J. Software Engineering & Applications, 2009, 2: 96-102
doi:10.4236/jsea.2009.22014 Published Online July 2009 (www.SciRP.org/journal/jsea)

Adaptive Motion Segmentation for Changing Background

Yepeng Guan[1,2]

[1]School of Communication and Information Engineering, Shanghai University, Shanghai, China; [2]Key Laboratory of Advanced Displays and System Application, Ministry of Education, 149 Yanchang Rd., Shanghai, China.
Email: ypguan@shu.edu.cn

Received December 25th, 2008; revised April 12th, 2009; accepted May 4th, 2009.

ABSTRACT

Segmentation of moving objects efficiently from video sequence is very important for many applications. Background subtraction is a common method typically used to segment moving objects in image sequences taken from a statistic camera. Some existing algorithms cannot adapt to changing circumstances and require manual calibration in terms of specification of parameters or some hypotheses for changing background. An adaptive motion segmentation method is developed according to motion variation and chromatic characteristics, which prevents undesired corruption of the background model and does not consider the adaptation coefficient. RGB color space is selected instead of introducing complex color models to segment moving objects and suppress shadows. A color ratio for 4-connected neighbors of a pixel and multi-scale wavelet transformation are combined to suppress shadows. The mentioned approach is scene-independent and high correct segmentation. It has been shown that the approach is robust and efficient to detect moving objects by experiments.

Keywords: *Motion Segmentation*, *Background Update*, *Background Subtraction*, *Motion Variation*, *Shadow Suppression*

1. Introduction

Moving objects segmentation is an important topic in computer vision applications, including video conferences, vehicle tracking, and three-dimensional object identification, and has been actively investigated in recent years [1]. The most widely adopted approach for moving object segmentation with a fixed camera is based on background subtraction. A background (called as background model also) is computed and evolved frame by frame. A reliable background model has to account for background at each time instant. Mistake in labeling foreground and background points could cause wrong update of the background model. A particularly critical situation occurs whenever moving objects stop for a long time and become a part of the background. When these objects start again, a ghost is detected in the area where they stopped. This will persist for all the following frames, preventing the area to be updated in the background forever.

In addition, moving object segmentation is easily affected by shadow problem. Researchers try to select an optimal color space for shadow eliminating among a set of color spaces, such as *HSV*, *YCrCb*, *XYZ*, $L^*a^*b^*$, $L^*u^*v^*$, $C_1C_2C_3$, $l_1l_2l_3$, normalized *rgb* and so on, however, it remains open-ended how important is the appropriate color space, and which color space is the most effective [2]. Many approaches in literature have been developed so far. Some existing methods require manual calibration in terms of the specification of parameters which are related to the environment and the lighting conditions or make some hypotheses.

An approach to adaptive background updating and shadow suppressing is developed. RGB color space is selected instead of introducing complex color models to segment moving objects. Motion evaluation is introduced to prevent giving erroneous segmentation in those corresponding with an un-updated background model. A color ratio and multi-scale wavelet transformation are combined to suppress shadows. The main contribution of the proposal is that the developed approach is scene-independent and automatic background updating according to motion variations caused by moving objects. The second contribution is that when segmenting motion objects it does not require any complex supervised training or manual calibration in terms of the specification of parameters or makes any hypotheses. Experimental results from indoor and outdoor environments have shown

*Tel: +86 21 56331967; Fax: +86 21 56336908.

the developed approach is efficient and flexible during segmenting moving objects and suppressing shadows in applications.

The remainder paper is organized as follows. Section 2 briefly reviews some related previous works. In the next section, background model update would be discussed. Shadow suppression is described in Section 4. Experimental results from indoor and outdoor environments are given in Section 5 and followed by conclusions in Section 6.

