Distance Estimation with a Two or Three Aperture SLR Digital Camera

Seungwon Lee, Joonki Paik, and Monson H. Hayes

Graduate School of Advanced Imaging Science, Multimedia, and Film
Chung-Ang University
Seoul, Korea
mhh3@gatech.edu

Abstract. When a camera is modified by placing two or more displaced apertures with color filters within the imaging system, it is possible to estimate the distances of objects from the camera and to create 3-d images. In this paper, we develop the key equations necessary to estimate the distance of an object and discuss the feasibility of such a system for distance estimation in applications such as robot vision, human computer interfaces, intelligent visual surveillance, 3-d image acquisition, and intelligent driver assistance systems. In particular, we discuss how accurately these distances may be estimated and describe how distance estimation may be performed in real-time using an appropriately modified video camera.

1 Introduction

In many applications, such as robot vision, human computer interfaces, intelligent visual surveillance, 3-d image acquisition, and intelligent driver assistance systems, it is important to be able to estimate the distance of objects within the field of view of a camera or the relative distance between two or more objects. Depending on the system that is used, there are many different approaches for distance estimation, such as estimating the disparity of objects in stereo image pairs or using a time-of-flight camera.

In this paper, we consider the capture of stereo information and the estimation of the distances of objects using a standard SLR camera that has been modified by inserting two or three off-axis apertures into the camera lens. While such cameras have been used for autofocusing [1], multifocusing [2], and distance estimation [3], [4], here the focus is on the relationship between the location of objects in the image plane as a function of the location of the apertures, the resolution of the distance estimates that are produced with such a camera, the calibration of such a system, and their use in real-time estimation of the distance of objects from a video sequence.

2 Color Filter Aperture Cameras

In order to capture stereo image data in a manner that mimics the human visual system, one needs a pair of lenses that are separated some distance from each

J. Blanc-Talon et al. (Eds.): ACIVS 2013, LNCS 8192, pp. 25–36, 2013.

Fig. 1. An SLR camera with three off-axis apertures that are covered by red, green, and blue filters

other and that capture an image of the same scene at the same time. Dual lens or dual camera capturing systems have been around since the late nineteenth century, and today there is a variety of systems of varying complexity that capture stereo imagery. These range from cameras for the hobbyist, such as the Fujifilm *FinePix 3D* Digital Camera or lenses that turn a digital SLR cameras into a 3-d camera, such as the Loreo *3D Lens in a Cap* or the Panasonic Lumix lens, to high end systems for applications such as movie production.

A simple modification to the optics of a camera, however, will also allow for the capture of 3-D images and provide the ability to estimate the distance of objects within the scene of a camera. One such system is the multiple color filter aperture camera shown configured in Fig. 1 with three displaced apertures [2]. If the apertures are covered with different colored filters, such as red and cyan in a dual-aperture camera or red, green, and blue in a three-aperture camera, then each aperture will generate a separate image in one or more color planes of the camera. Since the apertures are displaced from each other with respect to the optical axis of the lens, a point on an object will be shifted by different amounts through the apertures where the amount of shift is a function of its distance from the camera. As shown in the following sections, this provides the means for estimating the distances of objects within the field of view of the camera.

2.1 Off-Axis Imaging

For an imaging system represented by a single lens with a focal length f and an aperture that is centered on the optical axis of the lens, Gauss' thin lens equation is

$$\frac{1}{v_0} + \frac{1}{z_0} = \frac{1}{f}$$

where v_0 is the distance of the image plane from the vertex of the lens and z_0 is the location of plane of focus of the lens [5]. However, if the aperture of the lens is not centered on the optical axis as illustrated in Fig. 2, then objects within the field of view of the camera will be shifted in the image plane, and the amount of

