Efficient Video Object Segmentation Using Adaptive Background Registration and Edge-Based Change Detection Techniques

Jiashu Zhang

Sichuan Province Key Lab of Signal and Information Processing, Southwest Jiaotong University
Chengdu, 610031, P. R. China

Abstract - This paper presents an automatic and efficient video object segmentation algorithm. Moving object extraction is carried out by adaptive background model and edge-based change detection techniques. Background is updated by use of pixel history and moving object mask. The connected component analysis and morphological filtering are employed to obtain accurate VOP. Finally, a novel object tracking window scheme is applied to improve the processing speed. Experimental results for three different types of MPEG-4 video sequences are shown to demonstrate the effectiveness of the proposed algorithm.

1. INTRODUCTION

Video object segmentation and tracking is relevant to information extraction in many multimedia applications such as video surveillance, traffic monitoring, and semantic annotation of video. And semantic object extraction from video sequences is an essential element in content-based multimedia services (MPEG-4/MPEG-7 standards). These multimedia standards enable content-based functionalities by using the video object plane (VOP) as the basic coding element. Each VOP contains the shape and texture information of a semantically meaningful object in the scene. Since the shape information of moving objects may not be available from video sequences, segmentation is inevitable for object-based coding schemes.

Several promising efficient and reliable video object (VO) segmentation algorithms that aim at providing object segmentation in real time have been reported [1] – [4]. Kim and Hwang [1] developed a novel automatic object segmentation algorithm based on inter-frame edge-change detection that incorporates spatial edge information into the motion detection stage. This approach overcomes the disadvantages of conventional change detection methods. This algorithm performs well for still background video sequences, but does poorly in situations such as still object, slowly moving object, and uncovered background. And manual initialization is required in certain video sequences. This is mainly due to the lack of an adaptive background model. Thus a semi-automatic VO extraction algorithm using object tracking was also proposed in [1] to cope with moving background sequences. However, these efforts are complicated and increase the processing time.

Chien et al. [3] proposed a fast VO segmentation algorithm that uses a background registration technique to construct reliable background information from the video sequence. They also employed a morphological gradient operation to reduce the shadow effect. But the segmentation result may degrade if the object has still parts and low texture, or the object’s color is similar to the background. Another efficient VO segmentation system integrates object-level knowledge into a statistical background model to improve moving object detection and background update [4]. It performs well in terms of shadow and ghost detection and removal during the moving object stopping and starting.

In this paper, we propose a novel and simple automatic VO segmentation algorithm that efficiently detects moving objects and deals with uncovered background and shadow effect. The rest of the paper is organized as follows. Section 2 describes the proposed algorithm in detail. The impact of the edge-based change-detection technique and adaptive background registration method is discussed. Experimental results on MPEG-4 video sequences are provided in Section 3, followed by conclusions in Section 4.

2. VIDEO OBJECT SEGMENTATION

The basic idea of our algorithm is to detect moving objects by object boundary rather than pixel difference. We develop an automatic VO segmentation algorithm for real-time applications. This algorithm consists of a reliable adaptive background model and fast edge-based detection schemes, and can be categorized into three parts: object-based adaptive background update, object boundary extraction, and post-processing.

2.1 Adaptive Background Registration

The goal of the background registration is to construct a reliable background model from the video sequence. Pre-stored image is the simplest method to obtain a background model, but this method cannot adjust for any changes in background. And statistical methods based on Gaussian model suffer from slow processing speed and pre-stored video information [5]. A reliable background model can be constructed using the background registration technique proposed in [3]. This method records the history of each pixel in video frames, and the most stable pixels are
becoming a part of background. A disadvantage of this approach is that it is based only on the pixel, not on the object. Hence the background may not be accurate when dealing with still parts of a slowly moving object. As shown in Fig. 1(b) by a box of white-border, a ghost appears on the back of the person.

The proposed adaptive background registration technique is described as follows. Background model update is incorporated with the video object plane. When we obtain the current moving object mask, we apply it to the background and mark those pixels of moving object as changed portion, thus any pixel of the moving object will not be classified as background.

