THE PHILIPPINE-AMERICAN Academy of Science & Engineering

Fuzzy Estimation of Depth Maps from Coded Aperture Imaging Using Color Filters

Cecille Adrianne Ochotorena[a*], Carlo Noel Ochotorena[a] and Elmer Dadios[b]

[a]ECE Department
De La Sale University, Manila, Philippines

[b]MEM Department
De La Salle University, Manila, Philippines

[*]E-mail: cecille.ochotorena@dlsu.edu.ph

ABSTRACT

Coded aperture photography has shown promise in a variety of applications, including that of depth extraction. By placing colored filters in place of the traditional iris in a lens system, the red, green, and blue color channels are optically translated resulting in a visible misalignment. Previous research in this area has successfully exploited the degree of misalignment to obtain depth estimates at various positions in the image. These estimates can then be used to separate foreground and background objects. However, due to the manner in which color sensors are constructed, the degree of misalignment of the color channels can be quite noisy which would likewise translate to noisy depth maps. This work proposes to calculate the depth maps using fuzzy rules to obtain cleaner estimates. With the availability of better estimates, a more reliable foreground-background separation may then be derived.

KEYWORDS: coded aperture; color filters; fuzzy logic; adaptive filter

1 INTRODUCTION

Depth information is of great interest in the fields of image processing and computer visionas it allows the creation of three-dimensional maps that are useful in charting unknown spaces. In the area of robotics, depth-awareness can be used to detect the presence of near and distant obstacles (Charalampous et al., 2015). This information can then be used to plan an obstacle-free path. The same is true in medical applications such as robotics-assisted surgery where precision in navigating through a given area is required (Ayache, 1995).

Given the significance of depth information, it is no surprise that an equally large amount of research has been done towards finding the correct depth information for a given scene. Following the observation that humans have a pair of eyes, which together allows for depth perception, many techniques have been developed to calculate the depth from stereoscopic images typically obtained using two cameras (Jones, Lee, Holliman, & Ezra, 2001; Barnard & Fischler, 1982) . Researchers have then extended this to multi-view imaging where more than two image sources are used to calculate the depth of a scene (Zamarin, Zanuttigh, Milani, Cortelazzo, & Forchhammer, 2010; Merkle, Smolic, Muller, & Wiegand, 2007) . While these works are generally successful in extracting depth, the use of more than one imaging sensor naturally increases the cost of deploying such systems. This leads to the more complex problem of monocular depth extraction.

Much like their stereoscopic and multi-view counterparts, there is also a visible research thrust towards depth calculation using single imaging sensors. A potential solution to the depth extraction problem is through the use of coded aperture imaging (Babacan et al., 2012; Marwah, Wetzstein, Bando, & Raskar, 2013; Levin,

Fergus, Durand, & Freeman, 2007; Veeraraghavan, Raskar, Agrawal, Mohan, & Tumblin, 2007) . In such systems, the traditional iris used as part of the optics in cameras is replaced with a fixed or variable pattern designed to extract information from the scene. In using a fixed coded aperture, various researchers have introduced changes in the blurring of the images, which can then be used to calculate the depth of various objects in a scene (Levin et al., 2007; Veeraraghavan et al., 2007). Variable coded aperture systems rely on the use of an electronically controlled mask such as a liquid crystal display (LCD) to introduce different patterns in place of the aperture (Babacan et al., 2012). Such techniques allow for the recovery of a complete four-dimensional light field, which can then be used to effectively calculate for the depth of the objects.

An interesting variation of the coded aperture approach is to use colored filters in place of the traditional aperture. This causes a depth-dependent misalignment of the red, green, and blue color channels, which can then be used as a measure for calculating the depth in the image. In the work by Bando, Chen, and Nishita (2008), the local alignment of the red, green, and blue pixels is calculated by taking a disparity measure of the neighborhood. In particular, this calculation requires the computation of covariance matrices for each pixel neighborhood and the eigenvalues associated with these matrices. Combined with graph cuts, the researchers demonstrated an effective way of dealing with depth estimation, which can be adapted for use with smaller sensors.

This work builds upon the work by Bando, Chen, and Nishita (2008) . In place of their disparity measure, a simpler gradient calculation is performed on the chrominance channels of the image. To refine the resulting gradients, a fuzzy filtering operation based on the bilateral filter is used on each of the energy maps. Finally, a depth map is calculated based on these refined energy maps. In the succeeding section, a brief discussion on the coded aperture topology and depth-dependent misalignment is provided. Following this, the calculation of the energy maps from the gradients is detailed along with how depth can be derived from these maps. Finally, the fuzzy bilateral operation is introduced in the last section of the discussion.

2 CODED APERTURE

An aperture, in an optical sense, describes an opening in the optical system, which may be used to control the amount of light entering the system. In cameras, a mechanical aperture is used to, likewise, regulate the amount of light reaching the imaging sensor. An intuitive way of analyzing this behavior is to consider the case of a pinhole camera as shown in Figure 1.