2. Related Works

Many works have been put forward in literature for moving objects detection. Background subtraction based moving objects in the scene are detected by the difference between the current frame and the background model. When the deviation is greater than some critical value, the pixel is considered as foreground (moving object) [3]. The most simple background model is the previous frame. The difference between the observed frame and the previous frame is thresholded to determine which pixel is background and which pixel is the foreground. Another way to model the background is to take the mean, median, or minimum and maximum values of the previous N pixel [4]. It needs to keep track of the pixel value history. To avoid this problem a first order recursive filter is used to update the background model [5]. It easily causes 'tailing' or 'ghosting'. A particularly critical situation occurs whenever the moving object stopped for a long time and became a part of the background. When these objects start again, a ghost is detected in the area where they stopped [6]. This will persist for all the following frames, preventing the area to be updated in the background image forever [7]. One method is by modeling each pixel as unimodal Gaussian distribution [8]. It fails to model background pixels that are subject to repetitive motions which have multiple background colors. To overcome these difficulties, a parametric background modeling is done by modeling each background pixel value as a mixture Gaussian distribution [9,10]. The parametric background model still lacks flexibility when dealing with non-static backgrounds, a highly flexible non-parametric technique is proposed for estimating background probabilities from recent samples over time using Kernel density estimation [11]. False detections due to fluctuating backgrounds are still not covered in the algorithm until now. Codebook technique is a different approach proposed for the background subtraction in [12]. One of drawbacks is that the algorithm cannot adapt to changing circumstance when the environment was not present in the training phase. Moving objects that stop moving and should be adopted into the background will get difficulties in the algorithm.

Shadows cause serious problems while segmenting moving objects, due to the misclassification of shadow-points as a foreground. Many works have been developed to suppress shadow [1,2,6,13,14,15,16,17,19,21,22]. By shadow suppression, the major problem is how to distinguish moving cast shadows from moving object points [15]. Cucchiara *et al.* [6] defined a shadow mask for each point resulting from motion segmentation. However, it often makes additional assumptions such as small changes in hue and saturation, necessary a prior knowledge of the bands of the changes in the value channel. Moreover choice of some parameters is less straightforward and for now is done empirically. Tattersall *et al.* [16] proposed adaptive shadow identification through automatic parameter estimation based on the above method. The single variable parameter is only used. However, much additional assumptions must be made also in [16]. Salvador *et al.* [17] proposed invariant color features to detect cast shadows through using chrominance color components. It is found that several assumptions are needed regarding the reflecting surfaces and the lightings. In outdoor scene, shadows will have a blue color cast due to the sky, while the lighting regions have a yellow cast (sunlight), hence the chrominance color values corresponding to the same surface point may be significantly different in shadow and sunlit regions [18]. Cavallaro *et al.* [19] proposed the normalized *rgb* space to detect shadows. It is known that the practical application of normalized *rgb* suffers from a problem inherent to the noise at low intensities which would result in unstable chromatic components [20]. Texture analysis can be potentially effective in solving the problem. Heikkila *et al.* [21] proposed a texture-based method for modeling the background and detecting moving objects from a video sequence. Each pixel is modeled as a group of adaptive local binary pattern histograms that are calculated over a circular region around the pixel. Because of the huge amount of different combinations, it must be done more or less empirically to find a good set of parameter values. Spagnolo *et al.* [22] proposed a ratio-based algorithm to detect shadows with an empirically assigned ratio threshold. This ratio-based algorithm considering the ratio between only two adjacent pixels considerably shortens the computation time, but it easily misclassifies shadows as objects, because ratio magnitude of shadows may have a similar magnitude value.

3. Background Model Update

The main idea of the proposed approach is to update pixels according to motion variations caused by moving objects, which has been proven to be more reliable and less sensitive to noise.

Assuming the background model $B^{t+1}(x, y)$ at t+1 time, extract possible moving regions P based on background subtraction (seen from Figure 1(c)). According to the fact that the variation of sensible motion target in the scene can be found out from the sequence, the difference between the two adjacent frames is used to extract moving regions R (seen from Figure 1(d)). Extract moving object regions P_1 from the regions P according to moving objects with the same connective characteristic as regions R (seen from Figure 1(e)).

