Background subtraction based on Gaussian mixture models using color and depth information

Conference Paper · January 2015

DOI: 10.1109/ICCAIS.2014.7020544

Young-Min Song
Gwangju Institute of Science and Technology

Jongmin Yu
Gwangju Institute of Science and Technology

Cheonwi Park
Gwangju Institute of Science and Technology

Background Subtraction based on Gaussian Mixture Models using Color and Depth Information

Young-min Song, SeungJong Noh, Jongmin Yu, Cheon-wi Park, and Byung-geun Lee, *Member*, *IEEE*

***Abstract*— In this paper, we propose a background subtraction (BGS) method based on the Gaussian mixture models using color and depth information. For combining color and depth information, we used the probabilistic model based on Gaussian distribution. In particular, we focused on solving color camouflage problem and depth denoising. For evaluating our method, we built a new dataset containing normal, color camouflage and depth camouflage situations. The dataset files consist of color, depth and ground truth image sequences. With these files, we compared the proposed algorithm with the conventional color-based BGS techniques in terms of precision, recall and F-measure. As a result, our method showed the best performance. Thus, this technique will help to robustly detect regions of interest as pre-processing in high-level image processing stages.**

I. INTODUCTION

Detection of moving objects is an essential process in visual surveillance. Among the various techniques for moving object detection, conventionally, many background subtraction (BGS) methods have been used for it. The basic BGS methods utilizing temporal medians of previous n frames [1] and statistical approaches employing the Gaussian mixture model [2] had been proposed. More recently, self-organizing map [3], [4] and multiple features based methods were devised [5], [6].

However, these BGS techniques have some fundamental limitations because they utilized human perception (i.e., visible light) based color spaces such as the red, green, and blue (RGB), the hue, saturation, and value (HSV) and the YUV where Y and UV represent luminance and chrominance, respectively. Basically, those methods are weak to color camouflage situations and also sensitive to illumination changes.

To handle the problems, other BGS approaches [7], [8], [9] have been proposed using a new type of information besides color information. Especially, the depth information from stereo cameras, Microsoft Kinect or time-of-flight (ToF) sensors have been used with color information. Harvile *et al*. [7] proposed the foreground segmentation method using the YUV color space with the additional depth values. The method in [7] tried to figure out the correlation between the YUV and depth with the covariance matrix. However, there are obvious limitations because the method in [7] uses the 4-channel feature vector by adding the depth value to a 3-channel color vector, and color and depth are totally different types of information. In state of the art researches, Fernandez-Sanchez *et al*. [9] and Camplani *et al*. [8] applied color and depth in each background model (e.g., $P(x_c)$, $P(x_d)$) not using color and depth in the same model (e.g., $P(x_{cd})$). They used background modeling based on the Gaussian mixture model (GMM) and codebook, respectively, and built the final classifier combining the color model and depth models. In our paper, we present a fast BGS method based on GMM using the color and depth. Background and foreground models are estimated similarly as proposed in [10]. For reasonably combining color and depth, a likelihood of background and foreground model was induced from the product of the likelihood of color and depth model. Moreover, for achieving fast real-time implementation, 3-channel color vectors were converted to 1-channel gray scale values. Nevertheless, our algorithm showed better performance than the conventional BGS techniques for dataset which we made. We used Microsoft Kinect to get color and depth information with the OpenNI software development kit but Kinect generated noisy data. Although there is no pre or post-processing for dealing with depth noise, our method decreases depth noise probabilistically.

II. PROPOSED METHOD

Our approach starts from the idea that so many BGS methods already have been devised but the color-based BGS techniques have definite limitations. They cannot segment the object from the same-colored background due to color camouflage and very sensitive to illumination changes. To solve these problems, other approaches [7], [8], [9] using depth information have been proposed. For example, in Harvile *et al*. [7], they extended each pixel vector $X = (Y,U,V)$'s dimension to be (Y,U,V,D) by adding depth (D) value and built the background model based on the Gaussian mixture model (GMM) using the 4-channel pixel vectors where a pixel corresponds to a GMM, and then the number of GMMs is equal to the number of pixels in an scene (i.e., $width \times height$). A GMM can be initialized based on by using a clustering technique such as the EM algorithm and k-means on the observed T frames. However, the depth information from the current devices such as Kinect, ToF sensors and stereo cameras still has a lot of noise, thus it is not reliable. Also, clustering spends much processing time to be used in real-time applications and assumes the situation without any moving object while initialization time. Thus, for

