

STEREO IMAGE CAPTURE AND DISTANCE ESTIMATION WITH AN SLR DIGITAL CAMERA

Seungwon Lee, Joonki Paik, and Monson Hayes
Chung-Ang University
Graduate School of Advanced Imaging Science
Seoul, South Korea
mhh3@gatech.edu

ABSTRACT

In this paper, we consider the use of multiple color filter apertures in an SLR digital camera to produce stereo images and to estimate the distance of objects within the field of view of the camera. Although such a configuration has been proposed for auto-focusing, multi-focusing, and distance estimation, here we develop the relationship between the amount of color-shifting that occurs with off-axis apertures with color filters, and derive a relationship that specifies the resolution of the distance estimates that are produced with this camera. We also discuss an approach that may be used for distance calibration, and consider the feasibility of using a two-color or a three-color filter aperture SLR camera in applications such as robot vision, human computer interfaces, intelligent visual surveillance, 3-d image acquisition, and intelligent driver assistance systems.

KEY WORDS

Distance, Estimation, Stereo, SLR, Resolution

1 Introduction

In a 1908 Session of the Academy of Sciences, Gabriel Lippmann posed the following question:

“Is it possible to create a photographic print in such a manner that it represents the exterior world framed, in appearance, between the boundaries of the print, as if those boundaries were that of a window opened on reality?”

His answer was:

“It appears that yes, we can request from Photography infinitely more than from the human hand. Here I will attempt to provide a solution to this problem.

Since that time, photography and, more generally, image and video capture, has advanced considerably, and has moved away from the capture of images on photographic prints to the capture of visual information digitally.

An important component for any approach to produce an image that will “appear as a window opened on reality” is the capture and display of 3-d information or the capture of the light field from which 3-d images may be generated. There are different approaches that may be used to

accomplish this goal, such as the use of stereo cameras that are designed to mimic the human visual system, plenoptic (light field) cameras that use a microlens array to capture the incoming light field [1,2], and cameras with modified apertures [3,4].

In many applications, such as robot vision, human computer interfaces, intelligent visual surveillance, 3-d image acquisition, and intelligent driver assistance systems, it is necessary to go beyond 3-d image capture and estimate the absolute distance of objects in a scene or the relative distance between two or more objects. Again, depending upon the system that is used, there are many different approaches that may be used, 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 image information and the estimation of the distances of objects using a standard SLR camera with two or three color filter apertures. Multiple Color filter Aperture (MCA) cameras have been used for autofocus [5], multifocusing [4], and distance estimation [6,7]. Here, the focus is on the relationship between the color-shift vectors as a function of the color filter configuration, the resolution of the distance estimates produced by such a camera, and the calibration of the system.

2 Multiple 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 by some distance from each other and that capture 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 different systems of varying complexity that capture stereo imagery. These cameras range from cameras for the hobbyist, such as the Fujifilm FinePix 3D Digital Camera or lenses that turn digital SLR cameras into 3-d cameras, 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 may allow for the capture of stereo (3-d) images and provide the ability to estimate the distance of objects within the field of

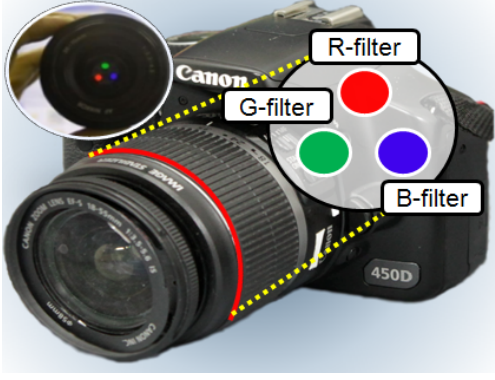


Figure 1: An SLR camera with a three aperture color filter array.

view of the camera. One such system is the multiple aperture camera with color filters shown in Figure 1. Depending on the configuration and the application, the camera may have two or three apertures. The effect of the apertures, which are displaced some distance away from the optical axis, is to shift the image of an object in the image plane by an amount that is a function of the distance of the object from the plane of focus of the camera. As a result, estimating this displacement allows one to estimate the distance of the object from the camera. Also, with the appropriate choice of colors for the apertures, the three color plane images captured by the Bayer array may be used to generate a stereo image pair for 3-d image rendering or viewing. The following section describes the effect of the off-axis apertures.

