Automatic Detection of Initial Object Region in Multiple Color-Filter Aperture Camera-Based Surveillance System

R.Raghu¹, K.Venkatesh², S.Jothiprasanth³, M.Pruthiv⁴

¹(UG Scholar, Dept.of.ECE, Christ the king Engineering college, Tamilnadu, India, raghuram006@gmail.com)
²(Asst. Professor, Dept.of.ECE, Christ the King Engineering College, Tamilnadu, India, venkanadasan@gmail.com)
³(UG Scholar, Dept.of.ECE, Christ the king Engineering college, Tamilnadu, India, dariprasanth@gmail.com)
⁴(UG Scholar, Dept.of.ECE, Christ the king Engineering college, Tamilnadu, India, pruthiv.saran@gmail.com)

Abstract—In this paper, we present an adaptive background generation method for automatic selection of initial object regions, which realizes simultaneous object detection and depth estimation using multiple color-filter aperture (MCA) camera. Since the conventional background generation method does not fit the depth estimation using the MCA camera, we propose a novel color-based background generation method which can reduce interference in the object region for stable depth estimation. For efficient estimation of color shifting vectors in the extracted object region, a simplified elastic registration (ER) algorithm is used. The proposed simplified method is essential factor to realize realtime depth estimation and tracking, which is the primary condition for consumer applications. Finally, the object distance is determined by using the relationship between the pre-specified distance transformation function and the estimated shifting vectors of the corresponding object region. Although traditional depth estimation methods generally use dual cameras for stereo vision, the proposed method uses only a single camera for both object detection and depth estimation. Experimental results show that the proposed MCA camera-based object detection system can be used in a variety of consumer surveillance systems such as intelligent transport systems, 3D-based cameras and advanced safety vehicles.

Index Terms—Adaptive background generation, surveillance system, depth estimation, computational camera.

1. INTRODUCTION

Object detection and depth estimation are important and challenging problems in consumer video analysis applications, such as surveillance systems, intelligent transport systems 3D-based digital compact cameras, and advanced safety systems for vehicles [1][2]. A two-dimensional (2D) image acquired by a conventional camera cannot provide sufficient information to overcome the occlusion problem and to estimate depth in video surveillance and robot vision. For solving these problems, estimation of three-dimensional (3D) depth information has been intensively studied for the past several decades. Most conventional depth estimation methods have relied on either multiple images, for example stereo vision or on additional cues such as shading, focusing, and motion. Stereo matching is a depth estimation method using binocular disparity generated by a stereo camera [3]. In spite of many advantages, it has a fundamental limitation that a pair of images of the same scene should be acquired by two cameras with both temporal and spatial synchronization. As an alternative to binocular systems, monococular methods have also been studied. Depth from defocus is a single camera-based depth estimation method that measures the amount of defocus blur from a pair of images with different focus settings on the same scene. However, this approach is limited to still photography because a fixed camera view is required for taking multiple defocused images [4]. Zhuo has used a single defocused image for depth estimation by considering the amount of blur measured by using the Gaussian gradient ratio at an edge as depth [5]. However, it has the ambiguity problem between defocus and motion blurs.

Recently, computational cameras have been developed that acquire information that cannot be obtained with a conventional digital camera, and therefore provide interesting new possibilities in consumer video devices. Computational cameras use a combination of unconventional optics and computations to produce the final image. The motivation for developing these cameras is to create new imaging functionalities that would be difficult to achieve using the traditional camera, such as enhanced fields of view, increased spectral resolution, and larger dynamic range [6]. A color shift model using a multiple color-filter aperture (MCA), which can be considered as a special optical device in a computational camera, provides depth information of objects at different distances according to the direction and amounts of relative shifts between color channels [7][8]. The MCA is inserted between the lens and the imaging sensor to provide geometric information of an object from the color shifting S. Lee et al.: Adaptive Background Generation for Automatic Detection of Initial Object Region 105 in Multiple Color-Filter Aperture Camera-Based Surveillance System property of the color filter array. Based on the MCA configuration, E. Lee has proposed a multi-focusing method using region-based depth estimation [9], and S. Lee has estimated the depth of an object by tracking the region-of-interest. Since these objects were manually selected in previous works, we present an object detection method using adaptive background generation [11] for single camera-based fully automatic object tracking and distance estimation. Given color video input and the depth information of an object in the previous frame we first detect the object region using adaptive background generation [11]. The adaptive background generation method generates the background image updated at the non-moving region from...
input image by estimating motion vectors. The proposed depth estimation and object detection framework is shown in Fig. 1. The object regions are detected using background subtraction between input and background images as shown in Fig1.

**Fig. 1. The proposed depth estimation and object detection framework.**

(MCA: multiple color-filtered apertures, ABG: adaptive background generation, SER: simplified elastic registration, CSVs: color shifting vectors)

Since the conventional background generation method does not fit the depth estimation using the color shifting model, we present a novel color-based background generation method which can reduce interference in the object region for depth estimation. Once the background has been estimated, the color shifting vectors of objects in the foreground are estimated. Because RGB color planes of the detected object region are misaligned among each other, shifting vectors between green-and-red (GR) and green-and-blue (GB) channels are estimated by using the simplified elastic registration algorithm.

