Article

Three-Dimensional Object Motion and Velocity Estimation Using a Single Computational RGB-D Camera

Seungwon Lee, Kyungwon Jeong, Jinho Park and Joonki Paik *

Image Processing and Intelligent Systems Laboratory Graduate School of Advanced Imaging Science, Multimedia, and Film Chung-Ang University, Seoul 156-756, Korea;
E-Mails: superlsw@gmail.com (S.L.); fjqmsa@gmail.com (K.J.); dkskzmffps@gmail.com (J.P.)

* Author to whom correspondence should be addressed; E-Mail: paikj@cau.ac.kr;
Tel.: +82-2-820-5300; Fax: +82-2-814-9110.

Abstract: In this paper, a three-dimensional (3D) object moving direction and velocity estimation method is presented using a dual off-axis color-filtered aperture (DCA)-based computational camera. Conventional object tracking methods provided only two-dimensional (2D) states of an object in the image for the target representation. The proposed method estimates depth information in the object region from a single DCA camera that transforms 2D spatial information into 3D model parameters of the object. We also present a calibration method of the DCA camera to estimate the entire set of camera parameters for a practical implementation. Experimental results show that the proposed DCA-based color and depth (RGB-D) camera can calculate the 3D object moving direction and velocity of a randomly moving object in a single-camera framework.

1. Introduction

Object tracking is a very popular research topic in the computer vision field because of its wide applications in video surveillance systems, intelligent driver assistant systems, robot vision, *etc.* [1,2]. Conventional object tracking methods provided only two-dimensional (2D) states of an object in the

image and therefore cannot deal with the occlusion and depth-related problems. For solving these problems, estimation of three-dimensional (3D) depth information has been intensively studied for the past several decades [3–6].

The traditional approach to 3D image acquisition or depth estimation involves the use of two cameras that capture two images of the same scene from different viewpoints [7,8]. Despite many advantages, the stereo vision system has a fundamental limitation that highly accurate calibration is required for both cameras. In addition, stereo cameras require both calibration and rectification steps for the alignment of two cameras. However, even after completion of the two steps, the rectification step should be performed again, if the stereo camera is influenced by various external factors that break the alignment between two cameras. As an alternative approach, Panasonic's LUMIX G 12.5 mm F12 lens can acquire two images with different viewpoints by dividing the imaging sensor into two regions. Although this system mimics the stereo vision in a single-camera framework, the reduced resolution is the fundamental disadvantage.

Another single camera-based depth estimation approach is the pupil plane coding method that modifies an aperture of a lens for coding geometric depth information into the image [9]. Recently, three apertures covered with red, green and blue filters have been used for the same purpose [10], where three color-filtered apertures (TCA) generate color shifts among channels depending on the distance of an object. Kim *et al*. proposed a multifocusing image restoration algorithm using the distance estimation in the color shifted image using the TCA [11]. Lee *et al*. proposed a simultaneous object tracking and depth estimation system using the TCA camera [12]. However, these TCA-based methods produce redundant pairs of disparity vectors and do not establish the theoretically complete relationship between color shifting and the real distance. Lee *et al*. proposed a novel configuration of dual off-axis color-filtered apertures (DCA) to remove the disparity redundancy of the color shifting vectors and to increase the size of an individual aperture to receive more incoming light [13]. The mathematical model of the DCA configuration with the relationship between color shifting values and the actual distance of the object were proposed in [14].

In this paper, we present a novel 3D object moving direction and velocity estimation method for robust object tracking using a DCA-based computational RGB-D camera. The incremental learning-based approach [15] is used for object tracking from the DCA camera, and the color shifting value (CSV), which is related to the distance of the object, is estimated in the region of the tracked object. The states of the object at the image coordinate are then transformed into the 3D parameters of the object using the estimated CSV. Finally, the direction and velocity of a moving 3D object are simultaneously calculated while tracking the object. We also present a calibration method of the DCA camera to estimate the entire set of camera parameters for real applications. Although this work shares the concept of multiple, color-filtered apertures in the previous work [11–14], the original contribution includes: (i) the incremental learning-based object tracking algorithm that is optimized for the DCA camera system; (ii) single camera-based simultaneous 3D object moving direction and velocity estimation; and (iii) a novel calibration method that can be used for a DCA camera system.

The paper is organized as follows. Section 2 presents the background of the DCA camera, and Sections 3 and 4 describe the object moving direction and velocity estimation for tracking and the DCA

camera calibration, respectively. Experimental results are provided in Section 5, and Section 6 concludes the paper.

