Robust and Accurate Segmentation of Moving Objects in Real-time Video

C. Rambabu and Woontack Woo

***Abstract*— Robust and accurate segmentation of moving object in real-time video is very important for object silhouette extraction in vision-based human computer interaction and video surveillance systems. However, the inherent problem of moving object segmentation based on background subtraction is to distinguish the changes from background disturbing effects such as noise and illumination changes. Therefore, the paper proposes an improved background subtraction scheme which is robust and accurate against noisy and changing illumination. The occlusion regions are detected based on the frame difference and background difference images. The moving shadows are eliminated very effectively by using the background statistical parameters. A queue-based connected component analysis method is introduced to isolate the moving object from background noise. Moreover, a pixel based background update is used to update the illumination changes. The proposed scheme has been implemented and evaluated regarding the segmentation quality and frame rate. From the experimental results it is known that the proposed method successfully extract object contours accurately against the illumination and noise changes.**

***Index Terms*— Moving Object, Illumination, Silhouette extraction, Moving Shadows.**

I. Introduction

FAST and accurate segmentation of moving objects in a video sequence is a basic step in most computer vision and video analysis applications, like perceptual human computer interfaces, virtual reality, video surveillance systems, etc. Accurate moving object segmentation will greatly improve the performance of object tracking, recognition, classification and activity analysis. Over the last decade, the segmentation of moving object from real-time video remains an open research problem. It has been a difficult problem to detect moving objects precisely, because of the illumination changes, small motions of background, moving shadows and occlusion. Several algorithms have been proposed for its solution.

The conventional moving object segmentation algorithms can be roughly classified into two categories as spatial homogeneity and change detection, according to their primary segmentation criterion. The region-based approaches [1] tends to track the object boundary more precisely than the other group. However, the computational complexity is very high because both the spatial segmentation and motion estimation are computationally intensive operations. On the other hand, the change detection-based approaches [2], [3] usually do not use of the spatial features and rely on the frame difference. However, the value of frame difference depends on the speed of object motion, so the quality of segmentation can not be maintained, if the speed of the object changes significantly in the sequence. Moreover, various change detection techniques have been developed based on the background subtraction [4]–[7]. Several algorithms have been proposed based on the environment conditions and application.

However, most of the techniques have been proposed under the restricted conditions (are motivated by specific applications). Moreover, they lack robustness in the presence of shadows, ghost effects, highlights, reflections, illumination changes, dark environment and if the foreground objects are similar to the background, and are not suitable for real-time applications. Hence, a robust and accurate algorithms can be developed based on the combination of multiple complementary features and selective update schemes. In this paper, we aim at formulating background subtraction method to segment the moving object in real-time video, robust against noisy data and varying illumination.

The paper is organized as follows. In the section II, the proposed background subtraction method is depicted. Section III presents the experimental results and discussion. Finally, Section IV outlines the conclusions and the possible future directions of this work.

II. Moving Object Segmentation

The section presents an improved version of background subtraction algorithm that preserves the accurate boundary of moving object in consideration of HSV color space. In general the RGB color space is not well behaved with respect to color perception, as a distance computed between two colors in RGB space does not reflect their perceptional similarity. In the present work, processing each color dimension independently improve the accuracy of the segmentation results. The HSV model, however, separates the intensity (V) from the chromatic components (H,S). The geometry of the HSV space makes more suitable for developing algorithms that rely on intensity measurements and on color information. Hence, the proposed method exploits the well-known HSV color space. Figure 1 shows the overall structure of the proposed algorithm composition. The proposed background subtraction scheme consists of five major functional components, namely, background modeling, occlusion detection, shadow elimination, post-processing, and background update.

