Accurate extraction of Moving Object Silhouette for Personalized Virtual Studio

C. Rambabu, Kiyoung Kim, Woontack Woo

GIST, U-VR Lab, South Korea, crambabu@gist.ac.kr

Abstract The paper proposes a hybrid method based on background subtraction criteria that preserves the boundary of moving object and also robust against noise and illumination changes. However, the inherent problem of background techniques is distinguishing the changes from disturbing effects such as noise and illumination changes. In the proposed method, the object regions are identified by fusing the results from the background difference and motion-based change detection criterion. The shadows and highlights are well detected by utilizing the normalized luminance and background difference in Hue and Saturation component. Moreover, the paper also introduces a novel connected component analysis procedure for detecting the object blob from the noise blobs. The proposed method has been implemented and evaluated regarding the segmentation quality and frame rate, it is known that the proposed method successfully extract the moving object silhouette accurately and robust against the disturbing effects.

1 Introduction

Fast and accurate extraction of moving object silhouette from real-time video is an important step in vision-based interactive systems. Accurate extraction of moving object will greatly improve the performance of object tracking, recognition, classification and activity analysis but become challenging research problem because of different disturbing effects like illumination changes, small motions of background, moving shadows, highlights and background occlusions. Several algorithms have been proposed for its solution under the restricted environment conditions but they are motivated by specific applications.

C. Wren, *et. al* [1] have been developed a system so called Pfinder that segments the human body from a static background and tracks it in real-time. The **Pfinder** uses a simple method, where the background pixels are modeled as Uni-modal Gaussian distribution and the foreground pixels are modeled with the mean and covariance, are updated recursively. In [2], the background is estimated by using the temporal median of *N* static background frames. A pixel is marked as foreground if the resultant background difference is greater than the pre-determined threshold. Fancois [3] has been proposed a system that uses HSV color space instead of RGB color space. In this system, the background is modeled as Gaussian distribution. In foreground detection, the current frame is subtracted from the mean model and then the resultant difference pixels are compared with the standard deviation model. However, Horprasert [4] has been introduced a new color model that separates the luminance component from the chrominance and several thresholds are determined to classify the pixel to be foreground, background, shadow or highlighted background. Moreover, in the W^4 system [5], the background is modeled by maximum and minimum intensity values, and the largest inter-frame absolute difference is observed in the background scene. The difference images are calculated with the current frame and the both maximum and minimum frames, are used to classify the foreground based on the inter-frame absolute difference image. Finally, D. Hong [6] has been developed a vision-based interface system based on the background subtraction technique that models the background with the well-known RGB and normalized *rgb* color models. The mean and variance of N static background frames are calculated over the each color component. Each color space has its own classification part in which the current frame is converted in each color space. Within the each color space, the pixels are classified in to four different categories, namely background, foreground with shadows, Background with shadow and foreground. These methods work well for the static background scene, but they may fail if the background pixels are multi-modally distributed.

Several background techniques exit in the literature; they have been proposed under the restricted environment conditions. However, they lack robustness in the presence of shadows, ghosts, highlights, reflections, illumination changes, dark environment and if the foreground objects are similar to the background. In order to get the accurate moving object silhouette, a hybrid technique can be developed by combining multiple complementary features.

In the present work, we aim at formulating a hybrid method based on background subtraction and motion information, that preserves the boundary of moving object and robust against the disturbing effects. In the proposed method, a linear recursive method is introduced that models the background from certain number of static background frames where the background pixels are modeled as uni-modal distribution. In order to obtain the accurate object boundary and track small motion in the background regions, we combine the results from motion-based change detection and background subtraction. The moving objects and the background regions with small motion are well identified. In general, change detection can not detect the inherent unwanted shadows and highlights in the foreground regions. The proposed method introduces a novel detection method that uses the normalized luminance (V), and background difference in Hue and Saturation components, to distinguish the moving shadows and highlights very effectively. Moreover, the proposed method derives a novel order-queue based method to find the object blob from the noise blobs. This procedure uses only 3 raster sans for detecting connected regions in a binary image. Nevertheless, the proposed method also introduces a robust update method that updates the dynamic background changes in the incoming images. Finally, we evaluate the results with the conventional methods regarding the segmentation quality and frame rate. The next sections give a detailed discussion about the proposed background subtraction method and simulation results.

