Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Image and Vision Computing 24 (2006) 411–423

www.elsevier.com/locate/imavis

# Moving object segmentation by background subtraction and temporal analysis

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

*Istituto di Studi sui Sistemi Intelligenti per l'Automazione—CNR, Via Amendola 122/D-I, 70126 Bari, Italy*

Received 20 February 2004; received in revised form 29 September 2005; accepted 3 January 2006

**Abstract**

In this paper, we address the problem of moving object segmentation using background subtraction. Solving this problem is very important for many applications: visual surveillance of both in outdoor and indoor environments, traffic control, behavior detection during sport activities, and so on. All these applications require as a first step, the detection of moving objects in the observed scene before applying any further technique for object recognition and activity identification.

We propose a reliable foreground segmentation algorithm that combines temporal image analysis with a reference background image. We are especially careful of the core problem arising in the analysis of outdoor daylight scenes: continuous variations of lighting conditions that cause unexpected changes in intensities on the background reference image. In this paper, a new approach for background adaptation to changes in illumination is presented. All the pixels in the image, even those covered by foreground objects, are continuously updated in the background model. The experimental results demonstrate the effectiveness of the proposed algorithm when applied in different outdoor and indoor environments.

*Keywords:* Moving object segmentation; Temporal analysis; Background updating

## 1. Introduction

Reliable detection of moving objects is an important requirement for many computer vision systems. In video surveillance applications, motion detection can be used to determine the presence of people, cars or other unexpected objects and then start up more complex activity recognition steps. Additionally, the segmentation of moving objects in the observed scenes is an important problem to solve for traffic flow measurements, or behavior detection during sport activities.

In the literature, the problem of moving object segmentation is discussed, identifying three different kinds of approaches: optical flow [1–4]; temporal differencing [5]; and background subtraction. In particular, backgrounding methods, using an opportune thresholding procedure on the difference between each image of the sequence and a model image of the background, are recognized by the scientific community as those that provide the best compromise between performance and reliability. In addition, they produce the most complete feature data and allow the recovery of the most reliable shapes of the segmented moving objects. Any motion detection system based on background subtraction, needs to handle a number of critical situations such as:

- gradual variations of the lighting conditions in the scene;
- small movements of non-static objects such as tree branches and bushes blowing in the wind;
- noise image, due to a poor quality image source;
- permanent variations of the objects in the scene, such as cars that park (or depart after a long period);
- movements of objects in the background that leave parts of it different from the background model (ghost regions in the image);
- sudden changes in the light conditions, (e.g. sudden clouding), or the presence of a light switch (the change from daylight to artificial lights in the evening);
- multiple objects moving in the scene both for long and short periods;
- shadow regions that are projected by foreground objects and are detected as moving objects.

* Corresponding author. Tel.: +39 80 5929442; fax: +39 80 5929460.

*E-mail addresses:* spagnolo@ba.issia.cnr.it (P. Spagnolo), spagnolo@ba.issia.cnr.it (T.D.' Orazio), leo@ba.issia.cnr.it (M. Leo), distante@ba.issia.cnr.it (A. Distante).

0262-8856/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2006.01.001

No perfect system exists. In the literature, a great number of works have been presented whose successes depend on their ability to solve the greatest number of the above problems as possible. In the following, without being exhaustive, we will cite some of them, trying to highlight the kind of problems that have been solved.

Basically, background approaches consist of two steps: the proper updating of a reference background model, and the suitable subtraction between the current image and the background model.

In the literature, many approaches for automatically adapting a background model to dynamic scene variations are proposed. Such methods differ mainly in the type of background model and in the procedure used to update the model. In completely static scenes, many works have modeled reasonably well the pixel intensity with a normal distribution, which has a mean color value and a variance about that mean. A simple adaptive filter has been used in [6], to update recursively the statistics of the visible pixels. In [7], the Kalman filter is used to model adaptively the background pixel according to known effects of the weather and the time of day on the intensity values. In [8], color and edge information have been used both for background modeling and for subtraction, using confidence maps to fuse intermediate results. These approaches work well with slight illumination changes, but cannot often handle either large, sudden changes in the scene or multiple moving objects in the scene.

Other important statistical approaches have been presented. In [9], the authors introduce a non-parametric background modeling, estimating the probability of observing a pixel intensity value based on a sample of intensity values for each pixel. The algorithm works well with small movements of vegetation, and uses color information to suppress the target's shadows. In [10], a background model dealing with small motions of background objects such as vegetation is proposed. Each pixel is represented by three values: its minimum and maximum intensity values, and the maximum intensity difference between consecutive frames observed during a training period. A more refined application of this algorithm, based on the use of a complex codebook model for the segmentation, is proposed in [11]. In [12], background maintenance is achieved by using a multi-layered approach that makes probabilistic predictions, processing images at the pixel, region and frame levels. It solves some of the problems mentioned above, such as sudden changes in illumination, but does not consider the problems of ghost regions or shadow detection.

The problem of determining whether a pixel belongs to the background or the foreground has also received great attention in the literature. Due to the noise in the image, the results can be unreliable if a simple threshold process is used. Therefore, many background subtraction algorithms have introduced noise measurements to extract moving objects. In [13,14], a mixture of Gaussian distributions has been used for modeling the pixel intensities, assuming that more than one process can be observed over time. Pixel values that do not fit the background distributions are considered foreground. Also in [15], a mixture of Gaussian classification model for each pixel is learned using an unsupervised technique. Noise measurements to determine foreground pixels, rather than a simple threshold, are also introduced in the statistical model proposed in [16]; each pixel is represented using a running average and standard deviation, maintained by temporal filtering. This approach uses a three-frames difference to handle objects that arrive or move from the scene to deal with the problem of variations in the background objects, but it has difficulty managing sudden changes in light conditions. In [17], the authors propose a simple background subtraction method based on logarithmic intensities of pixels. They claim to have results that are superior to traditional difference algorithms and which make the problem of threshold selection less critical.

In [18], a prediction-based online method for modeling dynamic scenes is proposed. The approach seems to work well, although it needs a supervised training procedure for the background modeling, and requires hundreds of images without moving objects.

Adaptive Kernel density estimation is used in [19], for a motion-based background subtraction algorithm. In this work the authors use optical flow for the detection of moving objects; in this way, they are able to handle complex background, but the computational costs are relatively high.

An interesting approach has been proposed recently in [20]. The authors propose to use spectral, spatial and temporal features, incorporated in a Bayesian framework, to characterize the background appearance at each pixel. Their method seems to work well in the presence of both static and dynamic backgrounds.

In [21], authors propose to use ratio images as the basis for motion detection; in this way, effects of illumination changes are smoothed out because static pixels change their value in a similar way, while the presence of a foreground object alters this uniformity in the ratio image. So, the problem of the threshold selection, usually related to the difference image, now is shifted onto the ratio image, even though the authors propose an automatic procedure based on histograms to address this.

In [22], the authors suggest the use of two or more cameras for disparity verification making their system invariant to rapid changes in illumination and also removing occlusion shadows. The use of two views is also the starting point of the approach proposed in [23], where the authors discuss at length false and missed detections due to particular geometric considerations.

The information from the neighborhood of each pixel is used in [24], to separate static pixels from moving ones; a double background model is considered to deal with slow and fast changes in illumination.

The analysis of previous methods reveals a common approach namely to update the intensity values of pixels belonging to the estimated background model. In this way these methods update only static points, selected according to the difference between the intensity value of the current image and the corresponding value of the background model. While this choice is effective in situations where objects move continuously and the background is visible for a significant part of the time, it could not be robust in scenes with many objects that move slowly. An interesting attempt to realize a more complex updating procedure, based on motion evaluation about the observed scene, has been proposed in [25]. The main drawback of this algorithm is that it is not general enough, being heavily conditioned by several heuristic choices.

