ROBUST MOVING OBJECTS SEGMENTATION BY BACKGROUND SUBTRACTION

P. Spagnolo, M. Leo, T. D'Orazio, A. Distante

Istituto di Studi sui Sistemi Intelligenti per l'Automazione - C.N.R.
Via Amendola 166/5, 70126 Bari, ITALY
{spagnolo,leo,dorazio,distante}@ba.issia.cnr.it

ABSTRACT

The problem of detecting moving objects is very important in many application contexts such as people detection and recognition, visual surveillance both in indoor and outdoor environments, and so on. In this paper we propose a motion detection algorithm based on background subtraction and shadow removing. The main idea is to implement a fast and reliable approach for motion detection, able to extract the moving objects without their own shadows, in a single-step algorithm. It is based on the correlation between regions selected from the reference image and the current one. In addition, the proposed algorithm is able to efficiently update the reference model: unlike traditional background updating algorithms, our approach works well on every pixel in the background image, even if covered by a foreground object, in order to have always a consistent reference image. The experiments, performed on real image sequences acquired both in indoor and outdoor environments with natural and artificial lights, demonstrate the effectiveness of the proposed algorithm.

1. INTRODUCTION

In the last years, motion detection has attracted great interest from computer vision researchers due to its promising applications in many areas, firstly visual surveillance. Video surveillance systems seek to automatically identify people, objects, or events of interest in different kinds of environments. The first step of these systems is to identify moving objects and also static objects that are different from a reference background.

In literature, motion detection for object segmentation has been treated by several papers. We can resume them saying that there are three conventional approaches to moving target detection: optical flow [1, 2]; temporal differencing [3]; and background subtraction [4, 5, 6]. Optical flow can be used to detect moving targets even in presence of camera motion, but the relative computational methods are very complex and cannot be coded into real-time algorithms without specialized hardware. Temporal differencing is very adaptive to dynamic environments but, generally, does not allow all relevant features and the true shape of the moving objects to be obtained. Usually, using this approach, only some boundary regions of the objects are detected, while the interior regions are detected as static. The last category, background subtraction, is the most used in literature. These methods implement a model of the background and compare the current image with this reference one. In this way the foreground objects present in the scene are detected, and the most reliable shapes of the moving objects are recovered.

The approach proposed in [5] models the background as a textured surface, each point of which is associated with a mean color and a variance about that mean. A threshold has been used for partitioning the background pixels into visible and occluded pixels. In [4] a background model that handle the small motions of background objects, such as vegetation, has been proposed; this algorithm is not able to cope with sudden changes in indoor luminance conditions. Both these algorithm use a simple threshold for the detection of moving pixels; this drawback has been avoided in [6,7], where each pixel is represented using a running average and a standard deviation maintained by temporal filtering. Unfortunately, these approaches do not solve the problem of updating the regions covered by foreground objects. All these approaches don't cope with the possible presence of light switches, typical in indoor contexts, as remarked in [8] and [9]. The only approaches able to handle this situation are proposed in [9, 10]: they are based on the presence of discontinuities in the training set. This constraint can alter the results of the algorithm when moving object is not much different from the background model. All these algorithm detect moving objects with their own shadows: so, a shadow removing algorithm must be invoked, in order to extract the correct shape of the objects, as indicated in [11]. The main drawback of such approaches is their excessive sensitiveness to natural and artificial light condition changes, slight movements of background objects (such as vegetation), and the presence of shadows. Removing shadows is very important to have a good segmentation of the detected objects. Shadows occur when objects partially or completely occlude direct light from a light source. We can interpret shadows in the image, and the effect they have on the pixels in the scene, as a semi-transparent region in which the scene reflectance undergoes a local attenuation. Many works have been done on background modeling but a general and valid solution has not been pointed out.

