SINGLE IMAGE-BASED DEPTH ESTIMATION USING DUAL OFF-AXIS COLOR FILTERED APERTURE CAMERA

Seungwon Lee, Nahyun Kim, Kyungwon Jung, Monson H. Hayes, and Joonki Paik

Image Processing and Intelligent Systems Laboratory
Graduate School of Advanced Imaging Science, Multimedia, and Film
Chung-Ang University, Seoul, Korea
superlsw@gmail.com, skgus1433@gmail.com, fjqmsa@gmail.com, mhh3@gatech.edu,
and paikj@cau.ac.kr

ABSTRACT

In this paper, we present a dual off-axis color filtered aperture (DCA)-based computational imaging system for depth estimation in a single-camera framework. The DCA has one primary (red) color and its complementary (cyan) color filtered apertures located off the optical axis to generate misalignment between color channels depending on the distance of a region-of-interest from the camera. Disparity of color shifting values (CSVs) between color channels in the image acquired by the DCA-based imaging system is estimated using L_1-norm minimization of energy functional considering both brightness and gradient constancies and only translation of x-axis. The two data terms both red to green (R-G) and red to blue (R-B) are combined in a single data term. The proposed imaging system can be implemented by simply inserting an appropriately resized DCA into any general optical system. Experimental results show that the proposed DCA can estimate the distance of the scene from the camera in the single-camera framework.

Index Terms— Depth estimation, computational camera, optical flow, 3D reconstruction, 3D image acquisition

1. INTRODUCTION

A variety of different approaches have been developed to estimate the depth map or reconstruct a 3D scene from one or more 2D images. This problem has broad applications in areas such as robot vision, human computer interfaces, intelligent visual surveillance, 3D image acquisition, and intelligent driver assistance systems. Traditional approaches to recover depth information from 2D images have relied on either stereo vision or cues such as shading, focusing, and motion. Stereo vision is one of the most well-known depth estimation approach using binocular disparity between two images having different viewpoints [1]. In spite of its many advantages, it requires the use of a pair of cameras.

Recently, computational imaging systems have been developed to obtain information that is not readily available or possible to obtain with a conventional imaging system [2]. Computational cameras are imaging systems that combine unconventional optics with signal processing to perform specialized functions. These cameras alter the rays entering the camera through special optics, and the light field that is imaged by the sensors is then processed to create the desired information. These cameras have been developed for such functions and applications as re-focusing, scene segmentation, and 3D computer vision.

One type of computational imaging system modifies the geometrical structure of the aperture. An example is the coded aperture camera that intentionally generates specific image blur using the coded aperture, and the coded aperture image is then processed to create a depth map. Levin has proposed frequency property-based coded apertures that efficiently adjust the defocus blur. The computational camera with the coded apertures is able to estimate the depth of the scene in a single image [3].

Another approach is to use a camera that has dual apertures that are placed off-center from the optical axis. The effect of the dual apertures is to shift the projected object points on the image plane as a function of the distance of an object from the camera. Dou has proposed an off-axis aperture camera using the aperture size as a variable to estimate depth for 3D shape reconstruction. However, this camera has limited freedom and requires a complicated camera configuration [4].

Since color images consist of red, green, and blue color channels, Maik considered the use of three color filtered apertures (TCA) using shifting property of three off-axis apertures [5]. Kim proposed a multifocusing image restoration algorithm using the distance estimation in the color shifted image using TCA [6], and Lee has proposed full depth map estimation using depth interpolation from the sparse depth map using the TCA camera [7]. Since these approaches use the region-based segmentation or the depth interpolation, the performance of these depth estimation is slow and inaccurate.

In this paper, we present a dual off-axis color filtered aperture (DCA) configuration using a primary (red) color filter and a complementary (cyan) color filter in a single-camera framework. The two color filter apertures are displaced horizontally with respect to each other along the x-axis and, as a result, when an object is imaged with the DCA camera, the object in the green and blue channels will be shifted along the x-axis with respect to the object in the red channel with the amount of the shift depending on the distance of the object from the camera.