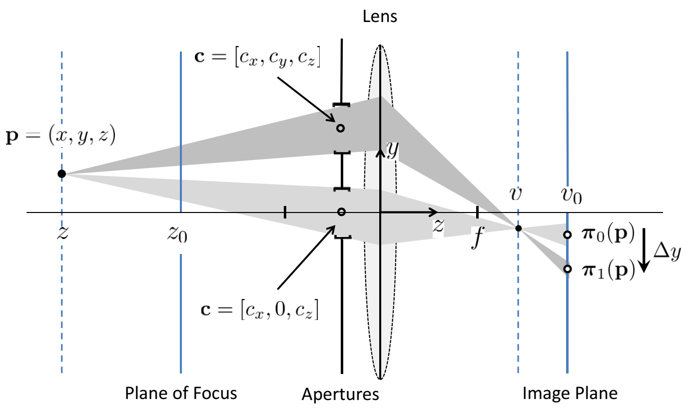


Fig. 2. Off-Axis Imaging: The effect of an off-axis aperture on the projection of a point onto the image plane

the shift will be a function of the distance of the object from the camera. More specifically, suppose that the center of the aperture is located at $\mathbf{c} = (c_x, c_y, c_z)$. If $\mathbf{p} = (x, y, z)$ is a point to the left of the lens and $\boldsymbol{\pi}(\mathbf{p}) = (\pi_x(v_0), \pi_y(v_0))$ is the projection of this point onto the image plane at v_0, then [6], [7].

$$\pi_x(v_0) = -v\frac{x}{z} + \left(1 - \frac{v_0}{v}\right)\frac{c_x z - c_z x}{z - c_z} \tag{1}$$

$$\pi_y(v_0) = -v\frac{y}{z} + \left(1 - \frac{v_0}{v}\right)\frac{c_y z - c_z y}{z - c_z} \tag{2}$$

Note that when $\mathbf{p}$ is in the plane of focus at z_0, then $v = v_0$ and, independent of the location of the aperture, the projection will be at

$$\boldsymbol{\pi}(\mathbf{p}) = -\frac{v}{z}(x, y)$$

which is the same as the perspective projection of $\mathbf{p}$ for a pinhole camera. However, when $\mathbf{p}$ is not in the plane of focus, then the projection will depend on the location of the aperture and the distance of the point $\mathbf{p}$ from the lens. In addition, the point $\mathbf{p}$ will generate a blur disk around the projected point $\boldsymbol{\pi}(\mathbf{p})$ with a diameter b that is approximately [8]

$$b \approx d\,\frac{|z - z_0|}{z_0}\frac{f}{|z - f|} \tag{3}$$

where d is the diameter of the aperture.

2.2 Image Shifting Due to Aperture Displacements

When a camera is configured with two or more apertures, then each aperture will, in general, project points in the object plane to different points in the image plane. More specifically, suppose that one aperture is at $\mathbf{c}_1 = (c_x, c_y, c_z)$ and another is displaced a distance Δy along the y-axis to $\mathbf{c}_2 = (c_x, c_y + \Delta c_y, c_z)$. From Eq. (2) it follows that the projections of the point $\mathbf{p} = (x, y, z)$ shown in Fig. 2 will be a distance Δy away from each other along the y-axis in the image plane, where

$$\Delta y = \left(1 - \frac{v_0}{v}\right) \frac{z}{z - c_z} \Delta c_y \tag{4}$$

Note that if $\mathbf{p}$ is in the plane of focus, then $v = v_0$ and the projected points will will be the same. However, when $z > z_0$ (the point $\mathbf{p}$ is at a distance greater than the plane of focus), then $v < v_0$ and $\Delta y < 0$. On the other hand, when $z < z_0$ (the point $\mathbf{p}$ is closer to the lens than the plane of focus), then $v > v_0$ and $\Delta y > 0$. Since

$$1 - \frac{v_0}{v} = 1 - \frac{z_0}{z} \frac{z - f}{z_0 - f} = \frac{f}{z} \frac{z_0 - z}{z_0 - f} \tag{5}$$

then substituting this relationship into Eq. (4) gives

$$\Delta y = f \frac{z_0 - z}{(z_0 - f)(z - c_z)} \Delta c_y \tag{6}$$

If $z \gg c_z$ and $z \gg f$, then

$$\Delta y \approx f \left(\frac{1}{z} - \frac{1}{z_0}\right) \Delta c_y \tag{7}$$

By symmetry, if the apertures are separated by a distance Δc_x along the x-axis, then there will be an equivalent relationship for the distance Δx between the two projected points along the x-axis.