In this paper we propose a background subtraction-based algorithm for foreground objects segmentation. It is based on the analysis of the correlation exhibited by the reference image (background) and the current one. In this way, it is possible to detect only effective moving objects, without their own shadows, because these regions present the same texture both in the current image and in the background. In addition, small movements in vegetation will be handled, like also gradual change in the light conditions, typical for outdoor environments, and sudden small luminance changes, due to light switches in indoor context. Moreover, the proposed algorithm will be able to automatically update the reference image, in order to have always the most reliable background image. So, an innovative updating algorithm will be presented, able to update all pixels in the image, even if covered by foreground objects.

In the rest of the paper, firstly the background model, insensitive to the presence of shadows, is presented (section 2); then, a subsystem for the background updating is explained (section 3). Finally, the experimental results obtained on real image sequences acquired both in indoor and outdoor sites are reported (section 4).

2. BACKGROUND SUBTRACTION

Foreground object segmentation is a fundamental step of visual surveillance systems: the results of this step are the inputs for the subsequent processing (object recognition, motion analysis, activity recognition…). So it is very important to correctly extract the moving objects. This makes necessary to develop very reliable motion detection algorithms, that should be adaptive to luminance variations and able to reduce the number of false alarms due to noise. All traditional background subtraction algorithms detect objects with their own shadows; so, they need a further shadow removing algorithm to obtain the correct shape of the moving objects. In addition, they generally use a pixel based approach to detect motion:

$$\left| \boldsymbol{I}(\boldsymbol{x}) - \boldsymbol{B}(\boldsymbol{x}) \right| > \sigma \qquad (1)$$

where B(x), I(x) are respectively the reference and the current values for the pixel x, and σ is a suitable noise threshold. As reported in [12], with this kind of approaches the results are poor due to the presence of noise and gray level similarity between the background and the moving objects. This sensitiveness to noise requires the implementation of a filtering algorithm in order to remove small agglomerate of pixels.

In this paper we propose a motion detection approach that makes the system less sensitive to noise. A pixel is detected as moving not only by comparing its value in two different images, but also evaluating its relationship with neighborhood pixels. In this way, the great part of spurious noise pixels are not detected. In addition, with this texture control, only effective moving objects are detected, without their own shadows.

We have chosen to model the background by means of the average values exhibited by each pixel during a supervised training period, in which no foreground objects are present in the scene.

To obtain only foreground objects, a motion detection algorithm based on the evaluation of the correlation measurement between neighborhood pixels has been implemented. In particular, to decide if the examined pixel belongs to the background or foreground, a small window around it is selected. Firstly, we have implemented a standard correlation measurement:

$$\boldsymbol{C} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{M}(\boldsymbol{i},\boldsymbol{j})\boldsymbol{N}(\boldsymbol{i},\boldsymbol{j})}{\sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{M}(\boldsymbol{i},\boldsymbol{j})^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{N}(\boldsymbol{i},\boldsymbol{j})^2}} \qquad (2)$$

where M and N are two n-size windows selected around the examined pixel. Even if this algorithm produces interesting results, we have chosen to implement a different approach to define if a pixel has to be classified as foreground or background. The main idea is that we can classify a pixel as static or moving if its relationship with the neighborhood pixels is substantially unchanged in the current image and in the reference one. Then we have compared the ratios between it and an adjacent point of corresponding pixels in both current and background images as follows:

$$D = \begin{cases} \left| \dfrac{I(i,j)}{I(i,j+1)} - \dfrac{B(i,j)}{B(i,j+1)} \right| & \text{if } \mathrm{j} < \#\mathrm{col} \\[2ex] \left| \dfrac{I(i,j)}{I(i+1,j)} - \dfrac{B(i,j)}{B(i+1,j)} \right| & \text{if } \mathrm{j} = \#\mathrm{col} \end{cases} \qquad (3)$$

If the difference D of (3) is less than 0.9 (a threshold experimentally selected), then the pixel (i,j) is marked as static, otherwise it is a moving point.

