the red colored aperture closed.

Since matrix **A** is determined from the color characteristics of the filters, the right and left aperture images can be obtained from the RGB images by Equation 9.

5 Results and Discussions

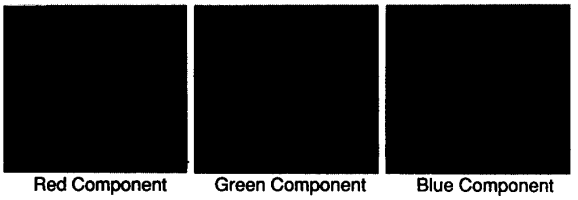

Figure 6: RGB Images of the Tilted Plane

An image of a tilted plane with a sine grating texture shown in Figure 5 is taken by the system. The RGB images of the tilted plane are shown in Figure 6, and the right and the left aperture images obtained by Equation 9 are shown in Figure 7(a),(b). Figure 7(c) shows the disparity between the images (from the right aperture image to the left aperture image). From this disparity data, range is calculated by using Equation 5. A surface plot of the estimated range is shown in Figure 8. Figure 9 shows the mean and standard deviation of the range estimate as a function of the object distance from the camera.

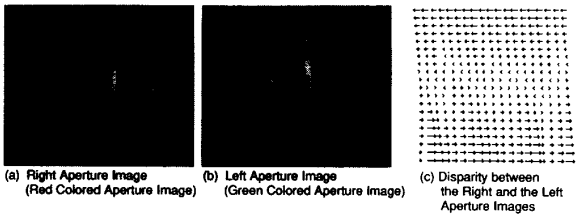

Figure 7: Estimated Right and Left Aperture Images and Disparity between the Estimated Images

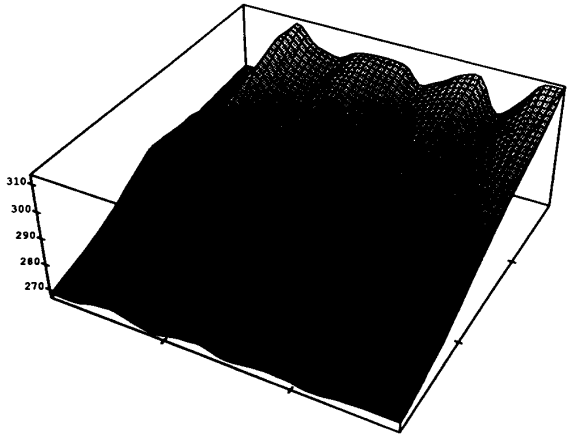

Figure 8: Surface Plot of the Range Estimation

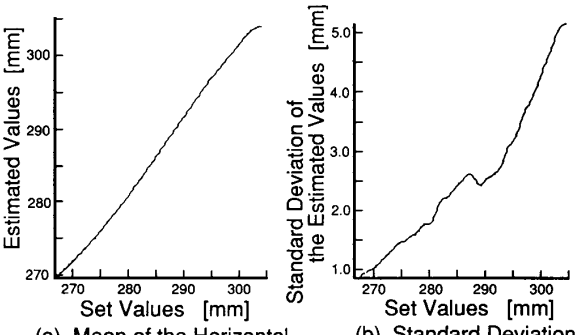

Figure 9: Result of the Range Estimation

The error of the estimated range shown in Figure 9 (b) is mainly caused by the range estimate resolution and the disparity extraction algorithm in dealing with images that are highly blurred.

Equation 5 can be rewritten:

$$disparity(x,y) = \frac{b \cdot D}{range(x,y)} - \frac{b \cdot D - f \cdot D}{f}. \qquad (12)$$

Equation 12 shows that this system is not fit for long range estimation because of the problem of the disparity resolution. *Disparity* will asymptotically become the second term with increasing range. The differentiation of the *disparity* can be calculated:

$$\Delta(disparity) = -\frac{b \cdot D}{range^2} \cdot \Delta(range). \qquad (13)$$

In Equation 13, we see that the disparity resolution will rapidly decay with increase in the range. For example, under the condition of the experiment, $\delta(disparity)/\delta(range)$ equals -0.26 for $range = 270mm$, -0.21 for $range = 300mm$, yet -0.12 for $range = 400mm$.

The result is not only influenced by the range disparity resolution but also by the image quality. Here, we use contrast as one measure of image quality to discuss quality of the images obtained. Contrast is defined as:

$$Contrast \equiv \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \qquad (14)$$

where I_{max} and I_{min} are maximum and minimum intensity of an image, respectively. In Figure 7 (a) and (b), there is more contrast in upper region than in the lower region. In these images, contrast corresponding to the range around 270mm (upper region) is almost 1.0, while around 300mm (lower region) is about 0.8. Variance of the extracted disparity increases in the lower region as seen in Figure 7(c). We see that the range estimate error increases with decreasing image quality. This problem can be minimized by using smaller apertures and better lenses.

Another design consideration is the crosstalk between the colored apertures. This is minimized by choosing color filters that have minimum frequency overlap, and by choosing the filters such that they closely match the color characteristics of the color camera.

The essence of this method is to apply colored filters to two differently placed apertures to obtain views from two viewpoints. This has the advantage that it requires no moving parts and the two views are obtained simultaneously.

One major disadvantage of this system is its color response to subjects of different color resulting in images of different scale and contrast as discussed earlier. Contrast differences can be minimized by using bandpassed images, since contrast mainly manifests as a low spatial frequency component in the image. In our experimental implementation, we use edge information obtained from bandpassed images to minimize contrast differences before applying the least-squares disparity extraction algorithm. Using bandpassed images does not minimize scale differences in the pair of images, so a more effective edge detection method and disparity extraction algorithm need to be considered in situations where the images have large scale differences.

6 Conclusions

A simple single-eye range camera can be easily constructed by introducing two displaced apertures with red and green filters on the main lens. A small range sensor system can be realized by adding displaced colored apertures to a color camera.

Acknowledgements

The authors wish to express their gratitude to John Wang, Graduate Student at the MIT Media Laboratory for his helpful discussions and experimental assistance.

References

[1] A. Pentland: A New Sense for Depth Field, IEEE Trans. on PAMI, Vol. PAMI-9, No. 4, pp. 523-531, 1987

[2] A. Pentland, T. Darrel, M. Turk, and W. Huang: A Simple, Real-Time Range Camera, Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 256-261, 1989

[3] E. H. Adelson and J. Y. A. Wang: Single Lens Stereo with a Plenoptic Camera, IEEE Trans. on PAMI, Vol. PAMI-14, No. 2, pp. 99-106, 1992

[4] B. D. Lucas and T. Kanade: An Iterative Image Registration Technique with an Application to Stereo Vision, Proc. Image Understanding Workshop, pp. 121-130, 1981

[5] L. H. Quam: Hierarchical Warp Stereo, Readings in Computer Vision, pp. 80-86, Los Altos, CA, Morgan Kaufmann Publishers, 1987 (ISBN 0-934613-33-8)

[6] J. R. Bergen, P. J. Burt, R. Hingorani, and S. Peleg: Computing Two Motions from Three Frames, Technical Report, David Sarnoff Research Center[1], April 1990

[1] Subsidiary of SRI International, Princeton, NJ 08543-5300