This work was supported by the Center for Integrated Smart Sensors funded by the Ministry of Science, ICT & Future Planning as Global Frontier Project (CISS-2013M3A6A6073718)

Y. Song, S. Noh, and J. Yu is with the School of Information and Communications, Gwangju Institute of Science and Techonology, 123 Cheomdangwagi, Buk-gu, Gwangju 500-712, South Korea sym@gist.ac.kr

C. Park and B. Lee is with the School of Mechatronics, Gwangju Institute of Science and Techonology, 123 Cheomdangwagi, Buk-gu, Gwangju 500-712, South Korea cwpark@gist.ac.kr

reliably combining color with depth, we did not add depth to a color vector but designed two probabilistic background models corresponding to color and depth based on GMM and denoised depth image with the model. Our algorithm follows two basic steps. The first is background modeling and the second is background subtraction using the model.

A. Background Modeling

Stauffer and Grimson [10] initialized the background models using the recent history of t frames, where as we initialized the background models using the first K frames in a pixel-wise level. K is the number of the Gaussian models in a GMM and X_i indicates a particular pixel located on (x,y) in the i-th frame I as in (1).

$$\{X_1,...,X_K\} = \{I(x,y,i) : 1 \leq i \leq K\} \tag{1}$$

Our initialization method can save clustering time. Each pixel has a corresponding mixture of K Gaussians. Heuristically, K is chosen from 3 to 5. However, we didn't add update procedures. Then, the probability function at the given pixel value at time t is as follows:

$$P(X_t) = \sum_{i=1}^{K} \omega_i \cdot \eta(X_t, \mu_i, \sigma_i^2) \tag{2}$$

where X_t is the pixel observed at time t, ω_i means the weight related to the i-th Gaussian distribution with mean μ_i,t and standard deviation σ_i^2. η is a Gaussian probability density function as in

$$\eta(X_t,\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{(X_t-\mu)^2}{2\sigma^2}} \tag{3}$$

in which μ_i is the pixel value in the i-th image sequence and σ_i is the user parameter. We set it from 10 to 20 randomly. After all the models are initialized, we should separate the background and foreground models as follows:

$$B = argmin_b(\sum_{k=1}^{b} \omega_k > T) \tag{4}$$

where the first B distributions are selected as the background models when the sum from ω_1 to ω_b exceeds a threshold T, and remaining distributions are the foreground models. We applied gray scale color (0-255) and depth (0-255) channels on these Gaussian distributions for a single variable. Each channel builds each Gaussian probability function. Statistical meaning of these functions is the likelihoods on the given pixels. Let the color model and the depth model be $P(c_t)$ and $P(d_t)$ as follows:

$$P(x_t) = P(c_t) \cdot P(d_t) \tag{5}$$

Then, we will introduce how to subtract background using this probabilistic model in the next section.

B. Background Subtraction

In this section, we explain background/foreground segmentation with the probabilistic background models. First, the color model B_c and depth model B_d classify the observed pixels c_t and d_t at time t, respectively, as shown in Fig. 1. A pixel finds a matched Gaussian model using the Euclidian distance :

$$\|x_t - \mu_i\| \geq k \tag{6}$$

where k is a constant threshold identical to 2.5. A survey about background modeling using mixture of Gaussians [11] introduced user parameter values in detail for implementing [10]. They presented Mahalanobis distance to classify a pixel. However, we used the Euclidean distance since gray scale color and depth values are one dimension from 0 to 255 in which there is no correlation. A pixel is classified as one of the following 3 cases.

Case1: If a match is found and the pixel is classified as background, a matched model is the background model.
Case2: If a match is found and the pixel is classified as foreground, a matched model is the foreground model.
Case3: If no match is found, the pixel belongs to foreground.