2.1 Off-Axis Imaging

For an imaging system represented by a single lens with a focal length of f and an on-axis aperture centered at $\mathbf{c} = (0, 0, 0)$ that coincides with the vertex of the lens, if the image plane of the camera is at a distance v_0 to the right of the vertex, then Gauss' thin lens equation [8] says that the plane of focus will be at a distance z_0 where

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

Thus, an object at a distance

$$z_0 = \frac{fv_0}{v_0 - f}$$

from the lens will be in focus in the image plane at v_0 . Similarly, if an object is located at a distance z from the lens, it will be in focus if the image plane is at

$$v = \frac{fz}{z - f}$$

However, if the image plane is at v_0 , then the object will be blurred, and the circle of confusion, which is the diameter

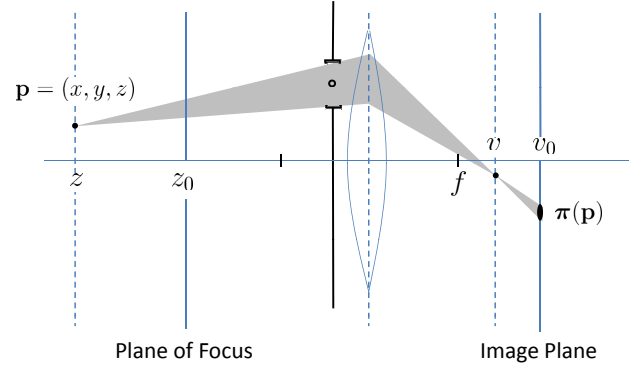


Figure 2: Off-Axis Imaging: The effect of an off-axis aperture on the projection of a point onto the image plane

of the blur, is approximately [9]

$$b = d \frac{z - z_0}{z_0} \frac{f}{z - f} \quad (1)$$

where d is the diameter of the aperture.

Now consider the imaging system illustrated in Fig. 2 where the center of the aperture has been shifted to $\mathbf{c} = (c_x, c_y, c_z)$. Let $\mathbf{p} = (x, y, z)$ be some point on an object to the left of the lens, and let $\pi(\mathbf{p}) = (\pi_x(v_0), \pi_y(v_0))$ be the projection of this point onto an image plane located at v_0 . The (x, y) -coordinates of $\pi(\mathbf{p})$ are given by [9]

$$\pi_x(v_0) = -v \frac{x}{z} + \left(1 - \frac{v_0}{v}\right) \frac{c_x z - c_z x}{z - c_z} \quad (2)$$

$$\pi_y(v_0) = -v \frac{y}{z} + \left(1 - \frac{v_0}{v}\right) \frac{c_y z - c_z y}{z - c_z} \quad (3)$$

Note that if the point $\mathbf{p}$ is in the plane of focus at z_0 , then $v/v_0 = 1$ and, for any aperture location $\mathbf{c}$, the projection will be at

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

which is the same as the perspective projection of $\mathbf{p}$ for a pinhole camera. However, when the point $\mathbf{p}$ is not in the plane of focus, then the projection will depend on the location of the aperture. In addition, the point $\mathbf{p}$ will generate a blur disk around the projected point $\pi(\mathbf{p})$ with a diameter given approximately by Eq. (1).

2.2 Relative Shift Versus Object Distance

In order to see how a camera with two or more color filter apertures may be used to generate a stereo image pair or to estimate the distance of an object in the scene, it is necessary to look at how projected points in the image plane move as the aperture is shifted away from the optical axis. Therefore, consider two apertures, one that is centered at $\mathbf{c}_1 = (c_x, c_y, c_z)$ and the other that is moved a distance Δc_y along the y -axis away from the first aperture to $\mathbf{c}_2 = (c_x, c_y + \Delta c_y, c_z)$. As seen from Eqs. (2) and (3),

this shift results in a shift along the y -axis of the projection $\pi(\mathbf{p})$ of a point $\mathbf{p}$ by an amount