Experimental results have shown that using correlation measurements proposed in (2) and (3) good results in foreground objects segmentation have been obtained. So, in order to reduce computational time (a fundamental requisite for a surveillance system that should work in real time) we have chosen to use the simpler and faster algorithm based on the correlation measurement proposed in (3). This kind of approach is able to extract only the effective moving foreground objects, without their own shadows, because shadow regions present the same texture both in reference and current image. In addition, the percentage of noise pixel is substantially reduced with respect to that obtained with other conventional pixel-based background subtraction algorithm. Finally, the great problem of sudden changes in luminance conditions, typical of indoor environments, has been avoided, because the texture of regions remains unchanged if the light source changes suddenly.

3. BACKGROUND UPDATING

Any background subtraction approach is sensitive to variations of the illumination. Each algorithm needs a reliable background model image consistent at each time instant with the current scene luminance condition. Many related works use background subtraction for determining the foreground regions that should not be considered by the background updating process. So, once the background points have been detected in the current image, only the pixels corresponding to the background model are updated, while all the other points, masked by foreground regions, remain unchanged. In particular, in case of objects moving slowly, as a person staying in the same region for a long period of time, this can invalidate the results. The proposed approach allows all the pixels of the background to be updated, even if they correspond to points covered by foreground objects. The main idea of our approach is that the intensity variation of each pixel of the background model is not estimated by referring only to the corresponding pixel on the current image, but considering the variations exhibited by all the pixels with the same intensity value. In this way, if a pixel is covered by a foreground object, it can be updated, accordingly to the variations observed at the other background pixels with the same intensity value. The updating value $\boldsymbol{B}^t(x,y)$ is estimated by evaluating the photometric gains as follows:

$$\Lambda^t(x,y) = \frac{I^{t+1}(x,y)}{B^t(x,y)} \qquad (4)$$

and averaging on all the pixels $\{(x,y)\}$ having the same intensity value $\boldsymbol{B}^t(x,y) = \boldsymbol{b}_i$:

$$\mu(\boldsymbol{b}_i) = \frac{1}{N(b_i)} \sum_{\{(x,y)\in I^{t+1} | B^t(x,y)=b_i\}} \Lambda^t(x,y) \qquad (5)$$

where $\{b_i\}_{i=1\ldots n}$ are the n different intensity values that a pixel can assume, and $N(b_i)$ is the number of pixels in the background image $\boldsymbol{B}^t(x,y)$ with intensity value b_i. The iterative updating rule becomes:

$$\boldsymbol{B}^{t+1}(x,y) = \alpha * \boldsymbol{B}^t(x,y)\mu(\boldsymbol{B}^t(x,y)) + (1-\alpha) * \boldsymbol{B}^t(x,y) \qquad (6)$$

In this way the artefacts due to an obsolete background obtained when an object moves after a long period of time, can be avoided.

4. EXPERIMENTAL RESULTS AND CONCLUSIONS

The experiments have been performed on real image sequences acquired with a static TV camera Dalsa CA-D6 with 528 X 512 pixels; the frame rate selected is 20Hz. The processing is performed with a Pentium IV, with 1,5 GHz and 128 Mb of RAM. We have chosen to test algorithm on two sequences acquired in outdoor and indoor environments. In particular, indoor sequences have been acquired in presence of sudden light changes, due to light switches, with a walking person in the room. Outdoor sequences have been acquired in an archeological sites, where typical illegal activities were simulated. The characteristic of the test sequences are reported in table 1.

Table 1: The characteristics of the test sequences.

	Frame	Frame rate	Size
Archeological site (outdoor)	17 622	10 frames/sec.	384 x 288
Laboratory (indoor)	20 948	10 frames/sec.	528 x 512

The results obtained applying the proposed motion detection algorithm are very encouraging. In fig. 1, some images obtained during the elaboration are plotted. In the first column the original grey level images are illustrated. The results after the background subtraction step are reported in the second column. The first three rows refers to image acquired in an archeological site, while in the last two rows the results obtained in an indoor context after switching off the light source are illustrated. As it can be seen, the foreground objects are correctly detected without their own shadows.