2. Dual Off-Axis Color-Filtered Aperture-Based Camera

Recently, computational imaging systems have been widely used for obtaining additional information, while traditional imaging systems only acquire intensity and color information. A computational imaging system projects rays, which is altered by specially-designed optics, in the light field of the scene onto the image sensor using novel optics and the correspondingly developed image processing algorithms. These systems can produce a new type of image that is potentially useful for re-focusing, scene segmentation and 3D computer vision. Computational imaging systems can be broadly classified into six categories, as shown in Figure 1 [9].
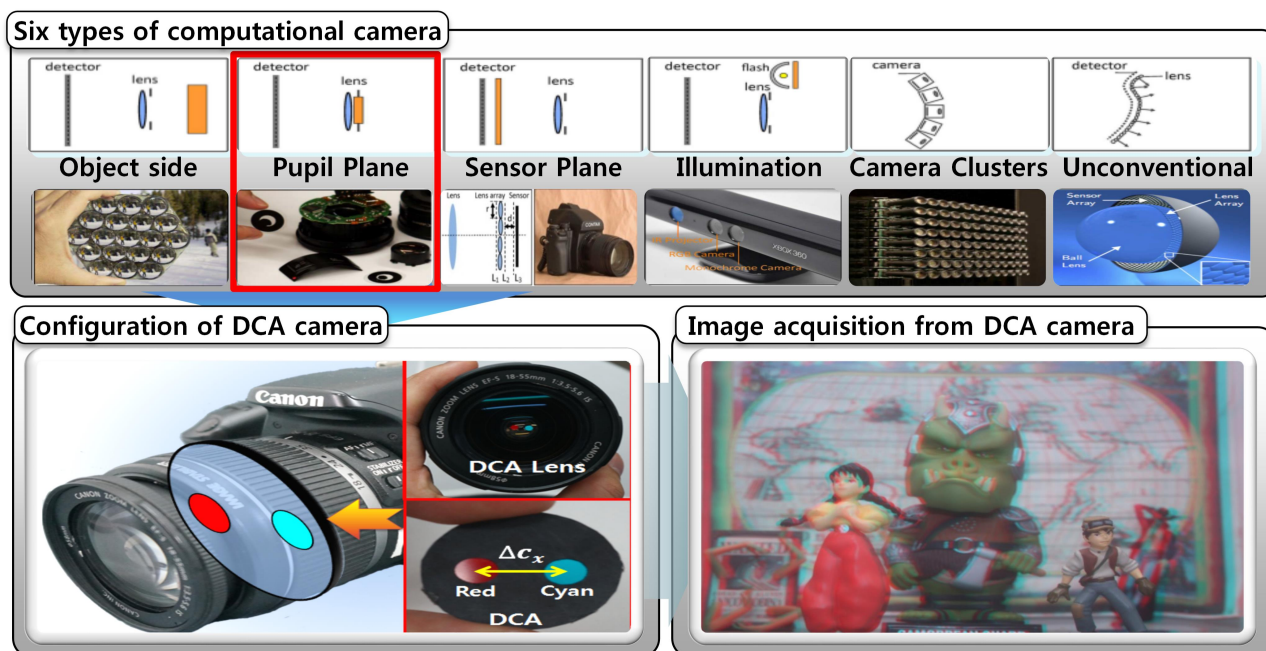

Figure 1. Six types of computational cameras and the proposed dual off-axis color-filtered aperture (DCA) camera [9].

The pupil plane coding places a specially designed pattern at the aperture in front of the lens to easily configure the computational camera for depth information. By incorporating the concept of pupil plane coding, we design a novel computational imaging system by simply inserting an appropriately resized DCA into any general optical system in the single-camera framework.

The aperture of an optical system is the opening device that adjusts the amount of light entering the image sensor. The center of the aperture in the traditional imaging system is generally aligned with the optical axis of the lens. If there are two off-axis apertures, 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 the plane of focus. The distance between two separated convergence patterns depends on the distance of the object

from the camera. It is noted that we cannot estimate the distance between two projected regions if they are mixed together.
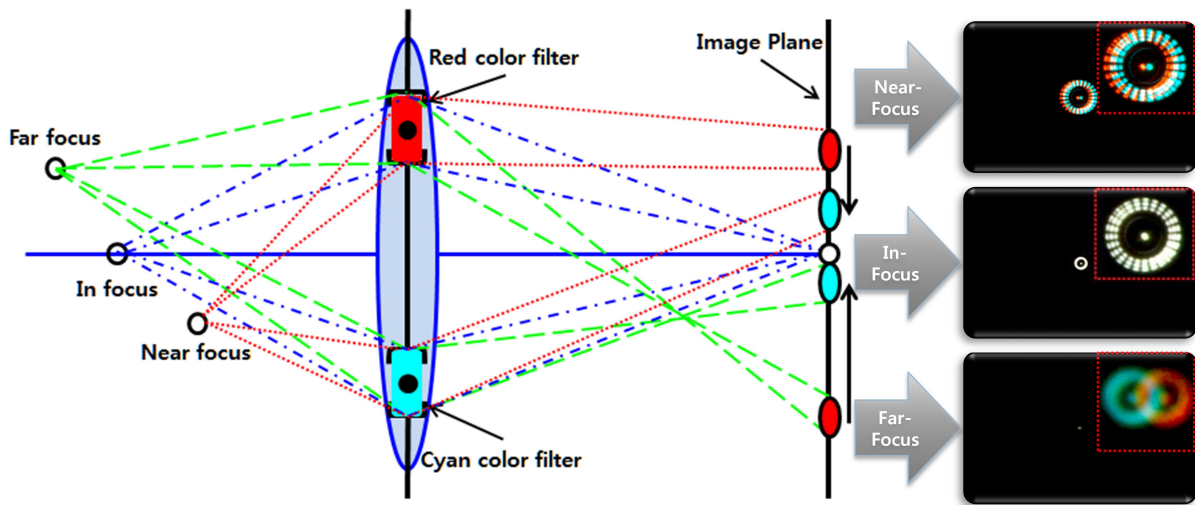

Figure 2. The convergence patterns of objects at three different distances in the DCA configuration.