A. Background Modeling

The goal of the background modeling is to train the background from a certain number of static background frames(with no video objects), witch has less noise and

This work was supported by the **Brain Korea 21 Project** in 2006, **ETRI OCR** and **Frontier** project
The authors are with the GIST U-VR Lab, South Korea. (email:{crambabu, wwoo}@gist.ac.kr)

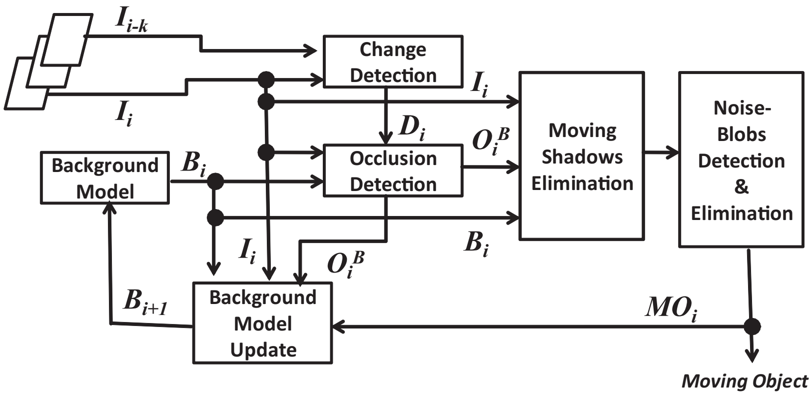

Fig. 1. Proposed Background Subtraction Approach

random illumination change, and will be used in background subtraction step. A background pixel is modeled by Gaussian distribution, characterized by its mean μ and its standard deviation σ. During the training the statistical parameters are updated by using recursive linear update technique.

Let I_i denote the input background frame at time 'i'. The pixel distributions in the first frame is unknown. So, initialize the mean μ_0 with the values of the first frame and set the variances to zero. Then, the distributions are updated from the next consecutive background frames. The mean μ_i for background frame at time 'i' can be defined as

$$\begin{aligned} \mu_0 &= I_0; \quad i = 0 \\ \mu_i &= (1-\alpha)\mu_{i-1} + \alpha I_i; \quad i \geq 1 \end{aligned} \tag{1}$$

Similarly, the variance σ_i^2 for background frame at time 'i' can be defined as

$$\begin{aligned} \sigma_1^2 &= (I_1 - \mu_0)^2; \quad i = 1 \\ \sigma_i^2 &= (1-\alpha)\sigma_{i-1}^2 + \alpha(I_i - \mu_i)^2; \quad i \geq 2 \end{aligned}$$

Where α is the learning rate of the linear model. The mean of the pixels at (x,y) in the N continuous frames can be formulated as

$$B(x,y) = [\mu_i(x,y) \; | i = N] \tag{2}$$

Similarly, the N frame standard deviation can be obtained by

$$\sigma_N(x,y) = \left[\sqrt{\sigma_i^2}(x,y) \; | i = N\right] \tag{3}$$

Where $B = < B^H, B^S, B^V >$ and $\sigma_N = < \sigma_N^H, \sigma_N^S, \sigma_N^V >$. The recursive linear update method utilize fewer number of frames and less memory space to train the background while compared to the conventional N frame averaging method.

B. Occlusion Detection

Background subtraction is an efficient way to discriminate moving objects from the still background. The basic idea is to subtract the current frame from the background model witch is acquired before the objects move in. In this step, first, current video frame is compared with the reference background model. Then, generate a confidence normalization map for each channel using thresholds and decide the pixel characteristic based on confidence map. The higher the maximum at a pixel, the more confidence, So, that the pixel belongs to the foreground. Finally, a binary mask can be obtained by using the logical OR operation between the background subtracted mask and the change detection mask.

It is intuitive that the change caused by a moving object can be large while the change caused by noise and varies only around the mean value of the corresponding pixel in the background frames. However, a generic background point will have a small variance, while a point in an moving object will have a higher variance value. Hence, the background variance can be used as a threshold to decide whether the pixel belongs to the background or occluding region. The moving object segregation is performed for every pixel $I_i(x,y)$, as follows: at each pixel of a given current frame, the pixel level change detection is performed by computing the Mahalanobis distance δ from the gaussian background model (vide eq. 2).

$$\delta_i(x,y) = |I_i(x,y) - B(x,y)| \tag{4}$$

Where $\delta = < \delta^H, \delta^S, \delta^V >$ and $I = < I^H, I^S, I^V >$.

The above operation is applied for each HSV color channel, resulting in three difference images. Next, perform a confident thresholding step for every channel using a threshold $\beta\sigma_N$, derived from the background model (vide eq. 3).

