

# Automatic Monitoring of Forbidden Areas to Prevent Illegal Accesses

M. Leo, T. D'Orazio, A. Caroppo, T. Martiriggiano, and P. Spagnolo

CNR- ISSIA via Amendola 122/D-I,  
70126 Bari, Italy  
{leo, dorazio, caroppo, martiriggiano, spagnolo}@issia.ba.cnr.it

**Abstract.** Surveillance systems that automatically detect illegal behaviors performed by unaware people have a wide range of applications: security, health-care, conservation of cultural heritage and so on. In particular monitoring public areas such as museums and archaeological sites is a challenging problem that has to be solved in order to avoid irreparable damages to historical heritage. In this paper a system able to check by common digital RGB cameras unexpected accesses to forbidden areas in a public museum is presented. The reliability of the proposed framework is shown by large experimental tests performed in the Messapic Museum of Egnathia (Italy) .

## 1 Introduction

Archaeological sites and museums, scattered across the world, keep physical remains of past human activity and they testify thousands of years of endeavors to develop culture, science and civilization.

Unfortunately archaeological finds are fragile and non-renewable therefore it is very important to plan adequate precautionary measures to avoid irreparable damages to the cultural heritage caused by visitors.

In general the preservation of the archaeological heritage is insured by a lot of strict rules that visitors have to observe. The monitoring of visitor behaviors is either performed by watchman or by traditional surveillance systems consisting of cameras, storage devices, video monitors and security personnel.

Both solutions are tedious as security staff or watchman need to identify specific and unusual events from a large number of very common and repetitive events. A solution of these problems might be a visual system able to automatically detect illegal behaviors performed by visitor in archaeological site or museums.

Automatic recognition of human behaviors is one of the most exciting and challenging problem in the computer vision research field.

Open literature proposes different methods for understanding human actions and good review can be found in [1,2,3]. Basically, works in recognition of behaviors of unaware humans can be classified into three categories. The first one consists of methods that perform recognition of human behaviors by detecting and tracking the hands and/or eventually some other major body components like arms, head or legs [4,5]. These algorithms, besides, being very complex and time consuming, require a very narrow field of view and a considerable amount of prior knowledge that generally is impossible to get in unaware human context as surveillance.

The second category consists of methods that perform recognition of human behaviors by using space-temporal information of the human body configuration (posture analysis) [6,7,8,9,10]. These algorithms, generally, are not view invariant and are based on the assumption that human activities can be deduced just from human body posture; the reliability of these approaches has been proved only for a limited number of human activities, very often in contrived scenarios.

Finally, the third category of algorithms for the recognition of human behaviours, to which our work belongs, uses the analysis of the spatial and temporal properties of human motion (position in the scene, moving velocity and so on) [11,12,13,14].

In this paper we address the problem of automatically detect in real time violations of forbidden areas. In some cases this problem can be solved using a simple window in the image and monitoring the optical flow to detect access violations. Nevertheless, this kind of approaches gives good results only when the camera observes the scene in a proper direction and avoid the problems of the perspective projection. But in many real contexts such as the surveillance of large areas, the simple motion detection is not significant unless it is combined with a position information on the ground plane. Besides the presence of shadows can alter the moving blob shape and gives rise to false alarms. For this reason we propose a three steps algorithm: at first the RGB digital images acquired by non-professional cameras are used to detect moving objects; then shadow regions are removed by temporal photometric gain analysis and finally, for each detected moving object, the approximate 3D localization is obtained by homographic transformations. Whenever the 3D position of a moving object falls in a forbidden area the system automatically detects the event and provides an alarm signal.

The rest of the paper is organized as follow: an overview of the proposed system is provided in section 2 where motion detection, shadow removing and 3D localization algorithms will be detailed; section 3 presents, instead, the experimental results obtained on the real image sequences acquired by IEEE 1394 cameras in the Messapic Civic Museum of Egnathia (Brindisi, Italy).

## 2 System Overview

The proposed system works on the RGB images acquired by common digital cameras. The acquired images are firstly processed by a motion detection algorithm performed through background subtraction. In this phase the background is automatically built and updated by temporal statistical analysis. After motion detection a shadow removing procedure is performed on each image in order to discard shadow regions that, generally, alter the shape of the moving objects. After shadow removing, moving points are merged in regions on the basis of their spatial relationship. Finally the centre of the bounding box of each moving region is considered and the corresponding coordinates on the ground plane are extracted by homographic projection.