Two maps of pixel history are constructed. The first map, stable history map (SM), defines the background pixel. It represents the number of times a pixel is stable in consecutive frames.

\[
SM(x, y) = \begin{cases} 
SM(x, y) + 1, & \text{if } |I_n(x, y) - I_{n-1}(x, y)| < T_p \\
0, & \text{otherwise}
\end{cases}
\]

where \(I_n\) is the \(n\)th frame in the video sequence, \(T_p\) and \(T_f\) are the pre-defined thresholds. The initial value for each pixel in SM is set to zero. If a pixel is in the VOP, it is marked as unstable and set its value to 0. The second map is a difference history map (DM), which represents the number of times a pixel is significantly different from the background in consecutive frames. It is the condition for a still object becoming a part of background.

\[
DM(x, y) = \begin{cases} 
DM(x, y) + 1, & \text{if } |I_n(x, y) - I_{n-1}(x, y)| > T_p \\
0, & \text{otherwise}
\end{cases}
\]

The initial value for each pixel in DM is 0. If the pixel belongs to the VOP, its value increases by 1. Based on the information from both maps and taking the still object and uncovered background situation into account, the background (BG) is adaptively updated frame-by-frame by:

\[
BG(x, y) = \begin{cases} 
I_n(x, y), & \text{if } (SM(x, y) \geq T_p \text{ and } DM(x, y) > T_f) \\
BG(x, y), & \text{otherwise}
\end{cases}
\]

\[
OB(x, y) = \begin{cases} 
255, & \text{if } |I_n(x, y) - BG(x, y)| \geq T_p \\
0, & \text{otherwise}
\end{cases}
\]

The proposed fast object boundary extraction algorithm generates the edge information, \(ED\), of the current frame by

\[
ED(x, y) = \max \{|I_n(x + 1, y - 1) - I_n(x - 1, y + 1)|, |I_n(x + 1, y + 1) - I_n(x - 1, y - 1)|, |I_n(x, y + 1) - I_n(x, y - 1)|, |I_n(x + 1, y) - I_n(x - 1, y)|, |I_n(x, y + 1) - I_n(x, y - 1)|
\]

and the updated edge information

\[
ED(x, y) = \begin{cases} 
255, & \text{if } ED \geq T_p \\
0, & \text{otherwise}
\end{cases}
\]

Then, using the updated background information from the object-based adaptive background registration step, we can define the object boundary (OB) as

\[
OB(x, y) = \begin{cases} 
255, & \text{if } |I_n(x, y) - BG(x, y)| \geq T_p \\
0, & \text{otherwise}
\end{cases}
\]
The proposed method is able to correctly extract moving objects under situations of moving object containing still parts and head-and-shoulder type sequences without manual initialization [1]. In this work, values of the thresholds $T_p$, $T_f$, and $T_e$ are set to 32, 10, and 16, respectively.

2.3 Post Processing

After object boundary detection, we are ready to extract the VOP. First, object boundary is enhanced, and morphological filter is applied to remove small holes and gaps. Then moving region is configured using a searching window to reduce processing data in the next frame. This scheme improves the processing speed significantly.

It is known that broken lines and holes may occur on the edge map after background subtraction, as shown in Fig. 3(a) and 3(f). We use smooth filtering to enhance the edge link and reduce hole size. But this process broadens the boundary. Thus we employ morphological filters to remove pixels not belonging to the actual object boundary. Smooth filtering is slow due to the convolution operation, and morphological filtering with large structuring element size is time-consuming. Hence, in our approach, we first use an average smoothing filter, but no convolution operation, to reduce hole size and repair the gap. The object boundary is enhanced by

$$OB(x, y) = \begin{cases} 
255, & \text{if } \sum_{i=1}^{x+2} \sum_{j=1}^{y+2} OB(i, j) > T_{ob} \\
0, & \text{otherwise}
\end{cases}$$

where $T_{ob}$ be the pre-defined threshold. The morphological filtering with a small structuring element size kicks in to refine the OB. After this step, most holes and broken lines have been eliminated, and the remaining noise can be removed by connected component analysis. First we assign all pixels temporary labels and record equivalence. Then we replace each temporary label by the lowest label in its equivalence class. All pixels with the same label are classified as a region. Since object regions are usually much larger than noise regions, regions smaller than a threshold are removed from the image. Object filling is applied after connected component analysis (see Fig. 3(e)).