2.3 Converting Image Shifts from Millimeters to Pixels

If c_y, z, and z_0 are expressed in meters in Eq. (6), then the change in the location of the projection, Δy, will also be in meters. To express Δy in pixels, it is necessary to know what type of sensor is used in the camera. For a camera with an $N_1 \times N_2$ array of pixels and an image sensor that is $W \times H$ mm in size, then the distance between two pixels (in mm) will be

$$\alpha = \sqrt{\frac{W \cdot H}{N_1 \cdot N_2}} \quad \text{mm} \tag{8}$$

and the expression for Δy, measured in pixels, becomes

$$\Delta y = \frac{f}{\alpha} \frac{z_0 - z}{(z_0 - f)(z - c_z)} \Delta c_y \tag{9}$$

A plot of Δy versus z using Eq. (9) is shown in Figure 3 for a 10 megapixel camera (3872×2592) with an APS-C sensor of size 25.2×16.7 mm, a 150 mm lens, a plane of focus that is set to 100 meters, and an aperture shift of 28 mm. Note that the amount that a point moves in the image plane for each meter it moves in the object plane increases significantly as the object gets closer to the camera, a relationship that is well-known in stereo imaging.

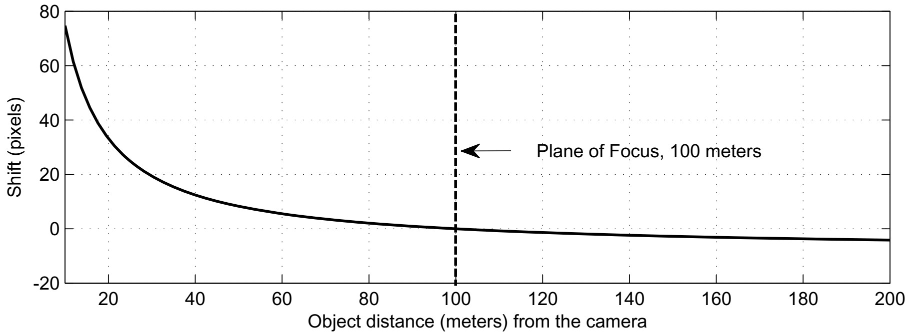


Fig. 3. The amount of shift that occurs in the image plane of a 10 megapixel camera as a function of the distance of an object from the camera for an aperture displacement of 28 mm with a plane of focus set at 100 meters.

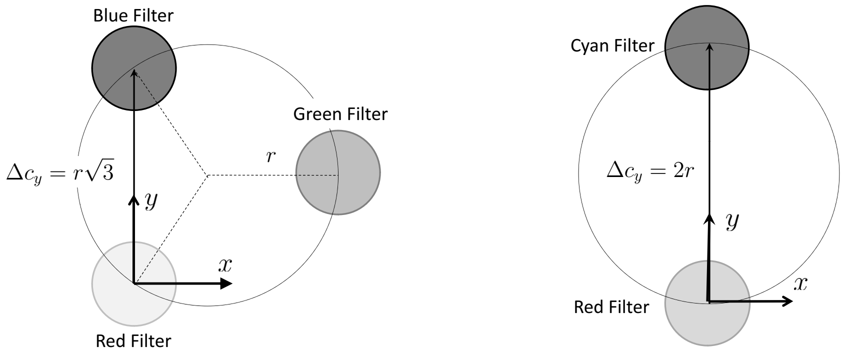
3 Distance Estimation

Equations (1) and (2) show how a point $\mathbf{p} = (x, y, z)$ in the object plane will be projected onto the image plane with an off-axis aperture. Equation (9) shows how much a projected point will move along the y-axis when the aperture is moved a distance Δc_y along the y-axis. In the following subsections, we describe how Eq. (9) may be used to estimate the distance of an object using a multi-aperture camera, discuss the camera calibration that is required, and examine the resolution of the distance estimates that are produced using such a camera.

3.1 Color Channels and Aperture Geometry

In most digital color cameras, a color filter array is placed over the pixel sensors to capture color information. The most common is the Bayer array consisting of red, green, and blue filters that generate three channels of color data. Therefore, if each color channel is imaged through a different off-axis aperture with a color filter that is matched to the color of the pixel sensor filter, then objects in the red, green, and blue channels will be shifted with respect to each other and the amount of the shift will be a function of the distance the object is from the camera. Thus, by finding these color shifts, the distances of objects from the camera may be estimated. Consider, for example, the three-aperture geometry shown in Fig. 4(a) where the red, green, and blue filtered apertures are moved radially a distance r away from the optical axis [2].