In this paper, we propose a new method for background subtraction that applies the updating procedure to all the pixels in the background image including those covered by a foreground objects. The main novelty of this work is the evaluation of the intensity variation of each pixel of the background model by the consideration of the variations exhibited by all the pixels in the image with the same intensity value. In this way, the algorithm is unaffected by the problems of erroneous detection of static points and of ghost regions in the image when a foreground object moves after a long period of time.

The segmentation of moving regions is achieved by combining results of background subtraction and temporal image analysis. Also in [26,16], pixel difference between successive frames was used to handle moving objects. However, we propose a new approach that uses the radiometric similarity between corresponding regions of successive frames to evaluate effective moving points, and also between the current temporal image and the background reference image to segment foreground objects, making our approach more robust against noise and able to avoid false detection of small movements of vegetation. Sudden light variations are addressed by introducing a further step of motion evaluation that prevents our system from giving erroneous segmentations in those frames that correspond to an un-updated background model.

The paper is structured as follows. In Section 2, a brief overview of the proposed approach is given. In Section 3, the temporal image analysis is described. Background subtraction and background updating are described in Sections 4 and 5. The motion evaluation step used to detect sudden light change is introduced in Section 6. Experimental results from outdoor and indoor environments are illustrated in Section 7. Finally, some conclusions are given in Section 8.

## 2. System overview

The first step of every scene analysis system is the segmentation of foreground objects from the background. This task is a crucial prerequisite for the effectiveness of the global system. A foreground segmentation algorithm should be able to cope with a number of the critical situations as mentioned above. In particular, it should deal with continuous and sudden light changes, temporal and permanent variation in background objects, and obviously the presence of noise. Moreover, a fundamental constraint is that the algorithm has to make the system substantially independent of the presence of foreground objects and their size and velocity.

In Fig. 1, the system overview is represented. The approach we propose for the segmentation of a moving foreground combines a temporal image analysis with a background subtraction approach. The system is composed of several steps in order to obtain the correct shapes of moving objects in every image of the sequence.

First of all, a temporal image analysis provides an image $I_m^t$ containing moving regions by comparing at each time $t$ two consecutive frames. The image $I_m^t$ is used in the background subtraction step, which compares the moving points to a reference background image $B^t$ evaluated at each time $t$. The radiometric similarity between corresponding regions is used both in the temporal image analysis and in the background subtraction process to, respectively, detect moving pixels and to segment foreground points. The background model $B^t$ is updated using the current image $I^t$ and evaluating the average photometric gain on all the pixels that have the same intensity value. The updating procedure is applied to all the pixels of the model, including those covered by foreground objects. Finally, a motion evaluation step is applied to the images $I_m^t$ and $I_m^{t-1}$ in order to cope with special situations such as sudden light changes. In the following sections, we describe the details of these steps.

## 3. Temporal image analysis

The goal of this step is to classify the points in each image as moving or static. At each instant $t$, the acquired image $I^t$ has to be compared with the image $I^{t-1}$ acquired previously at time $t-1$.

Classical frame difference methods, used in many approaches proposed in the literature [1], have the problem of producing resulting images that can be greatly corrupted by spot noise if threshold selection is not optimal. Many authors [1,4] suggest the use of a filtering process in order to delete these spot regions and obtain more meaningful images.

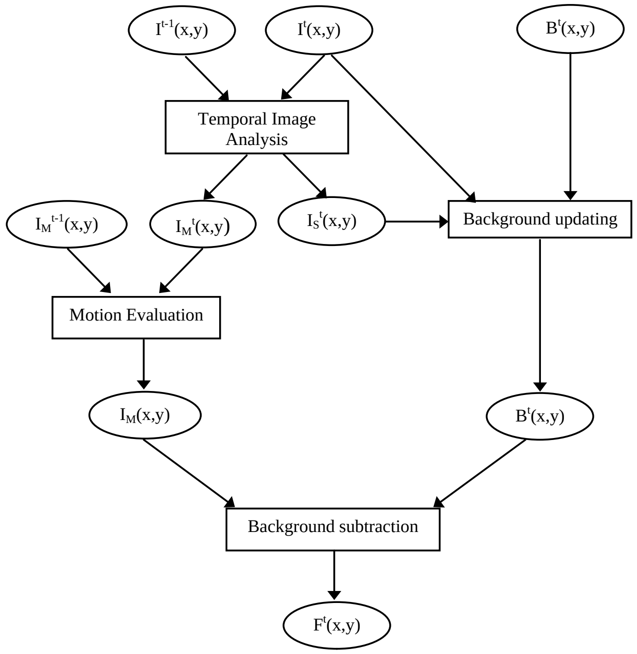

Fig. 1. System overview.

Starting from these considerations, we decided to estimate the similarity between the values that a pixel $(x,y)$ assumes in two consecutive frames $I^t(x, y)$ and $I^{t-1}(x,y)$ using the radiometric similarity $R(I^t(x,y), I^{t-1}(x,y))$ of their small neighborhoods determined by:

$$R(I^t(x,y), I^{t-1}(x,y)) = \frac{m[W(I^t(x,y))W(I^{t-1}(x,y))] - m[W(I^t(x,y))]m[W(I^{t-1}(x,y))]}{\sqrt{v[W(I^t(x,y))]v[W(I^{t-1}(x,y))]}} \tag{1}$$

where $m[W]$ and $v[W]$ represent the mean and the variance of the pixel intensities evaluated in the window $W$, respectively. The radiometric similarity varies in the range (0,1) and is estimated on a window $W$, centred on the pixel being compared between the successive images $((I^t(x,y)\ I^{t-1}(x,y))$. By using neighboring points in a window to compare corresponding points, we have chosen to give a local interpretation at the concept of difference, rather than a pixel-based one. In this way, the algorithm becomes more robust against noise: the effect of a single noise pixel is limited by other pixels in the window $W$. The window size has to be a reasonable trade-off between the ability to smooth agglomerates of noise pixels when large windows are used and the possibility to erroneously detect static points on the edge of moving objects.

We have decided to consider as static two points having a radiometric similarity value less than a given threshold $\sigma_S$; greater values identify moving pixels. Formally:

$$I_M(x,y) = \begin{cases} 1 & \text{if } R(x,y) < \sigma_S \\ 0 & \text{otherwise} \end{cases} \qquad I_s(x,y) = \begin{cases} 1 & \text{if } R(x,y) > \sigma_s \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The choices of the threshold $\sigma_S$, and the window size $W$ are made experimentally learning the proper values from different reference images. In this way, the results of the temporal analysis are less sensitive to the threshold selection, since the noise effects are smoothed out by the local analysis of the radiometric similarity.

In Fig. 2a and b, two successive images taken from a sequence in an outdoor context are shown. The image $I_m$ of moving pixels obtained with the temporal analysis is shown in Fig. 2c. Some small moving regions have been detected in the background, mainly due to the movement of vegetation. Also some static regions inside moving objects have been erroneously detected. The following step of background subtraction will tackle these problems. In Fig. 2d, the image obtained with a classical frame difference is shown. The improvement obtained with our method, using the radiometric similarity between regions to compare corresponding pixels in consecutive images and detect moving points, is clear.