Although DCA camera provides depth-dependent shifting among color channels, conventional matching algorithms only considered brightness constancy cannot easily estimate the spatial disparity because color channels have different intensity levels. To address this problem, disparity of color shifting values (CSVs) between color channels in the image acquired by the DCA camera is estimated using L_1 minimization of energy functional considering both brightness and gradient constancies and only translation of x-axis. The L_1 minimization of the energy functional can remove the outlier at the boundary in the discrete signals. The two data terms both red to green (R-G) and red to blue (R-B) are combined in a single data term.

Experimental results show that the proposed method can not only estimate efficiently the depth of the scene but also correct color misalignment in a color shifted image. The DCA-based imaging system can be used for various photographic applications such as multifocusing, refocusing, and depth-based image segmentation. This system can be also applied to video processing applications such as video surveillance systems, intelligent driver assistant systems, and robot vision.

2. DUAL OFF-AXIS COLOR FILTERED APERTURES

The aperture of an optical system is an opening that adjusts the amount of the light that is focussed onto the image sensor. The center of the aperture in a traditional imaging system is aligned with the optical axis of the lens. The convergence pattern of an object that is located at in-focus position, A_i, on the image plane will form a point as shown in Fig. 1(a). However, when the object moves away from the plane of A_i to a far-focus position, A_f, then the convergence pattern on the image plane will form a circle of confusion (COC), whose radius depends on the distance of the object as shown in Fig. 1(b).

If the camera is reconfigured to have two off-axis apertures as shown in Fig. 2, then the convergence pattern of the projected point on the image plane is divided into the two projected points if the object is not located at A_i. As the object moves from A_f to A_i, the two projected points converge to a single point in the image plane. As the object continues moving towards the near-focus position, A_n, the converged projected point begin to diverge from each other in the opposite direction. In both cases, the distance between two separated convergence patterns depends on the distance of the object

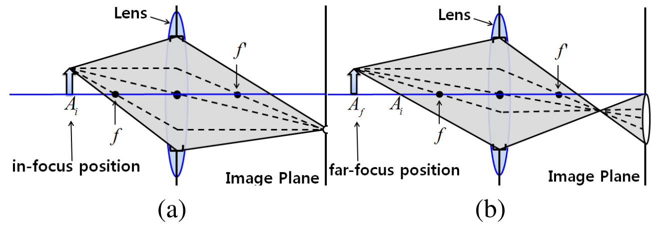


Fig. 1. A conventional imaging system with an aperture aligned with the optical axis. Image formation of an object at the (a) in-focus position and (b) far-focus position.

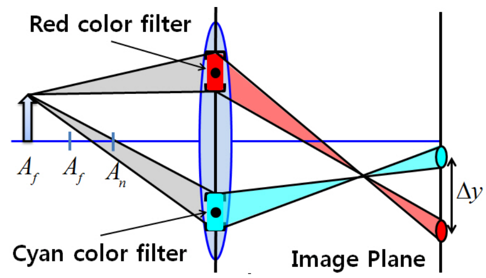


Fig. 2. Dual off-axis apertures configuration with color filters.

from the camera. When a red color filter is placed over one of the apertures and a cyan (blue plus green) color filter is placed over the other aperture, then the the two convergence patterns will be captured separately by the red sensors and by the green and blue sensors, respectively, in the Bayer array. Therefore, the color shift model can provide the geometric disparity of misalignment. The geometric disparity can be estimated from the amount of color deviation between the two projected points as shown in Fig. 3. For example, the point of the object at the in-focus position has the convergence point which does not have color misalignment. However, if the object goes farther than the in-focus position, color misalignment occurs in the image on the sensor. On the other hand, if the object comes closer than the in-focus position, color misalignment occurs in the opposite direction to the far-focus position. Thus, the distance of the object can be estimated using color shifting value (CSV) that corresponds to the length of the amount of misalignment between color channels from the color shifted image.