The original elastic registration (ER) algorithm proposed by Priyawarthy in [12] has considered intensity variation in the form of contrast and brightness parameters as well as local affine motions between two medical images. For efficient estimation of color shifting vectors in the extracted object region, the ER algorithm can be simplified under assumption that the object has only translation components in the affine matrix and that the same color shifting occurs in the entire region. In this paper, we present a novel color shifting vector estimation method that uses the simplified elastic registration and convergence of GR and GB shifting vectors. The proposed simplification method for the original elastic registration is the key factor to enable real-time depth estimation and tracking, which is the primary condition for consumer applications. Finally, the object distance is determined by using the relationship between the prespecified distance transformation function and the estimated shifting vectors of the corresponding object region. Experimental results show that the proposed MCA camerabased object detection system can be used in a variety of consumer surveillance systems such as intelligent transport systems, 3D-based digital compact cameras, and advanced safety vehicles. The paper is organized as follows. Section 2 presents principle of the CSM. Sections 3 and 4 respectively present detection of object region using adaptive background generation and reliable depth estimation using color-based background subtraction. Cooperative object detection and depth estimation system is proposed in section 5. Section 6 presents experimental results, and section 7 concludes the paper.

**II. PRINCIPLE OF THE COLOR SHIFT MODEL: A REVIEW**

The aperture of an optical system is the opening that adjusts the amount of light entering the camera. The center of an aperture is generally aligned with the optical axis of the lens, and the convergence pattern on the image plane will form either a point or a circular region depending on the distance of the object from the plane of focus. When the center of the aperture is not aligned with the optical axis, the convergence point will be shifted away from the optical axis by an amount that is a function of the distance of the object from the plane of focus of the camera.

For the complete color shift model using the MCA camera, we use three apertures with red (R), green (G), and blue (B) filters, which form the regular triangle whose center is at the optical axis of the camera as shown in Fig. 2. The specific location of the light collection depends on the distance between the lens and the object.

The main advantage of the color shift model is that it provides additional depth information which can be estimated from the direction and amount of color deviation from the optical axis as shown in Fig. 3. For example, an object at the in-focus position does not have color misalignment. Fig. 3(b) shows the in-focused imaging result. However, if the object goes farther than the in-focus position, color misalignment occurs in the image on the sensor as shown in Fig. 3(a). 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 as shown in Fig. 3(c).

**III. AUTOMATIC DETECTION OF OBJECT REGIONS USING ADAPTIVE BACKGROUND GENERATION**

To detect objects in the MCA camera image for object tracking and depth estimation, we use the adaptive background generation framework proposed in [11]. The first step in this method is the estimation of motion vectors using the optical flow. Let \( f(x, y) \) and \( f(x, y) \) be the current and the previous image frames, respectively, then optical flow each pixel \((x, y)\) is obtained by minimizing the Euclidean distance expressed as where \((dx, dy)\) that minimizes \(D(x, y)\) represents the displacement of the pixel at \((x, y)\), and the size of the searching window is \( (2w+1) \times (2w+1) \). If \( D(x, y) < \) \( T \), where \( T \) is a pre-specified threshold for the Euclidean distance, then the pixel \((x, y)\) is assumed to be within the background, and the background image is updated at the corresponding pixel as follows, \( Bf \) and \( 1 - Bf \) respectively...
represent background images at time $t$ and $t \square 1$, and $\square$ is the mixing ratio in the range $[0,1]$. Because depth is estimated using the color shifting property of the MCA camera, the original background generation method does not fit the color shifting-based depth estimation algorithm. We consider the color information to interference among color channels at the boundary of objects. Given the background in the $t$-th frame, the proposed color-based object region detection is performed.

Fig. 4 shows the result of object detection, where the detected region has color shifting among RGB color channels in the neighborhood of object boundaries.

$$f \overset{G}{x \ y} \overset{f}{C \ x \ y} \overset{v}{G \ y}$$ with respect to horizontal and vertical directions, respectively. The approximated errors can be rewritten in the vector form. Since $E(v)$ is a quadratic function of $v$, a closed-form solution for the vector $v$ that minimizes this error may be found by differentiating with respect to $v$ and setting the result equal to zero. The result is a set of linear equations that need to be solved for $v$. Finally, these error functions can be combined into one vector function determined as

$$G = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

where $\square$ represents the region of an object. Note that the argument of $EGR(\cdot)$ is not $(\cdot) GR \overset{G}{GR} \overset{G}{GR} \overset{(\cdot)}{\square y}$, but $(\cdot) GB \overset{GB}{GB} \overset{GB}{GB} \overset{\square x y}$ that minimizes (6). Therefore, Periaswamy used a first-order truncated Taylor series approximation of the errors such as

where $C$ will generally be invertible if the corresponding region is sufficiently large and the image has sufficient content. We can further simplify (10) using the position property of the MCA system. If we configure the same horizontal axis of green and blue channels, then $\overset{GB}{GB} \overset{GB}{GB} \overset{\square x y}$ is equal to zero. So, $v$ can be formed as the single parameter $\overset{GB}{GB} \overset{GB}{GB} \overset{\square x y}$ using the trigonometric property and the angle of the color-filter aperture.