![Fig. 3. VOP extraction process, (a) object boundary, (b) boundary enhancement, (c) noise removal and object filling, (d) no morphological processing, (e) with morphological processing, (f) without post-processing.](image)

Finally, a searching window scheme is employed with the aim of reducing the processing time [8]. Since moving objects only move within a certain area in the image, we can find this region and process those pixels in this region in the next frame cycle, instead of the whole image. Fig. 4 shows some results of such process. Its merit is shown in Fig. 5.

3. EXPERIMENTAL RESULTS

The proposed algorithm was applied to ‘Hall Monitor’, ‘Miss America’, ‘Akiyo’, and ‘Parking Lot Monitor’ [4]. The video is reconstructed and playback by updated key backgrounds and extracted video objects. Accuracy of extracted moving objects and video background can be evaluated by PSNR and the matching error, which is defined by [9]

$$E_m = \frac{1}{NM} \sum_{k=1, l=1}^{K, L} | T_{kl} - O_{kl} |$$

where $E_m$ is the matching error and $T_{kl}$ is the binary value of the hand-drawn target mask image which can be extracted manually by using any image-editing tool. $O_{kl}$ is the binary value of the extracted object image, and $NM$ is the image size. PSNR is used to compare the similarity between reconstructed frames and original frames. Fig. 5 shows these performance measure results for ‘Hall Monitor’. In Fig. 5(a), the matching error is less than 1% in each frame. This indicates good spatial accuracy of the proposed segmentation algorithm. Playback results are satisfactory as seen from the PSNR results in Fig. 5(b). Figs. 5(c) and 5(d) show that the processing time is around 0.13 sec for a CIF frame and 0.04 sec for a QCIF frame with a 700-MHz Pentium III PC.

![Fig. 4. Search windows (‘Hall Monitor’, Frame: 22, 47, 86)](image)

![Fig. 5. Performance evaluation for ‘Hall Monitor’ video, (a) matching error, (b) PSNR of reconstructed background, and the processing speeds in (c) CIF format, (d) QCIF format.](image)
The parking lot monitoring video is used to measure the reactivity to background update [4]. It considers the situation when the background reflects changes from a car that starts to move out of the parking lot after having initially been part of the background. The car starts to back out at frame 65 (Fig. 6(a)). At frame 100, the car still covers the area closely to where it was parked. This presents the separation from its forming ghost (Fig. 6(c)). When the car moves out of the parked location completely, our algorithm is able to obtain the correct background update (Fig. 6(f)). Such results cannot be achieved by statistical methods. Statistical approaches usually have slow response to background changes (correct background update after 40 frames in this example). The proposed algorithm is able to eliminate ghost effectively and has a rapid response to background changes. In addition, moving object can be extracted correctly.

‘Miss America’ and ‘Akiyo’ video sequences are typical head and should type videos with no initial background and very small object movement. Most approaches for such videos need human interaction for segmentation [1, 2, 6]. Here we take the first frame as the initial background and consider the pixel difference between current frame and background and the edge difference in consecutive frames as the object boundary. Correct segmentation can be achieved with the VOP scheme described in Section 2 (Figs. 7 and 8).

4. CONCLUSIONS

In this paper, an automatic VOP segmentation algorithm has been presented. Moving object extraction was performed by adaptive background and edge change detection. The proposed approach achieves background accuracy, automatic shadow removal, and fast processing speed. Results show that our algorithm is suitable for indoor and outdoor monitoring, and video sequence without initial background.

ACKNOWLEDGMENT

This work is supported in part by Sichuan Youth Science and Technology Foundation (Grant No. 03ZQ026-33)

REFERENCES