REAL TIME SMART SURVEILLANCE USING MOTION ANALYSIS

Marco Leo, P. Spagnolo, T. D'Orazio, P. L. Mazzeo and A. Distante
Institute of Intelligent Systems for Automation
Via Amendola 122/D-I 70126 Bari, Italy

Keywords: Smart surveillance, Motion analysis, Homographic Transformations, Edge Detection.

Abstract: Smart Surveillance is the use of automatic video analysis technologies for surveillance purposes and it is currently one of the most active research topics in computer vision because of the wide spectrum of promising applications. In general, the processing framework for smart surveillance consists of a preliminary and fundamental motion detection step in combination with higher level algorithms that are able to properly manage motion information. In this paper a reliable motion analysis approach is coupled with homographic transformations and a contour comparison procedure to achieve the automatic real-time monitoring of forbidden areas and the detection of abandoned or removed objects. Experimental tests were performed on real image sequences acquired from the Messapic museum of Egnathia (south of Italy).

1 INTRODUCTION

Smart surveillance is the use of automatic video analysis technologies in video surveillance applications. The aim is to develop intelligent visual equipment to replace the traditional vision-based surveillance systems where human operators continuously monitor a set of CCTV screens for specific event detection. This is not only quite a tedious activity, but with increased demand for area coverage, the continuous surveillance quickly becomes unfeasible due to the information overload for the human operators.

Current literature proposes different smart surveillance systems to measure traffic flow, monitor security-sensitive areas such as banks, department stores and parking lots, detect pedestrian congestion in public spaces, compile consumer demographics in shopping malls, etc. In (Wu & Huang,1999), (Cedras & Shah,1995), (Gravila,1999), (Aggarwal & Cai,1999), (Hu, Tan, Wang & Maybank, 2004) excellent surveys on this subject can be found .

Nearly every visual surveillance system involves a preliminary motion analysis step to segment regions corresponding to moving objects from the rest of an image (Haritaoglu, Harwood, Davis, 2000),(Wren et al.,1997),(Remagnino, Shihab, Jones,2004) ,(Dee & Hogg, 2004),(Collins et al., 2000), (Mittal & Davis,2003), (Bobick & Davis, 2001).

In this paper a motion analysis approach is coupled with semantic paradigms to achieve automatic smart surveillance of a public museum. In particular two problems are addressed: the monitoring of forbidden areas and the detection of abandoned or removed objects. In both cases the system has to automatically detect the unexpected event and to send an alarm containing a label of the detected anomaly (access violation, removed object or abandoned object).

This work is aimed towards the design of a reliable and automatic surveillance system to ensure a more efficient protection of the archaeological heritage of the considered sites.

The rest of the paper is organized as follows: section 2 details the algorithmic steps of the proposed methodology; section 3 reports experimental results and finally computational factors are discussed.

2 OVERVIEW OF THE SYSTEM

The first step of the whole procedure is a complex preprocessing phase which extracts the binary shapes (without shadows) on which the following algorithms have to work (Spagnolo, D'Orazio, Leo,

Leo M., Spagnolo P., D'Orazio T., L. Mazzeo P. and Distante A. (2007).
REAL TIME SMART SURVEILLANCE USING MOTION ANALYSIS.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - IU/MTSV*, pages 527-530