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

From this, note that when the point $\mathbf{p}$ is in the plane of focus of the optical system, $\mathbf{p} = (x, y, z_0)$, then $v = v_0$ and the projection will be the same for both apertures. However, if $z > z_0$ (the point $\mathbf{p}$ is at a distance greater than the plane of focus), then $v < v_0$ and $\Delta y < 0$, which means that the projection in the image plane moves in a direction *opposite* to the direction of movement of the aperture. Similarly, if $z < z_0$ (the point $\mathbf{p}$ is closer to the lens than the plane of focus), then $v > v_0$ and $\Delta y > 0$, which means that the projection in the image plane moves in the same direction as the aperture. Since

$$1 - \frac{v_0}{v} = 1 - \frac{z_0}{z} \frac{z - f}{z_0 - f} = \frac{f}{z} \frac{z_0 - z}{z_0 - f} \quad (5)$$

substituting this into Eq. (4) we have

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

which gives the amount that the object in the image plane moves, Δy , for a given aperture displacement, Δc_y , as a function of the distance the point $\mathbf{p}$ is away from the plane of focus of the lens, $z_0 - z$. If we assume that $z \gg c_z$ and $z \gg f$, then the amount that the projection moves is approximately

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

By symmetry, if the aperture is moved along the x -axis from $\mathbf{c}_1 = (c_x, c_y, c_z)$ to $\mathbf{c}_2 = (c_x + \Delta c_x, c_y, c_z)$, then there will be a shift in the projection along the x -axis in the image plane, and the amount of the shift will be

$$\Delta x = f \frac{z_0 - z}{(z_0 - f)(z - c_z)} \Delta c_x \quad (8)$$

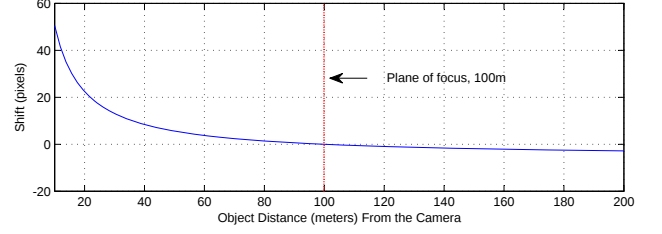
Again assuming that $z \gg c_z$ and $z \gg f$, the projection moves approximately

$$\Delta x \approx f \left(\frac{1}{z} - \frac{1}{z_0} \right) \Delta c_x \quad (9)$$

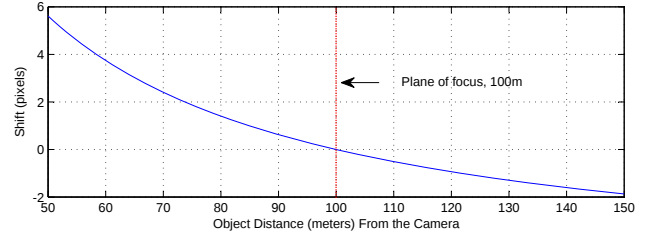
2.3 Conversion of Shift Vectors from Millimeters to Pixels

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. In order to express Δy in pixels, it is necessary to know what type of sensor is used in the camera. For an N megapixel camera with a ccd array that is $W \times H$ mm in size, assuming a standard rectangular array of square sensors, the distance between two pixels (in mm) is

$$\alpha_x = 2^{-10} \sqrt{\frac{W \cdot H}{N}} \quad (10)$$



(a) Object at a distance from 10m to 200m.



(b) Object at a distance from 50m to 150m.

Figure 3: A plot of the amount of shift versus the distance of an object from the camera when the plane of focus is set at 100 meters with an aperture displacement of 28mm.

and the expression for Δy , measured in pixels, becomes

$$\begin{aligned} \Delta y &= \frac{f}{\alpha_x} \frac{z_0 - z}{(z_0 - f)(z - c_z)} \Delta c_y \\ &\approx \frac{f}{\alpha_x} \left(\frac{1}{z} - \frac{1}{z_0} \right) \Delta c_y \end{aligned} \quad (11)$$

For example, for a 10 megapixel camera (3872×2592) with an APC-C sensor (25.2×16.7 mm), the pixel size is

$$\alpha_x = 0.0063 \text{ mm}$$

and Δy (in pixels) becomes

$$\Delta y \approx 157.85 f \left(\frac{1}{z} - \frac{1}{z_0} \right) \Delta c_y \quad (12)$$

A plot of Δy versus z using the approximation given in Eq. (12) is shown in Figure 3(a) assuming a 150 mm lens, a plane of focus that is set to 100 meters, and a displacement of 28 mm between the two apertures. Shown in Figure 3(b) is an expanded view showing the amount of shift for objects that are within 50 meters of the plane of focus. Note that as a point $\mathbf{p} = (x, y, z)$ moves a distance Δz along the z -axis, the amount that its projection moves in the image plane increases significantly as the z becomes small (the object is close to the camera). This, of course, is well-known from stereo imaging principles.

