ROBUST OBJECT SEGMENTATION USING ADAPTIVE THRESHOLDING

Xiaxi Huang and Nikolaos V. Boulgouris

Department of Electronic Engineering
King's College London
United Kingdom

ABSTRACT

This paper proposes a robust object segmentation algorithm that tackles problems that arise in environments in which the foreground and background colours are similar and there is light reflection in the shadow areas. The proposed algorithm bases its efficiency on a novel RGB colour detection process with adaptive threshold and edge detection, which are combined in order to obtain a foreground map containing objects and shadows. Subsequently, edge information is used to remove the shadow regions that are out of the edge bounding boxes. Finally, a post-processing procedure is applied and, for offline detection, a temporal filter is included in order to retrieve the misclassified pixels. The experimental results demonstrate the superior performance of our algorithm in comparison to other existing methods.

Keywords: adaptive thresholding, Sobel edge detection, confidence map, shadow removal

1. INTRODUCTION

The extraction of moving objects from video sequences is a key operation for content-based video coding, multimedia content description, and intelligent signal processing. Several algorithms have been proposed for object detection in video surveillance applications [1]. For colour images, a popular approach is based on the assumption that each pixel of the background is a realization of a random variable with Gaussian distribution (SGM–Single Gaussian Model) [2]. The mean and covariance of the Gaussian distribution are independently estimated for each pixel. In order to adapt the frequent colour change of the same background pixel, a Multiple Gaussian Model (MGM) is used to represent the distribution of the background pixels [3]. But there are still some serious drawbacks of these methods, e.g., mistakenly subtracting objects that have colours similar to the background colour and misclassifying shadows as foreground.

In [4], Jabri *et al.* presented a novel background subtraction method that combines both colour and edge information to improve the quality and reliability of the results. In [5], chromaticity information was used for the elimination of shadows. In [6], Cucchiara *et al.* used some shadow properties in HSV color space to distinguish shadows from moving objects. The above two methods generated good results in outdoor environment, but had poor performance for indoor shadows, which are usually accompanied by light reflection. In [7] and [8], edge information was used in order to cope with indoor shadows. However, those methods needed very clear boundaries of the moving objects, which are hard to be obtained when the difference of the foreground and background colours is very small.

In this paper, we present a robust algorithm for the detection of foreground objects that have similar colours with the background and with their light-reflected shadows. The main contribution of this paper is the use of a novel adaptive thresholding detection and an improved shadow removal method based on edge information in order to deal with complicated indoor object segmentation problems. The new approach is compared with some popular methods for object segmentation. The experimental results show that the proposed method generally outperforms the other methods in the comparison.

The paper is organized as follows. Section 2 describes our foreground detection method. Section 3 describes the shadow removal and the post-processing operations of the proposed algorithm. In Section 4, the proposed system is experimentally evaluated and compared with other existing methods. The conclusions are given in Section 5.

2. PROPOSED ALGORITHM

2.1. System Outline

The proposed system is outlined in Fig. 1. First, the background image is updated based on the background information of the previous frame. Second, an initial block-size change detection is applied to roughly obtain foreground areas. Then, a novel RGB colour detection with adaptive threshold and edge detection are combined to obtain a foreground map containing objects and shadows. A novel shadow removal process using edge information and a post-processing procedure are subsequently applied in order to obtain the final results.

2.2. Background Updating

To take into account slow illumination changes, which occur in long-term tracking, the background image should be subsequently updated. In many background subtraction approaches (e.g. [3], [4]), all pixels are updated for each frame. The drawback of this method is that a stop-moving object will soon be misclassified as background. To avoid this, we only update the pixels that are classified as background in the previous frame.

$$B^t(\phi) = \alpha I^{t-1}(\phi) + (1-\alpha) B^{t-1}(\phi) \tag{1}$$

where $I^t(\phi)$ is a three-dimensional vector representing the intensity values of the three colour channels at image position ϕ, which is classified as a background pixel after the object segmentation of the previous frame. The parameter α is the learning rate. In our experiments we used $\alpha = 0.05$.