In order to separate two regions, we covered apertures using two different color filters that generate geometric disparity between two color images. For the complete color shift model using the DCA camera, we use two apertures with red (R) and cyan (C) filters, whose centers are located on the same line crossing the optical axis. The color shift model can provide the geometric disparity of misalignment, which can be estimated from the amount of color deviation between the pair of two projected points, as shown in Figure 2. Thus, the distance of the object can be estimated using the color shifting value (CSV) that corresponds to the length of the amount of misalignment between color channels from the color-misaligned image.

3. Object Moving Direction and Velocity Estimation with Tracking

The object region should be continuously tracked to estimate the 3D motion of the object while the position and scale of the object change. We use the incremental learning-based method for robust object tracking [15].

The statistical object tracking problem is usually defined as the Bayesian inference with a hidden Markov model. Given the state of an object at time t, denoted as $\mathbf{o}_t$, cumulated observations up to time t, denoted as $Y_{1:t}$, the Bayesian filter updates *a posteriori* probability $p(\mathbf{o}_t|\mathbf{y}_{1:t})$ with the following rule,

$$p(\mathbf{o}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{o}_t) \int p(\mathbf{o}_t|\mathbf{o}_{t-1})p(\mathbf{o}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{o}_{t-1} \tag{1}$$

where $p(\mathbf{y}_t|\mathbf{o}_t)$ represents the observation model as a likelihood term that measures the similarity between the observation at the estimated state and the given model and $p(\mathbf{o}_t|\mathbf{o}_{t-1})$ the transition model as a prior term, which predicts the next state $\mathbf{o}_t$ based on the previous state $\mathbf{o}_{t-1}$. With the posterior probability $p(\mathbf{o}_t|\mathbf{y}_{1:t})$, we obtain the maximum *a posteriori* (MAP) estimate over N samples at time t as:

$$\mathbf{o}_t^{MAP} = \arg\max_{\mathbf{o}_t^n} p(\mathbf{o}_t^n | \mathbf{y}_{1:t}), \text{for } n = 1, ..., N \tag{2}$$

where $\mathbf{o}_t^{MAP}$ denotes the best configuration, which can explain the current state with the given model. The observation model is updated by the incremental principle components analysis (PCA) algorithm proposed by Ross, and then, the current state $\mathbf{o}_t^{MAP}$ is determined by the distance metric in this work [15]. The transition model is formulated by a random walk with a Gaussian distribution.

Let the 2D state of the object $\mathbf{x}_t$ at time t in the image coordinates denote $\mathbf{x}_t = \{x_t, y_t, w_t, h_t\}$, where x_t and y_t indicate the center position of the object, w_t and h_t represent the width and the height of the object, respectively. In order to convert the image coordinate system into the 3D camera coordinate system, we need to know the distance Z of the object and camera intrinsic parameters, including the focal lengths f_x, f_y and the principal point s_x, s_y.

Because two disparities between red and green and red and blue are the same with respect to the horizontal direction, the distance measure combines two energy functions as:

$$E(\Delta x) = \sum_{x,y \in \Omega} \left(I^r(x,y) - I^g(x+\Delta x, y)\right)^2 \\ + \sum_{x,y \in \Omega} \left(I^r(x,y) - I^b(x+\Delta x, y)\right)^2 \tag{3}$$

where Δx represents the CSV, Ω the extracted object region and I the image acquired by the DCA camera. Superscripts r, g and b respectively represent the red, green and blue color channels. Since this error function is nonlinear, it cannot be analytically minimized. To simplify the minimization, we approximate this error function using a first-order truncated Taylor series expansion [16] as:

$$E(\Delta x) = \sum_{x,y \in \Omega} \left(I_t^{rg}(x,y) - \Delta x I_x^{rg}(x,y)\right)^2 \\ + \sum_{x,y \in \Omega} \left(I_t^{rb}(x,y) - \Delta x I_x^{rb}(x,y)\right)^2 \tag{4}$$

where $I_t^{rc} = I^r - I^c$, for $c \in \{g, b\}$, and I_x^{rc} represents the first derivative of $(I^r + I^c)/2$ using the Sobel operator in the horizontal direction. Since $E(\Delta x)$ is a quadratic function of Δx, a closed-from solution for minimizing the energy can be found by differentiation with respect to Δx and setting the result equal to zero as:

$$\Delta x = \sum_{x,y \in \Omega} \frac{\left[I_x^{rg}(x,y)I_t^{rg}(x,y) + I_x^{rb}(x,y)I_t^{rb}(x,y)\right]}{\left[(I_x^{rg}(x,y))^2 + (I_x^{rb}(x,y))^2\right]} \tag{5}$$

A more accurate estimation of Δx can be performed by a Gaussian pyramid-based iterative coarse-to-fine approach for accommodating large color shifting values between color channels [17]. In [13], Lee *et al.* estimated the dense depth map in the entire image. However, the proposed method estimates only one CSV by minimizing the error function in the object region. If the object region is large enough to contain meaningful features, it is almost always reliable.
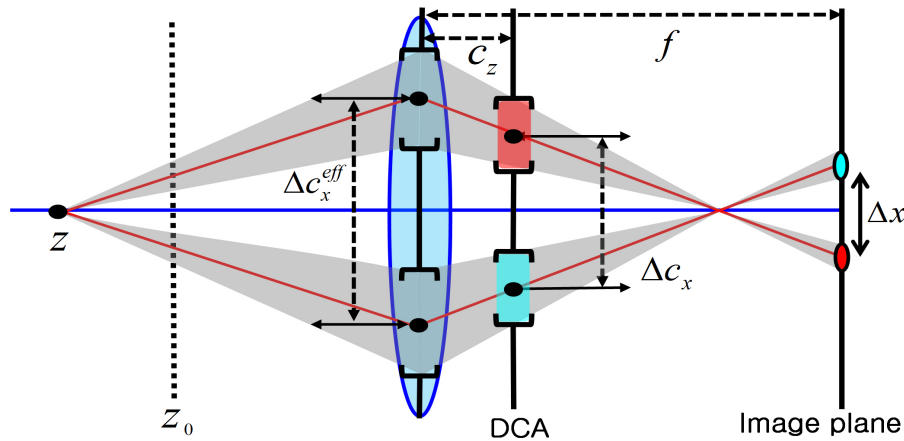
Figure 3. The configuration of the DCA camera.