3. DCA CAMERA-BASED DEPTH ESTIMATION

In order to estimate the distance of an object within an image, it is necessary to estimate the color shift vector (CSVs) between the red and green (R-G) channels or between the red and blue (R-B) channels. The early optical flow algorithms assume brightness constancy between a pair of images. The images acquired by the DCA camera, however, do not satisfy the brightness constancy property because the images in the three color channels generally have different intensities. Brox has proposed an optical flow estimation algorithm that

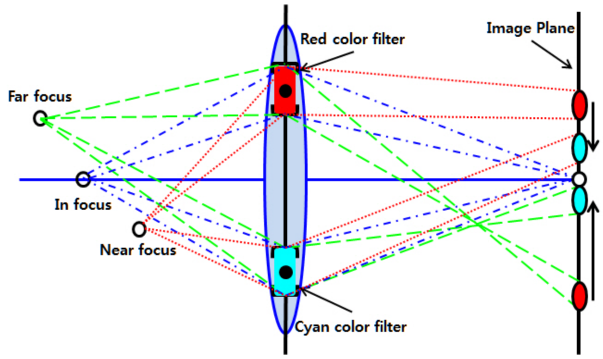


Fig. 3. The convergence patterns of objects at three different distances in the DCA configuration.

involves the use of a "data" function that involves the minimization of an energy function that includes a gradient constancy assumption as well as brightness constancy [8]. Adapting this function to the DCA camera image where optical flow only needs to be estimated along the x-axis, the energy function for the R-G data term is given by

$$E_{data}^{rg}(u) = \int_{\Omega} \psi\Big(\left|I^r(x+u,y) - I^g(x,y)\right|^2\Big)\, dxdy + \int_{\Omega} \psi\Big(\left|I_x^r(x+u,y) - I_x^g(x,y)\right|^2\Big)\, dxdy, \quad (1)$$

where u is the shift between the red and green channels, ψ is a concave function, and $I^r(x)$ and $I^g(x)$ are the R and G color channel images, and $I_x^r(x)$ and $I_x^g(x)$ are the partial derivatives of the R and G color channels with respect to x, respectively. It should be noted that although u is a function of x and y and should be written as $u(x,y)$, to simplify notation the dependence on position is omitted. As in [8], we use $\psi(s^2) = \sqrt{s^2+\varepsilon^2}$ where $\varepsilon = 0.001$, which results in a *modified* L_1 minimization of the energy. A similar definition for the R-B data term, $E_{data}^{rb}(u)$, is formed by replacing $I^g(x)$ with $I^b(x)$ in Eq. (1). Finally, a piecewise smoothness constraint is applied by including a penalty term of the form [8]

$$E_{smooth}(u) = \int_{\Omega} \psi\big(u_x(x,y)^2 + u_y(x,y)^2\big) dxdy, \quad (2)$$

where $u_x(x,y)$ is the partial derivative with respect to x of the color shift map. Thus, the total energy function that is to be minimized is the weighted sum of the two data terms and the smoothness term,

$$E(u) = E_{data}^{rg}(u) + E_{data}^{rb}(u) + \alpha E_{smooth}(u), \quad (3)$$

where $\alpha > 0$ is regularization parameter.

The color shift function, $u(x,y)$, is found by minimizing the total energy function $E(u)$. However, minimizing $E(u)$ is non-trivial because $E(u)$ is highly nonlinear. Therefore, the data terms are first linearized using a first-order Taylor series expansion. For example, the linearization of $E_{data}^{rg}(u)$ gives

$$E_{data}^{rg}(u) \approx \int_{\Omega} \psi\Big(\left|I^{rg}(x,y) + uI_x^r(x,y)\right|^2\Big)\, dxdy + \int_{\Omega} \psi\Big(\left|I_x^{rg}(x,y) + uI_{xx}^r(x,y)\right|^2\Big)\, dxdy, \quad (4)$$

where

$$I^{rg}(x,y) \triangleq I^r(x,y) - I^g(x,y),$$

and $I_x(\cdot)$ and $I_{xx}(\cdot)$ are the first and second order partial derivatives of $I(\cdot)$ with respect to x. Performing the linearization of the data term $E_{data}^{rb}(u)$ yields a similar expression with $I^r(x,y)$ replaced with $I^b(x,y)$ and $I^{rg}(x,y)$ replaced with $I^{rb}(x,y)$.