2.3. Initial Block-size Mask

The aim of this step is to roughly determine the foreground areas, so that lower thresholds can be used in those areas at the next RGB

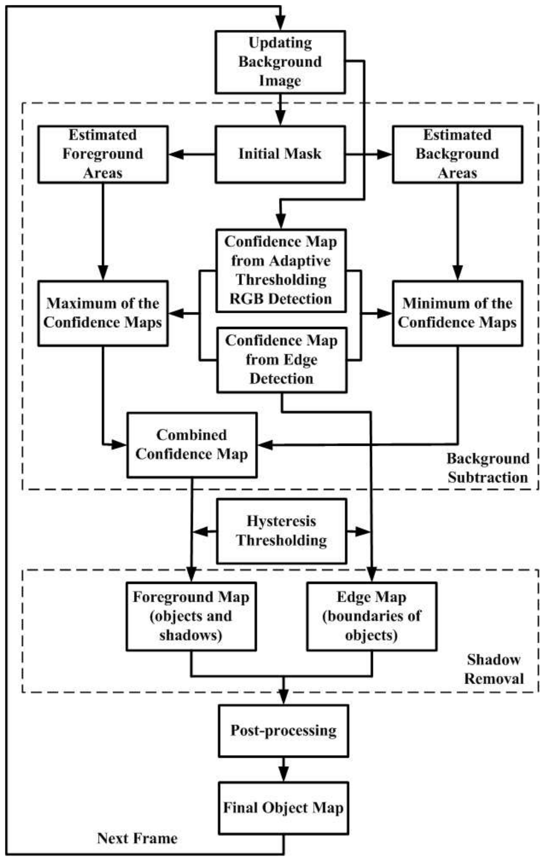

Fig. 1. Outline of the proposed algorithm

change detection step, in order to detect foreground pixels with similar colours to the background. We calculate the average difference between the current and background frame in a block

$$d_{avg} = \frac{1}{\Phi} \sum_{\phi \in \Phi} |I(\phi) - B(\phi)| \tag{2}$$

then we sum up the differences of three colour channels and threshold it with T. For the size of the frames in our experiments is $384 * 288$, we firstly use an initial block size of $\Phi_0 = 32 * 32$. The blocks with larger difference are assumed to contain foreground pixels, we then divide these 'foreground blocks' into smaller sub-blocks

$$\Phi_n = \Phi_{n-1}/4 \qquad (n = 1, 2, 3) \tag{3}$$

and apply the above detection within the sub-blocks, with a higher threshold

$$T_n = T_{n-1} + T_{added} \qquad (n = 1, 2, 3) \tag{4}$$

For preciseness, we iterate this subdivision three times, then apply a median filter for the smallest size ($\Phi_3 = 4 * 4$) blocks to obtain a foreground blocks' map. Since we initially calculate the average difference within blocks sized $\Phi_0 = 32 * 32$, the edge pixels of the objects might be in the blocks which contain mostly background pixels, and the edges will likely be missed. To avoid missing foreground information, we use another initial block size $\Phi_0' = 24 * 24$ to go through the above steps again to obtain another map, and then combine the two maps to obtain a initial foreground map (Fig. 2(b))

$$Map_{block} = Map_{\Phi} \cup Map_{\Phi'} \tag{5}$$

2.4. Colour Change Detection with Adaptive Threshold

For the accurate detection of foreground objects, a pixel-wise method is appropriate. However, when using basic colour change detection methods, regardless of the colour space, the use of a fixed threshold achieves poor adaptation. Specifically, using a high threshold, foreground pixels with small colour differences will be misclassified, while using a lower threshold will result in much unremovable noise. When using a single or a mixture of Gaussian models, the threshold for each pixel is a fixed multiple of its variance, in which case only temporal factors are considered whereas spatial factors are ignored. Therefore, when foreground colours are similar to their surrounding background colours, methods based on mixtures of Gaussians do not achieve satisfactory performance. In our algorithm, we propose a novel method for threshold adaptation. Specifically, we define the threshold as

$$T_{adapt} = k_1 \left(\sigma_{curr}^2 + k_2 \sigma_{diff}^2 \right) / \mu_{diff} \tag{6}$$

where σ_{curr}^2 is the local variance in the current frame I, σ_{diff}^2 and μ_{diff} are the local variance and local mean value in the difference frame D ($D = |I - B|$). For pixel ϕ_0:

$$\mu_{diff}(\phi_0) = \frac{1}{\Psi} \sum_{\phi \in \Psi} D(\phi) \tag{7}$$

$$\sigma_{diff}^2(\phi_0) = \frac{1}{\Psi} \sum_{\phi \in \Psi} |D(\phi) - \mu_{diff}(\phi_0)|^2 \tag{8}$$