## 4. Background subtraction

The core of each motion detection system is the part of background subtraction that effectively extracts the correct shape of moving objects. In this paper, we have used the image $I_m$ of moving points detected by the temporal image analysis and the reference background image $B^t$ evaluated at the time $t$. Radiometric similarity as introduced in Eq. (1) is used again for selecting from the moving points in $I_m$ those that are different from the model background image $B^t$. The resulting binary foreground image is obtained as follows:

$$F^t(x,y) = \begin{cases} 1 & \text{if } R(I_M(x,y), B^t(x,y)) < \sigma_s \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

In Fig. 2e, the resulting image obtained applying (3) is shown. The use of small windows around each point to compare the values of the pixel intensities in the couple of images $(I_m, B^t)$ has solved the problem of static holes inside moving objects and also small movement of vegetation. In Fig. 2f, the image obtained with the classical background subtraction, which compares locally pixels of the current image $I^t$ with the background model $B^t$ is shown. A large number of noise points have been detected. In order to use this kind of image, related works in literature introduce a further step of noise cleaning, with the problem of choosing the proper filtering levels and the corresponding risk of eliminating also some true moving pixels. The result obtained with the classical background subtraction using the image $I_m^t$ and the background model $B^t$ is shown in Fig. 2g. The results are better than the ones shown in Fig. 2f in terms of spot noise, but some holes

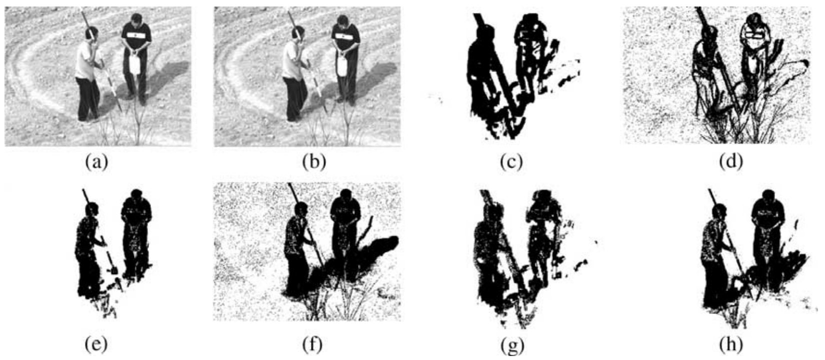

Fig. 2. (a) and (b) two images of the sequence; (c) $I_m$ obtained with our temporal image analysis; (d) image obtained with classical frame difference; (e) our foreground image obtained with $I_m$ and $B^t$; (f) image obtained with classical background subtraction between $I^t$ and $B^t$; (g) image obtained with classical background subtraction between $I_m$ and $B^t$; (h) our foreground image obtained with $I^t$ and $B^t$.

inside moving objects are still detected. Also the use of radiometric similarity between current image $I^t$ and the background model $B^t$ has been tested, the resulting image is shown in Fig. 2h. The results are similar to those obtained with (3) and shown in Fig. 2e, meaning that the use of radiometric similarity is actually effective. The only difference observable is that some moving points have been detected in shadow regions, while our approach has removed them, because of the use of $I_m^t$. However, the usefulness of temporal image analysis $I_m^t$ in (3) will be much clear when viewing the results shown in Fig. 3 where a sequence acquired in a parking area is analyzed.

The use of $I_m$, instead of the current image $I^t$, in the comparison with the background model solves the problem of ghost regions in the image. In the presence of a background object that moves away from the scene, a conventional algorithm continuously detects the presence of such an object in its old position. The image $B^t$ of the background model is shown in Fig. 3a. In Fig. 3b and c, two images acquired at the time $t$ and $t+1$ are shown. In this case, the car moves after a long period of staying in the image. The resulting image obtained with our background subtraction algorithm but using the current image $I^t$ and $B^t$ is reported in Fig. 3d, while the result obtained using $I_m^t$ instead of $I^t$ is shown in Fig. 3e. The ghost region of Fig. 3d, due to the change with respect to an old background, is demonstrated by plotting the difference image between Fig. 3d and f. The black area is usually found with all the approaches that do not use the temporal image analysis. With the proposed approach, which uses radiometric similarity and temporal image analysis, only effective moving objects are detected (Fig. 3e), reducing the number of false alarms due to temporal (vegetation) motion or permanent (static objects that move in the scene) variations in the background model.

## 5. Background updating

A reliable background model image has to account for varying light conditions at each time instant. This imposes the updating of the reference background image on the basis of the background observed in each image of the sequence.

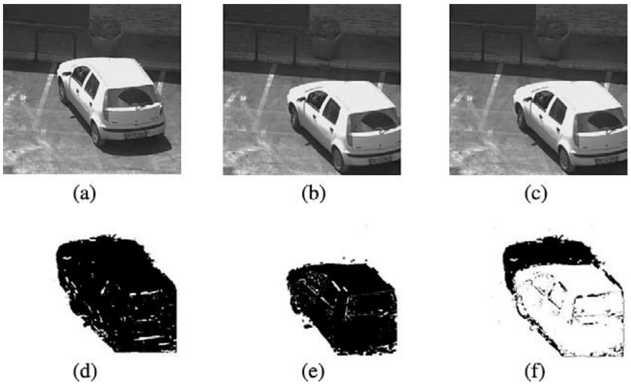

Fig. 3. (a) Background image $B^t$; (b) and (c) two images of the sequence; (d) foreground image obtained with our approach using $I^t$ and $B^t$; (e) our Foreground image obtained using $I_m^t$ and $B^t$; (f) image difference between (d) and (e).

Most previous approaches use background subtraction for determining the foreground regions that should not be considered by the background updating process; so, a mistake in labeling foreground and background points could cause the wrong updating of the background model. In this way, the pixels covered by foreground objects are not updated. If in the scene there are objects that move slowly, the corresponding background pixels are not modified with the others for a long period of time. As a consequence, when objects move away, the background subtraction algorithm produces a great number of artefacts due to an inconsistent background model.

In this work, we consider all the pixels of the image in the background model, even those corresponding to points that are covered by foreground regions and are therefore at that moment not visible. The main idea of the proposed approach is to update each pixel according to the variations exhibited by all the pixels in the image with the same intensity value.

For each pixel the photometric gain is evaluated as follows:

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

where $B^{t-1}(x,y)$ is the background model at the step $t-1$, and $I^t(x,y)$ is the current image.

The photometric gains measured on all the static pixels having the same intensity value $B^{t-1}(x,y)=b_i$ are considered for the evaluation of a mean photometric gain as follows:

$$\mu(b_i) = \frac{1}{N(b_i)} \sum_{\{(x,y)\in I_S^t | B^{t-1}(x,y)=b_i\}} \Lambda^{t-1}(x,y) \tag{5}$$

where $\{b_i\}_{i=1\ldots n}$ are the $n$ different intensity values that a pixel can assume, and $N(b_\mathrm{i})$ is the number of pixels in the background image $B^{t-1}(x,y)$ with intensity value $b_i$. Eq. (5) is evaluated on all the possible intensity values of the image.

The updating rule for the background image is defined as follows:

$$B^t(x,y) = B^{t-1}(x,y)\mu(B^{t-1}(x,y)) \tag{6}$$

In order to validate the effectiveness of the proposed updating algorithm, we have plotted in Fig. 4 the mean photometric gains for all the pixel intensities and the corresponding variances. Two different background models have been used to evaluate Eq. (4): the updated background image obtained at the step $t-1$ with Eq. (6) and an old background image taken 30 s before the current image $I^t$. From Fig. 4b, we can observe that in both cases the variances are very low, meaning that the pixels with the same intensity scattered through the entire image are varying in the same way. For this reason, we can generalise this trend and use the mean photometric gain to update all the pixels, including those covered by foreground objects. In Fig. 4a, these mean photometric gains are plotted: the values are close to 1 when the updated background model is used, means that our updating procedure works well. On the contrary, an old background produces mean photometric gains greater than 1: luminance is increasing in the image, but always in the same way.