2 Proposed Method

In this section, we present a hybrid background subtraction method that preserves the accurate boundary of moving object in the consideration of HSV color model. In this work, we exploit the advantages of well-known HSV color space, each color channel processed independently in order to improve the accuracy of segmentation results. Figure 1 shows the overall structure of proposed algorithm composition. The proposed method composition consists of five sequential

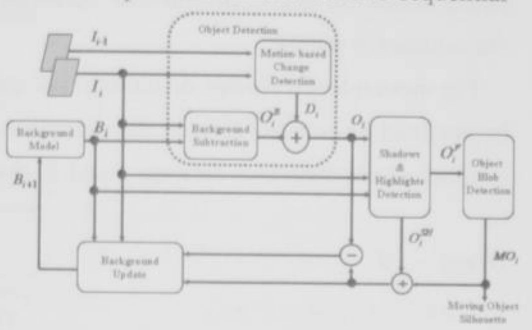

Figure 1 Proposed Background Subtraction Method

stages. First, we determine the background model from certain number of static background frames, before the objects move in. In the modified background modeling step, the background pixels are modeled as Uni-modal Gaussian distribution N (μ, σ^2), characterized by their mean and variance, are updated recursively during the background training period. Then, we perform the core object detection step, where the moving objects are well distinguished from the static background scene. In order to get the accurate boundary and track small motion in the background regions, we fuse the results from motion-based change detection and background subtraction stages. In the background subtraction step, first, the incoming frame is compared with the background model by using the predetermined threshold image, is obtained from the background noise (standard deviation of the background pixels), then, the binary confident map is generated for each channel, higher the maximum at pixel, the more confident so that the pixel belonging to the foreground. Similarly, we perform motion-based change detection where the difference between the current and previous frame, is compared with the mean of the difference image by using the normalized standard deviation of the frame deference. Finally we perform the logical OR between the results from both the stages. Next we perform the shadow and highlight detection step where the shadow and highlight pixels are detected from the foreground mask by using the normalized luminance (V), and background subtraction of Hue and Saturation component. In the subsequent step, we introduced a novel object blob detection procedure for detecting the blobs in an object binary mask O_t is acquired from the previous stage. Finally, we update the background model with incoming images. The modified update is different for pixel positions which are detected as belongs to foreground, as ghosts and part of the background. The entire procedure will be repeated for every incoming frame. The detailed description of the individual functional blocks is given next.

2.1 Background Model

The main goal of the background modeling is to train the background from a certain number of static background frames (with no video objects), which has less noise and random illumination changes. In the proposed approach, a background pixel is modeled by Uni-modal Gaussian distribution, characterized by its mean

μ and its standard deviation σ, are linearly updated. The mean μ_i for background frame at time 'i' can be defined as

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

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

$$\begin{aligned} \sigma^2_i &= (I_i - \mu_0)^2 & i &= 1 \\ \sigma^2_i &= (1-\alpha)\sigma^2_{i-1} + \alpha(I_i - \mu_i)^2 & i &\geq 2 \end{aligned} \tag{2}$$

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

$$B(x,y) = \left[\mu_i(x,y) \middle| i = N\right] \tag{3}$$

Similarly, the N frame standard deviation can be obtained by

$$\sigma_N(x,y) = \left[\sqrt{\sigma_i^2(x,y)} \middle| i = N\right] \tag{4}$$

Where $B = \langle B^H, B^S, B^V \rangle$ and $\sigma_N = \langle \sigma^H_N, \sigma^S_N, \sigma^V_N \rangle$. This method utilizes less number of frames to model the background.