In the case of color value c_t, after the pixels at time t are classified as background or foreground, the pixel values are computed by using the inequality for the matched Gaussian distribution as follows:

$$\theta \cdot \eta(c_t, \mu_i, \sigma_i^2) \geq maximum\,pixel\,value \tag{7}$$

in which θ is a constant to scale the Gaussian probability density function values. For example, we assume that the minimum pixel value is 0 and the maximum is 255. Then, θ is equal to 10,000. In the case 1, when a pixel is matched to a background model, if the inequality is satisfied, the pixel has 0. Otherwise, the pixel has $\theta \cdot \eta$. In the case 2, when a pixel is matched to a foreground model, if the inequality is satisfied, the pixel has 255. Otherwise, the pixel has 255-$\theta \cdot \eta$. In the case 3, the value becomes 255 because there is no a background or foreground model to decide the probability of the pixel. Eventually, the BGS results R_c consists of pixel values 0 to 255 where the higher probability $P(c_t)$ is closer to 0 (background) as shown in Fig. 1(a). In the case 1 and the case 3 for the BGS results R_d based on B_d, the probabilistic pixel values are allocated in the same way as R_c. However, in the case 2, the depth-based results have 255 regardless of whether the inequality is satisfied or not. This is for filtering the false positive regions. Fig. 1(b) shows an example of R_d. Thus, the final BGS results R_{final} as seen in Fig. 1(c) are computed as follows:

$$\begin{aligned} &if\ \ R_c(x,y) * R_d(x,y) > maximum\,pixel\,value \\ &then\ \ R_{final}(x,y)\ \ is\ \ foreground \\ &else\ \ then\ \ R_{final}(x,y)\ \ is\ \ background \end{aligned} \tag{8}$$

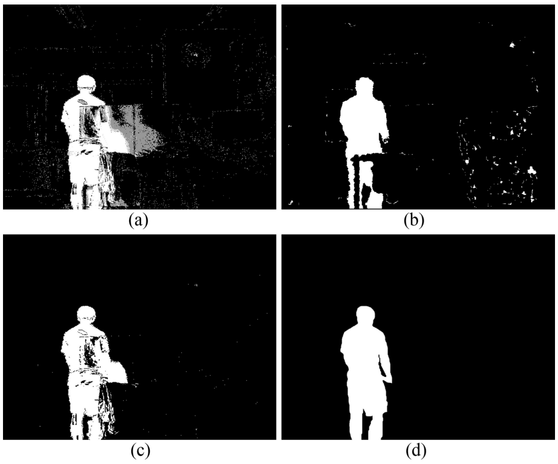

Fig. 1. BGS results based on (a) color model B_c (b) depth model B_d (c) proposed method (d) ground truth

C. Depth Image Denoising

Our method can handle the noise of depth image. We found the phenomenon that $R_c(x,y)$ is rightly classified as background at a location (x,y), where false foreground detection occurs in $R_d(x,y)$ for depth noise. Thus, we designed the pixel classification method as seen in (8). Even if there are false positive regions in R_d caused by depth noise, the most of those regions belongs to background in R_{final} as in Fig. 1(c).

III. Dataset

We built a new dataset for evaluating BGS algorithm using color and depth and comparing it with the conventional color-based BGS techniques. The dataset contains 3 categories: 1) Normal situation, 2) color camouflage and 3) depth camouflage that has 1, 2, and 1 scenes, respectively. Basically, we focused on solving the color camouflage problem so made one more scene in the color camouflage category. Every scene consists of color, depth and ground truth image sequences as shown in Fig. 2.

1) Normal Situation: There is not any intended color or depth camouflage situation in the given frames.

2) Color Camouflage: Intended color camouflage regions exist, where an object of interest has the same color with background.

3) Depth Camouflage: Intended depth camouflage regions exist, where an object of interest has the same depth (i.e., distance from depth sensor) with background.

IV. Experimental Results

We introduce evaluation results of our algorithm by comparing it with other color-based conventional BGS techniques [1], [2], [3], [4], [5], and [6], using our dataset. We used three main measures in terms of precision, recall, and F-measure [13] as follows:

$$Precision(P) = \frac{TP}{TP+FP} \tag{9}$$

$$Recall(R) = \frac{TP}{TP+FN} \tag{10}$$

$$F-measure = \frac{2PR}{P+R} \tag{11}$$

in which TP, FP and FN denotes true positive rate, false positive rate and false negative rate, respectively. F-measure indicates the harmonic mean of precision and recall.