In [14], the relation between Δx and the object distance Z has been derived as:

$$\Delta x = f^2 \frac{Z_0 - Z}{(Z_0 - f)(Z \cdot f - (Z - f)c_z)} \Delta c_x \tag{6}$$

where f represents the focal length, Δc_x the distance between the two off-axis apertures, c_z the distance of the DCA away from the lens and Z_0 the plane of focus. Figure 3 shows the configuration of the DCA camera. If we assume that both Z and Z_0 are sufficiently larger than the focal length, then the movement of the projection is approximately computed as:

$$\Delta x \approx f \left(\frac{1}{Z} - \frac{1}{Z_0} \right) \frac{f}{f - c_z} \Delta c_x \tag{7}$$

where $f/(f - c_z)$ determines the distance between the two equivalent apertures, $\Delta c_x^{eff} = (f/(f - c_z)) \cdot \Delta c_x$. If c_z is zero, Δc_x^{eff} is equal to Δc_x.

If the parameters in Equation (7) are expressed in millimeters, then Δx is also determined in millimeters. In order to express Δx in pixels, Δx has to be multiplied by the distance α_x between two pixels, which is given as:

$$\alpha_x = S_h / M_h \tag{8}$$

where S_h represents the size of an image sensor with respect to the horizontal direction and M_h the size of the image resolution with respect to the horizontal. f is redefined as $f_x = f / \alpha_x$, and then, Equation (8) can be rewritten as:

$$\Delta x \approx f_x \frac{\Delta c_x^{eff}}{Z} - f_x \frac{\Delta c_x^{eff}}{Z_0} \tag{9}$$

Equation (9) is similar to the relationship between the disparity and the depth in the stereo vision defined as $\Delta D = f_x \frac{B}{Z}$, where B represents the baseline, except theoffset given as $f \frac{\Delta c_x^{eff}}{z_0}$. Solving (9) for Z yields:

$$Z = \frac{f_x Z_0 \Delta c_x^{eff}}{f_x \Delta c_x^{eff} + Z_0 \Delta x} \tag{10}$$

Given Z, the 2D states of the object x_t and y_t are transformed into 3D camera coordinate as:

$$X_t = (x_t - s_x) \cdot Z_t/f_x \;\; \text{and} \;\; Y_t = (y_t - s_y) \cdot Z_t/f_y \tag{11}$$

where s_x and s_y are the principal points with respect to horizontal and vertical directions, respectively. The 2D states of the object w_t and h_t are transformed into 3D camera coordinates as:

$$W = w \cdot Z/f_x \;\; \text{and} \;\; H = h \cdot Z/f_y, \tag{12}$$

3D object moving direction $\mathbf{D}$ and velocity V are then calculated as:

$$\mathbf{D} = \{\dot{X} = X_t - X_{t-k}, \dot{Y} = Y_t - Y_{t-k}, \dot{Z} = Z_t - Z_{t-k}\} \tag{13}$$

$$V = \sqrt{\dot{X}^2 + \dot{Y}^2 + \dot{Z}^2} \tag{14}$$

where $\dot{X}, \dot{Y}$ and $\dot{Z}$ represent the amount of the object moving with respect to the x-, y- and z-coordinates, respectively. k represents the number of frames per second. Finally, the proposed 3D state of the object is defined as $\dot{\mathbf{o}}_t = \{\mathbf{o}_t, \mathbf{O}_t, \mathbf{D}_t, V_t\}$, where $\mathbf{O}_t = \{X_t, Y_t, Z_t\}$.

In $\dot{\mathbf{o}}_t$, the state of the object to be used for object tracking is $\mathbf{o}_t$. After determining $\mathbf{o}_t^{MAP}$, the final 3D state of the object $\dot{\mathbf{o}}_t$ is then computed. Continuously estimated Z may have a minor oscillation due to the estimation error of the CSV. For robust estimation of Z, we use Kalman filtering to predict and compensate temporally changing CSV.