2.2 Object Detection

In this section, we present a object detection procedure where the moving objects are distinguished from the static background scene. In this step, first, current video frame is compared with the reference background model. Background subtraction 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 pre- determined background model,

$$\delta(x,y) = \left|I_i(x,y) - B(x,y)\right| \tag{5}$$

Where $\delta = \langle \delta^H, \delta^S, \delta^V \rangle$ and $I = \langle I^H, I^S, I^V \rangle$.

The above operation is performed 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. A pixel deemed to be considered as a foreground if the following condition is met:

$$O_i^B(x,y) = \begin{cases} 1 & \text{if} \begin{pmatrix} (\delta_i^H(x,y) > \beta\delta_N^H(x,y)) & \vee \\ (\delta_i^S(x,y) > \beta\delta_N^S(x,y)) & \vee \\ (\delta_i^V(x,y) > \beta\delta_N^V(x,y)) & \end{pmatrix} \\ 0 & \text{Otherwise} \end{cases} \tag{6}$$

Here β is the confident level, which controls the number of segmented regions. The β selection is purely based on the environmental condition, like outdoor or indoor.

The motion-based change detection mask can be obtained by thresholding on the normalized statistics of the difference between I_i and I_{i-k} that is,

$D_i(x,y) = 1$ if

$$\left|I_i(x,y) - I_{i-1}(x,y) - \mu_d\right| > \sigma_d T_d, \tag{7}$$

and zero, otherwise. Where μ_d and σ_d are mean and standard deviation of difference frame $I_i - I_{i-1}$. Finally we perform the logical OR between the results from both the stages. In order to get the accurate boundary and track small motion in the background regions, we fuse the results from motion-based change detection and background subtraction stages. The binary mask which representing the moving object regions can be generated as $O_i = D_i \wedge O_i^B$. This step well preserves the boundary of the moving object and also detects the ghosts, small in the background regions.

2.3 Shadows and Highlight Detection

In this section, we present a shadow and highlight detection criteria that identify the moving shadows and highlights in the scene by utilizing the normalized luminance value between the current and background frame, and the background difference in Hue and Saturation components. By analyzing the color properties of the shadows and highlights, they have similar chromaticity with the background but lower brightness for shadows and higher brightness for highlights, than those of the pixels in the background model. A foreground pixel $(x, y) \in O^{B}{}_{i}$ can be considered as shadowed, highlighted or object if the following conditions hold:

$$\begin{cases} O^{S}(x,y)=1 \ (\text{Shadow}) & \text{if} \begin{pmatrix} \left[\gamma \le I_i^{V}(x,y) \big/ B^{V}(x,y) \le \phi\right] \wedge \\ \left(\left|I_i^{S}(x,y)-B^{S}(x,y)\right| < \beta\sigma_k^{S}(x,y)\right) \wedge \\ \left(\left|I_i^{H}(x,y)-B^{H}(x,y)\right| < \beta\sigma_k^{H}(x,y)\right) \end{pmatrix} \\ O^{H}(x,y)=1 \ (\text{Highlight}) & \text{if} \begin{pmatrix} \left(\frac{1}{\phi} \le I_i^{V}(x,y) \big/ B^{V}(x,y) \le \frac{1}{\gamma}\right) \wedge \\ \left(\left|I_i^{S}(x,y)-B^{S}(x,y)\right| < \beta\sigma_k^{S}(x,y)\right) \wedge \\ \left(\left|I_i^{H}(x,y)-B^{H}(x,y)\right| < \beta\sigma_k^{H}(x,y)\right) \end{pmatrix} \\ O(x,y) \ (\text{Object}) & \text{Otherwise} \end{cases} \tag{8}$$

Where γ and ϕ are the thresholds for limiting the shadow & highlights, and β be the confident level, used to avoid a foreground point classified as shadow or highlight point. The aforementioned parameters are set based on the environment condition. The proposed method utilizes the background parameters to distinguish the shadows and highlights very efficiently.