The experimental results show that our approach is the best as seen in Fig. 3 in terms of precision and F-measure. However, in view of recall, our algorithm ranked 3^{rd}, because our method detected less camouflage regions than the 1^{st} algorithm. However, that does not mean that the proposed method is worse than the 1^{st} and the 2^{nd}, because they segmented some true background as foreground, the more false positive alarms occurred than ours, and thus ranked joint 4^{th} in recall.

V. Conclusions

We developed the BGS method based on the GMM using color and depth information to solve the limitations of color-based BGS, especially in color camouflage situation. Thus, we built the probabilistic background model for combining color and depth information. In addition, we made a new dataset for evaluating the BGS algorithm using color and depth. Using the dataset, we compared our method with the conventional color-based BGS methods [1], [2], [3], [4], [5] and [6] in terms of precision, recall and F-measure [13]. We found that there is no color noise in the depth noise regions, and made color and depth information-based results fill each other's false negative regions. On the other hand, the conventional color-based BGS methods cannot detect color camouflage regions at all. As a result, the proposed method not only showed the better performance than others but also decreased the depth noise. As our future work, our probabilistic background models will be extended to deal with multi-modal data such as those from thermal imaging cameras and night vision besides depth. Also, it can improve high-level applications (e.g., object detection, object tracking and action recognition) performance as pre-processing for detecting regions of interest robustly in color-based problems like dynamic color change, and low illumination in addition to color camouflage.

VI. Acknowledgments

The authors gratefully acknowledge reviewers' comments. Also, I would like to express special thanks to my beloved girlfriend Hyejin Yu who always cheers me up.

References

[1] R. Cucchiara, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," in the IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, October 2003, pp. 1337-1342.

[2] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel," in Pattern Recognition Letters, vol. 27, 2006, pp. 773-780.

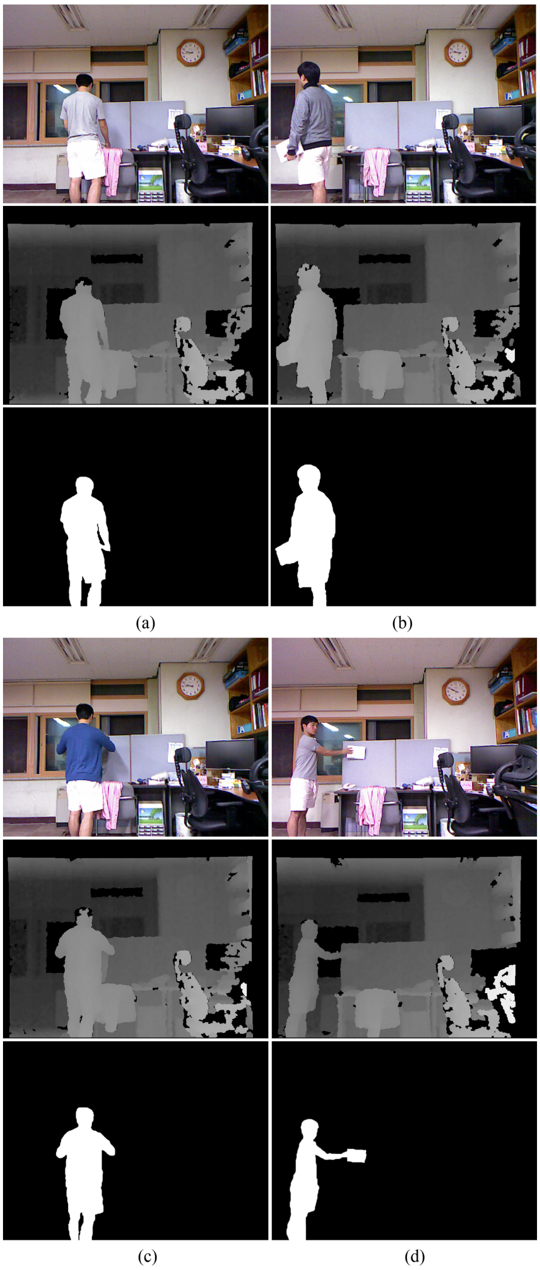

Fig. 2. The dataset samples of (a) color camouflage 1 (b) color camouflage 2 (c) normal situation (d) depth camouflage : In each situation (a), (b), (c) and (d), color, depth and ground truth images are on the top, middle, and bottom, respectively.

[3] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," in the IEEE Transactions on Image Processing, vol. 17, July 2008, pp. 1168-1177.