4. DCA Camera Calibration

For accurate depth estimation, the entire set of camera parameters, except Δx, should be calibrated. Calibration of the DCA camera consists of two steps: (i) estimation of the camera parameters, including the focal lengths, principal point and lens distortion coefficients; and (ii) estimation of the DCA parameters, including Z_0 and c_z.
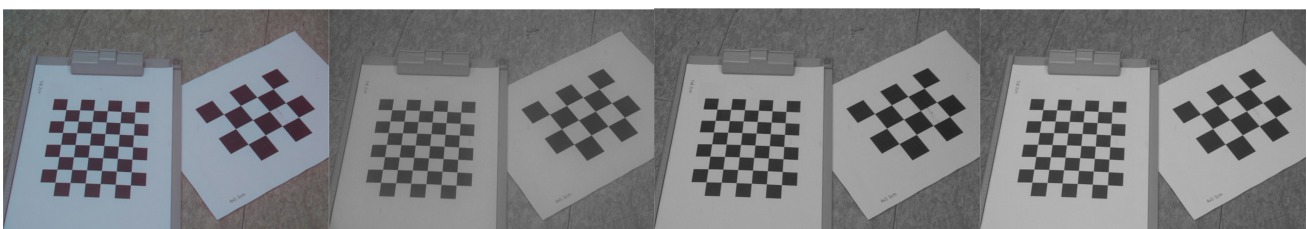
(a) color channel   (b) red channel   (c) green channel   (d) blue channel

Figure 4. Calibration patterns for estimating the camera parameters. (**a**) Color channel; (**b**) Red channel; (**c**) Green channel; (**d**) Blue channel.

Although each color channel has the same lens distortion coefficients, because the DCA camera uses a single lens, s_x and s_y have to be estimated in each color channel due to two off-axis apertures, as shown Figure 4. For the calibration of the camera parameters, we took the images using a checkerboard at different camera locations and used a calibration tool at each color channel.
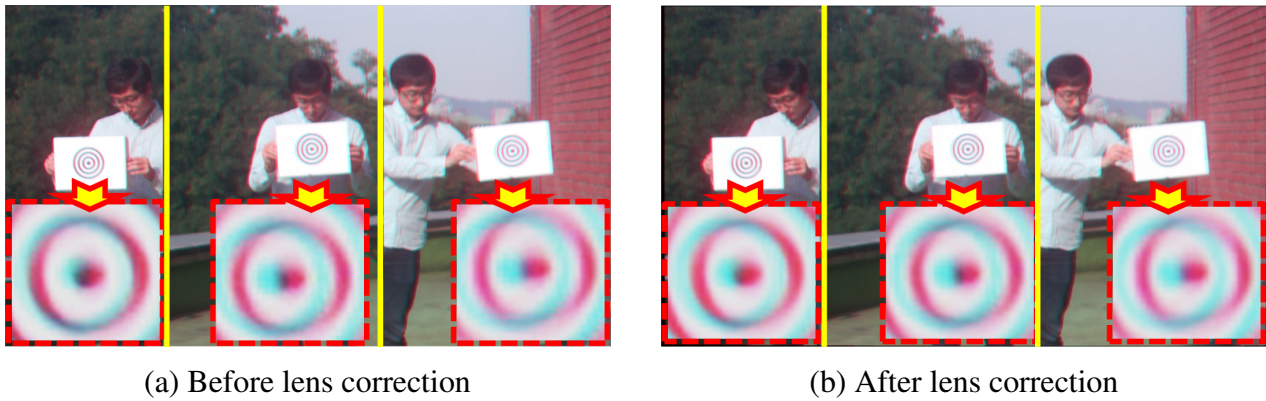
(a) Before lens correction   (b) After lens correction

Figure 5. The result of the lens distortion correction. (**a**) Before lens correction; (**b**) After lens correction.

Although the object is located at the same distance, each object has different CSVs before lens correction, as shown in Figure 5a. However, after lens correction, the objects located at the same distance also have the same CSVs, as shown in Figure 5b.
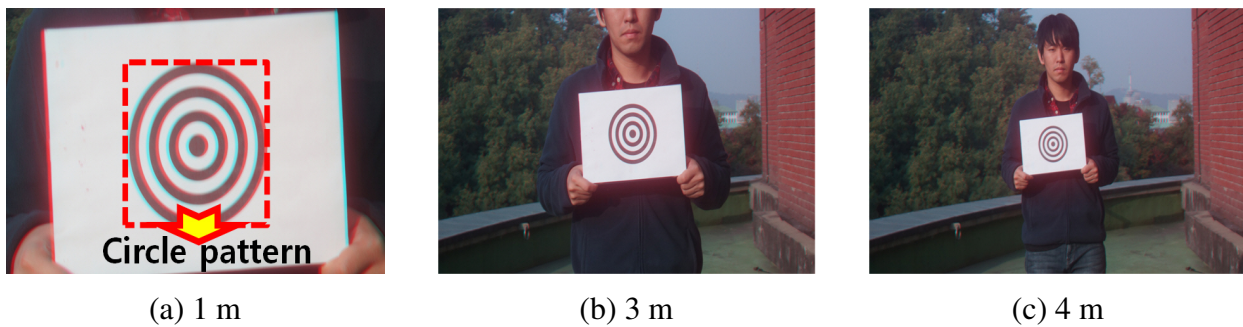

(a) 1 m   (b) 3 m   (c) 4 m

Figure 6. A concentric circle pattern for calibrating DCA parameters. (**a**) 1 m; (**b**) 3 m; (**c**) 4 m.