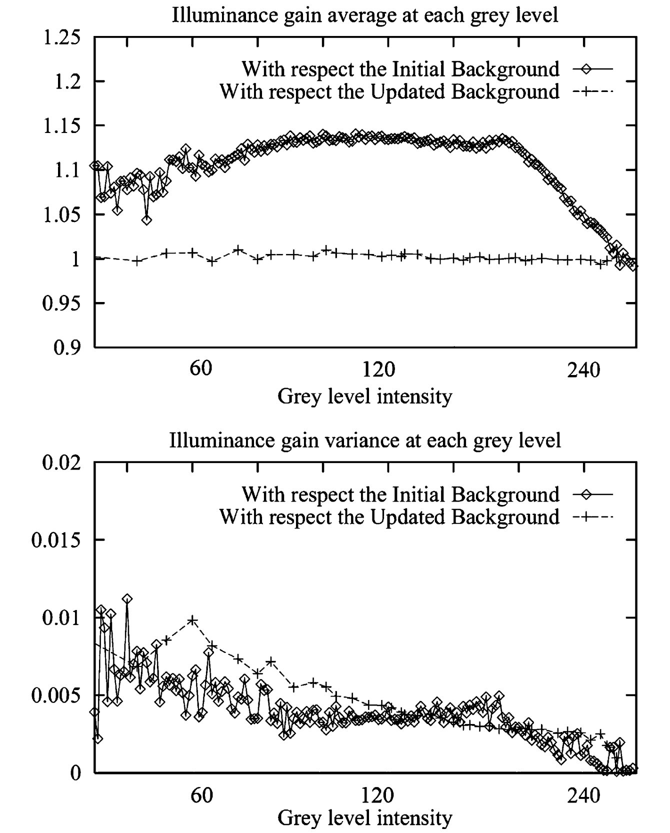

Fig. 4. The photometric gain exhibited by each different intensity value over the whole current image with respect to the initial background model (i.e. 30 s ago) and with respect to the background updated at the previous step.

The proposed background updating procedure reveals a number of advantages: it smooths and reduces the effects of noise in the image (sudden variations of a single pixel are not included in the background model since they are averaged out by the behaviors of the other pixels with the same intensity); it does not depend on the correct detection of static or moving points (all the pixels in the image are updated). When objects move slowly in the scene the corresponding background points are also updated using the general trend of similar pixels. In order to demonstrate the last point, in Fig. 5 we show some images of a sequence where a car arrives, stays for a while, and then leaves. We have applied our algorithm of background updating only to static points in the image. In this case, when the car enters in the image the corresponding background points are not updated. The results on the last two Fig. 5i and l have been compared with those obtained applying our background updating algorithm to all the points in the image. In Fig. 6c and 7c, some noise points are clearly visible in the region where the car was parked for a while. In Figs. 6b and 7b, the results obtained with the proposed approach show that the correct updating of all the points in the background model avoids the detection of spot points in the segmented foreground image.

## 6. Motion evaluation

Another typical problem of any motion detection system is its reliability in the presence of sudden changes in light conditions. This is typical of indoor environments (owing to light switches) but also of outdoor contexts, owing, for example, to a sudden cloud. In such cases, the system should be able to re-establish normal conditions as soon as possible. This kind of situation can be easily detected by monitoring continuously the variations between consecutive images of moving points.

The difference between the percentages of moving points in $I_m^t$ and $I_m^{t-1}$ has been evaluated as follows:

$$D^t = \frac{|\mathrm{Num}(I_m^t) - \mathrm{Num}(I_m^{t-1})|}{\dim(I)} \tag{7}$$

where Num() is the number of black points (i.e. moving points) in the image and dim() is the image dimension. In Fig. 8, the quantity $D^t$ has been plotted for three different sequences. The sequences are characterized by people moving in an office, soccer players on a pitch, and a parked car that moves away in an outdoor scene. In normal situations, continuous movements of people or objects in the scene provide values of $D^t$, which are always less than the peaks obtained when sudden light variations cause an increase in the number of moving points detected. In Fig. 8, the difference $D^t$, produced by a quick movement of the car, is less than the value 0.2 (this means that only 20% of the pixels moved in two successive images); on

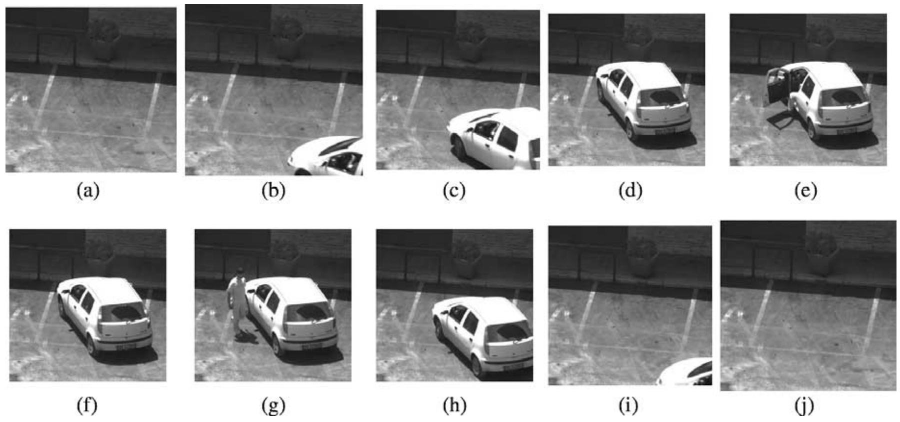

Fig. 5. Some images of a sequence in a parking area.

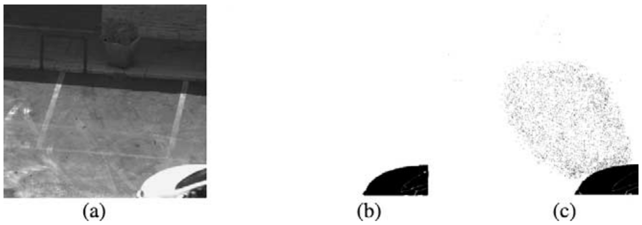

Fig. 6. (a) an image of the sequence, the car is going away; (b) the resulting image obtained with the proposed approach of background updating; (c) the image obtained with our approach but using only static points in the background updating procedure.

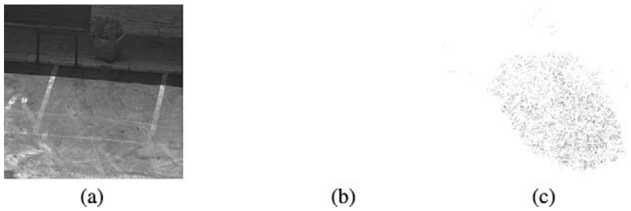

Fig. 7. (a) an image of the sequence, after that the car is gone away; (b) the resulting image obtained with the proposed approach of background updating; (c) the image obtained with our approach but using only static points in the background updating procedure.

the contrary, sudden light variations involve a percentage of moving pixels greater than 60%.

For this reason it is quite simple to fix a threshold value $\varepsilon$, used to evaluate the current motion in the scene as follows:

$$I_m = \begin{cases} I_m^t & D_t < \varepsilon \rightarrow \text{Normal Motion} \\ I_m^{t-1} & D_t \geq \varepsilon \rightarrow \text{Sudden Light Variation} \end{cases} \tag{8}$$

In the case of a normal situation, the current image $I_m^t$ is provided as input to the background subtraction step. If a sudden light variation occurs in the image $I_m^t$ is not meaningful and the use of $I_m^t$ in the background subtraction step would compromise the foreground segmentation, since it would be completely different from the background model $B^t$. In this case, the image $I_m^{t-1}$ is used in the background subtraction step even if it is not representative of the actual motion in the current image. Simultaneously, the background updating module includes the global light variation in the model; then, at the time $(t+1)$ the segmentation using Eq. (3) with $I_m^{t+1}$ and $B^{t+1}$ will produce a foreground image that is strictly representative of actual motion.

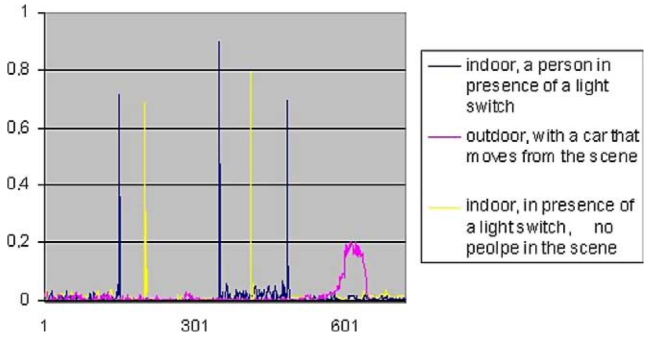

Fig. 8. The difference $D^t$ (normalized according to the image size) has been plotted for two different indoor sequences in presence of light switches, and for one outdoor sequence in presence of a sudden movement of a car.