Having linearized the data terms, the next step is to solve for the color shifting values, $u(x,y)$. To do this, we use Euler Lagrange equations to find the function $u(x,y)$ so that

$$\frac{\partial L}{\partial u} - \frac{\partial}{\partial x}\frac{\partial L}{\partial u_x} - \frac{\partial}{\partial y}\frac{\partial L}{\partial u_y} = 0,$$

where L is the integrand of the energy expression in Eq. (3), which gives

$$\begin{aligned} 0 &= \psi_1'(\cdot)\left[I^{rg}(x,y) + uI_x^r(x,y)\right] I_x^r(x,y) \\ &+ \psi_2'(\cdot)\left[I_x^{rg}(x,y) + uI_{xx}^r(x,y)\right] I_{xx}^r(x,y) \\ &+ \psi_3'(\cdot)\left[I^{rb}(x,y) + uI_x^r(x,y)\right] I_x^r(x,y) \\ &+ \psi_4'(\cdot)\left[I_x^{rb}(x,y) + uI_{xx}^r(x,y)\right] I_{xx}^r(x,y) \\ &- \alpha\psi_5'(\cdot)\left[u_{xx}(x,y) + u_{yy}(x,y)\right], \end{aligned} \quad (5)$$

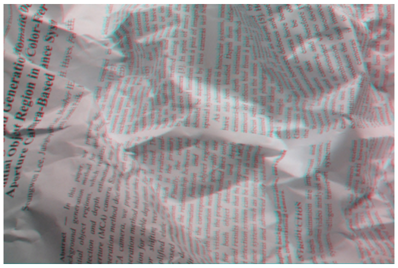
where

$$\begin{aligned} \psi_1'(\cdot) &= \psi'\big(I^{rg}(x,y) + uI_x^r(x,y)\big) \\ \psi_2'(\cdot) &= \psi'\big(I_x^{rg}(x,y) + uI_{xx}^r(x,y)\big) \\ \psi_3'(\cdot) &= \psi'\big(I^{rb}(x,y) + uI_x^r(x,y)\big) \\ \psi_4'(\cdot) &= \psi'\big(I_x^{rb}(x,y) + uI_{xx}^r(x,y)\big) \\ \psi_5'(\cdot) &= \psi'\big(u_x^2(x,y) + u_y^2(x,y)\big). \end{aligned}$$

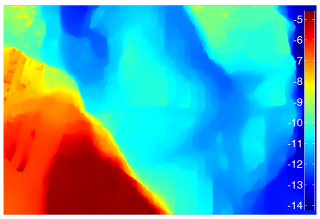
Discretizing this equation leads to non-linear equation that may be solved for u using coarse-to-fine warping techniques, outer and inner fixed iteration, and a fixed-point iteration [8].

4. EXPERIMENTAL RESULTS

To evaluate the feasibility of using a DCA camera for depth estimation, we used a Canon 450D digital camera with a dual aperture color filter attached to a Canon EF-S $18-55$mm lens. The resolution of the test images is 2136×1424. Shown in Fig. 4(a) is an image of a crumpled piece of paper that has been imaged using the DCA camera. The results of estimating the color shift map $u(x,y)$ using the approach presented



(a)

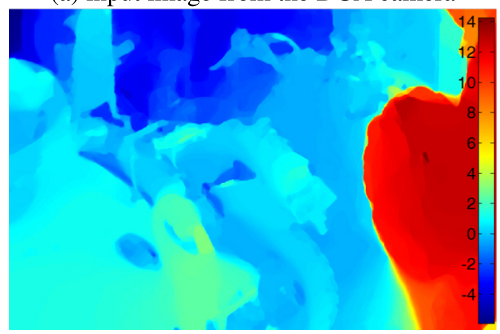


(b)

(c)

Fig. 4. Result of the color shifting value estimation; (a) input image from the DCA camera, (b) estimated color shift map, (c) reconstructed 3-D scene.