For solving Z as in Equation (10) from Δx, the unknown parameters Z_0 and c_z should be determined. We use a concentric circle pattern for calibrating the DCA camera and then take the image of the pattern at various distances, as shown in Figure 6. Because we know the real diameter of the circle, the distance of the concentric circle pattern can be calculated as:

$$Z = f_x \cdot \frac{W}{w} \tag{15}$$

where W represents the real diameter of the circle in meters and w represents the diameter of the circle in pixels. Rearranging Equation (10) yields:

$$Z \cdot Z_0^{-1} - \frac{\Delta x \cdot Z \cdot \alpha_x}{c_x \cdot f^2} c_z = 1 - \frac{\Delta x \cdot Z \cdot \alpha_x}{c_x \cdot f} \tag{16}$$

For solving this equation, we need to develop more than two linear equations. Therefore, we estimate Δx from the circle pattern of more than two images and then calculate Z using Equation (15). Using these values, Equation (16) can be developed as a matrix form as:

$$
\begin{bmatrix}
Z_1 & -\frac{\Delta x_1 \cdot Z_1 \cdot \alpha}{c_x \cdot f^2} \\
Z_2 & -\frac{\Delta x_2 \cdot Z_2 \cdot \alpha}{c_x \cdot f^2} \\
\vdots & \vdots \\
Z_n & -\frac{\Delta x_n \cdot Z_n \cdot \alpha}{c_x \cdot f^2}
\end{bmatrix}
\begin{bmatrix}
Z_0^{-1} \\
c_z
\end{bmatrix}
=
\begin{bmatrix}
1 - \frac{\Delta x_1 \cdot Z_1 \cdot \alpha}{c_x \cdot f} \\
1 - \frac{\Delta x_2 \cdot Z_2 \cdot \alpha}{c_x \cdot f} \\
\vdots \\
1 - \frac{\Delta x_n \cdot Z_n \cdot \alpha}{c_x \cdot f}
\end{bmatrix}
\tag{17}
$$

where Z_n and Δx_n represent the n-th object distance and the corresponding CSV, respectively. Finally, Z_0 and c_z can be solved by multiplying the pseudo inverse matrix. As a result of the calibrations of Z_0 and c_z, we are able to calculate the distance of the object from the estimated color shifting value, Δx, between color channels in the image using Equation (10).

In [14], Lee *et al.* estimated c_z by assuming that two objects are placed at Z_0 and Z_1, which should be given before the estimation process. It also needs manual identification of whether the object is placed at the in-focus position to determine Z_0. However, the proposed calibration method estimates the unknown parameters Z_0 and c_z using the least squares optimization method with the distance data acquired at different locations using the known size of the circle pattern. Because the proposed method just takes the image using the circle pattern and does not need to measure the distance of the object, it can better estimate the calibration parameters than [14].

5. Experimental Results

To demonstrate the feasibility of the DCA-based RGB-D camera for the distance estimation of the tracked object, we used a Sony NEX6 digital single lens reflected (DSLR) camera with a $18 - 55$ mm lens. The resolution of the video sequence is 1920×1080. The DCA lens is configured by the red and the cyan color filters, and the distance between the two apertures Δc_x is set to 6 mm.

Figure 7 shows comparison of the input data for the DCA calibration. As shown in the figure, the curve of the function using Equation (9) closely passes through the input data. In addition, Equation (7) is almost the same as its approximated version in (9).
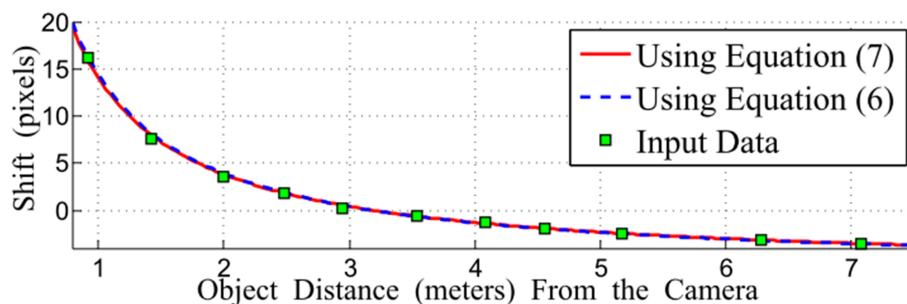

Figure 7. Plot of Δx *versus* the distance Z of an object from the camera ($f = 49.8$ mm, $z_0 = 3250$ mm, $c_z = -10$ mm).

Figure 8 shows the results of the proposed 3D state estimation of the tracked object. The images were taken of this object at distances that ranged from 5 m to 1 m, and cross pattern was moved at 1-m intervals.
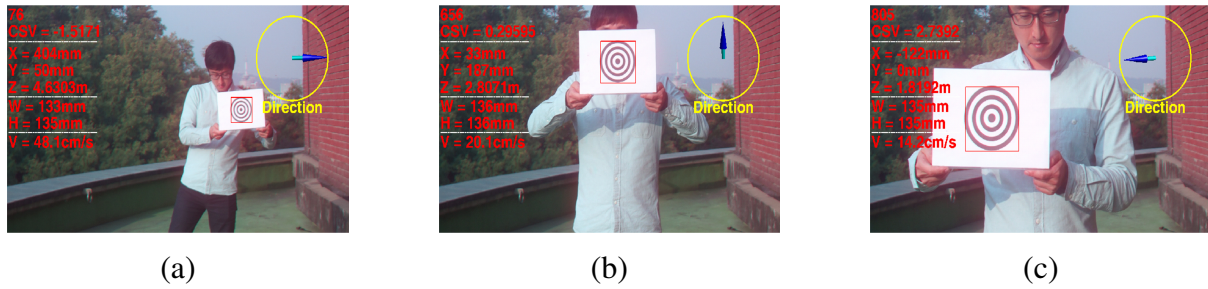
(a)   (b)   (c)

Figure 8. Estimation results of the 3D states of the object $\dot{\mathbf{x}}_t$ using the proposed method. (**a**) 76th frame; (**b**) 656th frame; (**c**) 805th frame.