## 7. Experimental results

The proposed method has been tested on a number of image sequences in different situations, in both outdoor and indoor environments. In particular, the sequences acquired in three different outdoor contexts are: an *archeological site*, where people move slowly in the scene; in a *parking area*, where a car arrives and moves away after a while; and in a *soccer stadium*, during a real football match where players move quickly. This choice of such different contexts was made to emphasize the great reliability and robustness of the proposed system. Indoor sequences were also considered to test our algorithm with fast illumination changes, obtained by turning off some fluorescent lights in our office.

The experiments were performed on a Pentium IV 2.0 GHz and the algorithm was implemented using a MS Visual C++ development tool. The processing time is strictly dependent on the quantity of moving points and on the image dimension. In order to give a general idea of the computational time, we have processed 1000 images for each sequence during which people or cars move in the scenes, and we have evaluated the average processing time. The results together with the image dimensions are reported in Table 1. The difference between the processing time in the parking area context and in the soccer context is due to the different kinds of current motion: in the parking area the number of moving points corresponding to cars is greater than the number of moving points corresponding to players in the soccer field. Moreover, these computational times are only indicative since no code optimisation has been introduced.

The proposed method requires the setting of two threshold values. The first one is $\sigma_S$ used in Eq. (3) and (4) to establish the similarity between corresponding points in both temporal analysis step and the background subtraction step. The value of $\sigma_S$ is learned experimentally from different training images: it is fixed equal to 0.9 and remains unchanged for all the experiments carried out. The value of the threshold $\varepsilon$ is chosen as 0.4, and represents the minimum percentage of expected moving points in successive frames when normal motion situations are in progress. Of course, this last threshold value depends on the focal length of the camera and can be set at the beginning of the experimental phase and then it remains unchanged for all the experiments. It is also necessary to evaluate the window size for the similarity measure. Different

Table 1
Average Processing time evaluated over 1000 frames for each sequence with different kinds of motion.

| Context | Image size | Average processing time (s) |
|---|---|---|
| Archeological site | 384×288 | 0.14 |
| Parking area | 266×256 | 0.106 |
| Soccer match | 266×256 | 0.099 |

experiments were executed comparing the results for window size of 3, 5 and 7 pixels. A small window size reduces the possibilities to smooth the noise effects, whereas a large window size gives rise to misdetection on the foreground object contours. In all three contexts considered, the best results were obtained by using windows of 5×5 pixels.

In Section 7.1, we show the results obtained on the three outdoor sequences; in Section 7.2, we show the results obtained in the case of a sudden illumination change in indoor sequences. In the last Section 7.3, we describe the possible generalization of the proposed algorithm to color images and multi-modal background models.

### 7.1. System evaluation

In Fig. 9, some images belonging to the first test sequence are shown. The results obtained with our algorithm are quite good, considering that the effects of the movement of vegetation have been greatly avoided and the shadows projected onto the ground (which in this case presents a low contrast) have been in part eliminated without any specific processing. The white static points detected inside the silhouettes are due to the color of the shirt being similar to the background color. Indeed, people wearing different shirts produce correct foreground segmentation. A simple connected component approach can be used to fill moving regions so that complete blobs can be extracted.

In Fig. 10, other results of applying our algorithm on some images of a parking area are shown. In this case there is no motion of background objects such as vegetation, but only global illumination changes, since the sequence has been acquired during a long period of time. The results show the robustness and reliability of the proposed algorithm to cope with this situation.

The results obtained on some images acquired during a soccer match with some players moving quickly, are shown in Fig. 11. In this case, the shadows are much more contrasted and the radiometric similarity used in Eq. (3) is not enough to eliminate them. A further step, based on texture similarity, is necessary for shadow removing, but this is beyond the scope of this paper (see [27], for more details on this subject).

The results illustrated in the previous figures give qualitative information about the effectiveness of the proposed algorithm. 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 [18]:

$$\mathrm{DR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \qquad \mathrm{FAR} = \frac{\mathrm{FP}}{\mathrm{TP}+\mathrm{FP}} \tag{9}$$

where TP (true positives) are the detected regions that correspond to moving objects; FP (false positives) are the detected regions that do not correspond to a moving object; and FN (false negatives) are moving objects not detected. In particular, we have manually segmented three gray level images, one for each context illustrated above. It should be noted that the manual segmentation includes only the effective moving objects, leaving out the shadows projected onto the ground. In Table 2, the results obtained on the three test sequences are shown. The DR parameter is always over 93%, demonstrating that the proposed system is very reliable, independent of the environmental context. In addition, the FAR parameter is under 4% in the first two test sequences

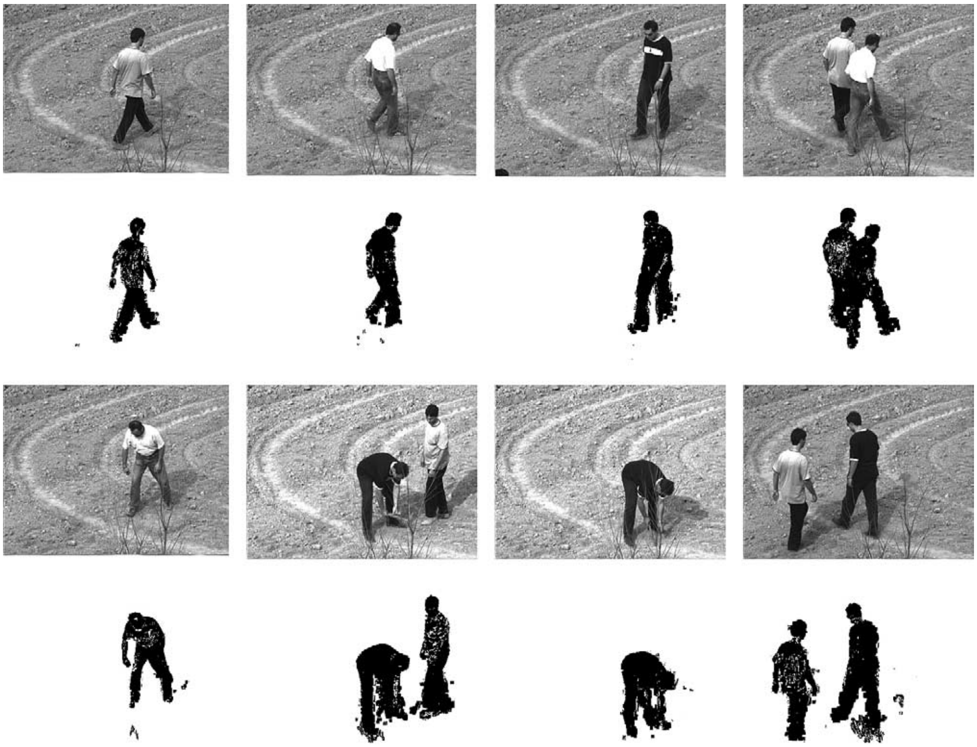
Fig. 9. The results obtained on some images acquired in an archeological site, with people moving around in the scene.