3 Distance Estimation with a Multiple Aperture Camera

Equations (2) and (3) specify how a point $\mathbf{p} = (x, y, z)$ in the object plane will be projected onto the image plane

with an off-axis aperture. Equations (6) and (7) show how much the projection of a point will shift in the image plane when the aperture is moved a distance Δc_y along the y -axis. Therefore, with two displaced apertures the relative shift between the two projections of a point $\mathbf{p}$ on an object onto the image plane may be used to estimate the distance of the object from the camera.

In a digital camera with a Bayer array, three images are naturally formed in the red, green, and blue color channels. If each color plane is formed using an aperture at a different location, then from the relative shifts of the projected points onto the image plane, it is possible estimate the distance of an object. One way to do this is with the Multiple Color Filter (MCA) camera that is described in [4]. The geometry of the apertures in the MCA camera is shown in Fig. 4 where each color aperture is moved radially away from the optical axis by an amount r . The color filters form an equilateral triangle, so the angle between the two lines that connect the optic center to two different apertures is 120° . Therefore, the distance between any pair of apertures is $\sqrt{3}r$. If an object at a distance z is imaged with the MCA camera, since the blue and red apertures are shifted along the x -axis, then the object in the blue channel will be shifted with respect to the object in the red channel along the x -axis. Therefore, if the correspondence of a point $\mathbf{p}$ can be made between the blue and red color channels, then the shift vector will provide the information that is necessary to recover the distance of the point $\mathbf{p}$ from the camera. Specifically, if the shift is Δx , then

$$\Delta x = f \frac{z_0 - z}{(z_0 - f)(z - c_z)} \Delta c_x \approx f \left(\frac{1}{z} - \frac{1}{z_0} \right) \Delta c_x$$

where $\Delta c_x = \sqrt{3}r$. Solving for z we have

$$z \approx f \frac{z_0 \Delta c_x}{f \Delta c_x - z_0 \Delta x}$$

Note that the distance may also be estimated from the color shift vectors between the red and green channels and the blue and green channels.

If z_0 is unknown, even though z cannot be determined, the relative distance between two points $\mathbf{p}_1$ and $\mathbf{p}_2$ may be found. Specifically, if one object is at a distance z_1 and another at z_2 , then

$$\frac{1}{z_i} = \frac{1}{z_0} + \frac{1}{f} \frac{\Delta y_i}{\Delta c_y}$$

and the difference in the reciprocal distance (along the z -axis) between these objects is

$$\frac{1}{z_2} - \frac{1}{z_1} = \frac{\Delta y_2 - \Delta y_1}{f \Delta c_y}$$

Another configuration [7] is to use two color filter apertures that are separated a distance $\pm \Delta c_y$ away from the optical axis with one being a red filter and the other being cyan (blue plus green). The advantages of this geometry is that (a) a 3-d image may be viewed using a pair of

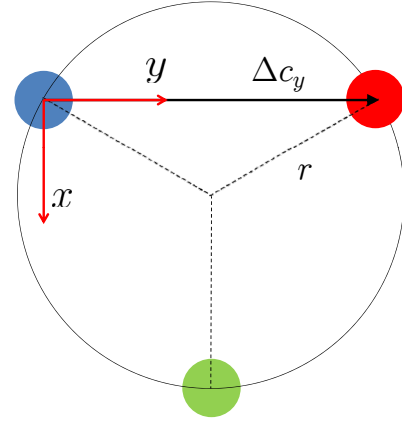


Figure 4: Configuration of the color filters in the MCA camera.

anaglyphic glasses and (b) the maximum possible distance between the apertures will be larger than between any two apertures in the MCA camera.