The following subsections explain the details of each step involved.

### 2.1 Motion Detection

The motion detection step has been achieved by implementing a novel background subtraction algorithm. It is essentially a variation of [15] with the advantages of

allowing the background modeling also in cases of motion in the initial image sequences, and also of being more robust in cases of multiple reflections in indoor contexts. The proposed motion detection approach is composed by three distinct phases: firstly, a model of the background needs to be created; then a background subtraction procedure is used to distinguish moving objects from static ones. Finally, an updating algorithm adapts the background to any variation in light conditions.

The background modeling algorithm implemented is very reliable because it does not require any assumption about the presence of moving objects in the scene.

It uses a sliding window (of  $N$  frames) whose first frame is assumed as first coarse background model, even if there are moving objects. Then, each frame of this window is compared with the coarse background: if a pixel value is similar (in all the three color channels) to the correspondent in the model image, mean value and standard deviation are evaluated for that point.

Practically, for each pixel, 6 parameters are considered:  $\mu_R, \mu_G, \mu_B, \sigma_R, \sigma_G, \sigma_B$  where  $\mu_n$  e  $\sigma_n$  represent respectively the mean value and the standard deviation in the  $n$ -th color band.

After checking all frames of the examined window, the statistical parameters are maintained only for those pixels with intensity values similar to the model for almost 90% of the whole considered window.

After this, a new sliding window is examined using as referring model the statistical parameters where maintained and the intensity values of the first image for those points in which the statistical parameters were rejected in the previous window.

This procedure is iterated until mean and standard deviation values have been maintained for all the pixels.

After the model construction, the system is able to automatically detect the presence of moving objects. For this purpose, a simple subtraction algorithm has been implemented. It is based on the evaluation of the difference between current image and the model; this difference is calculated for each color band. A pixel will be considered as a moving point if it differs more than two times from the relative variance at least in one color band. Formally, denoting with  $I_{OUT}$  the output binary image:

$$I_{OUT}(x, y) = \begin{cases} 1 & \text{if } |I_R(x, y) - \mu_R(x, y)| > 2 * \sigma_R(x, y) \vee \\ & |I_G(x, y) - \mu_G(x, y)| > 2 * \sigma_G(x, y) \vee \\ & |I_B(x, y) - \mu_B(x, y)| > 2 * \sigma_B(x, y) \\ 0 & \text{otherwise} \end{cases}$$

In order to make the system substantially insensible to variations in light conditions, an updating module has been implemented. The characteristics of the applicative context requires some specific constraints: in particular, objects that differ from the background image have always to be detected, that is they will be never included in the background model in order to maintain information about the presence of object removed from the scene until anomalous conditions will be restored.

So, the updating procedure starts from the output of the last algorithm, and only the pixels corresponding to static points ( $I_{OUT}(x, y) = 0$ ) will be updated. In detail, for each point, a weighted mean between the historic value and current value is carried out.

The parameter  $\alpha$  used for the updating can vary in  $[0,1]$  and smoothes the relative relevance of the current image instead of the background one.

$$\mu_R^{t+1} = \begin{cases} \alpha * \mu_R^t + (1 - \alpha) * I_R^t & \text{if } I_{OUT} = 0 \\ \mu_R^t & \text{if } I_{OUT} = 1 \end{cases}$$

## 2.2 Shadow Removing

After the background subtraction, in the resulting binary image many small clusters of pixels are still observable: a one-step filter removes blobs whose size is lower than a certain threshold. Finally, an image with only foreground objects is generated, where each object contains also its own shadows. The presence of shadows is a great problem for a motion detection system, because they alter real size and dimension of the objects and they make very difficult any following automatic scene interpretation attempt. This problem is mostly remarked in indoor contexts, where shadows are emphasized by the presence of many reflective objects; in addition shadows can be detected in every direction, on the floor, on the walls but also on the ceiling, so typical shadow removing algorithms, that assume shadows in a plane orthogonal with the human plane, cannot be used.