A pixel considered as a foreground if the following condition is met:

$$O_i^B(x,y) = \begin{cases} 1 & \text{If } (\delta_i^H(x,y) > \beta\sigma_N^H(x,y)) \vee \\ & (\delta_i^S(x,y) > \beta\sigma_N^S(x,y)) \vee \\ & (\delta_i^V(x,y) > \beta\sigma_N^V(x,y)) \\ 0 & \textbf{Otherwise}. \end{cases} \tag{5}$$

Here β is the confident level, witch controls the number of segmented regions. The β selection is purely based on the environmental condition, like outdoor or indoor. If the β value is less than one, then more the false foreground regions get generated, on contrary, more the β value leads to under segmentation. In the present work, the choice of confident level is done based on experimental experience.

The change detection mask can be obtained by thresholding on the normalized statics of the difference between I_i and I_{i-k}, that is, $D_i(x,y) = 1$ if

$$|I_i(x,y) - I_{i-k}(x,y) - \mu_d| > \sigma_d T_d,$$

and zero otherwise. Where μ_d and σ_d are the mean and the standard deviation of $I_i - I_{i-k}$. The resultant binary mask witch representing occlusion regions can be generated as $O_i = D_i \cup O_i^B$.

C. Shadow Elimination

This section describes an improved shadow elimination method based on the conventional HSV-based shadow detection algorithm [8]. The shadows can easily detected as video objects by background subtraction (vide equation 4). By analyzing the color properties of the shadows, have similar chromaticity with the background but lower brightness than those of the same pixels in the background model. Therefore, HSV color space is very suitable for shadow separation. Based on the HSV color space, a foreground pixel $(x, y)\ \epsilon\ O_i$ is considered as shaded background if the following three conditions hold:

$$O_i^S(x,y) = \begin{cases} 1 & \text{if}\,(\gamma \leq \frac{I_i^V(x,y)}{B^H(x,y)} \leq \phi) \\ & \wedge((I_i^S(x,y) - B^S(x,y)) < \tau\sigma_N^S(x,y)) \\ & \wedge(|I_i^H(x,y) - B^H(x,y)| < \tau\sigma_N^H(x,y)) \\ 0 & \textbf{Otherwise}. \end{cases} \quad (6)$$

Where γ and ϕ are the thresholds for limiting the shadows and τ be the confident threshold level witch is used to avoid foreground points as shadow points. The resultant foreground object mask, $O_i^F = O_i^{\wedge} O_i^S$.

D. Post-Processing

The confident thresholding and consecutive shadow detection elimination step usually produces noise holes on the object as well as over-segmentation in the background. So, a series of simple median filters can be used to elevate the salt and pepper noise, and fills the holes in regions of high confident regions (desired object) and remove isolated regions of low confident regions. However, the size of the filter that selected is proportional to the size of the noise exits in an image. The large sized noise regions can be eliminated by sequential median filtering with variable size. On contrary, we proposed a hybrid procedure for elimination of small noise blobs. It requires only three raster scans for entire computation.

The proposed post-processing step use a 3X3 median filter and ordered queue -based connected component labeling analysis for detecting the noise regions. It consists of three sequential processing steps. For each binary object image, Fist, median filter has been applied for smoothen the salt and pepper noise, and then, the connected component analysis step is performed for detecting the location and area of the regions. Finally, limit the regions by using a simple area-based thresholding method. The detailed description of the queue-based connected-component analysis procedure is given next.

1) Queue-based Connected-Component Analysis: In this procedure, a single scan has been employed, and a FIFO queue Q has been utilized as shown in figure 2. At this stage, the plateaus of pixels with the gray value 0 or 255 are detected and labeled by a raster scan of the input binary image.