4 Distance Resolution

In order to determine the accuracy of a distance estimate that is derived from the relative displacement of two projections, we find the derivative of Δy with respect to z . Differentiating Eq. (11) with respect to z , and again assuming that $z \gg f$ and $z \gg c_y$, we have

$$\frac{d}{dz} \Delta y \approx \frac{f}{\alpha_x} \frac{1}{z^2} \Delta c_y$$

This expression gives the number of pixels that an object moves in the image plane per meter that it moves in the object plane along the optical axis. Thus, we may define the *resolution* to be the inverse of this,

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

which is the distance in meters that an object must move to produce a shift in the image plane of one pixel. Note that resolution is inversely proportional to Δc_y , which is well-known from stereo imaging where an increase in the distance between two cameras in a stereo camera pair will result in better resolution (higher accuracy) in distance estimation. A plot of the resolution versus z is shown in Figure 5(a) when $\Delta c_y = 28$ mm, $f = 150$ mm, and $\alpha_x = 0.0063$ mm. Thus, we see that an object at 100 meters must move 15 meters in order to produce a shift of one pixel. Therefore, if we can estimate the movement of the object in the image plane to a *quarter pixel accuracy* then this would imply that the distance estimate of an object would be accurate to within ± 3.75 meters when the object is at a distance of 100 meters. Note that, by comparison, an object at a distance of 200 meters must move 60.3 meters to produce a one pixel shift whereas an object at 20

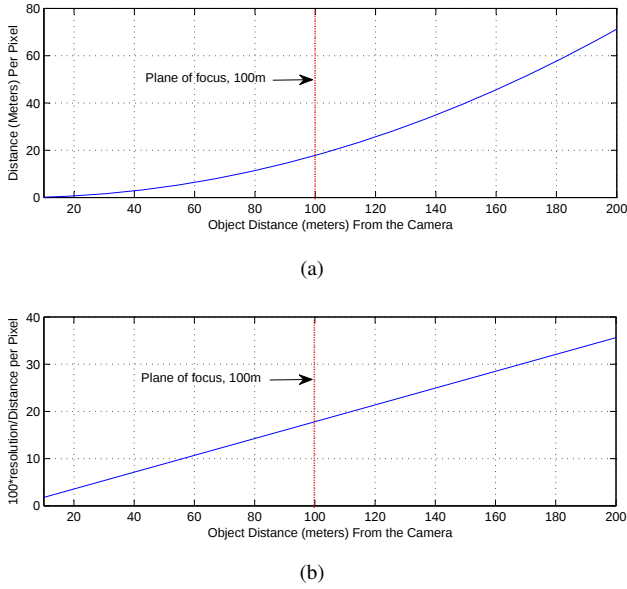


Figure 5: Resolution. (a) Distance resolution, number of meters per pixel in the image plane versus the distance of an object from the camera, and (b) the resolution as a percentage of the distance that is being estimated.

meters must move only 0.604 meter. Shown in Fig. 5(b) is a plot of $100 \times \text{Res}(z, \Delta c_y)/z$ which shows the accuracy of a distance estimate as a percentage of the distance that is being estimated. Thus, for example, at 100 meters a distance estimate will be accurate to within 15% for single pixel resolution whereas for an object at 40 meters it will be accurate to within 7.5%.

5 Calibration

The color shift vectors estimated from the MCA camera provide an estimate of the distance z^{-1} of the object relative to the inverse of the distance z_0^{-1} of the plane of focus, which is generally unknown. Therefore, if absolute distance estimates are required, then it is necessary to estimate z_0 or find a way to establish the scale along the x -axis of the curve in Figure 3. There are several ways that this can be done, and perhaps the most simple one is to use what we refer to as *single point calibration*. This approach works as follows. Since only one point on the Δy versus z curve is necessary in order to calibrate the camera and define the scale, an object may be placed at a known distance, z^* , from the camera. Once the color shift vector for this object has been determined, then the distance z_0 may be found by solving Eq. (7) for z_0 ,

$$z_0 = f \frac{z^* \Delta c_y}{f \Delta c_y - z^* \Delta y} \quad (13)$$

Since the resolution of the distance estimate is higher for shorter distances, an object such as a pole or a box with red, green, and blue stripes may be positioned a short distance