To prevent all these problems, correct shapes of the objects must be extracted: the system needs the implementation of a shadow removing algorithm.

The shadow removing approach here described starts from the assumption that a shadow is a uniform decreasing of the illumination of a part of an image due to the interposition of an opaque object with respect to a bright point-like illumination source. From this assumption, we can note that shadows move with their own objects but also that they have not a fixed texture, as real objects: they are half-transparent regions which retain the representation of the underlying background surface pattern. Therefore, our aim is to examine the parts of the image that have been detected as moving regions from the previous segmentation step but with a texture substantially unchanged with respect to the corresponding background. The algorithm looks for moving points whose attenuation values, at each color band, are similar; differently, moving points belonging to true foreground regions will have different attenuation values. In addition, these attenuation value will be lower than 1, because of the minor light that illuminates the shadow regions. Formally, we evaluate, for each moving point  $(x,y)$  the attenuation values  $S$  at each color band:

$$S_R(x,y) = \frac{I_R(x,y)}{B_R(x,y)} \quad S_G(x,y) = \frac{I_G(x,y)}{B_G(x,y)} \quad S_B(x,y) = \frac{I_B(x,y)}{B_B(x,y)}$$

where  $I_n(x,y)$  and  $B_n(x,y)$  are respectively the intensity value in the  $n$ -th color band of the pixels  $(x,y)$  in the current image and in the background image.

After this, pixels with an uniform attenuation will be removed:

$$I_{OUT}(x,y) = \begin{cases} 0 & \text{if } S_R(x,y) \equiv S_G(x,y) \equiv S_B(x,y) \wedge S_R(x,y), S_G(x,y), S_B(x,y) < 1 \\ 1 & \text{otherwise} \end{cases}$$

The output of this phase provides a motion image with the real shape of the moving objects, without any artifacts due to noise or shadows.

### 2.3 Localization of the Moving Objects in the Scene

After Motion Detection and Shadow Removing each pixel in the acquired image is labeled as belonging to the background or to some moving object. To perform the localization of the moving object in the 3D scene is, at this point, important to analyze the spatial relationship between moving pixels and aggregate them in uniform regions (connectivity analysis). The proposed system makes use of the 8 connectivity criterion: two pixels are part of the same object, regardless of whether they are connected along the horizontal, vertical, or diagonal direction. Each object is now localized in the 2D image plane but, due to the perspective distortion, it is not possible to determine its position in the 3D scene. To localize them in the 3D scene a further step must be introduced. For each detected moving region a point  $p$  is considered: it is obtained as the intersection of the vertical line passing through the center of the bounding box of the considered region and the lower side of the same bounding box.

To localize the point  $p$  in the 3D scene an homographic relationship between the image plane and the ground plane is introduced.

The relation between the generic point  $P(kx_i, ky_i, kz_i, k)$  belonging to the ground plane and its corresponding point  $p(u_i, v_i, 1)$  in the image plane is:

$$\mathbf{P} = \mathbf{M} \mathbf{p} \rightarrow \begin{bmatrix} kx_i \\ ky_i \\ kz_i \\ k \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \\ m_{41} & m_{42} & m_{43} \end{bmatrix} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \quad (1)$$

To get the position in the scene of the moving object detected in the image plane the 11 unknown items of the matrix  $\mathbf{M}$  have to be computed ( $m_{43}$  can be set to 1 considering that this is an homogenous linear system). The  $m_{ij}$  elements can be discovered considering 4 couples of points for which the coordinates both in the ground planes and in the image plane are a priori known.

## 3 Experimental Results

The experiments were performed in the Messapic Civic Museum of Eganthia (Brindisi, Italy). This museum have a lot of rooms containing important evidence of the past: the smallest archeological finds are kept under lock in proper showcases but the largest ones are exposed without protection. The areas next to the unprotected finds are forbidden to the visitors and a cord separates forbidden from allowed areas.

Sometime visitors step over the cord in order to touch the finds or to see them in more detail. In our experiment a IEEE 1394 camera was placed in the main room of the museum in order to monitor the behaviors of the visitors in the area where were placed some of the most important messapic archaeological finds.

The acquired images were sent to a laptop (Pentium III, 1200 Mhz, RAM 512, HD 30 Gb) where run the algorithms described in the previous section.