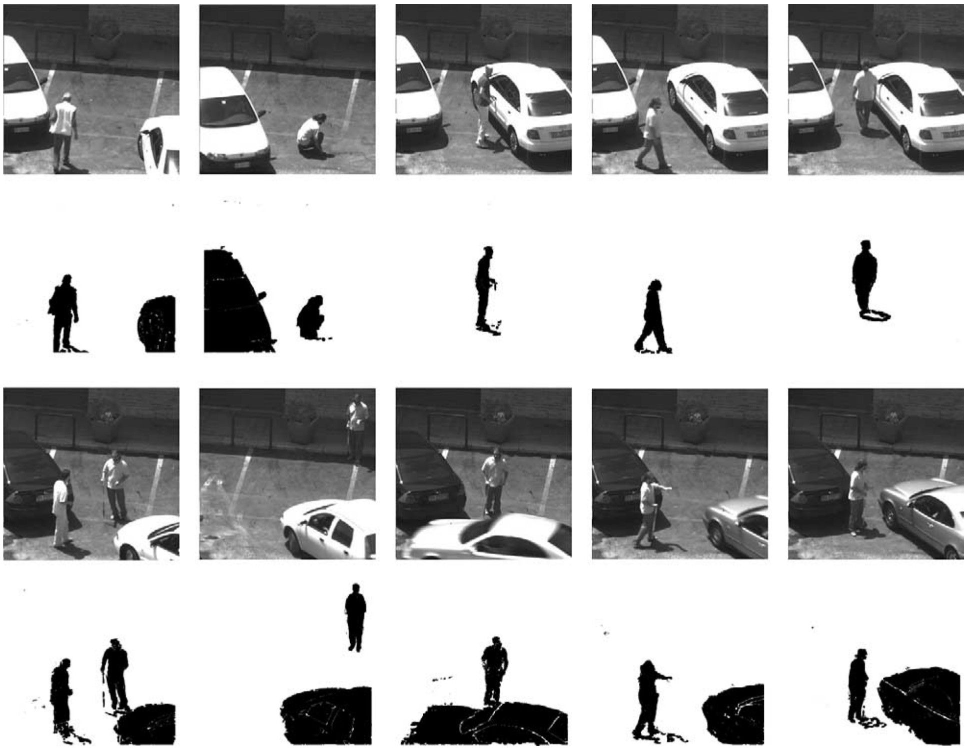

Fig. 10. Other examples show the results of applying our algorithm on some images of an outdoor scene containing people and car moving in a parking area.

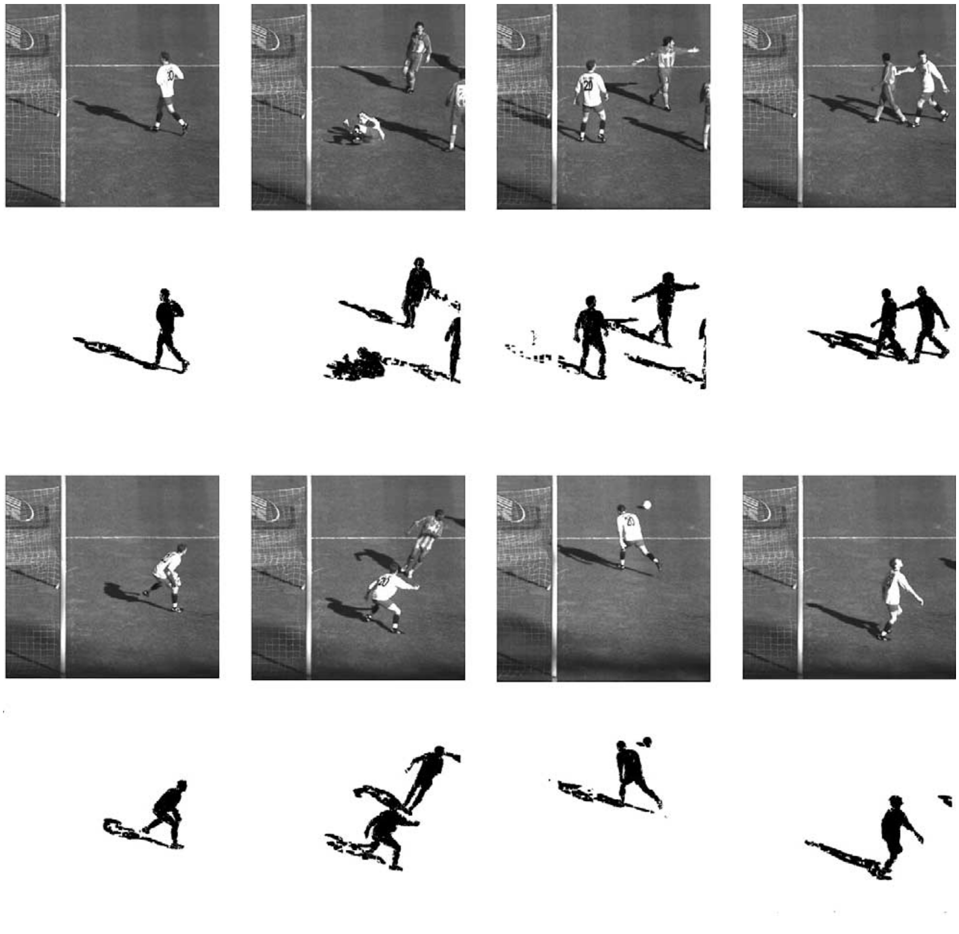

Fig. 11. The results obtained on some images acquired during a soccer match, with people moving quickly in the scene.

Table 2
Rates to measure the confidence

| Context | DR (%) | FAR (%) |
|---|---|---|
| Archeological site | 93.78 | 2.57 |
| Parking area | 96.24 | 2.11 |
| Soccer match | 97.11 | 11.69 |

while it is over 11% in the third one. As expected, these results are mainly due to the presence and the characteristic of shadows: in the first two situations (in the archeological site and in the parking area) shadows are more suffused, and have been partially removed by our algorithm; in the soccer context, shadows are more contrasted and so they have been segmented as moving objects.

Finally, in order to make the experimental results less sensitive to the effects of manual ground truth segmentation, we have used the evaluation method proposed in [28]. The goal of the perturbation detection rate (PDR) analysis described in this work is to measure the detection sensitivity of a background subtraction algorithm without assuming knowledge of the actual foreground distribution. This analysis is performed by shifting or perturbing the entire background distributions by values with fixed magnitude Δ, and computing an average detection rate as a function of contrast Δ. More details about this procedure can be found in this paper.

In our experiments, we have tested this technique in the three different experimental contexts presented above. The test set is given by 1000 points for each frame, 400 frames of the sequences for each context. So, for each Δ, 400×1000 perturbations and detection tests were performed. In Fig. 13, we have plotted the PDR graphs for the three test sequences.

It can be observed that the worst results have been obtained in the archeological site, where the critical conditions due to the presence of moving background objects decrease the performance. In this case, the pixel intensity variations, due to the movement of the vegetation, are amplified by the perturbation introduced, causing a decrease of the global detection ability. On the other hand, the results obtained in the other two contexts are very interesting, with a fast growth of the curve towards best performances, as already observed in Table 2.

### 7.2. Special situation: sudden illumination change

In this Section, we show the results obtained with our algorithm when sudden illumination changes occur. In these cases the use of temporal image analysis produces an incorrect segmentation of moving regions containing many regions corresponding to light variations with respect to the background model. In Fig. 12, one of the central fluorescent lights is turned off, and in the third frame a global darkening of the image is visible. The use of the image $I_m^t$ produces a segmented foreground image that is not significant as shown in the corresponding image of the second row. Some effects are still visible also in the successive image. The motion evaluation has been applied on the images of the third row of the figure. The sudden light variation does not have any effect: the shape of the person is successfully detected. The segmentation has been obtained, using the image $I_m^{t-1}$, at the same time as the background model is updated, including the global light variation. Also the following images (the fourth and the fifth images in the third row in Fig. 12) show our algorithm can cope with sudden illumination change.

### 7.3. Generalization of the algorithm

The proposed algorithm has been evaluated on grey level images; moreover, we supposed to have a simple background model, since the use of the radiometric similarity smooths out effects of slightly moving background objects. In order to assess if it is possible to generalize the proposed approach on color images, or when more complex background models are needed, further experiments have been carried out.