When the object moves slowly in situ, the extracted region P_1 may be incorrect (seen from Figure 2(e)). In order to overcome the problem, chromatic information is used to further extract moving region P_2 from P_1 according to moving objects with the same chromatic in P_1 as in R (seen from Figure 2(f)). According to the mentioned above, the background model is updated as follows,

$$B^{t+1}(x,y) = \begin{cases} I^{i+1}(x,y) \rightarrow B^{t+1}(x,y), & \text{if} \quad \arg XOR(P,P_2) \geq T \\ B^{t}(x,y), & \text{otherwise} \end{cases} \tag{1}$$

where

$$P_1 = connect(R,P), P_2 = chrom(P_1,R) \tag{2}$$

In (1), I^{t+1} is a current frame at t+1 time, "→" represents updating, Connect (R, P) represents selecting a region with the same connective characteristic as R from P, chrom (P_1, R) represents selecting a region with the same chromatic information in R as in P_1, XOR is a logical exclusive-or operator, T is a threshold value.

The proposed background update approach reveals some advantages. Firstly it smoothes and reduces the effects of noise in the image since sudden variations of a single pixel are not included in the background model. Secondly it does not consider the adaptation coefficient or the learning rate used in the existing literatures. Finally it does not depend on static or moving objects in the image. In order to demonstrate the last point, some results of two sequences where lena walks, and she bends in two scenes from http://www.tele.ucl.ac.be/~gaitanis/results/Human_Action_Video_Database/2Feet/ are given in Figures 1 and 2, respectively.

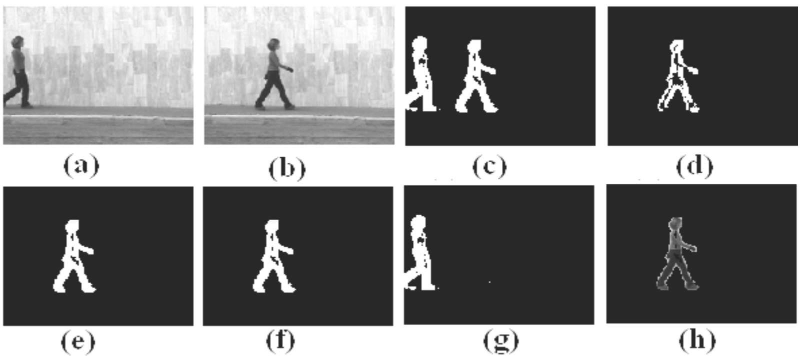

Figure 1. Lena walks in the scene and the extracted results. (a) First frame in the sequence. (b) The 25th frame. (c) Segmentation based on subtraction of (a) and (b). (d) Segmentation based on difference between the 24th and 25th frames. (e) Extracted object region by connectivity. (f) Extracted object region by chromatic consistency. (g) Detected varied background region. (h) Extracted final foreground region

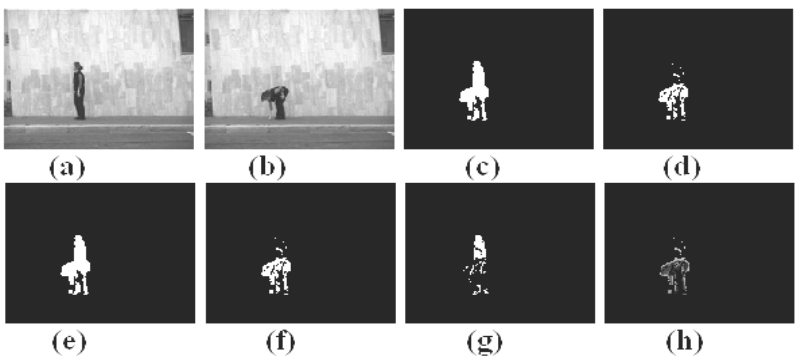

Figure 2. Lena bends in situ and the extracted results. (a) First frame in the sequence. (b) The 25th frame. (c) Segmentation based on subtraction of (a) and (b). (d) Segmentation based on difference between the 24th and 25th frames. (e) Extracted object region by connectivity. (f) Extracted object region by chromatic consistency. (g) Detected varied background region. (h) Extracted final foreground region

In the second example, some existing algorithms in the literature are difficult to correctly extract moving objects. In Figures 1 and 2, the ghosts are removed since their connective pixels do not contain any frame-difference pixels. From Figures 1(g) and 2(g), one can see that some varied background pixels are detected in the region no matter how lena walks or bends in situ. The results obtained with the approach show correct foreground segmentation and effective background updating from Figures 1 (h) and 2(h). It highlights that the aperture problem resulting from the adjacent frame-difference is overcome by combining region connectivity and the chromatic consistency.