The three apertures form an equilateral triangle, and the distance between each aperture is $r\sqrt{3}$. If an object at a distance z is captured by this camera, since the blue and red apertures are shifted along the y-axis, then the object in the blue channel will be shifted with respect to the object in the red channel along the y-axis by an amount given in Eq. (9). Therefore, if the correspondence between points in the blue and red images can be found for a point $\mathbf{p}$ on the



(a) A three aperture system. (b) A two aperture system.

Fig. 4. Placement of color filter apertures using (a) three apertures with red, green, and blue filters and (b) two apertures with red and cyan filters

object, then the difference in the locations of the projected points, Δy, provides information that is sufficient to estimate the distance of the point $\mathbf{p}$ from the camera. Specifically, solving Eq. (9) for z we have

$$z = \frac{z_0 f \Delta c_y + c_z \alpha \Delta y (z_0 - f)}{f \Delta c_y + \alpha \Delta y (z_0 - f)} \tag{10}$$

Note that the distance may also be estimated by finding the relative displacements of an object between the red and green channels or between the blue and green channels. Although these apertures are displaced by the same distance with respect to each other, the shifting in the image plane will be along different lines. In order to increase the accuracy of the estimate, the distance estimates produced from each pair of color channels may be averaged,

$$z = \tfrac{1}{3} \sum_{k=1}^{3} \frac{z_0 f \Delta c_y + c_z \alpha \Delta y_k (z_0 - f)}{f \Delta c_y + \alpha \Delta y_k (z_0 - f)} \tag{11}$$

A dual aperture geometry is shown in Fig. 4(b). In this case, one aperture is covered with a red filter and the other with a cyan (green plus blue) filter, and the distance between the apertures is $\Delta c_y = 2r$. Since both green and blue are passed through the cyan filter, object distances may be estimated by either finding the relative displacements of an object between the red and green color channels or between the red and blue channels, or between both pairs and averaging the two displacements.

3.2 Calibration

Before Eq. (10) may be used to estimate the distance of an object, it is necessary to determine the camera parameters f, α, Δc_y, and c_z. For a fixed focal length camera, f will be given in the lens specification. If a zoom lens is used, an additional step of calibration would be required. The value of α that converts shifts in millimeters to shifts in pixels may be determined from the image sensor specifications as discussed in Sect. 2.3.

It is assumed that c_z, the location of the apertures along the z-axis, is the same for all apertures. In this case, the value of c_z may be found using a simple calibration procedure as follows. First, the camera is focused on an object at a known distance z_0 from the camera, thereby setting the plane of focus to a given value. (Note that the object will be in focus when the images in the three color channels are perfectly aligned.) Then, with two additional objects at different but known distances, z_1 and z_2, the shifts between two color channels of each object are found. Assume, for example, that the shifts between the blue and red channels are Δy_1 and Δy_2 for the first and second object, respectively. From Eq. (9), it follows that the ratio of these shifts is

$$\frac{\Delta y_1}{\Delta y_2} = \frac{(z_0 - z_1)(z_2 - c_z)}{(z_0 - z_2)(z_1 - c_z)}$$

Therefore, solving for c_z we have

$$c_z = \frac{\dfrac{z_0 - z_1}{z_0 - z_2} z_2 - \dfrac{\Delta y_1}{\Delta y_2} z_1}{\dfrac{z_0 - z_1}{z_0 - z_2} - \dfrac{\Delta y_1}{\Delta y_2}}$$

Once c_z is known, then Eq. (9) may be used to solve for Δc_y, the displacement between the red and blue apertures in the three-aperture system or between the red and cyan apertures in the dual-aperture camera. More specifically, using the object at distance z_1 with shift Δy_1, and solving Eq. (9) for Δc_y gives

$$\Delta c_y = \frac{\Delta y_1}{f} \frac{(z_0 - f)(z_1 - c_z)}{z_0 - z_1}$$

To increase the accuracy of the estimate of c_z, multiple objects at distances $z_1, z_2, \ldots, z_n$ with displacements $\Delta y_1, \Delta y_2, \ldots, \Delta y_n$ may be used, pairwise, to form estimates $c_z(1), c_z(2), \ldots, c_z(m)$ and these estimates may then be averaged,

$$c_z = \frac{1}{m} \sum_{k=1}^{m} c_z(k)$$

to produce the final value of c_z. Similarly, multiple objects may be used to form estimates of Δc_y, and an average of these estimates used for Δc_y.