2.4 Post-Processing

In this section, we introduce an order-queue based connected component analysis procedure for detecting the moving object blob in a given binary image as shown in Figure below.

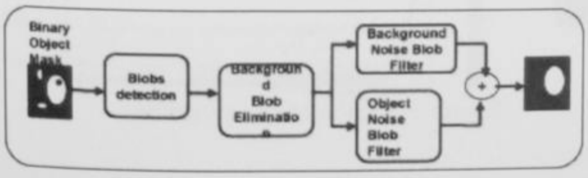

Figure 2 Procedure for Object blob detection

The proposed post-processing step consists of three sequential stages. First we detect the regions by performing the blob detection process, and then separate the background from blobs. Finally, we extract the moving object blob by using the area-based thresholding for blob filtering.

Blobs Detection: In the Blobs detection procedure, a single scan has been employed, and a FIFO queue Q has bee utilized. 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. For each pixel p which is not yet visited, its 4-connected or 8-connected neighborhood $q \in N_G(p)$, 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 De-queued from the queue Q one at a time and its label is assigned to the non-labeled neighboring pixel q ($q \in N_G(p)$ with same gray value. Then, labeled pixel q becomes a candidate, and it is inserted in the 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 labeled with the current label. This procedure requires only three raster scans for entire computation.

Blob Filtering: In this procedure, we utilize the area-based thresholds to separate the object blob from the noise blobs. Background noise blob filtering first detects the noise blobs in the background and then merges the detected noise blobs with the background. Similarly, the object noise blob filter eliminates the noise regions in the foreground. The final object MO_i can be obtained by performing the logical operation between the filtered outputs.

2.5 Background Update

This section describes the modified method that update the background model with the incoming images. The modified update scheme is different for pixel positions, are detected as belongs to foreground, as ghosts and part of the background:

$$B_{i+1}(x,y)=\begin{cases} B_i(x,y) & \text{if } (x,y)\in\left(MO_i\cup O_i^{SH}\right) \\ I_i(x,y) & \text{elseif } (x,y)\in\left(O_i-\left(MO_i\cup O_i^{SH}\right)\right) \\ \alpha B_i(x,y)+(1-\alpha)I_i(x,y) & \text{Otherwise} \end{cases}$$

Where B_{i+1} be the updated background and α be the learning rate. This method tends to track the dynamic changes in the background and robust against the noise and illuminations changes.

3 Experimental Results

The proposed hybrid background technique has been coded with OpenCV Library and implemented in Pentium 1.7GHz, and tested in real-time with the Sony CCD camera, 320X240, 30fps. Figures 3 and 4 illustrate the simulation results from the proposed method [7].

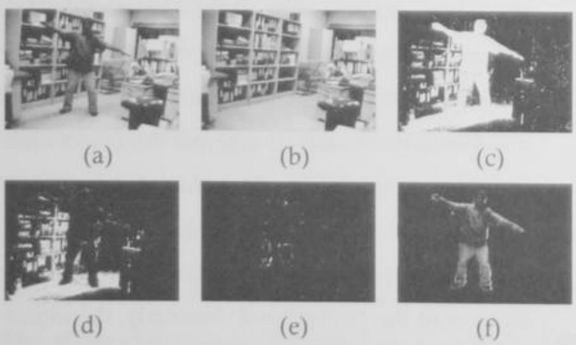

Figure 3 (a) the current frame, (b) mean of the background learnt from 300 frames, (c) segmented object, (d) shadows and illumination change, (e) highlights detection, and (f) final object silhouette