The label image is initialized with INIT value. For each pixel p which is not yet visited, its 4-connected or 8-connected

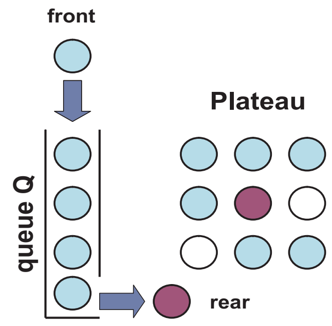

Fig. 2. Queue-based plateau detection

neighborhood ($q \epsilon N_G(p)$), where $N_G(p)$ is 8 or 4-connected neighbors of p) is inspected. Initialize the queue Q with p which has the gray value 0 or 255 and not yet visited. Pixel p is dequeued from the queue Q one at a time and its label is assigned to the non-labeled neighboring pixel q ($q \epsilon N_G(p)$) with same gray value. Then, labeled pixel q becomes a candidate, and it is inserted in the $FIFO$ queue. When the queue of candidates is emptied, each pixel in the same plateau get current label. Thus, each plateau is scanned in a breadth-first order, and the visited pixels are labelled with the current label.

2) Noise-blob Elimination: After the connected component analysis step, the area of each connected component regions is calculated by using histogram technique. Then, a simple area-based thresholding method is used to limit the noise regions. The final moving object can be determined by,

$$MO_i(x,y) = \begin{cases} 1 & \text{if}\, Histogram(Label(O_i^F)) > T_C \\ 0 & \textbf{Otherwise}. \end{cases} \quad (7)$$

E. Background Update

The background subtraction approaches are usually very sensitive to variations of the illumination; to elevate this problem the background model must be updated. In the proposed robust update, the background model is dynamically updated with incoming images. The update scheme is different for pixel positions witch are detected as belongs to foreground, as ghosts and part of the background:

$$B_{i+1}(x,y) = \begin{cases} B_i(x,y) & \textbf{\textit{if}}\ (x,y)\,\epsilon\,(\textit{Object} \vee \textit{Shadow}) \\ I_i(x,y) & \textbf{\textit{else if}}\ (x,y)\,\epsilon\,\textit{Ghost} \\ \alpha B_i(x,y) + (1-\alpha) I_i(x,y) & \textbf{\textit{Otherwise.}} \end{cases} \quad (8)$$

Where B_{i+1} is the updated background and α be the learning rate.

III. Experimental Results

The proposed background segmentation technique has been coded with OpenCV Library and implemented in Mobile Pentium 1.7GHz. Figure 3 illustrates the simulation results from the proposed background subtraction scheme. The simulation results show that the proposed technique is more robust and accurate than the conventional methods. We set the confident thresholds based on the experimental experience