For the three-aperture camera, the distance between the blue and green apertures and between the red and green apertures should be the same as the distance between the blue and red apertures. However, if necessary, these distances may be found using the same calibration procedure described above. The only thing that will change is that the shifts in the image plane will be in different directions.

3.3 Finding the Plane of Focus and Estimating the Color Shifts

Once the camera has been calibrated, Eq. (10) may be used to find the distance of an object from its displacement Δy in two color channels, provided that the plane of focus, z_0, is known.[1] Since z_0 is generally unknown and may change from one image to the next, it is necessary to find the plane of focus, and there are several ways that this may be done. One approach would be to set the plane of focus on an object that is a known distance, z_0, from the camera. Another approach would be to find the shift between the color channels of an object that is a known distance, z^*, from the camera, and solve Eq. (9) for z_0,

$$z_0 = f \frac{\alpha \Delta y(z^* - c_z) - z^* \Delta c_y}{\alpha \Delta y(z^* - c_z) - f \Delta c_y} \tag{12}$$

Once the plane of focus has been determined, the last step is to find the distance Δy between the projections of a point on an object whose distance is to be determined. This is equivalent to the stereo correspondence problem, and there are many approaches that may be used. Perhaps the simplest is to define a block of pixels around a projected point or an object of interest in one channel, and find the corresponding block in the other channel that maximizes the correlation between the two blocks. This approach is efficient for two reasons. First, only blocks along a given direction need to be searched since the shift is known to be in a direction that is defined by the geometry of the apertures. The shift between the red and blue channels in the three-aperture system, for example, is known to be along the y-axis. Secondly, if the distances of objects are known to lie within a given range, $z_{\min} \leq z \leq z_{\max}$, then this will place a limit the range of possible shifts, $(\Delta y)_{\min} \leq \Delta y \leq (\Delta y)_{\max}$. However, unlike typical stereo matching problems, the correspondence problem here is a bit more difficult because of the fact that when a block of pixels is separated into two color channels, and one channel is displaced with respect to the other, it is not always possible to find the correct disparity. Consider, for example, a block of 16×16 pixels with the upper half of the block being green and the lower half being red. When this block of pixels is separated into red and green color channels as illustrated in Fig. 5, there is an *apparent* shift $\Delta y = 8$ pixels even before one channel is shifted with respect to the other. The basic problem is that the three color channels of an image will generally have different intensities and, therefore, the brightness constancy property that is assumed in many disparity estimation approaches does not apply. Therefore, it is important to consider an approach that does not assume the constant brightness property, such as the elastic registration method proposed by Periaswamy [10]. Another approach that may be used is to identify key feature points of an object in the two channels and find the shift that does the best job of aligning the points [9]. Since the estimation of relative shifts of projected points in the image plane is not the focus of this paper, the reader is referred to the references for more details.

[1] For apertures not displaced along the y-axis, the shift will be estimated along the appropriate direction.

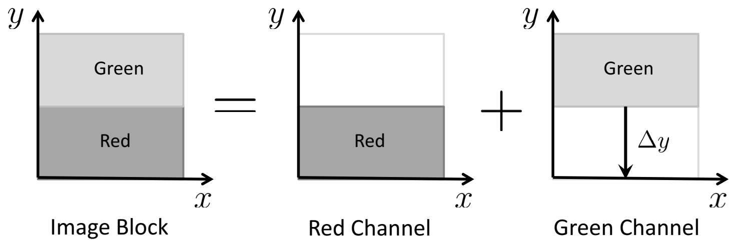


Fig. 5. Illustration of the difficulty in the correspondence problem when an image is separated into different color channels

