AN EFFECTIVE FOREGROUND/BACKGROUND SEGMENTATION APPROACH FOR
BOOTSTRAPPING VIDEO SEQUENCES

Han-Hui Hsiao and Jin-Jang Leou
Department of Computer Science and Information Engineering
National Chung Cheng University
Chiayi 621, Taiwan
{hhh95p, jjleou}@cs.ccu.edu.tw

ABSTRACT
In this study, an effective foreground/background segmentation approach for bootstrapping video sequences is proposed. First, a modified block representation approach is used to classify each block of the current video frame into one of the four categories, namely, “background,” “still object,” “illumination change,” and “moving object.” Then, a new background updating scheme is developed, in which the side-match measure is used to determine whether the background exposes. Finally, an improved noise removal and shadow suppression procedure using the edge information is used to enhance the final segmented foreground. Based on the experimental results obtained in this study, as compared with two comparison approaches, the proposed approach has better background modeling and foreground extraction results.

Index Terms—foreground/background segmentation, side-match measure, block representation, noise removal, shadow suppression

1. INTRODUCTION
Foreground/background segmentation, a basic process of a computer vision system, is to separate some interesting objects (the foreground) from the rest (the background) of each video frame in a video sequence [1]. Background subtraction is a popular foreground/background segmentation approach, which detects the foreground by thresholding the difference between the current video frame and the modeled background in a pixel-by-pixel manner [2]. The correctness of the modeled background is usually affected by the following factors: (1) illumination changes, (2) dynamic background: some “moving” objects, such as waving trees, fountains, and flickering monitors, are not interested for a vision-based surveillance system, (3) shadows: foreground objects often cast shadows, which are different from the modeled background [3].

For the popular Gaussian background model, Stauffer and Grimson [4] presented a pixel-wise representation using mixture of Gaussians (MoG) and pixel-wise background updating to update the intensity mean and variance of each pixel in real-time. Generally, a training period without foreground objects is required (non-bootstrapping) and some ghost (false positive) objects may be detected when some foreground objects change their motion status (static or moving) suddenly.

Recently, background subtraction methods focused on background initialization for bootstrapping video sequences, in which a training period without foreground objects is not available in some cluttered environments [5-7]. Two simple techniques for background initialization are the pixel-wise temporal mean and median filters over a large number of frames [5]. For the pixel-wise temporal median filter, it is assumed that for each pixel within the estimation period, the exposure of the background must be more than that of the foreground. For the block-wise strategy, Farin et al. [6] used a block similarity matrix to segment the input video frames into foreground and background regions, which contain the block-wise temporal differences between any video frame pair. Reddy et al. [7] proposed a block selection approach by the discrete cosine transform (DCT) among the neighboring blocks to estimate the unconstructed parts of the background. This approach is usually degraded by error propagation if some blocks in a video frame are erroneously estimated. Most background initialization approaches are computationally expensive and need large memories. Furthermore, one free background is usually obtained as its output during a “learning” duration. In this study, an effective foreground/background segmentation approach for bootstrapping video sequences is proposed.

2. PROPOSED FOREGROUND/BACKGROUND SEGMENTATION APPROACH
Fig. 1 shows the framework of the proposed video foreground/background segmentation approach for bootstrapping video sequences, which contains block-wise
background modeling and pixel-wise foreground extraction. In Fig. 1, the input includes the current (gray-level) frame \( I' \) and the previous (gray-level) frame \( I'^{-1} \) of a video sequence, and the output includes the modeled background frame \( B' \) and the segmented foreground frame \( F' \), where \( t \) denotes the frame number (index). Here, \( I'_{(x,y)} \), \( I'^{-1}_{(x,y)} \), \( B'_{(x,y)} \), and \( F'_{(x,y)} \) denote pixels \((x,y)\) of \( I' \), \( I'^{-1} \), \( B' \), and \( F' \), respectively, and the frame size is \( W \times H \) pixels. Each frame is divided into blocks of size \( N \times N \) pixels. Let the block index be \((i,j)\), where \( i = 0,1,2,...,(W/N)-1 \) and \( j = 0,1,2,...,(H/N)-1 \). Here, \( b^i_{(j)} = \{I'_{(i\times N+a, j\times N+b)}: a,b = 0,1,2,...,N-1\} \), \( b'^{-1}_{(j)} = \{I'^{-1}_{(i\times N+a, j\times N+b)}: a,b = 0,1,2,...,N-1\} \), and \( \bar{b}^i_{(j)} = \{b^i_{(j)}: a,b = 0,1,2,...,N-1\} \) denote blocks \((i,j)\) in \( I' \), \( I'^{-1} \), and \( B' \), respectively.

Fig. 1. The framework of the proposed video foreground/background segmentation approach.

2.1. Block Representation

As the illustrated example shown in Fig. 2, in the proposed block representation approach, each block of the current frame is classified into one of the four categories, namely, “background,” “still object,” “illumination change,” and “moving object.” In the proposed block representation approach (as shown in Fig. 3), motion vector estimation and correlation coefficient computation are used to perform block representation (classification).