A pinhole camera generally projects an inverted copy of the image onto a parallel plane. In the given figure, the original scene is depicted on the left side with the pinhole plane in the middle. The image is then projected onto the imaging plane on the right. As shown in Figure 1*b*, a shift in the lateral position of the pinhole would likewise cause a shift in the projected image. When another object is placed in the scene, the new object is similarly projected onto the imaging plane as described by Figures 1*c* and 1*d*. As the new object is placed in front of the old object, the foreground object occludes the view of the background object. As with the previous case of a single object, a lateral shift in the position of the pinhole would result to a shift in the projection. However, due to the differences in depth, it becomes clear that the foreground and background objects do not experience the same amount of translation. This leads to the notion of depth-dependent translation.

$$I = \sum_{x} \sum_{y} \tag{1}$$

As the projections experience depth-dependent translations, the sum of these projections end up as the average of their respective neighborhood resulting to a blurred image. In combination with lenses, these translations can be adjusted such that a specific depth experiences no translations at all. Thus, the resulting projection for objects at this depth portrays a clear image of the object while those closer or farther appear blurred.

In the system proposed by Bando, Chen, and Nishita (2008) , colored filters are placed in place of the mechanical aperture in the manner shown in Figure 2. The effect of these filters can be intuitively analyzed

as placing three distinct apertures corresponding to red, green, and blue colors at different positions. It naturally follows that the resulting image would experience depth-dependent translations as well. However, unlike a pure aperture, the use of color filters result to translations of the individual color channels instead. This leads to the notion of depth-dependent misalignment of the color channels. Such misalignments may then be used to extract the depth information from the images captured using the color-coded aperture.

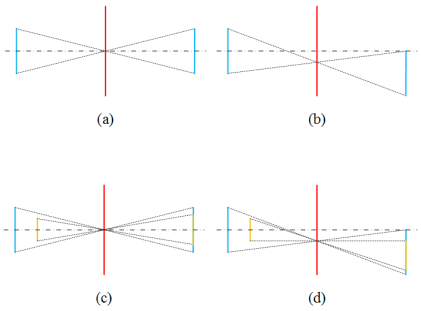

Figure 1: Top view of the projection through a pinhole aperture. (a) Single object through a centered aperture. (b) Single object through a laterally shifted aperture. (c) Two objects through a centered aperture. (d) Two objects through a laterally shifted aperture.

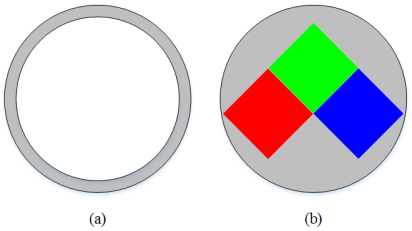

Figure 2: Aperture designs. (a) This demonstrates the traditional aperture. The gray area is completely opaque while the white region is unobstructed. (b) In the coded aperture topology, colored filters replace the unobstructed region of a traditional aperture.

3 MEASURES FOR ALIGNMENT

Once the sensor on the imaging plane has captured an image, any depth information from the original scene is effectively flattened onto a two-dimensional plane. In the case of the color-coded aperture system described previously, however, this depth information becomes encoded as translations of the color channels. To recover the depth for a pixel located at the coordinates (x,y), one has to find the solution to the following expression:

$$\{\hat{u}_r,\hat{v}_r,\hat{u}_b,\hat{v}_b\} = \underset{u_r,v_r,u_b,v_b}{\arg\min} \sum_{u_r}\sum_{v_r}\sum_{u_b}\sum_{v_b} E(x,y,u_r,v_r,u_b,v_b) \tag{2}$$

In essence, this amounts to testing all possible translation values (u_r,v_r) and (u_b,v_b) for the red and blue channels while keeping the green channel fixed until the best match for the three color channels can be found. If we assume that the aperture is correctly aligned, it becomes apparent that the vertical translation for the blue and red channels are practically identical while their horizontal translations are negations of each other assuming the green channel is taken as a reference point. With these assumptions, the translations can then be simplified to the expressions below.

$$u = u_r = -u_b \tag{3}$$

$$v = v_r = v_b \tag{4}$$

These adjustments allow us to reformulate the problem as such:

$$\{\hat{u},\hat{v}\} = \underset{u,v}{\arg\min} \sum_{u}\sum_{v} E(x,y,u,v) \tag{5}$$

While most of the earlier discussion focuses on simplifying the search, another important aspect of the problem itself is the energy function. In previous literature, the alignment of the color channels was calculated using a disparity measure.

$$d = \frac{\lambda_r \lambda_g \lambda_b}{\sigma_r^2 \sigma_g^2 \sigma_b^2} \tag{6}$$

The denominator takes the covariances σ_r, σ_g, σ_b, along the original red, green, and blue axes respectively. Similarly, by utilizing the eigenvalues λ_r, λ_g, and λ_b, the covariances are also calculated along the principal axes of the data. This measure was derived from the notion that colors in natural images cluster around color lines leading to small eigenvalues for two of the principal components. On the other hand, misalignment in the color channels would cause the colors in a local patch to spread leading to larger eigenvalues.