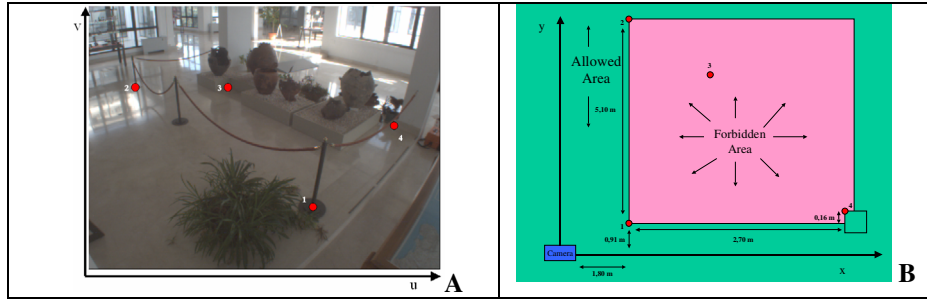
In figure 1A it is possible to see a frame acquired by the camera where the 4 red markers indicate the point of the ground plane chosen to discover the parameters of the homographic projection. The matrix  $M$  obtained by solving the homogeneous liner system (1) for the aforesaid 4 point is:

$$M = \begin{pmatrix} 0.0599709 & -0.0300762 & 9.030553 \\ -0.0493247 & -0.0313738 & 43.0326 \\ 0 & 0 & 0 \\ 0.000152512 & 0.0313815 & 1 \end{pmatrix}.$$

In figure 1B the plan of the acquired area is reported: the green color indicates allowed areas whereas pink color indicates forbidden areas. The red points correspond to the red reference points in figure 1A.





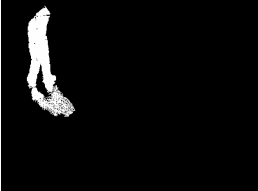
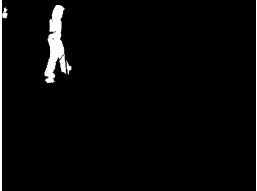




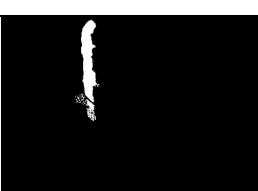
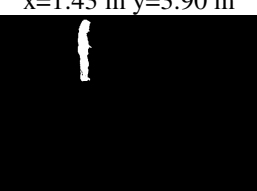
The room was monitored for about 1 hour (10 frame/sec). During the experiment several visitors came to the room but nobody of them went inside the forbidden areas. Some illegal accesses were performed by some actors.

In the figures 1A and 1B are furthermore pointed out the reference coordinate systems for both the image plane and the ground plane; onto the image plane the measure unit is the “pixel coordinate” whereas onto the ground plane is the “meter”.



**Fig. 1.** **A)** A frame acquired by the camera where the 4 red markers indicate the point of the ground plane chosen to discover the parameters of the homographic projection. **B)** the plan of the acquired area: the green color indicates allowed areas whereas pink color indicates forbidden areas. Red point correspond to the red point in figure 1A. Each figure reports also the reference coordinate systems used in the experiment.

In figure 2 the column A shows some frames extracted meanwhile a person steps over the cord and access to the forbidden area, whereas the column B shows the relative images containing the moving points detected before the shadow removing step and finally the column C shows the results obtained after shadow removing. The relative position of the moving person onto the image plane and onto the ground plane are respectively reported in columns A and C. By comparing the position of the moving person onto the ground plane with the boundary lines of the forbidden area the system detected that in the third and fourth rows the person is performing an illegal access and it sent an alarm to avoid dangerous interaction of the transgressor with the finds.

A	B	C
		
u=66 v=234		x=0.74 m y=3.88 m
		
u=98 v=219		x=1.08 m y=3.97 m
		
u=134 v=211		x=1.43 m y=3.90 m
		
u=211 v=160		x=2.83 m y=4.55 m

**Fig. 2.** The column **A** (on the left) shows some frames extracted meanwhile an actor performed the step over the cord; the column **B** shows the corresponding people segmentation results before shadow removing step and finally, the column **C** shows the relative images containing only the moving objects (after shadow removing step).