The denominator and the numerator are $1 \times 1$ matrices. This approximation results in the estimation of a single parameter without matrix inversion. As discussed in [12], a Gaussian pyramid-based iterative coarse-to-fine approach is used in order to accommodate large displacements between the target and the source image.

Fig. 6(a) shows an MCA image containing two objects with different distances. Fig. 6(b) shows the result of the registered object A with $v \square (12.2107, 0.7205)$. Fig. 6(c)

Fig. 5. (a) Geometric configuration of the MCA and (b) color shifting property based on the green channel.

IV. RELIABLE DEPTH ESTIMATION USING COLOR-BASED BACKGROUND SUBTRACTION

To estimate the depth of an extracted object, image registration among RGB color channels is performed to estimate the color shift vectors between GR and GB channels. Here, we modify the elastic registration algorithm proposed by Periaswamy that geometrically registers a source image to a target image using an affine matrix [12]. Since we are concerned with estimating the color shift vectors of a single homogeneous object, the displacement can be globally modeled by a simple translation. Another property of the MCA is that three apertures are located at vertices of the equilateral triangle as shown in Fig. 5(a). Using these properties, the original elastic registration algorithm is simplified to can reduce the computational cost without sacrificing the accuracy of the depth estimation. More specifically, in the $i$-th object region the color shift model between GR and GB channels, minimizing the following quadratic error functions.
shows the result of the registered object B with \( \nu_1 \) and \( \nu_2 \) (12.0700, 0.3205).

Two color shift vectors enable to estimate the relative distance of objects A and B.

V. COOPERATIVE DETECTION AND DEPTH ESTIMATION

In this section, we present a relative distance transformation function generation method by combining the detected initial object region and color shift model-based depth estimation. Using color shifting vectors we can estimate the relative depth of a detected object. In this work we estimate the distance of an object using the normalized distance transformation function, for which we acquire video of a walking human using the MCA camera in the range between 17 and 49 meters as shown in Fig. 7.

Figs. 7(a)-(c) show the results of tracking a walking person.

The detected object is registered using the estimated color shift vectors as shown in Two components in the color shift vector, \( GB_x \) and \( GB_y \), are simultaneously estimated, and the estimated values are shown in Fig. 8. As the person comes close to the focusing position (approximately 21 meters in this experiment) of the camera, the shifting vector converges to zero. If the person passes the focusing position and keeps approaching to the camera, the shifting vector diverges as shown in Fig. 8(a).

Fig. 8(b) shows GR and GB distance transformation function quantized by one meter.

VI. EXPERIMENTAL RESULTS

For the experiment, we captured three test sequences of size 640x480 and of rate 30 frames per second using the MCA camera. Three test sequences containing color shift model-based channel misalignment are called CSM-A, CSM-B, and CSM-C sequences, respectively. To evaluate the accuracy in depth estimation, we compare the result with the ground truth distance. Fig. 9(a) shows that in Multiple Color-Filter Aperture Camera-Based Surveillance System the estimated distance is almost identical to the ground truth since CSM-A sequence was taken along the same direction to the distance measurement sequence. On the other hand Fig. 7(c) shows that estimation error increases as the object comes closer to the camera, since CSM-C sequence was taken from the different view of the distance measurement sequence. The result of depth estimation using the proposed algorithm is shown in Fig. 10, where three different sequences were tested. Fig. 6 shows experimental results of the distance estimation. Each detected object using background subtraction is registered by using simplified elastic registration algorithm with estimated color shifting vectors as shown in Fig.4(c). An object at 21m distance in the last row of Fig.3 does not have color shifting because it is in-focus.

VII. CONCLUSION

In this paper, we addressed an adaptive background generation method for automatic selection of initial object region for simultaneous object detection and depth estimation using multiple color-filter aperture camera. Since the conventional background generation method does not fit the depth estimation under the color shifting model in the MCA camera, a novel color-based background generation method was proposed for reducing interference in the object region for stable depth estimation. Using the simplified elastic registration algorithm the object’s distance can be determined by using the relationship between the pre-specified distance transformation function and the estimated shifting vectors of the corresponding object region.

A traditional depth estimation method has generally uses multiple cameras for stereo vision. However, the proposed method used only a single camera for object detection as well as depth estimation. Experimental results showed that the proposed MCA camera-based object detection system can be used in a variety of consumer surveillance systems such as intelligent transport systems, 3D-based digital compact cameras, and advanced safety vehicles.

REFERENCES