While the previous researchers have determined that disparity provides a good measure of the misalignment in the color channels, computing the eigenvalues for the local neighborhood of each pixel is quite an expensive task. Considering this calculation is performed for each candidate translation of a pixel, the computational costs rapidly grow. In this work, an alternative energy measure is given using simple gradient calculations.

One key observation with the coded aperture system is that the misalignment in colors causes rapid changes in the color content of an image patch. When mapped onto a suitable color space, these changes become more observable. For this work, it was found that the *YCbCr* color space provided a good separation between the intensities contained in the luminance channel and the color information in the chrominance components. In particular, taking the well-known Sobel filter and applying it to both the *Cb* and *Cr* components find the energy (Duda & Hart, 1973) .

$$E_b(x,y,u,v) = \sqrt{\left(C_b(x+u,y+v) * \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}\right)^2 + \left(C_b(x+u,y+v) * \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}\right)^2} \tag{7}$$

$$E_r(x,y,u,v) = \sqrt{\left(C_r(x+u,y+v) * \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}\right)^2 + \left(C_r(x+u,y+v) * \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}\right)^2} \tag{8}$$

These two measures are then consolidated by taking the maximum energy for a given pixel location.

$$E(x,y,u,v) = \max(E_r(x,y,u,v), E_b(x,y,u,v)) \tag{9}$$

Given this simpler measure for energy, the calculation of the per-pixel translation can then be performed easily. The pixel depth can then be calculated as shown below.

$$d = \sqrt{\hat{u}^2 + \hat{v}^2} \tag{10}$$

It should be noted that depth in this sense is the relative displacement from the position of the green channel and is not the physical distance of the objects. The latter can only be calculated given full knowledge of the optical system but in most cases, this type of depth is not necessary.

4 FUZZY BILATERAL FILTER

While most energy measures including disparity are very effective along image edges and textures, they tend to suffer along smooth regions. As these energy measures are taken locally, there is no knowledge of the more distant neighborhood. To avoid this, some form of filtering must be performed on the energy functions to take into account the neighboring energy costs. One potential solution to this is to apply a Gaussian filter. In this case, however, the energy near the edges of an image would spread causing inaccurate estimates

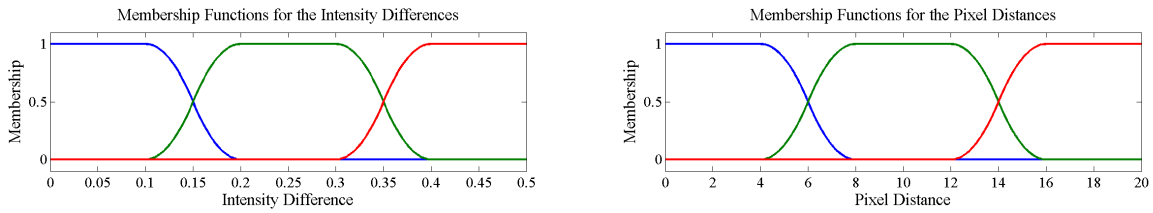

Figure 3: Membership functions used in the fuzzy bilateral filter.

Figure 4: Test results using the images from Bando, Chen, and Nishita (2008) showing the original image, the depth estimates using disparity measures, and the depth estimates using proposed algorithm.

of depth. Another approach is to use the popular bilateral filter, which places weights on both distance and intensity (Tomasi & Manduchi, 1998) . Formally, the bilateral filter is described with the following mathematical structure.

$$\hat{I}(x,y) = \sum_{x_0}\sum_{y_0} w_d(x,y,x_0,y_0)\, w_i(x,y,x_0,y_0)\, I(x,y) \tag{11}$$

$$w_d(x,y,x_0,y_0) = \exp\left(\frac{\sqrt{(x-x_0)^2+(y-y_0)^2}}{2\sigma_d^2}\right) \tag{12}$$

$$w_i(x,y,x_0,y_0) = \exp\left(\frac{|I(x,y)-I(x_0,y_0)|}{2\sigma_i^2}\right) \tag{13}$$

As the bilateral filter is clearly dependent on the difference in pixel intensities, energy maps cannot effectively be used for this. Instead, the green channel is used in the proposed algorithm for calculating all the weights and the same weights are applied to each energy map. Furthermore, a non-linear response curve is obtained by using fuzzy rules in place of the Gaussian function used in the traditional bilateral filter. In particular, the proposed algorithm uses probabilistic membership functions shown in Figure 4.
The fuzzy rules used in the system are shown in the rule matrix in Table 1. It should be noted that min-max functions are used for the and-or operations in the system. Finally, defuzzification is performed using the center-of-singletons method for rapid calculation. The mapping between the output classes and the singleton values are shown in Table 2.