Figure 9 shows the 3D trajectory of the moving object in the camera coordinate system. The trajectory without Kalman filtering has minor oscillations due to the estimation error. However, the trajectory with Kalman filtering smoothly changes without oscillation.

Figure 10 shows the estimation error of the distances and diameters of the concentric circle. This result can provide a key indicator for accuracy of the depth estimation. If the estimated diameter is similar to the real value, 140 mm, the estimated 3D distance is also accurate. As shown in Figure 10a, the proposed method robustly estimates the diameter of the circle with an error less than ± 10 mm.
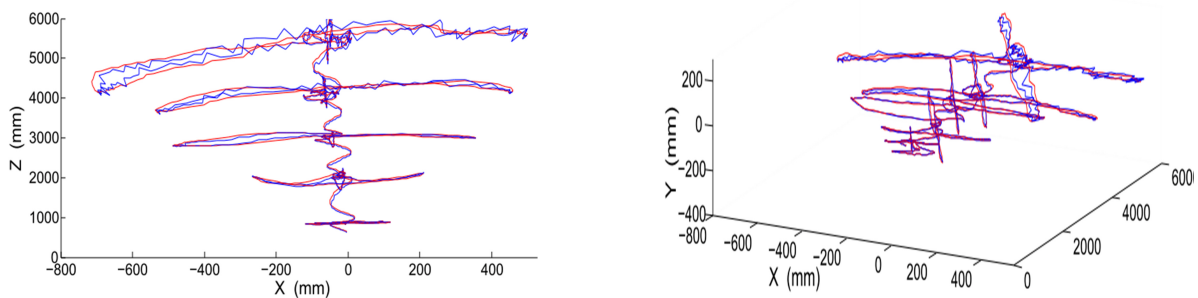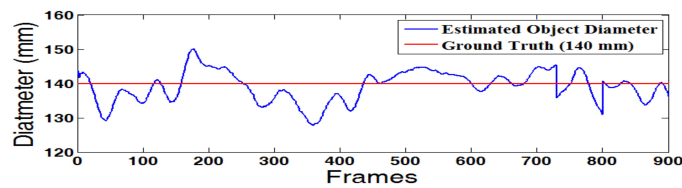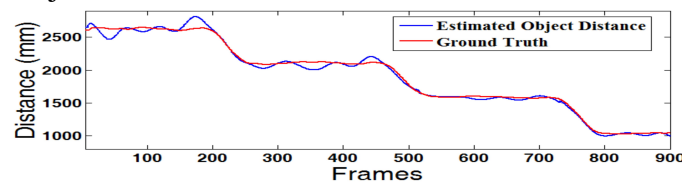

Figure 9. 3D trajectory of the moving object (red, with Kalman filtering; blue, without Kalman filtering).

(a) Object diameter estimation error of the concentric circle

(b) Object distance estimation error of the concentric circle

Figure 10. Comparison of the object diameter and distance estimation with the ground truth. (**a**) Object diameter estimation error of the concentric circle; (**b**) Object distance estimation error of the concentric circle.

Figure 11 shows the results of the estimated 3D states of the tracked object in another scene. The object having complex patterns was also successfully tracked with the proposed 3D states of the object.
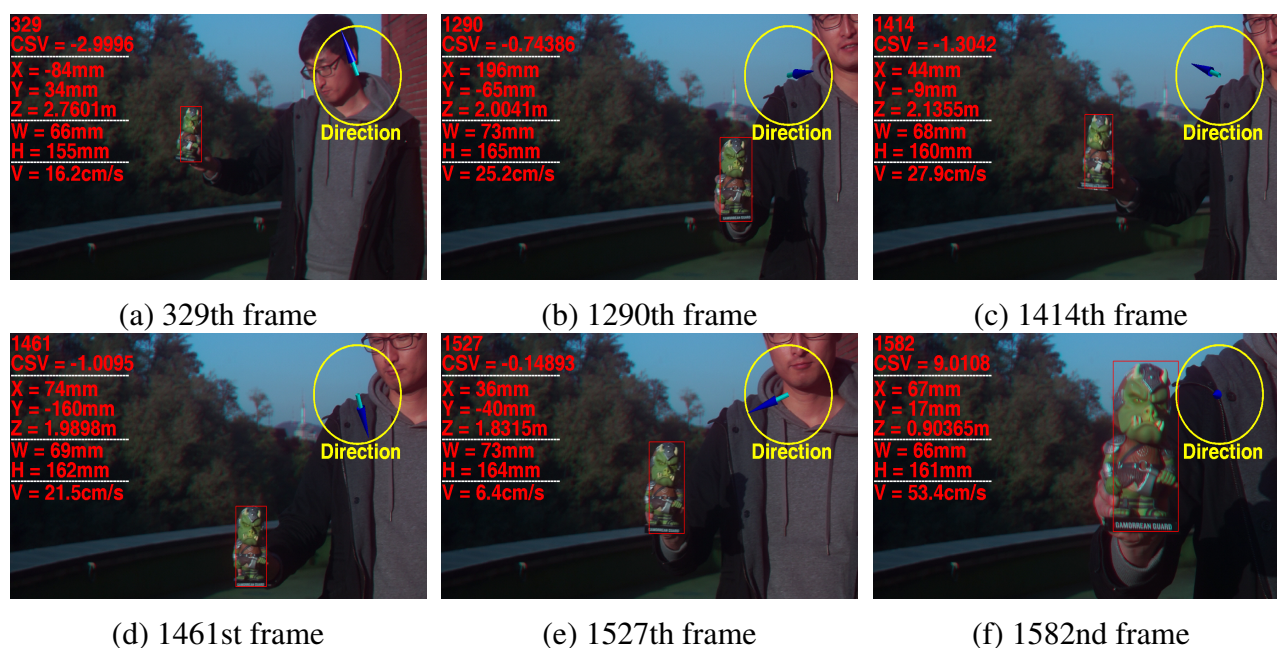