$$\sigma_{curr}^2(\phi_0) = \frac{1}{\Psi} \sum_{\phi \in \Psi} |I(\phi) - \mu_{curr}(\phi_0)|^2 \tag{9}$$

where Ψ is the local neighbour area of ϕ_0. In order to create a confidence map [4], for each pixel in the difference frame, we set two pairs of k_1 and k_2 in (6), to get two thresholds T_1 and T_2 ($T_1 > T_2$). Then we use the method in [4] to create confidence maps in three colour channels respectively, and take the maximum value of the three channels to create a confidence map of colour change detection $CMap_{colour}$. As seen in Fig. 2(c), due to the denominator in (6), pixels within changing areas have lower thresholds, that makes sure foreground pixels with small $D(\phi)$ (e.g., white clothes of the person in front) will not be missed. However, much noise will be generated. Due to the numerator in (6), most of the noise are separated from the foregrounds, that makes them easy to be removed in the later stages.

2.5. Edge Detection

In the proposed algorithm, edge detection will be involved for the extraction of the foreground map and the removal of shadows. First, the Sobel mask is used for the calculation of the edge gradient

$$G = \sqrt{G_x^2 + G_y^2} \tag{10}$$

where G_x and G_y are the horizontal and vertical differences in the difference frame D. Then, an edge confidence map $CMap_{edge}$ (Fig. 2(d)) is created.

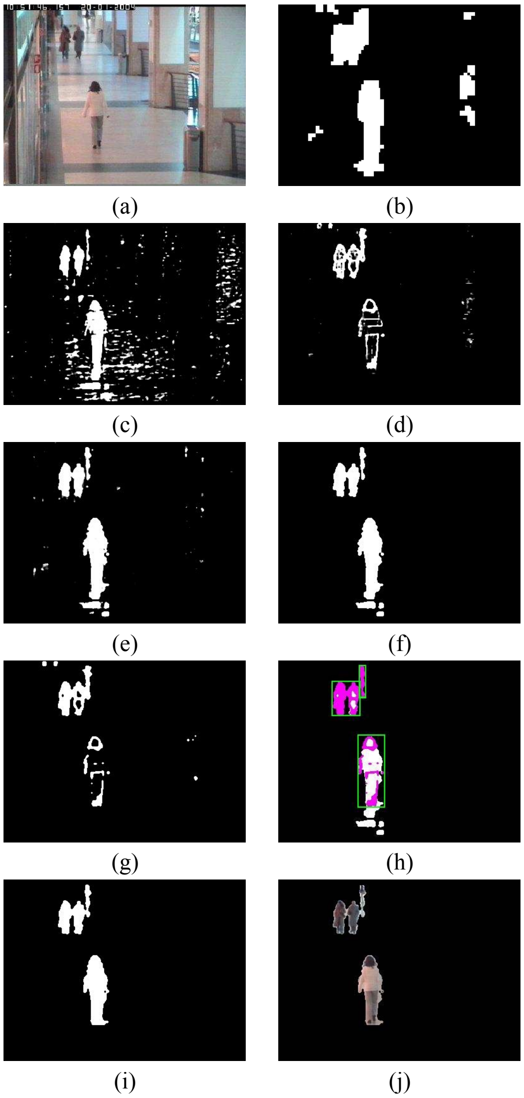

Fig. 2. (a) Original image, (b) Initial block-size map, (c) Confidence map of RGB change detection with adaptive threshold, (d) Confidence map of Sobel edge detection, (e) Combined confidence map, (f)Foreground map before shadow removal, (g) Hysteresis threholding result for (d), (h) Foreground map after shadow removal, (i) Binary map of extracted objects, (j) Colour image of extracted objects.

2.6. Combination

Based on the method presented in [4], we propose an improved method, by taking into account the initial mask, in order to combine the two confidence maps. In the estimated foreground areas, we take the maximum value of $CMap_{colour}$ and $CMap_{edge}$, while in the estimated background areas, we take the minimum value of $CMap_{colour}$ and $CMap_{edge}$, to retrieve the missing edge information.

$$CMap_1 = max\left(CMap_{colour}, CMap_{edge}\right) \cap Map_{block} \quad (11)$$

$$CMap_2 = min\left(CMap_{colour}, CMap_{edge}\right) \cap \overline{Map_{block}} \quad (12)$$

$$CMap = CMap_1 \cup CMap_2 \quad (13)$$

After obtaining $CMap$ (Fig. 2(e)), a hysteresis thresholding step [9] is applied to remove false positives by eliminating all components that are not connected to a 100% confidence region (Fig. 2(f)).