3.4 Distance Resolution

To determine the accuracy of a distance estimate using a multiple aperture camera, we may differentiate Eq. (9) with respect to z as follows,

$$\frac{d}{dz}\Delta y = \frac{f}{\alpha}\frac{c_z - z_0}{(z - c_z)^2(z_0 - f)}\Delta c_y$$

which gives the number of pixels that the projection of an object in the image plane will move for each meter that the object moves along the z-axis. The *resolution* of a distance estimate may then be defined to be the magnitude of the inverse of this derivative,

$$\text{Res}(z, \Delta c_y) = \left|\frac{d}{dz}\Delta y\right|^{-1} = \frac{\alpha}{f}\frac{(c_z - z)^2|z_0 - f|}{|c_z - z_0||\Delta c_y|} \quad \text{meters/pixel}$$

which is the distance in meters that an object must move to produce a shift of one pixel in the image plane. Assuming that $z \gg f$ and $z \gg c_y$, the resolution is approximately

$$\text{Res}(z, \Delta c_y) \approx \frac{\alpha}{f}\frac{z^2}{|\Delta c_y|} \quad \text{meters/pixel}$$

Note that the resolution is inversely proportional to Δc_y, a relationship that is well-known in stereo imaging. Specifically, with a pair of cameras, as the distance between the camera lenses increases, the disparity increases, which implies that a more precise distance measurement may be found. A plot of the resolution as a function of z is shown in Figure 6 for a 10 megapixel camera with an APS-C sensor of size 25.2×16.7 mm, and a 150 mm lens, when the plane of focus is set to 100 meters and the apertures are separated by a distance of 28 mm. Note that an object at 100 meters must move fifteen meters to produce a shift of one pixel, whereas an object at 20 meters must move only 0.604 meter.

4 Two Aperture versus Three Aperture Cameras

In Section 3, two different multiple aperture geometries were presented. The first consists of three color filtered apertures (red, green, and blue) that are moved

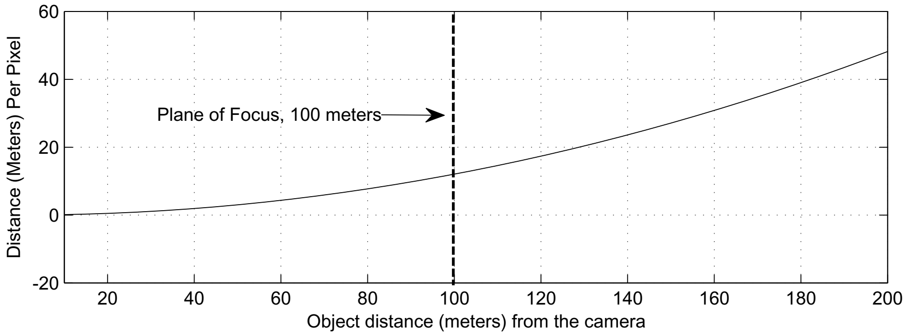


Fig. 6. Distance resolution. The number of meters an object must move as a function of z to produce a change of one pixel in Δy for an aperture displacement of 28 mm.

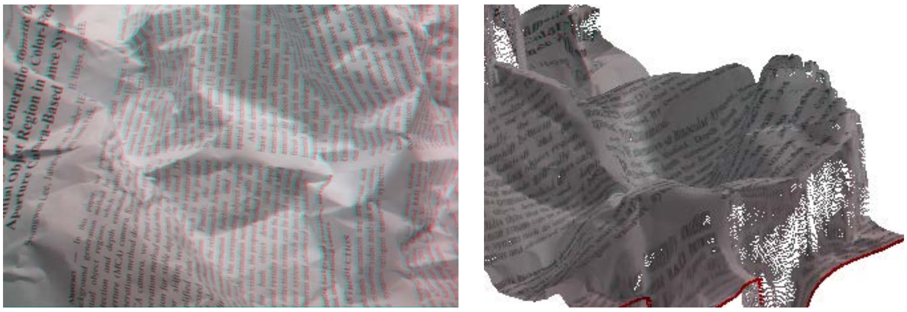
radially a distance r from the optical axis at angles of 120° with respect to each other. The second has two color filtered apertures (red and cyan) that are moved in opposite directions a distance r away from the optical axis. Both geometries may be used to estimate the distances of objects from the camera, but each one has its own advantages and disadvantages. For example, a three aperture camera produces lower resolution distance estimates than a dual-aperture camera due to the fact that the distance between each pair of apertures in a three-aperture camera is smaller than the equivalent dual-aperture camera. For example, if the maximum distance between two apertures in a dual-aperture camera is $2r$, then the maximum distance will be $r\sqrt{3}$ in a three aperture camera. However, with a three-aperture system, three independent distance estimates may be found from each pair of apertures (red and green, red and blue, and blue and green), whereas for the dual-aperture camera. only two independent estimates may be found - one from the shift between the red and blue color channels, and one from the red and green color channels (recall that the cyan filter passes both blue and green). Another advantage of the three-aperture geometry is that disparities along two orthogonal axes may be estimated, whereas for the dual-aperture camera the disparity along only one axis may be found. This has implications for objects that are aligned with the axis of the dual-aperture camera such as an extended wall or fence. Finally, an interesting feature of the dual-aperture camera is that it may be used to create a 3-D image that may be viewed using a pair of anaglyphic glasses [7].