[4] L. Maddalena and A. Petorisino, "A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection," in Neural Computing and Applications, vol. 19, March 2010, pp. 179-186.

[5] Jian Yao and Jean-Marc Odobez, "Multi-Layer Background Subtraction Based on Color and Texture," in Proc IEEE Conf on Comp Vision and Pattern Recognition, June 2007.

[6] SeungJong Noh and Moongu Jeon, "A new framework for background subtraction using multiple cues," in Asian Conference on Computer Vision, 2012.

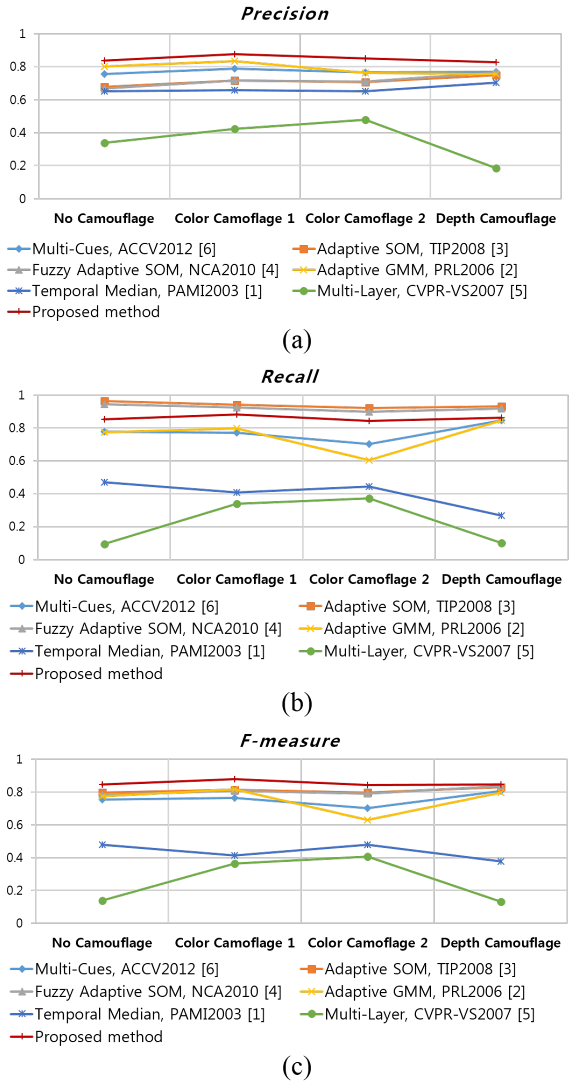

Fig. 3. Evaluation results in terms of (a) Precision (b) Recall (c) F-measure by comparing proposed method with [1], [2], [3], [4], [5] and [6]

[7] Harville, M., Gordon, G., and Woodfill, J. "Foreground segmentation using adaptive mixture models in color and depth", in Proceedings of the IEEE Workshop on Detection and Recognition of Events in Video, IEEE Computer Society, Los Alamitos, CA, USA, pp. 311, 2001

[8] Camplani, M., and Salgado, L. "Background foreground segmentation with RGB-D Kinect data: An efficient combination of classifiers," in Journal of Visual Communication and Image Representation, 25(1), 122136, 2014.

[9] Fernandez-Sanchez, E. J., Rubio, L., Diaz, J., and Ros, E. "Background subtraction model based on color and depth cues," in Machine Vision and Applications, 25(5), 1211-1225, 2013.

[10] Stauffer C, Grimson W. "Adaptive background mixture models for real-time tracking," in Proc IEEE Conf on Comp Vision and Pattern Recognition, 1999, 246-252.

[11] Bouwmans, T., Baf, F. El, and Vachon, B. "Background modeling using mixture of gaussians for foreground detection-a survey," in Recent Patents on Computer Science 1, 3, 219-237, 2008.

[12] Andrews Sobral, Laurence Bender, Donovan Parks, Jian Yao and Jean-Marc Odobez, SeungJong Noh, "A background subtraction library," in GitHub, Available: https://github.com/andrewssobral/bgslibrary

[13] David M.W., "Evaluation: from precision, recall and f-measure to ROC, informedness, markedness and correlation," in International Journal of Machine Learning Technology, vol. 2, 2011, pp. 37-6.