3. SHADOW REMOVAL AND POST-PROCESSING

In many indoor environments, due to the soft, sometimes coloured, illumination, and the light-reflective floor, the colour and the luminance of the shadows are very similar to those of the objects. In such conditions, it is very hard to distinguish shadows from objects by using the colour information. Furthermore, the edges between the shadow areas and their surrounding background appear to be smooth. Actually, in most practical cases, there are no edges in shadow areas. In order to exploit this fact, we combine the foreground map, which is obtained at the former step, and the edge map to cut out the shadow areas.

Initially, a hysteresis thresholding step is applied to the edge confidence map $CMap_{edge}$, to obtain a map with connected edge information (Fig. 2(g)). In the sequel, a bounding box (marked in Fig. 2(h)) is automatically set to contain all of the edge pixels (marked in Fig. 2(h) and Fig. 2(h)) that are in the same connected foreground region (marked as white in Fig. 2(f) and Fig. 2(h)). Finally, the pixels that are out of the bounding boxes are considered to be shadows and removed. After removing shadows, morphological operations and a global median filter are applied to remove the noise and tune the boundaries in the final map (Fig. 2(i)).

Furthermore, for offline detection, some temporal filters could be applied:

$$Map_{TF}^t = Map^t \cap \left(Map^{t-1} \cup Map^{t+1}\right) \quad (14)$$

$$Map_{TF}^t = Map^t \cup \left(Map^{t-1} \cap Map^{t+1}\right) \quad (15)$$

By application of (14), spurious pixels will be eliminated, and some missed foreground pixels can be retrieved by application of (15). Also, both above purposes can be achieved using:

$$Map_{TF}^t = median\left(Map^{t-1}, Map^t, Map^{t+1}\right) \quad (16)$$

4. EXPERIMENTAL RESULTS

For the experimental evaluation of the proposed algorithm, several sequences from the CAVIAR database were used. The results are compared with those generated by popular existing methods, e.g., the mixture of Gaussians model (MGM) [3] and the HSV colour space model [6].

As seen, in Fig. 3 and Fig. 4, compared to the MGM and the combination of MGM and HSV methods, our object segmentation algorithm achieves superior performance when the foreground and background colours are very similar. In Figs 3(e),(f), and Figs 4(e),(f), many pixels of the clothes are misclassified as background because the colour differences between the current frame and the background frame ($D(\phi)$) are very small. At the same time, some shadow pixels are misclassified as foreground because the distance $D(\phi)$ of those pixels is larger than the fixed threshold. In Figs 3(g),(h), and Figs 4(g),(h), the HSV method mistakenly removes some object pixels of the shadowed clothes, which have similar characters to shadow pixels (i.e. low luminance), and extracts some shadow pixels on the light-reflecting floor, which have similar characters to object pixels (i.e. high luminance). In contrast, as shown in Figs 3(c),(d), and Figs 4(c),(d), our approach generates more robust results, since all of the objects, including the white clothes pixels are correctly detected as foreground, and the shadows around the people are successfully classified as background.

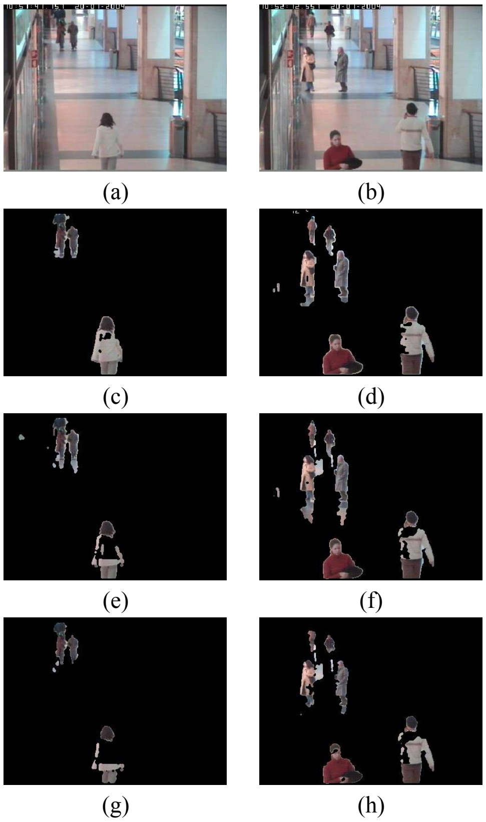

Fig. 3. (a),(b): Original images, (c),(d): Foreground maps created by the proposed algorithm, (e),(f): Foreground maps created by mixture Gaussian model, (g),(h): Foreground maps created by MGM plus HSV method.