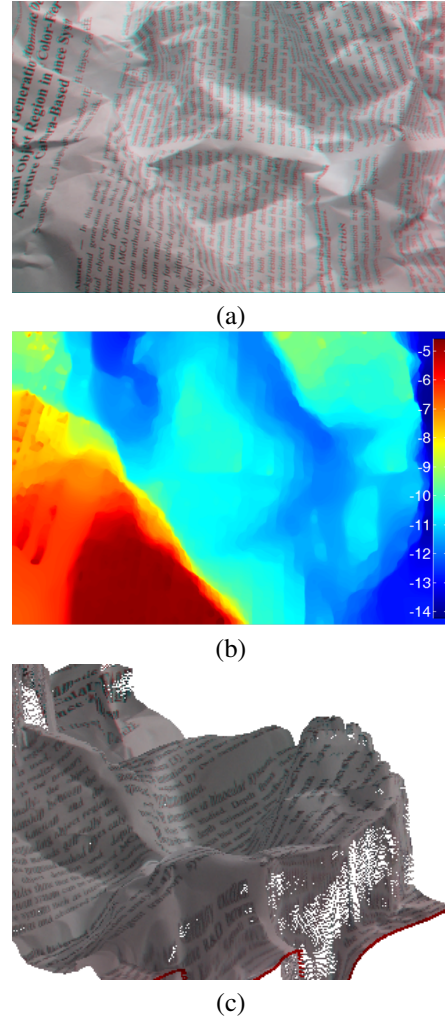


Figure 6: (a) The image of a crumpled piece of paper using a dual color filter array camera, (b) the estimated color shift map, and (c) a 3-d depth map.

from the camera, such as $z^* = 20$ meters when the camera is focussed at approximately 100 meters. Estimating the color shift vector for this object and substituting this value into Eq. (13) completes the calibration.

6 Example

Using a pair of apertures, one red and one cyan, the image of a crumpled piece of paper is shown in Fig. 6(a) and an estimate of the shifting that occurs in the image plane is illustrated in Fig. 6(b). The shifts are estimated using an L_1 minimization of an energy functional that considers both the brightness and gradient constancy and assumes a single-axis translation of objects in the image plane [7]. The reconstruction of a 3-depth map is shown in Fig. 6(c). With this example, we see that it is feasible to estimate a depth map or to estimate the distance of objects using a color filter array camera, but the resolution of the distance

estimates is limited, and may only be feasible for estimating distances of objects that are within a few hundred meters of the camera.

7 Conclusions

In this paper, we have looked at modifying an SLR camera by adding a color filter aperture to the lens, which creates a displacement of objects in the image plane as a function of the distance of the object from the camera. The focus was on the relationship between distance and the amount of shift that occurs in the image plane, and a discussion of the distance resolution that is possible for a given aperture separation. An example was given to demonstrate that, in some cases, such a camera might provide a simple and effective way to either create a 3-d image or to estimate the distances of objects.

References

1. E. Adelson and U. Wang, "Single Lens Stereo with a Plenoptic Camera," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 99-106, Feb. 1992.
2. C. Zhou and S. Nayar, "Computational Cameras: Convergence of Optics and Processing," *IEEE Trans. on Image Processing*, vol. 20, no. 12, pp. 3322-3340, Dec. 2011.
3. A. Levin, R. Fergus, R. Durand, and W. Freeman, "Image and Depth from a Conventional Camera with a Coded Aperture," *ACM Trans. Graphics*, 26, 3, Article 70, July 2007.
4. S. Kim, E. Lee, M.H. Hayes, and J. Paik, "Multifocusing and Depth Estimation Using a Color Shift Model-Based Computational Camera," *IEEE Trans. on Image Processing*, vol. 21, Issue 9, pp. 4152-4166, Sept. 2012.
5. K. Koh, J.G. Kuk, B. Jin, W. Choi, and N. Cho, "Autofocus Method Using Dual Aperture and Color Filters," *Journal of Electronic Imaging*, vol. 20, no. 3, July 2011.
6. Y. Amari and E.H. Adelson, "Single-Eye range Estimation by Using Displaced Apertures with Color Filters," *Proc. 1992 Int. Conference on Industrial Electronics, Control, Instrumentation, and Automation*, vol. 3, pp.1588-1592, vol.3, Nov 1992.
7. S. Lee, N. Kim, K. Jung, M. Hayes, and J. Paik, "Single Image-Based Depth Estimation Using Dual Off-Axis Color Filtered Aperture Camera," *Proc. 2013 Int. Conf. on Acoustics, Speech, and Sig. Proc.*, Vancouver, Canada, May 2013.
8. E. Hecht, *Optics*, Addison-Wesley, 2001.
9. Q. Dou and P. Favaro, "Off-axis aperture camera: 3D shape reconstruction and image restoration," *IEEE Conf. on Computer Vision and Pattern Recognition*, p. 1-7, 2008.