Table 1: Fuzzy Rule Matrix

Intensity	**Pixel Distance**		
	Near	**Mid**	**Far**
Near	Very High	High	High
Mid	High	Medium	Low
Far	Low	Very Low	Very Low

Table 2: Output Singleton Membership

Fuzzy Label	**Weight**
Very Low	0.00
Low	0.25
Medium	0.50
High	0.75
Very High	1.00

5 RESULTS

To demonstrate the performance of the proposed algorithm, samples obtained from the the previous researchers were run through the system. For the purpose of testing, the energy maps were obtained from the images using an *11x11* search window for possible translations for a total of *121* energy maps. Each of these maps were then subjected to the fuzzy bilateral filter based on the fixed green channel from their respective input images. The lowest energy was then obtained for each pixel and the associated depth for these energy values were calculated.

For the purpose of comparison, the disparities for each of the *121* translations were also calculated using Bando, Chen, and Nishita's method (2008) using a *5x5* neighborhood for every pixel. Similarly, the minimum disparity was found along with the corresponding depth estimates. The resulting depth maps are shown in Figure 5. Given these results, it is apparent that the use of the simple Sobel filter on chrominance channels in combination with fuzzy bilateral filtering is effective in finding the relative depth of objects in the coded aperture image. The reference depth maps obtained using disparity measures exhibits a large amount of noise compared to the proposed technique. It should be noted, however, that the disparity measures can be improved by applying filtering techniques as well. In doing so, most of the noise in the depth estimates may be eliminated at the cost of significantly increasing the complexity of the algorithm.

6 CONCLUSION

In dealing with depth extraction, coded aperture techniques have been demonstrated to be viable alternatives to multi-camera setups. In this paper, a simplified technique at processing the obtained coded aperture images is presented. Building on existing literature, the work presented here uses simple gradient-based energy maps which, when used together with a proposed fuzzy bilateral filtering framework, results to a more computationally-efficient approach to depth estimation. For faster applications, the use of binary-masked filtering may also warrant further investigation.

REFERENCES

Ayache, N. (1995). Medical computer vision, virtual reality and robotics. *Image and Vision Computing*, 313.

Babacan, S., Ansorge, R., Luessi, M., Mataran, P., Molina, R., & Katsaggelos, A. (2012, Dec). Compressive light field sensing. *Image Processing, IEEE Transactions on*, *21*(12), 4746-4757.

Bando, Y., Chen, B.-Y., & Nishita, T. (2008). Extracting depth and matte using a color-filtered aperture. *ACM Transactions on Graphics*, *27*(4), 134.

Barnard, S. T., & Fischler, M. A. (1982, December). Computational stereo. *ACM Comput. Surv.*, *14*(4), 553–572.

Charalampous, K., Kostavelis, I., Boukas, E., Amanatiadis, A., Nalpantidis, L., Emmanouilidis, C., & Gasteratos, A. (2015). Autonomous robot path planning techniques using cellular automata. In G. C. Sirakoulis & A. Adamatzky (Eds.), *Robots and lattice automata* (Vol. 13, p. 175-196). Springer International Publishing.

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New Yotk: John Willey & Sons.

Jones, G. R., Lee, D., Holliman, N. S., & Ezra, D. (2001). *Controlling perceived depth in stereoscopic images* (Vol. 4297).

Levin, A., Fergus, R., Durand, F., & Freeman, W. T. (2007, July). Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.*, *26*(3).

Marwah, K., Wetzstein, G., Bando, Y., & Raskar, R. (2013). Compressive Light Field Photography using Overcomplete Dictionaries and Optimized Projections. *ACM Trans. Graph. (Proc. SIGGRAPH)*, *32*(4), 1–11.

Merkle, P., Smolic, A., Muller, K., & Wiegand, T. (2007, Sept). Multi-view video plus depth representation and coding. In *Image processing, 2007. icip 2007. ieee international conference on* (Vol. 1, p. I - 201-I - 204).

Tomasi, C., & Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Proceedings of the sixth international conference on computer vision* (pp. 839–). Washington, DC, USA: IEEE Computer Society.

Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A., & Tumblin, J. (2007, July). Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph.*, *26*(3).

Zamarin, M., Zanuttigh, P., Milani, S., Cortelazzo, G. M., & Forchhammer, S. O. (2010). A joint multi-view plus depth image coding scheme based on 3d-warping. In *Proceedings of the 1st international workshop on 3d video processing* (pp. 7–12). New York, NY, USA: ACM.