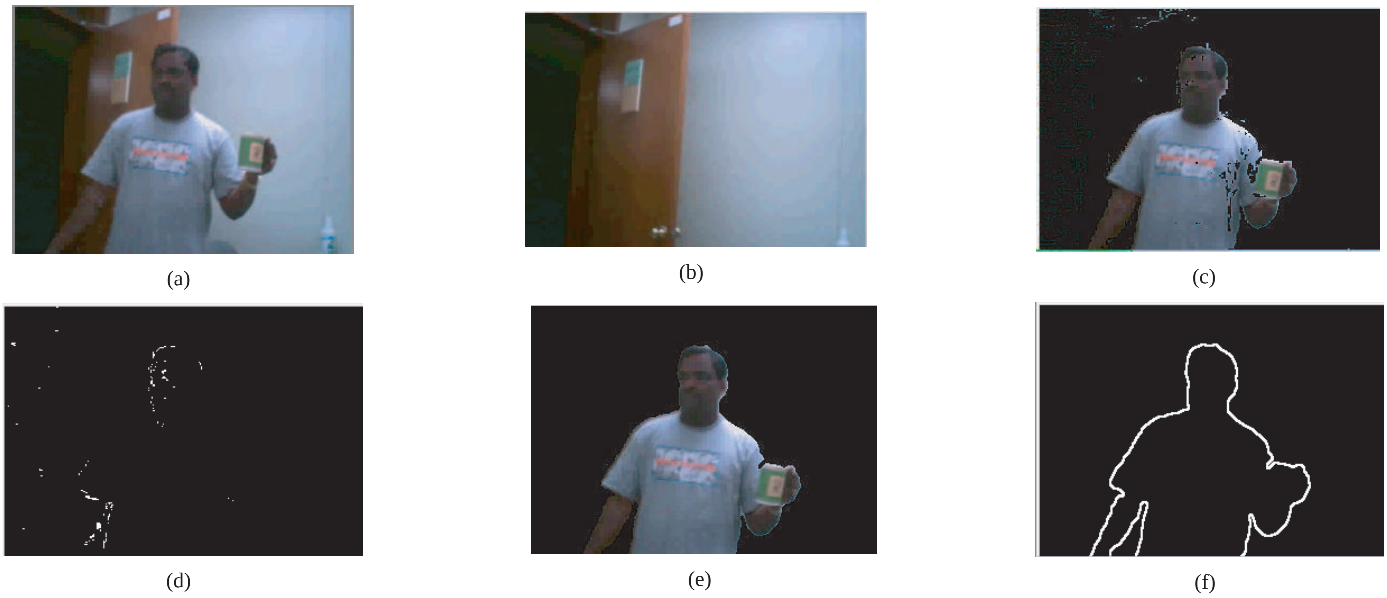

Fig. 3. (a) The current frame; (b) Background model; (c) Segmented object(α=0.5 and β=3.5); (d). Shadow elimination; (e). Noise Blob elimination; (f). Object Silhouette

and environment conditions. The segmented moving object is evaluated with the ground truth based on the pixel based measuring scheme [9]. It is identified that the proposed algorithm outperformed with the segmentation quality is around 91.8% with the ground truth, while compared to the conventional schemes [4]–[6]. Moreover, the proposed scheme consumed 0.05 seconds per frame and utilized less memory space while compared to the conventional techniques.

IV. Conclusion

A robust and accurate background subtraction method using HSV color model is presented to adapt noisy and various illumination conditions. A statistical background model is firstly setup by obtaining the means and variances of each pixel's color components from the first N frames without video objects. Then, the confident level thresholding is used to detect the occluded regions and moving shadows are eliminated very effectively by using the background statistical parameters. A queue-based connected component analysis method is introduced to isolate the moving object from background noise. Moreover, a pixel based background update is used to updating the illumination changes. The proposed scheme has been implemented and evaluated regarding the segmentation quality and frame rate. From the experimental results, we conclude that the proposed method successfully extract object contours accurately against the illumination and noise changes. The future work of this paper includes object silhouette refinement and tracking.

References

[1] D. Wang, "Unsupervised video segmentation based on watersheds and temporal tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 539–546, September 1998.

[2] S.-Y. M. Shao-Yi Chien and L.-G. Chen, "Efficient moving object segmentation algorithm by using background registration technique," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 7, pp. 577–587, July 2000.

[3] S. M. L. Di Stefano and M. Mola, "A change-detection algorithm based on structure and colour," in *IEEE Conference on Advanced Video and Signal Based Surveillance*, July 2003, pp. 252 – 259.

[4] J. B. Ming Zhao and C. Chen, "Robust background subtraction in HSV color space," in *Proceedings of SPIE: Multimedia Systems and Applications*, vol. 4861, December 2002, pp. 325–332.

[5] D. Hong and W. Woo, "A background subtraction for a vision-based user interface," in *Proceedings of ICICS-PCM*, Singapore, January 2003, pp. 1B3.3.1–5.

[6] G. A. P. Spagnolo, M. Leo and A. Distante, "A supervised approach in background modelling for visual surveillance," in *The fourth International conference on Audio- and Video-Based Biometric Person Authenticati*, ser. Lecture Notes in Computer Science, vol. LNCS 2688, AVBPA 2003 Guildford, UK, June 2003, pp. 592–599.

[7] J. M. S. Hu and B. F. Buxton. (2005) A real-time tracking system developed for an interactive stage performance. v5-25.pdf. [Online]. Available: http://www.enformatika.org/data/v5/

[8] M. T. A.Prati, I. Mikic and R. Cucchiara, "Detecting moving shadows: algorithms and evaluation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 7, pp. 918 – 923, July 2003.

[9] D. H. T. H. Chalidabhongse, K. Kim and L. Davis, "A perturbation method for evaluating background subtraction algorithms," in *Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS 2003)*, Nice, France, October 2003.