In figure 3 the benefits of using the proposed approach is evident: the figure A on the left shows a visitor that stays behind the limit of the forbidden area but he seems very close to the limit due to the perspective projection onto the image plan. In this case every approach based only on motion detection could wrongly detect an access violation and send a false alarm. The proposed approach instead detects the real position of the visitor and it is able to label this situation as normal. In the figure B on the right the visitor is inside the limit of the forbidden area. In this case the position estimation indicates the access violation and an alarm could be provided.

The proposed algorithm was tested on a large number of normal activities and on 15 forbidden situations that have been recorded during the 1 hour monitoring. The

system succeeded in detecting all the access violations. Of course the systems can fail every time the first two steps of the algorithm produce a not precise people segmentation. In fact the position evaluation, as explained in the previous section, depends on the exact estimation of the people shape: for instance the presence of shadows can modify the point  $p$ , obtained as intersection between the vertical line passing through the bounding box center and the lower side of the same bounding box.

Future work will be addressed to monitoring different activities such as the people trajectories evaluation, the interaction between people, the number of people who look at a each museum piece.



**Fig. 3.** Two critical situations: **A)** a visitor stays behind the limit of the forbidden area but he seems very close to the find due to the perspective projection onto the image plan; the proposed system avoid the error of the perspective perception and classifies as normal this behaviors. **B)** A visitor is inside the limit of the forbidden area. In this case the position estimation indicates the access violation and an alarm can be provided.

## References

1. Ying Wu, Thomas S. Huang "Vision-Based Gesture Recognition: A Review" *Lecture Notes in Computer Science* 1739, 1999, pp.103-114
2. C. Cedras, M. Shah "Motion based recognition: a survey", *Image and Vision Computing*, vol.13, n.2, pp. 129-155, 1995
3. D.M. Gravila "The visual analysis of human movement: a survey" *Computer Vision Image Understanding*, vol.73, n.1, pp.82-98, 1999
4. Jezekiel Ben-Arie, Zhiqian Wang, Purvin Pandit, Shyamsundar Rajaram "Human Activity Recognition Using Multidimensional Indexing", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 24, Issue 8 (August 2002) pp. 1091 – 1104
5. H. Ren, G. Xu "Human Action Recognition in a Smart Classroom" *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 2002
6. A. Galata, N. Johnson, D. Hogg, "Learning Variable-Length Markov Models of Behavior", *Computer Vision and Image Understanding*, 81, pp. 398-413 (2001)
7. A. Elgammal, V. Shet, Y. Yacoob, and L. S. Davis "Learning Dynamics for Exemplar-based Gesture Recognition" in the *proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR03)*, Madison, Wisconsin, June 16-22, 2003



8. R.T. Collins, A.J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, "A System for Video Surveillance and Monitoring", *Technical Report CMU-RI-TR-00-12*, Carnegie Mellon University, 2000.
9. C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland "Pfinder: Real-time tracking of the human body" *IEEE Transactions on PAMI*, 19(7):780–785, 1997.
10. B. Boulay and F. Bremond and M. Thonnat, "Human Posture Recognition in Video Sequence" *In the Proc. Joint IEEE International Workshop on VS-PETS, Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 2003, pp. 23-29
11. P. Remagnino and G.A. Jones, "Classifying Surveillance Events from Attributes and Behavior" *in the Proceeding of the BMVC*, Sept. 10-13, Manchester, pp. 685-694, 2001.
12. C. Stauffer and W.E.L. Grimson, "Learning Patterns of Activity Using Real-Time Tracking" *IEEE transactions on PAMI*, vol. 22, n.8, pp. 747-757, August 2000.
13. D. Ayers, M. Shah, "Monitoring human behavior from video taken in an office environment", *Image and Vision Computing* 19 (2001) pp.833-846
14. Nair, V. and Clark, J.J., "Automated Visual Surveillance Using Hidden Markov Models", *in the proc. of the 15th Vision Interface Conf.*, Calgary, pp 88-92, May 2002
15. T. Horprasert, D. Harwood, and L.S. Davis, "A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection" *Proc. IEEE ICCV'99 FRAME-RATE Workshop*, Kerkyra, Greece, September 1999