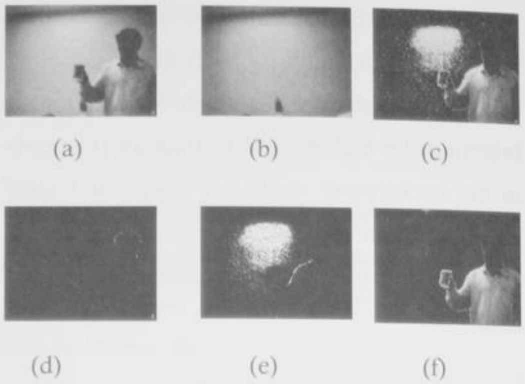

Figure 4 (a) the current frame, (b) mean of the background learnt from 300 frames, (c) segmented object, (d) shadows and illumination change, (e) highlights detection, and (f) final object silhouette

The segmented moving object is evaluated with the ground truth based on the pixel based measuring scheme. The performance of proposed scheme depends on a set of parameters, are mainly the confident threshold β and background learning rate α. We set the confident threshold β based on the Receiver Operating Curve (ROC) as shown in figure 5. The performance of the proposed scheme for different thresholds is shown in Figure 6. From the Table, it is identified that the proposed method outperformed with the segmentation quality is around 94 % while compared to the Normalized RGB-based background subtraction method [6].

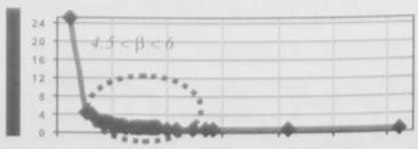

Figure 5: Region of Convergence

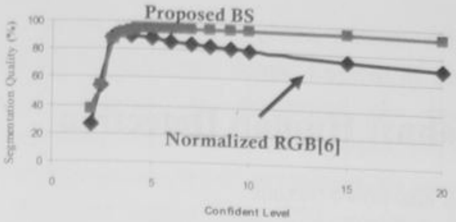

Figure 6: Performance of Proposed and NRGB [6]

Moreover, the proposed method consumed 0.05 seconds per frame and utilized less memory space, compared to the conventional technique [6].

Table 1Performance of object detection Algorithms

Performance			
Algo.	False Positive %	False Negative %	Seg. Quality (%)
NRGB[6]	2 – 3.5	1.5 – 5%	75 – 85%
Proposed	0.5 –1.5	0.3 – 1.0	90 –94%
Time per Frame in Sec.			
NRGB[6]	0.3sec.		
Proposed	0.05sec.		

The proposed method has been tested in the Personalized Virtual studio (VR @ Home) for personal broadcasting at home.

4 Conclusions

A hybrid background subtraction method using HSV color model is presented to adapt noisy and various illumination conditions. The moving regions are well identified by fusing the results from background difference and motion-based change detection criterion. The paper introduced a methodology for detecting the shadows and highlights and also proposed a novel connected component analysis method to isolate the moving object blob from background noise. From the experimental results, we conclude that the proposed method successfully extract object silhouette more accurately against the illumination and noise changes.

References

[1] C. Wren, A. Azabayejani, T.D. and Pentland, A. '' Pfinder: Real-time tracking of the human body", IEEE Transactions on Pattern Anal. & Machine Intell. 19, 780 –785(1997)

[2] Cutler, R. and Davis, L.,'' View-based detection and analysis of periodic motion," In proceedings of International Conference on Pattern Recognition, pp. 495-500. Brisbane, Australia (1998).

[3] Francois, A.R. and Medioni, G.G.'' Adaptive color background modeling for real time segmentation of video streams", In proceedings of International Conference on Imaging Science, Systems, and Technology, pp. 227 –232. Las Vegas, NA (1999).

[4] T. Horprasert, D.H. and Davis, L.,'' A statistical approach for real time robust background subtraction and shadow detection", In IEEE Frame Rate Workshop (1999).

[5] Haritaoglu,'' W4: real-time surveillance of people and their activities", IEEE Transaction on Pattern Anal. & Machine Intell. 25, 809-830 (2000).

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

[7] T. H. Chalidabhongse K. Kim, D.H., Davis, L.,''A perturbation method for evaluating background subtraction algorithms," Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS 2003), Nice, France (2003)