(a) input image from the DCA camera



(b) estimated color shift map.

Fig. 5. Result of color shifting value estimation for a complex outdoor scene.

in Section 3 are shown in Fig. 4(b). The scale that is shown on the right indicates the amount of the shift, where the magnitude of the shift is a function of the distance an object is from the plane of focus. Positive numbers indicate that the object is in the near-focus plane, whereas negative numbers correspond to an object in a far-focus plane. Thus, red corresponds to objects that are close whereas blue corresponds to objects that are distant. The color misalignments in Fig. 4(a) may be corrected to form a normal image by inversely shifting the both G and B color channels using estimated color shift map. The proposed method takes about 13 second in processing 2136×1424 resolution image.

Another example is given in Fig. 5, which shows the results of depth estimation for a complex outdoor environment. Note that due to the L_1-norm minimization in the energy functional, the depth discontinuities are preserved in the occluded regions.

5. CONCLUSION

In this paper, we presented a dual off-axis color filtered aperture computational imaging system that is capable of creating a depth map of an image. The DCA camera has a pair of color filter apertures, one that is a primary color with the other being its complement. Here, we use red as the primary and cyan as the complement. The estimation of the color shift map was estimated using a modified L_1 minimization of an energy functional that assumes constancy both in brightness as well as the gradient of the brightness. Since the color shifting occurs only along one-axis, optical flow estimation is simplified over more general optical flow estimation problems.

Most traditional imaging systems for three-dimensional (3D) depth estimation have relied on either multiple images or additional cues such as shading, focusing, and motion. However, the DCA-based imaging system estimated the depth information using the color shifting property of dual apertures from a single image. The proposed imaging system was implemented by simply inserting an appropriately resized DCA into any general optical system.

The proposed imaging system can be used for various photographic applications such as multifocusing, refocusing, and depth-based image segmentation. This system is also applied to video processing applications such as video surveillance systems, intelligent driver assistant systems, and robot vision.

6. ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through National Research Foundation (NRF) of Korea funded by the Ministry of Education, Science and Technology (2009-0081059) and by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2012 (00045420-1).

7. REFERENCES

[1] D. Schastein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithm," *International Journal of Computer Vision*, vol. 47, no. 1, pp. 7-42, 2002.

[2] C. Zhou and S. Nayar, "Computational cameras: convergence of optics and processing," *IEEE Trans. Image Processing*, vol. 20, no. 12, pp. 3322-3340, December 2011.

[3] A. Levin, R. Fergus, F. Durand, and W. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graphics*, vol. 26, no. 3, pp. 70-79, 2007.

[4] Q. Dou and P. Favaro, "Off-axis aperture camera: 3D shape reconstruction and image restoration," *Int. Conf. Computer Vision and Pattern Recognition*, pp. 1-7, 2008.

[5] V. Maik, D. Cho, J. Shin, D. Har, and J. Paik, "Color shift model-based segmentation and fusion for digital autofocusing," *Journal of Imaging Science and Technology*, vol. 51, no. 4, pp. 368-379, 2007.

[6] S. Kim, E. Lee, M. Hayes, and J. Paik, "Multifocusing and depth estimation using a color shift model-based computational camera," *IEEE Trans. Image Processing*, vol. 21, no. 9, pp. 4152-4166, September 2012.

[7] S. Lee, J. Lee, M. Hayes, A. Katsaggelos, and J. Paik, "Single camera-based full depth map estimation using color shifting property of a multiple color-filtered aperture," *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pp. 801-804, May 2012.

[8] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," *European Conf. on Computer Vision, LNCS*, vol. 3024, pp. 25-36, 2004.