5 Examples

Shown in Fig. 7 are two frames from video sequences that were captured using a three-aperture camera and used to estimate the distance of an object (person) from the camera. Although this is an example of some preliminary results, our current and future work is focused on real-time distance estimation, methods

(a) (b)

Fig. 7. Frames from a video sequence used to estimate the distance of objects from the camera using a three aperture system



(a) (b)

Fig. 8. (a) The image of a crumpled piece of paper using a dual color filter array camera, (b) the estimated 3-D depth map

for evaluating and improving the accuracy of the distance estimates, and incorporating a Kalman filter to help in the distance estimation. Another example is shown in Fig. 8(a) where a dual-aperture camera with red and cyan colored filters is used to form an image of a crumpled piece of paper. Using a pair of anaglyphic glasses, a 3-D image of the crumpled piece of paper may be viewed (it is best to expand the image to a larger size for best viewing). An estimate of the shifting that occurs in the image plane is illustrated in Fig. 8(b) and the reconstruction of a depth map of the image is shown in Fig. 8(c) [4].

6 Conclusions

In this paper, we have considered the modification of an SLR camera by adding a two or three color filter aperture into the lens of a camera that create displacements of objects in the image plane that are a function of the distance of the object from the camera. The focus was on the relationship between the distance of an object and the amount of shift that occurs in the image plane, and a discussion of the resolution that is possible for a given aperture separation. Two examples were given to demonstrate that such a camera might provide a simple and effective way to either create a 3-D image or to estimate the distances of objects within the field of view of the camera.

References

1. Koh, K., Kuk, J.G., Jin, B., Choi, W., Cho, N.: Autofocus Method Using Dual Aperture and Color Filters. Journal of Electronic Imaging 20(3) (July 2011)
2. Kim, S., Lee, E., Hayes, M., Paik, J.: Multifocusing and Depth Estimation Using a Color Shift Model-Based Computational Camera. IEEE Trans. on Image Processing 21(9), 4152–4166 (2012)
3. Amari, Y., Adelson, E.H.: Single-Eye range Estimation by Using Displaced Apertures with Color Filters. In: Proc. 1992 Int. Conf. on Industrial Electronics, Control, Instrumentation, and Automation, vol. 3, pp. 1588–1592 (November 1992)
4. Lee, S., Kim, N., Jung, K., Hayes, M., Paik, J.: Single Image-Based Depth Estimation Using Dual Off-Axis Color Filtered Aperture Camera. In: Proc. 2013 Int. Conf. on Acoustics, Speech, and Sig. Processing, Vancouver, Canada (May 2013)
5. Hecht, E.: Optics. Addison-Wesley (2001)
6. Dou, Q., Favaro, P.: Off-axis aperture camera: 3D shape reconstruction and image restoration. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1–7 (2008)
7. Lee, S., Paik, J., Hayes, M.: Stereo Image Capture and Distance Estimation with an SLR Digital Camera. In: Proc. 15th IASTED International Conference on Signal and Image Processing, Banff, Canada (July 2013)
8. Bae, S., Durand, F.: Defocus Magnification. Eurographics 2007 26(3) (2007)
9. Maik, V., Cho, D., Shin, J., Har, D., Paik, J.: Color shift Model-Based Segmentation and Fusion for Digital Auto-focusing. J. Imaging Science Tech. 51, 368–379 (2007)
10. Periaswamy, S., Farid, H.: Elastic Registration in the Presence of Intensity Variations. IEEE Trans. on Medical Imaging 32(7), 865–874 (2003)