(a) 329th frame (b) 1290th frame (c) 1414th frame

(d) 1461st frame (e) 1527th frame (f) 1582nd frame

Figure 11. Estimation results of the proposed 3D states of the object $\dot{x}_t$ using the proposed method. (**a**) 329th frame; (**b**) 1290th frame; (**c**) 1414th frame; (**d**) 1461st frame; (**e**) 1527th frame; (**f**) 1582nd frame.

As a result, the proposed method can estimate the 3D states of an object with a single object motion. However, it cannot accurately estimate the 3D states of multiple moving objects with possible occlusions or fade-in/-out, because the object tracking method may fail to track the object.

6. Conclusions

In this paper, we proposed a novel 3D object direction and velocity estimation method for object tracking using the DCA-based computational RGB-D camera. We estimated the amount of the color shifting value as the disparity by minimizing the error function. After the proposed DCA camera calibration, the 2D states of the object were converted into the 3D camera coordinates using an approximated mathematical model of the relationship between color shifting values and the actual distance of the object. Finally, the 3D object moving direction and velocity are calculated by the temporal changes of the object.

Based on the experimental results, the DCA camera-based object tracking system can successfully estimate the three-dimensional direction and velocity of a randomly-moving object. More accurate depth estimation with the extended range will be possible using an improved sub-pixel interpolation-based registration method in future research.

Supplementary Materials

Supplementary materials can be accessed at: http://www.mdpi.com/1424-8220/15/1/995/s1.

Acknowledgments

Author Contributions

Seungwon Lee initiated the research and designed the experiments; Kyungwon Jeong and Jinho Park performed experiments; Joonki Paik wrote the paper

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Gade, R.; Moeslund, T. Thermal tracking of sports players. *Sensors* **2014**, *14*, 13679–13691.
2. Li, X.; Guo, R.; Chen, C. Robust pedestrian tracking and recognition from FLIR video a unified approach via sparse coding. *Sensors* **2014**, *14*, 11245–11259.
3. Le, A.; Jung, S.; Won, C. Directional joint bilateral filter for depth images. *Sensors* **2014**, *14*, 11362–11378.
4. Fernandez-Sanchez, E.; Diaz, J.; Ros, E. Background subtraction based on color and depth using active sensors. *Sensors* **2013**, *13*, 8895–8915.
5. Blanco, C.; Mantecon, T.; Camplani, M.; Jaureguizar, F.; Salgado, L.; Carcia, N. Foreground segmentation in depth imagery using depth and spatial dynamic models for video surveillance applications. *Sensors* **2014**, *14*, 1961–1987.
6. Zhu, Y.; Fujimura, K. A bayesian framework for human body pose tracking from depth image sequences. *Sensors* **2010**, *10*, 5280–5293.
7. Dhond, U.; Aggarwal, J. Structure from stereo-a review. *IEEE Trans. Syst. Man Cybern.* **1989**, *19*, 1498–1510.
8. Schastein D.; Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithm. *Int. J. Comput. Vision* **2012**, *47*, 7–42.
9. Zhou, C.; Nayar, S. Computational cameras: Convergence of optics and processing. *IEEE Trans. Image Process.* **2011**, *20*, 3322–3340.
10. Maik, V.; Cho, D.; Shin, J.; Har, D.; Paik, J. Color shift model-based segmentation and fusion for digital autofocusing. *J. Imaging Sci. Technol.* **2007**, *51*, 368–379.
11. Kim, S.; Lee, E.; Hayes, M.; Paik, J. Multifocusing and depth estimation using a color shift model-based computational camera. *IEEE Trans. Image Process.* **2012**, *21*, 4152–4166.

12. Lee, S.; Lee, J.; Paik, J. Simultaneous Object Tracking and Depth Estimation Using Color Shifting Property of a Multiple Color-Filter Aperture Camera. In Proceedings of the IEEE Conference Acoustics, Speech, and Signal Processing, Praha, Czech Republic, 22–27 May 2011; pp. 1401–1404.

13. Lee, S.; Kim, N.; Jung, K.; Hayes, M.; Paik, J. Single Image-Based depth Estimation Using Dual off-Axis Color Filtered Aperture Camera. In Proceedings of the IEEE Conference Acoustics, Speech, and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 2247–2251.

14. Lee, S.; Hayes, M.; Paik, J. Distance estimation using a single computational camera with dual off-axis color filtered apertures. *Opt. Express* **2013**, *21*, 23116–23129.

15. Ross, D.; Lim, J.; Lin, R.; Yang, M. Incremental learning for robust visual tracking. *Int. J. Comput. Vision* **2008**, *77*, 125–141.

16. Lucas, B.; Kanade, T. An Iterative Image Registration Technique with an Application to Stereo Vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence, Vancouver, BC, Canada, 24–28 August 1981; pp. 674–679.

17. Periaswamy, S.; Farid, H. Elastic registration in the presence of intensity variations. *IEEE Trans. Med. Imaging* **2003**, *22*, 865–874.