Motion vector estimation is applied between \( I' \) and \( I'^{-1} \) using a block matching algorithm so that a block in \( I' \) is determined as “static” or “moving.” In this study, the sum of absolute differences (SAD) is used as the cost function for block matching between block \( b^i_{(j)} \) in \( I' \) and the corresponding block in \( I'^{-1} \) and the search range for motion estimation is set to \( \pm N/2 \). For a block in \( I' \), if the minimum SAD for motion vector \((u,v)\), \( D_{mv}(u,v) \), is smaller than 90% of the SAD for the null-vector \((0,0)\), \( D_{mv}(0,0) \), the block is determined as a “moving” block; otherwise, it is a “static” block [6].

On the other hand, the correlation coefficient \( C_B(i,j) \) between block \( b^i_{(j)} \) in \( I' \) and the modeled background block \( \bar{b}^i_{(j)} \) in \( B'^{-1} \) is computed as

\[
C_B(i,j) = \frac{\sum (b^i_{(j)} - \mu_{b^i_{(j)}})(\bar{b}^i_{(j)} - \mu_{\bar{b}^i_{(j)}})}{\sqrt{\sum (b^i_{(j)} - \mu_{b^i_{(j)}})^2 \times \sum (\bar{b}^i_{(j)} - \mu_{\bar{b}^i_{(j)}})^2}},
\]

where \( \mu_b \) is the mean of the pixels in block \( b \). As shown in Fig. 3, based on \( C_B(i,j) \) and the threshold \( T_{cb} \), a “static” block can be further classified into either a “background” block (if \( C_B(i,j) \geq T_{cb} \)) or a “still object” block (otherwise), and a “moving” block can be further classified into either an “illumination change” block (if \( C_B(i,j) \geq T_{ic} \)) or a “moving object” block (otherwise). That is, four different block representations can be obtained.

2.2. Background Updating

Based on block representation, each modeled background block \( \bar{b}^i_{(j)} \) in \( B' \) can be updated as follows.

(a) Background. The modeled background block \( \bar{b}^i_{(j)} \) in \( B' \) is updated by

\[
\bar{b}^i_{(j)} = \alpha \cdot \bar{b}^{i-1}_{(j)} + (1 - \alpha) \cdot b^i_{(j)},
\]

where \( \alpha \), the updating weight, is empirically set to 0.9 in this study.

(b) Still object. The modeled background block \( \bar{b}^i_{(j)} \) in \( B' \) is updated by

\[
\begin{align*}
\bar{b}^i_{(j)} = \bar{b}^{i-1}_{(j)}, \\
\bar{b}^i_{(j)} = b^i_{(j)},
\end{align*}
\]

if \( Count_{(j)} \geq T_{sl} \), otherwise.

Fig. 3. The flowchart of the proposed block representation approach.
where \( \text{Count}_{(i,j)} \) is the number of times that \( b_{(i,j)}' \) in \( I' \) is successively determined as a “still object” block previously, and \( T_{\text{still}} \) is a threshold to determine when a “still object” block will learn to be a “background” block.

(c) Illumination change. The modeled background block \( \mathbf{b}_{(i,j)}' \) in \( B' \) is similarly updated by Eq. (2).

(d) Moving object. The modeled background block \( \mathbf{b}_{(i,j)}' \) in \( B' \) is updated by

\[
\mathbf{b}_{(i,j)}' = \begin{cases} 
\mathbf{b}_{(i,j)}', & \text{if } SM(b_{(i,j)}') < SM(\mathbf{b}_{(i,j)},) \\
\mathbf{b}_{(i,j)}, & \text{otherwise},
\end{cases}
\]

(4)

where \( SM(b_{(i,j)}') \) and \( SM(\mathbf{b}_{(i,j)},) \) denote the side-match measures for \( b_{(i,j)}' \) embedded in \( B' \) and that for \( \mathbf{b}_{(i,j)} \) in \( B' \) [8]. \( SM(b_{(i,j)}') \) is defined as the sum of squared differences between the boundary of the embedded \( b_{(i,j)}' \) and the boundaries of the four neighboring blocks in \( B' \), i.e.,

\[
SM(b_{(i,j)}') = \sum_{x=0}^{N-1} (B_{(i,j),N-1-x}' - I_{(i,j),N-1-x})^2 + \sum_{x=0}^{N-1} (B_{(i,j),N+x-1}' - I_{(i,j),N+x-1})^2 + \sum_{y=0}^{N-1} (B_{(i,j),N-1-y}' - I_{(i,j),N-1-y})^2 + \sum_{y=0}^{N-1} (B_{(i,j),N+y-1}' - I_{(i,j),N+y-1})^2
\]

(5)

SM(\mathbf{b}_{(i,j)}) is defined as the sum of squared differences between the boundary of \( \mathbf{b}_{(i,j)} \) in \( B' \) and the boundaries of its four neighboring blocks in \( B' \), can be similarly defined.