4. Shadow Suppression

Since RGB color camera system is one of the most popular color spaces, and all colors are seen as a variable combination of the three primaries in the RGB color space, RGB color space is selected to eliminate shadows in the paper. Before segmenting image, start by applying a smoothing operator both to the background image $B(x, y)$ and the current frame $F(x, y)$.

Calculate a color ratio difference between $B(x, y)$ and $F(x, y)$ for 4-connected neighbors of a pixel and for all three color components as:

$$p^n_{k\in R,G,B}(x,y,i,j)=\frac{I^F_k(x,y)}{I^F_k(x+i,y+j)}-\frac{I^B_k(x,y)}{I^B_k(x+i,y+j)}(i=\pm1,\ j=\pm1, i\neq j, n=1,2,3,4) \quad (3)$$

where $I^F_k(x,y)$ is the intensity of $F(x, y)$ for the kth color component at location (x, y), $I^F_k(x+i,y+j)$ is the intensity of its neighbor color components; $I^B_k(x,y)$ is the intensity of $B(x, y)$ for the kth color component at location (x, y), and $I^B_k(x+i,y+j)$ is the intensity of its neighbor color components.

Since moving shadow makes the region covered by itself darker than the background and it has similar chromaticity, a suitable threshold Th is applied to the color ratio difference (3) to obtain a candidate foreground region $SP^n(x,y)$ as:

$$SP^n(x,y)=\begin{cases}1 & if \quad p^n_k\geq Th\\ 0 & otherwise\end{cases} \quad (4)$$

If the distortion distribution of p^n_k is assumed to be a Gaussian one, we can threshold the distortion by $\sigma_{p^n_k}$, where $\sigma_{p^n_k}$ is a standard deviation of p^n_k. However, it is found from experiments that the distribution of p^n_k is not a Gaussian one, but is has a very sharp peak at zero. The standard deviation of p^n_k is used to select the threshold Th, and (4) can be rewritten as:

$$SP^n(x,y)=\begin{cases}1 & if \quad p^n_k\geq \sigma_{p^n_k}\\ 0 & otherwise\end{cases} \quad (5)$$

The value of color ratio mentioned above may not uniquely highlight the property of a particular material, and there may be instable for particular values. To obtain a robust segmentation result, the results from the color ratio and multi-scale wavelet transformation method proposed in [23] are combined.

5. Experimental Results

To confirm the effectiveness of the above proposed method, we have conducted experiments with different indoor and outdoor video sequences.

Figure 3. Detected foregrounds for indoor circumstances. The first column is the first frame of the sequences, the second column is the current frames (67#, and 300#, from top to bottom, respectively), and the third column is the detected foreground masks with light grey pixels superimposed on the original image

Figure 4. Detected foregrounds in outdoor circumstances. The first column is the first frame of the sequences, the second column is the current frames (69#, and 84#, from top to bottom, respectively), and the third column is the detected foreground masks with light grey pixels superimposed on the original image

In the experiments, the threshold T used in (1) is chosen as 100, which represents the minimum number of expected background updating in consecutive frames. The threshold T can be chosen lower in which more background pixels can be updated. The selection of lower T means more time to process background updating and some noises may drift into updating. The background reference image used is the first frame of sequence no matter objects enter the field of view before captured or not.

The sequences with different indoor conditions including the MPEG-4 test sequence *Hall Monitor*, aton project test sequence *intelligent room* from http://cvrr.ucsd.edu/aton/shadow have been used to test the developed algorithm. Some results are given in Figure 3.

In Figure 3, the results obtained by the developed approach are quite good, considering that the contrast between the object and background is very low.

In order to test the proposed approach further in outdoor conditions, the sequences with different outdoor conditions including *campus* and *highway* from the *ATON* project are used. The choice of such different scenes is made to emphasize the reliability and robustness of the proposed approach in outdoor circumstances. Some results are shown in Figure 4.

In Figure 4, the results show the robustness of the proposed algorithm to cope with different outdoor circumstances.