In presence of RGB images, the whole procedure described in sections 3–5 was applied for each color band, and the results

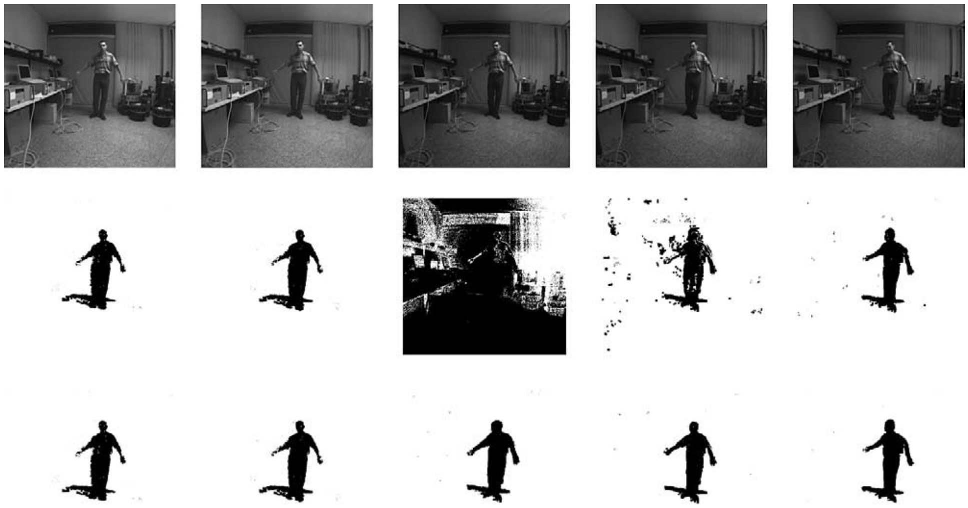

Fig. 12. The results obtained on some indoor images. In the first row, the original image: in the third frame a sudden light change is visible. In the second row, the results obtained with our algorithm without motion evaluation, while in the last row the result obtained with the proposed approach.

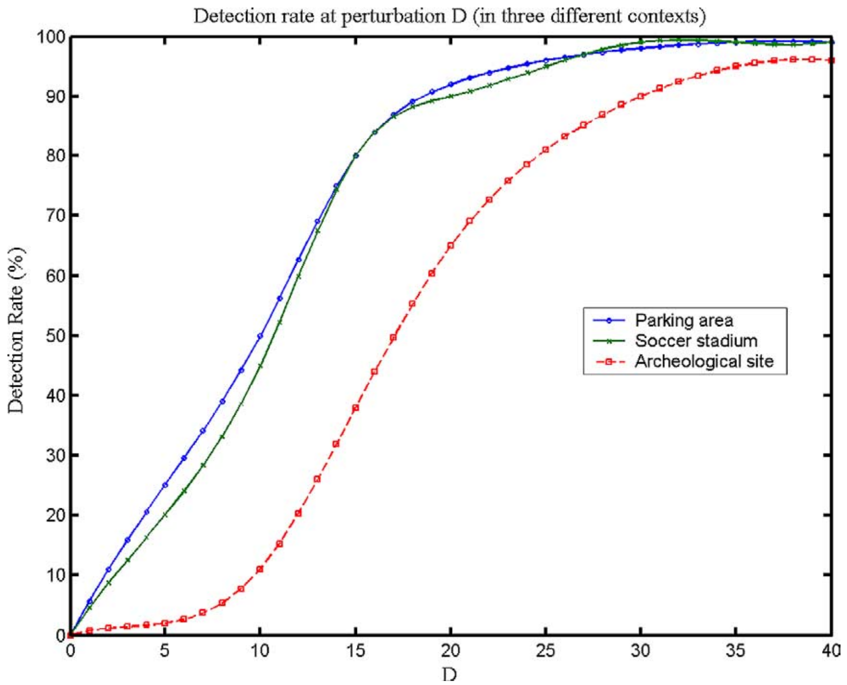

Fig. 13. The Perturbation Detection Rate for the three test sequences, 'Archaeological site', 'Soccer Stadium' and 'Parking Area'. It can be noted that the best performance has been obtained in the last two contexts, while, the worst results have been reported in the archaeological site, for the presence of moving vegetations.

obtained for the three bands were combined by means of an 'OR' logic operator. In Fig. 14, we present the experimental results obtained by applying our approach to a color sequence. The algorithm can be generalized to color images since the results are comparable to those obtained on grey level images.

The proposed approach does not require a complex supervised training phase; it compensates for the simplicity of the uni-modal background model by using a more articulated detection procedure. However, in some contexts, it could be necessary to use multi-modal background models for managing more complex situations, such as large movements of vegetation.

In this case, the background model was described by a set of Gaussians; in the following, we will use the notation $B_i^t$ to indicate the $i$th component of the background in terms of mean value, and $V_i^t$ to indicate the standard deviation for the same model.

The evaluation of the radiometric similarity between two consecutive frames $I^t(x,y)$ and $I^{t-1}(x,y)$ illustrated in (1) remains unchanged, whereas the same radiometric similarity between the motion image and the background, proposed in (3), has now to be applied to each background model:

$$F^t(x,y) = \begin{cases} 1 & \text{if } R(I_{\mathrm{M}}^t(x,y), \quad B_i^t(x,y)) < \sigma_{\mathrm{S}} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

A pixel is now considered as foreground if (10) is verified at least for one of the background models.

For the detection phase, the difference between the formulations (3)–(10) is not very relevant. On the other hand, the background modeling updating procedure is more complex since now we have more information to handle. Firstly, in (4) we have to consider the presence of several background

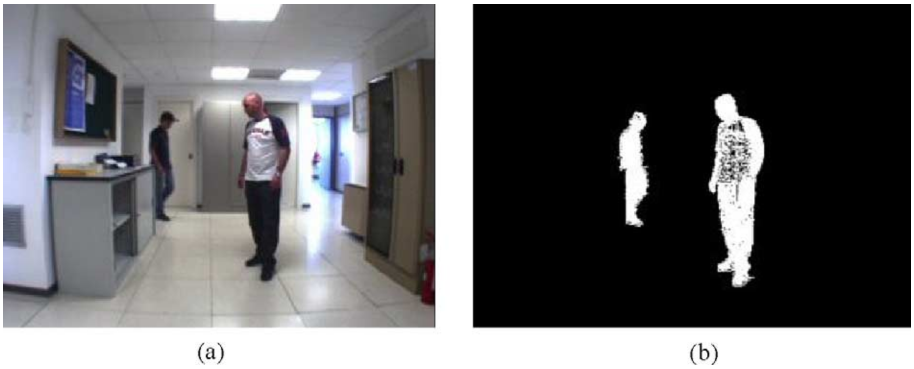

Fig. 14. (a) Original color image from an indoor sequence; (b) the result of the segmentation algorithm obtained by applying the improvements proposed in section 7.3 to the standard procedure described in sections 2–5.

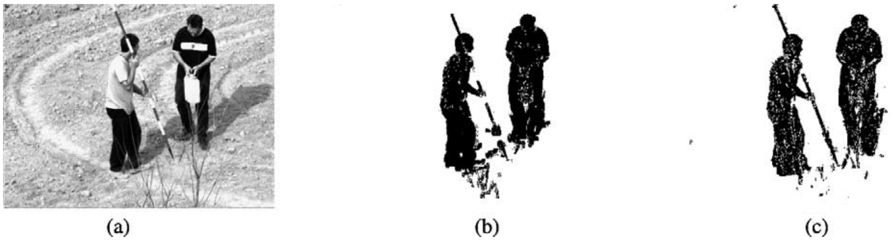

Fig. 15. (a) Original grey level image of the sequence; (b) the result of the standard algorithms proposed in sections 2–5 for foreground segmentation; (c) the result obtained using a multi-modal background model, as suggested in section 7.3.

models, so the photometric gain now will have to be evaluated for each point and for each of these models as follows:

$$\Lambda^{t-1}(x,y)=\frac{I^t(x,y)}{B_i^{t-1}(x,y)} \tag{11}$$

All these values will be averaged and used for the evaluating of the adjustment parameter introduced in (5). It is important to note that now the information about the standard deviation are used for the evaluation of $\mu(b_i)$. In particular, the pixels that are used in (11) to evaluate the parameter for the grey level $b_i$ are all pixels that in each background model differ from $b_i$ by at most two times the relative standard deviation. So, the new formulation of (5) becomes:

$$\mu(b_i)=\frac{1}{N(b_i)}\sum_{\{(x,y)\in I_S^t||B_j^{t-1}(x,y)-b_i|<2*V_j^{t-1}\}}\Lambda^{t-1}(x,y) \tag{12}$$

This setting parameter is used again for the adjustment of the overall updating parameters, as proposed in (6). The new formulation is

$$B_i^t(x,y)=B_i^{t-1}(x,y)\mu(B_i^{t-1}(x,y)) \tag{13}$$

In the same way also the standard deviation values are updated.