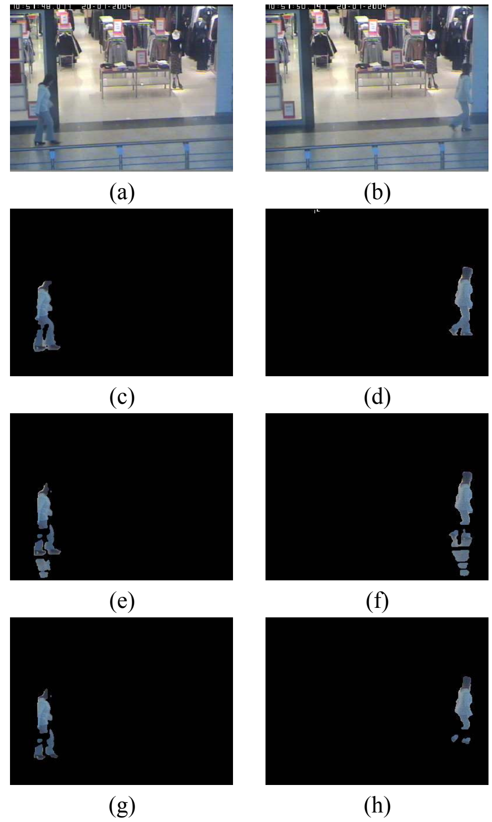

Fig. 4. (a),(b): Original images, (c),(d): Foreground maps created by the proposed algorithm, (e),(f): Foreground maps created by mixture Gaussian model, (g),(h): Foreground maps created by MGM plus HSV method.

5. CONCLUSIONS

In this paper, we proposed a robust object segmentation method for video sequences. First, an initial mask is obtained to roughly estimate the foreground areas. Then, adaptive colour change detection and edge detection are combined to create a confidence map, which contains foreground objects and their shadows. Finally, shadow removal is applied and the final foreground map is generated.

Compared with the popular MGM object segmentation method and the HSV shadow removal method, the proposed method achieves more robust performance, especially in situations of similar foreground and background colour or in presence of shadows with reflected light. Considering that the proposed algorithm does not involve future frames, it can be used in real-time processing applications. Furthermore, if it is used offline, a temporal filter can be applied to further improve the performance of the algorithm.

6. REFERENCES

[1] J. Nascimento and J. Marques, "Performance evaluation of object detection algorithms for video surveillance," *Multimedia, IEEE Transactions*, vol. 8, no. 4, pp. 761–774, Aug. 2006.

[2] T. Darrell C. R. Wren, A. Azarbayejani and A. P. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 780–785, Jul. 1997.

[3] W. Eric C. Stauffer and L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 8, pp. 747–757, Aug. 2000.

[4] H. Wechsler S. Jabri, Z. Duric and A. Rosenfeld, "Detection and location of people in video images using adaptive fusion of color and edge information," *Pattern Recognition, 2000. Proceedings. 15th International Conference*, vol. 4, pp. 627–630, Sep. 2000.

[5] Z. Duric S. J. McKenna, S. Jabri and A. Rosenfeld, "Tracking groups of people," *Computer Vision and Image Understanding*, , no. 80, pp. 42–46, 2000.

[6] M. Piccardi R. Cucchiara, C. Granan and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Trans. PAMI.*, vol. 25, pp. 1337–1342, 2003.

[7] S. Y. Ma S. Y. Chien and L. G. Chen, "Efficient moving object segmentation algorithm using background registration technique," *IEEE Trans. CSVT.*, vol. 12, pp. 577–585, 2002.

[8] Z. Liu D. Xu, J. Liu and X. Tang, "Indoor shadow detection for video segmentation," *IEEE International Conf. Multimedia and Expo (ICME)*, vol. 1, pp. 41–44, Jun. 2004.

[9] J. F. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, Nov. 1986.