The above results show qualitative information about the effectiveness of the developed approach. It is necessary to quantitatively evaluate the performance of the method with a ground-truth.

The main goal of the proposed method is not accurate detection or discrimination of shadow pixels, but the improvement of moving object detection because accurate object detection is crucial for further applications. The performance of moving object detection is measured in term of correct segmentation rate (*csr*) and false segmentation rate (*fsr*) as:

$$csr = 2\frac{TP \cap GT}{TP + GT},\ fsr = \frac{FP + FN}{2GT} \qquad (6)$$

where *TP* (true positive) is the number of correctly detected foreground pixels, *GT* (ground-truth) is the ground-truth for the foreground, *FP* (false positive) is the number of background falsely marked as foreground pixels, *FN* (false negative) is the number of foreground pixels falsely classified as background ones.

Using the *csr* and *fsr* measurements, the total correct segmentation will have 100% by *csr* while the total agreement with the ground-truth will have 0% by *fsr*.

The available sequence *intelligent room* and its ground-truth data are available from http://cvrr.ucsd.edu/aton/shadow. In our work, two recent proposed methods including the invariant color features (ICF) proposed in [17] and the ratio map (RM) developed in [22] are quantitatively evaluated together, and the results are shown in Figure 5 for a comparison.

The symbols in the legend of Figure 5 refer to object extraction results: ICF [17], RM [22], and the proposed method. The mean values of *csr* corresponding to the plots of Figure 5 are the following: RM 0.84, ICF 0.86, proposed 0.98. The mean values of *fsr* are the following: RM 3.42, ICF 4.11, proposed 1.61.The proposed method outperforms the investigated methods with the best object detection over time.

6. Conclusions

This paper has presented an approach to adaptive motion segmentation and shadow suppression. The ghosts are detected and removed by the developed background update function, which prevents undesired corruption of the background model and does not consider the adaptation coefficient or the learning rate used in the literature. By

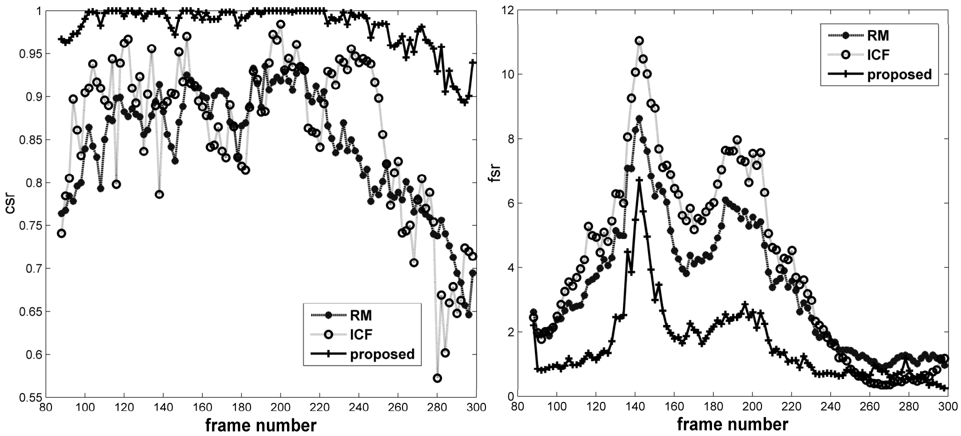

Figure 5. Comparison of video object segmentation rate for the test sequence intelligent room

comparison, it has been shown that the proposed method outperforms the investigated methods and is robust and efficient to detect moving objects during coping with different indoor or outdoor circumstances.

7. Acknowledgements

This work in part is supported by the National Natural Science Foundation of China (Grant No. 60872117).

REFERENCES

[1] W. Zhang, X. Z. Fang, X. K. Yang, and Q. M. J. Wu, “Moving cast shadows detection using ratio edge,” IEEE Transactions on Multimedia, Vol. 9, No. 6, pp. 1202–1214, September 2007.

[2] C. Benedek and T. Sziranyi, “Study on color space selection for detecting cast shadows in video surveillance,” International Journal of Imaging Systems and Technology, Vol. 17, No. 3, pp. 190–201, October 2007.