2.3. Initial Segmented Foreground Detection

Based on the modeled background frame \( B' \), as an illustrated example shown in Fig. 4, the initial (binary) segmented foreground frame \( \hat{F}' \) can be obtained as

\[
\hat{F}'(x,y) = \begin{cases} 
1, & \text{if } [(I'(x,y) - B'(x,y)] \geq T_{\text{sf}}, \\
0, & \text{otherwise},
\end{cases}
\]

(6)

where \( T_{\text{sf}} \) is a threshold.

![Fig. 4. An illustrated example of initial segmented foreground detection.](image)

2.4. Noise Removal and Shadow Suppression

As shown in Fig. 4, \( \hat{F}' \) usually contain some fragmented parts (noises) and shadows. To obtain the precise foreground frame \( F' \), a noise removal and shadow suppression procedure is proposed, which combines the shadow suppression approach in [9] and the edge information extracted from \( I' \) with \( \hat{F}' \) being the (binary) operation mask.

Let \( \hat{F}'_s \) be the saturation component of the current frame represented in the HSV color space and \( \hat{F}'_b \) be the gradient image of \( I' \) using the Sobel operator with \( \hat{F}' \) being the (binary) operation mask. The segmented foreground frame \( F' \) is defined as

\[
F'(x,y) = \begin{cases} 
1, & \text{if } (\hat{F}'_s(x,y) \cap (\hat{F}'_b(x,y) \geq \sigma_{E})) \\
0, & \text{otherwise},
\end{cases}
\]

(7)

where \( \cap \) and \( \cup \) denote the logical AND and OR operators, respectively, \( \sigma_{E} \) is a standard deviation of \( \hat{F}'_s \) and \( T_E \) is a threshold. Here, \( T_{\text{sf}} \) and \( T_E \) are empirically set to 0.05 and 0.45, respectively. Fig. 5 shows an illustrated example.

![Fig. 5. An illustrated example for noise removal and shadow suppression: (a) \( \hat{F}'_s \geq \sigma_{E} \), (b) \( \hat{F}'_s \cap (\hat{F}'_b \geq \sigma_{E}) \), (c) \( \hat{F}'_b \geq T_E \), (d) \( F' \).](image)

3. EXPERIMENTAL RESULTS

Ten test bootstrapping video sequences, selected from the ATON, CAVIAR, and PETS2006 benchmark datasets, are used in this study, whereas only the processing results for the bootstrapping video sequence, “Highway,” are shown in the paper. The video frames are 320×240 in size and the processing block is 16×16 in size, i.e., \( N = 16 \). To evaluate the performance of the proposed approach, two comparison approaches, namely, Reddy background estimation (Reddy) [7] and self-organizing background subtraction (SOBS) [10], are implemented in this study. In Reddy, only the gray-level component is employed, whereas in SOBS, the H, S, and V components are employed. Here, \( T_{\text{sf}} \) in Eq. (3) is set to 10 for the 15fps outdoor video sequence “Highway.”

Fig. 6 illustrates frames 1, 20, 40, 60, 80, and 100 of the sequence “Highway” (a), and the corresponding modeled background frames, i.e., \( B' \), \( B'^{20} \), \( B'^{40} \), \( B'^{60} \), \( B'^{80} \), and \( B'^{100} \), by Reddy (b), SOBS (c), and the proposed
approach (d). Figs. 7-9 illustrate frames 20, 40, and 80 of the sequence “Highway” (a), and the corresponding segmented foreground frames, i.e., $F^{20}$, $F^{40}$, and $F^{80}$, by Reddy (b), SOBS (c), and the proposed approach (d), respectively. In Figs. 7-9, the contents within red rectangles indicate the ghost (false positive) objects.

As shown in Fig. 6(b), the Reddy approach is usually degraded by error propagation if some blocks in a video frame are erroneously estimated. As shown in Fig. 6(c), for SOBS, the modeled background frame $B'$ may preserve some foreground objects in $I$, resulting in some ghost objects in $F'$. For the bootstrapping video sequence “Highway,” because the proposed approach can obtain the modeled background frame “completely” after $F^{37}$, the proposed approach can obtain the good segmented foreground frames without ghost objects after $F^{37}$.

![Fig. 6. Some background modeling results of the sequence “Highway” (a) by Reddy (b), SOBS (c), and the proposed approach (d).](image)

![Fig. 7. The foreground extraction results of frame 20 of the sequence “Highway” (a) by Reddy (b), SOBS (c), and the proposed approach (d).](image)

![Fig. 8. The foreground extraction results of frame 40 of the sequence “Highway” (a) by Reddy (b), SOBS (c), and the proposed approach (d).](image)

![Fig. 9. The foreground extraction results of frame 80 of the sequence “Highway” (a) by Reddy (b), SOBS (c), and the proposed approach (d).](image)

4. CONCLUDING REMARKS

In this study, an effective foreground/background segmentation approach for bootstrapping video sequences is proposed, in which a modified block representation approach, a new background updating scheme, and an improved noise removal and shadow suppression procedure are employed. Based on the experimental results obtained in this study, as compared with Reddy [7] and SOBS [10], the proposed approach has better background modeling and foreground extraction results.

5. REFERENCES