In order to have a quantitative estimation of the error, we have characterized the Detection Rate (DR) and the False Alarm Rate (FAR), as proposed in [12]:

$$DR = \frac{TP}{TP + FN} \qquad FAR = \frac{FP}{TP + FP} \qquad (7)$$

where TP (true positive) are the detected regions that correspond to moving objects; FP (false positive) are the detected regions that do not correspond to a moving object; and FN (false negative) are moving objects not detected. In table 2 we can se the results obtained on the two test sequences.

Table 2. Rates to measure the confidence

Test sequence	DR (%)	FAR (%)
1	97,46	3,7
2	97,31	1,6

We can note that the FAR parameter is under 4% in the first test sequence (outdoor environments, more sensitive to changes in luminance conditions) and even under 2% in the second test sequence (indoor context, light conditions more controlled). As a future work, we are effecting more intensive experimental tests, both in indoor and outdoor contexts, to evaluate the robustness and reliability of the system in different conditions.

9. REFERENCES

[1] S. Fejes, L.S. Davis, "What can projections of flow fields tell us about the visual motion", in *Proc Intern. Confer. on Computer Vision ICCV98*, 1998, pp. 979-986.

[2] L.Wixson, M.Hansen, "Detecting salient motion by accumulating directional-consistent flow", in *Proc. of Intern. Conf. on Comp. Vis.*, 1999, vol II, pp 797-804.

[3] C.Anderson, P.Burt, G. Van Der Wal, "Change detection and tracking using pyramid transformation techniques", in *Proc. of SPIE – Intell. Robots and Comp. Vision* Vol. 579, pp.72-78, 1985.

[4] I.Haritaoglu, D.Harwood, L.Davis, "A Fast Background Scene Modeling and Maintenance for Outdoor Surveillance", *ICPR*, pp.179-183, Barcelona, 2000.

[5] C.Wren, A.Azarbayejani, T.Darrell, A.Pentland, "Pfinder: Real-time tracking of the human body", *IEEE Trans. on Patt. An. and Mach. Intell.* 19(7): pp.780-785, 1997.

[6] T.Kanade, T.Collins, A.Lipton, "Advances in Cooperative Multi-Sensor Video Surveillance", *Darpa Image Underst. Work.*, Morgan Kaufmann, Nov. 1998, pp. 3-24.

[7] H. Fujiyoshi, A. J. Lipton, "Real-time human motion analysis by image skeletonisation", *IEEE WACV*, Princeton NJ, October 1998, pp.15-21.

[8] B. Xie, V. Ramesh, T. Boult, "Sudden Illumination Change Detection Using Order Consistency", *Workshop on Statistical Methods in Video Processing (in conjunction with ECCV2002*), June 2002.

[9] K. Toyama, J. Krumm, B. Brumitt, B. Meyers, "Wallflower: Principles and Practice of Background Maintenance", *International Conference on Computer Vision*, September 1999, Corfu, Greece.

[10] O. Javed, Z. Rasheed, O. Alatas and M. S., "KNIGHTM:a Real Time Surveillance System for Multiple Overlapping and Non-Overlapping Cameras", The fourth *Intern. Conference on Multimedia and Expo (ICME 2003)*, Baltimore, Maryland, 2003.

[11] R. Cucchiara, C. Grana, M. Piccardi, & A. Prati, "Detecting objects, shadows and ghosts in video streams by exploiting color and motion information", *Proc. ICIAP 2001*, pp.360-365, 2001.

[12] E.H. Jaraba, C.O. Urunuela, and J. Senar, "Detected motion classification with a double-background and a Neighborhood-based difference," *Patt. Recogn. Letter*, Elsevier, Location, pp. 2079-82, 2003(24).

Fig. 1: experimental results obtained in different applicative contexts