[3] T. Horprasert, D. Harwood, and L. S. Davis, “A statistical approach for real-time robust background subtraction and shadow detection,” Proceedings of IEEE International Conference on Computer Vision Frame-Rate Workshop, pp. 1–19, September 1999.

[4] I. Haritaoglu, D. Harwood, and L. S. Davis, “W4: real time surveillance of people and their activities,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, pp. 809–830, August 2000.

[5] J. Heikkila and O. Silven, “A real-time system for monitoring of cyclists and pedestrians,” Proceedings of 2nd IEEE Workshop on Visual Surveillance, pp. 74–81, July 1999.

[6] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, “Improving shadow suppression in moving object detection with HSV color information,” Proceedings of IEEE International Conference on Intelligent Transportation Systems, pp. 334–339, August 2001.

[7] P. Spagnolo, T. D'Orazio, M. Leo, and A. Distante, “Moving object segmentation by background subtraction and temporal analysis,” Image and Vision Computing, Vol. 24, No. 5, pp. 411–423, May 2006.

[8] C. Wren, A. Azabayejani, T. Darrell, and A. Pentland, “Pfinder: Real-time tracking of the human body,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 780–785, July 1997.

[9] C. Stauffer and W. E. L. Grimson, “Adaptive background mixture models for real-time tracking,” Proceedings of Computer Vision and Pattern Recognition, Vol. 2, pp. 246–252, June 1999.

[10] Z. Zivkovic and F. van der Heijden, “Efficient adaptive density estimation per image pixel for the task of background subtraction,” Pattern Recognition Letters, Vol. 27, No. 7, pp. 773–780, May 2006.

[11] A. Elgammal, D. Harwood, and L. S. Davis, “Nonparametric model for background subtraction,” Proceedings of 6th European Conference on Computer Vision, Vol. 2, pp. 751–767, June 2000.

[12] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, “Real-time foreground-background segmentation using codebook model,” Real-Time Imaging, Vol. 11, No. 3, pp. 172–185, June 2005.

[13] J. Stander, R. Mech, and J. Ostermann, “Detection of moving cast shadows for object segmentation,” IEEE Transactions on Multimedia, Vol. 1, No. 1, pp. 65–76, March 1999.

[14] I. Mikic, P. C. Cosman, G. T. Kogut, and M. M. Trivedi, “Moving shadow and object detection in traffic scenes,” Proceedings of International Conference on Pattern Recognition, Vol. 1, pp. 321–324, September 2000.

[15] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, “Detecting moving shadows: Formulation, algorithms and evaluation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 7, pp. 918–924, July 2003.

[16] S. Tattersall and K. Dawson-Howe, “Adaptive shadow identification through automatic parameter estimation in

video sequences," Proceedings of Irish Machine Vision and Image Processing, pp. 57–64, September 2003.

[17] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast shadow segmentation using invariant color features," Computer Vision and Image Understanding, Vol. 95, No. 2, pp. 238–259, August 2004.

[18] E. A. Khan and E. Reinhard, "Evaluation of color spaces for edge classification in outdoor scenes," Proceedings of International Conference on Image Processing, Vol. 3, pp. 952–955, September 2005.

[19] A. Cavallaro, E. Salvador, and T. Ebrahimi, "Detecting shadows in image sequences," Proceedings of IEEE Conference on Visual Media Production, pp. 165–174, March 2004.

[20] M. Kampel, H. Wildenauer, P. Blauensteiner, and A. Hanbury, "Improved motion segmentation based on shadow detection," Electronic Letters on Computer Vision and Image Analysis, Vol. 6, No. 3, pp. 1–12, December 2007.

[21] M. Heikkila and M. Pietikainen, "A texture-based method for modeling the background and detecting moving objects," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 4, pp. 657–662, April 2006.

[22] P. Spagnolo, T. D. Orazio, M. Leo, and A. Distante, "Advances in shadow removing for motion detection algorithms," Proceedings of 2nd International Conference on Vision Video Graphics, pp. 69–75, July 2005.

[23] Y. P. Guan, "Wavelet multi-scale transform based foreground segmentation and shadow elimination," The Open Signal Processing Journal, Vol. 1, No. 6, pp. 1–6, November 2008.