In Fig. 15, we propose the experimental results obtained on the sequence acquired in the archaeological site. In Fig. 15a, an original image is presented, in Fig. 15b the result obtained with the basic implementation of our algorithm is illustrated, while in Fig. 15c the results obtained with the improved multi-modal version, just explained, is presented. For this last experiment, we have chosen to use two background models. Predictably, the results improve in terms of false alarms, because of the presence of a more refined multi-modal background that can handle the movements in the background objects. On the other hand, the detection rate slightly decreases for real foreground objects, and this is a crucial point of all algorithms based on multi-modal approaches, as exhaustively reported in [28].

## 8. Conclusions

In this work, a system for moving object segmentation is presented. The separation of foreground moving objects from the background can be very useful in many contexts such as video surveillance, traffic flow measurement, behavior detection, and so on.

We have designed a robust algorithm for foreground segmentation, which combines background subtraction with temporal image analysis. The use of radiometric similarity between regions to compare pixels, both in the temporal image analysis and in the background subtraction has been proven to solve the problems of small movement of vegetations, gradual variations of light conditions, and also ghost elimination when background objects moves in the scene. A novel way to update all the pixels in the background model using the mean photometric gain allows our system to cope with multiple objects moving slowly in the scene for long periods. Sudden changes in luminance conditions have also been addressed using motion evaluation between successive frames.

The results obtained in different outdoor and indoor contexts show the robustness and reliability of the proposed algorithm. The method may mistakenly detect objects that are motionless in the image, and it is not able to eliminate shadows especially when they are highly contrasted on the background. However, the last problem is beyond the scope of this paper, and can be solved using a further step of shadow removing that considers the transparency property of segmented regions. Moving objects that enter the scene and then appear motionless seem quite improbable except for some cases such as parked cars or left objects that are slowly included into the background model.

## Acknowledgements

The authors would like to thank Neil Owens for his helpful suggestions and clarifications during the preparation of this paper. This work was developed under MIUR grant (Ref. D.M. n. 1105, 2/10/2002) 'Tecnologie Innovative e Sistemi Multisensoriali Intelligenti per la Tutela dei Beni Culturali'.

## References

[1] S. Fejes, L.S. Davis, Detection of independent motion using directional motion estimation, Technical Report, CAR-TR-866, CS-TR 3815, University of Maryland, August 1997.

[2] S. Fejes, L.S. Davis, What can projections of flow fields tell us about the visual motion, ICCV'98, 4–7 January, Bombay, India.

[3] S. Fejes, L.S. Davis, Exploring visual motion using projections of flow fields, DARPA'97, 2–5 February, Virginia, USA.

[4] L. Wixson, M. Hansen, Detecting salient motion by accumulating directional-consistent flow, ICCV'99, 20–27 September, Corfù, Greece.

[5] N. Paragios, R. Deriche, Geodesic active contours and level sets for the detection and tracking of moving objects, IEEE Transactions on Pattern Analysis and Machine Interface 22 (3) (2000) 266–280.

[6] C. Wren, A. Azarbayejani, T. Darrell, A. Pentland, Pfinder: real-time tracking of the human body, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 780–785.

[7] D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, S. Russel, Towards robust automatic traffic scene analysis in real-time, in: Proceedings of the International Conference on Pattern Recognition, Israel, November 1994.

[8] S. Jabri, Z. Duric, H. Wechsler, A. Rosenfeld, Detection and location of people in video images using adaptive fusion of colour and edge information, ICPR 2000, 3–8 September, Barcellona, Spain, pp. 627–630.

[9] A. Elgammal, D. Harwood, L. Davis, Non-parametric model for background subtraction, in: IEEE Frame Rate Workshop, 21 September, Corfù, Greece, 1999.

[10] I. Haritaoglu, D. Harwood, L. Davis, W4: real-time surveillance of people and their activities, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8) (2000) 809–830.

[11] K. Kim, T.H. Chalidabhongse, D. Harwood, L.S. Davis, Real-time foreground–background segmentation using codebook model, Real-Time Imaging 11 (3) (2005) 172–185.

[12] K. Toyama, J. Krumm, B. Brumitt, B. Meyers, Wallflower: principles and practice of background maintenance, in: Proceedings of the Seventh International Conference on Computer Vision, 20–27 September, Corfù, Greece, 1999, pp. 255–261.

[13] E. Grimson, C. Stauffer, R. Roman, L. Lee, Using adaptive tracking to classify and monitoring activities in a site, in: Computer Vision and Pattern Recognition Conference, 23–25 June, Santa Barbara, CA, USA, 1998, pp. 22–29.

[14] C. Stauffer, W.E.L. Grimson, Learning patterns of activity using real-time tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8) (2000) 747–757.

[15] N. Friedman, S. Russel, Image segmentation in video sequences: a probabilistic approach, in: Uncertainty in Artificial Intelligence, 1–3 August, Providence, RI, USA, 1997.

[16] T. Kanade, T. Collins, A. Lipton, Advances in cooperative multi-sensor video surveillance Darpa Image Understanding Workshop, Morgan Kaufmann, Los Altos, CA, 1998. pp. 3–24.

[17] Quen-Zong Wu, Bor-Shenn Jeng, Background subtraction based on logarithimc intensities, Pattern Recognition Letters 23 (2002) 1529–1536.

[18] A. Monnet, A. Mittal, N. Paragios, V. Ramesh, Background modeling and subtraction of dynamic scenes, in: Proceedings of the International Conference on Computer Vision, 2003, pp. 1305–1312.

[19] A. Mittal, N. Paragios, Motion-based background subtraction using adaptive kernel density estimation, in: Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR), 2004, pp. 302–309.

[20] L. Li, W. Huang, I.Y.H. Gu, Q. Tian, Statistical modeling of complex backgrounds for foreground object detection, IEEE Transactions on Image Processing 13 (11) (2004).

[21] Q.Z. Wu, H.Y. Cheng, B.S. Jeng, Motion detection via change-point detection for cumulative histograms of ratio images, Pattern Recognition Letters 26 (5) (2005) 555–563.

[22] Y. Ivanov, A. Bobick, J. Liu, Fast lighting independent background subtraction, in: Proceedings of the IEEE Workshop on Visual Surveillance, 2 January, Bombay, India, 1998, pp. 49–55.

[23] S.N. Lim, A. Mittal, L.S. Davis, N. Paragios, Fast illumination-invariant background subtraction using two views: error analysis, sensor placement and applications, in: Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR), 2005, pp. 1071–1078.

[24] E. Herrero-Jaraba, C. Orrite-Urunuela, J. Senar, Detected motion classification with a double-background and a neighborhood-based difference, Pattern Recognition Letters 24 (2003) 2079–2092.

[25] T. Thongkamwitoon, S. Aramvith, T.H. Chalidabhongse, Non-linear learning factor control for statistical adaptive background subtraction algorithm, Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS'05), May 23–26, Kobe, Japan, 2005, pp. 3785–3788.

[26] R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, A system for video surveillance and monitoring, CMU-RI-TR-00-12, Robotic Institute, Carnegie Mellon University, May 2000.

[27] A. Branca, G. Attolico, A. Distante, Cast shadow removing in foreground segmentation, in: Proceedings of the International Conference on Pattern Recognition, Quebec, Canada, 2002.

[28] T.H. Chalidabhongse, K. Kim, D. Harwood, L.S. Davis, A perturbation method for evaluating background subtraction algorithms, Proceedings of the Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS 2003), October 11–12, Nice, France, 2003.