Distante, 2006) ,(Spagnolo, D'Orazio, Leo, Distante, 2005).
The following step is to aggregate pixels belonging to the same moving object in order to build a higher logical level entity named region or blob. Detected regions are the input of a color based probabilistic tracking procedure. The main aim of the tracking procedure is to analyze temporally the displacements of each moving region in order to manage overlapping or occlusion when the following decision making procedures could otherwise be misleading .
The tracking procedure exploits appearance and probabilistic models, suitably modified in order to take into account the shape variations and the possible region of occlusion (Cucchiara, Grana & Tardini, 2004). Using the procedures outlined each object is localized in the 2D image plane and is temporally tracked. Tracking information is the input to the two procedure dealing with the automatic recognition of suspicious human behaviors.
The first procedure deals with the problem of detection of forbidden area violation.
This procedure consists of two steps: firstly the 3D localization of moving regions is obtained using an homographic transformation (Hartley,R., Zisserman, A., 2003); then object positions on the ground plane are compared with those labeled as forbidden in the foregoing calibration procedure. If a match occurs the algorithm generates an alarm.
The second procedure deals with the problem of recognition of abandoned and removed objects.
In the literature usually these two issues are not distinguished, and they are dealt with in a similar way. So, detecting an abandoned/removed object becomes a tracking problem, with the aim of distinguishing moving people from static objects left/removed by human people (see (Connell, 2004) and (Spengler & Schiele, 2003) for good reviews).
In this work, instead, the goal is to distinguish between these two cases: so a classic tracking problem now becomes a pattern recognition problem. The reliability of the algorithm is strictly related to the ability to find/not find correspondences between patterns extracted in different images.
The approach implemented starts from the segmented image at each frame. If a blob is considered as static for a certain period of time (we have chosen to consider a blob as static if its position does not change for 5 seconds, but this value is arbitrary and does not affect the algorithm), it is passed to the module for removed/abandoned discrimination. By analyzing the edges, the system is able to detect the type of static regions as abandoned object (a static object left by a person) and removed object (a scene object that is moved). Primarily, an edge operator is applied to the segmented binary image F^t to find the edges of the detected blob. The same operator is applied to the current gray level image I^t.
To detect abandoned or removed objects a matching procedure of the edge points in the resulting edged images is introduced. To perform edge detection, we have used the Susan algorithm (Smith, 1992), which is very fast and has optimal performances. The matching procedure physically counts the number of edge points in the segmented image that have a correspondent edge point in the corresponding gray level image. Additionally, a searching procedure around those points is introduced to avoid mistakes due to noise or small segmentation flaws. Finally if the matching measurement $M^t{}_{FI}$ is greater than a certain value th_a experimentally selected, it means that the edges of the object extracted from the segmented image have correspondent edge points in the current grey level image and it is labeled as an abandoned object by the automatic system. Otherwise, if $M^t{}_{FI}$ has a small value, typically less then a given threshold th_r, it means that the edges of the foreground region do not match with edge points in the current image, so it is labelled by the automatic system as an object of the background that has been removed. For values of $M^t{}_{FI}$ between these two thresholds the system is not able to decide on the nature of the object.

3 EXPERIMENTAL RESULTS

The experiments were performed in both the Messapic Civic Museum of Eganthia.
The museum has many 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 around the unprotected finds are no-go zones for visitors and are defined with cord. Only a visual control can ensure that visitors don't step over the cord in order to touch the finds or to see them in more detail.
The proposed framework was tested to detect forbidden entry into protected areas of the museum and to recognize removed and abandoned objects in the monitored areas.

In our experiment IEEE 1394 cameras were placed in the main room of the museum. The acquired images were sent to a laptop where the algorithms described in the previous sections were processed.
The room was monitored for about 3 hours (30 frame/sec) during visiting hours: several visitors came to the room but none went inside the forbidden areas or touched the archeological finds. In this experimental phase no false positives were found, that is the system never gave the alarm in an improper way.
After closing time some people performed illegal behaviors in order to validate the capability of the system to automatically detect them. A set of 29 sequences were recorded collecting 15 forbidden area violations, 8 abandoned objects, and 6 removed objects.
Misclassification of human behaviors did not occur even in non trivial conditions. In particular, during the experimental phase, different people entered the scene at the same time and the sunlight shone through the large window with continuous changes of illumination conditions. The procedure for abandoned and removed object recognition did not fail even considering that texture in the static areas of the scene was not uniform and, in theory, this could cause false detections (due to possible edge matching between the contour of the removed object in the segmented image and the edge of the texture in the background).
In figure 1 column on the left shows a frame extracted while a person steps over the cord into the forbidden area, column in the middle shows the relative image containing the moving points detected before the shadow removing step and finally column on the right shows the results obtained after shadow removing. The relative position of the moving person on the image plane and onto the ground plane are also reported.
By comparing the position of the moving person on the ground plane with the boundary lines of the forbidden area the decision making procedure detected that in the third and fourth rows the person is performing an illegal access and sends an alarm.
In figure 2 the correct detection of a removed object is shown; the three images show a person approaching the finds and stealing a piece of an ancient vessel. In the second row the processed images are shown: notice that a red rectangle (red rectangles indicate removed object whereas blue rectangles indicate abandoned objects) has been positioned around the area where the removed object was.

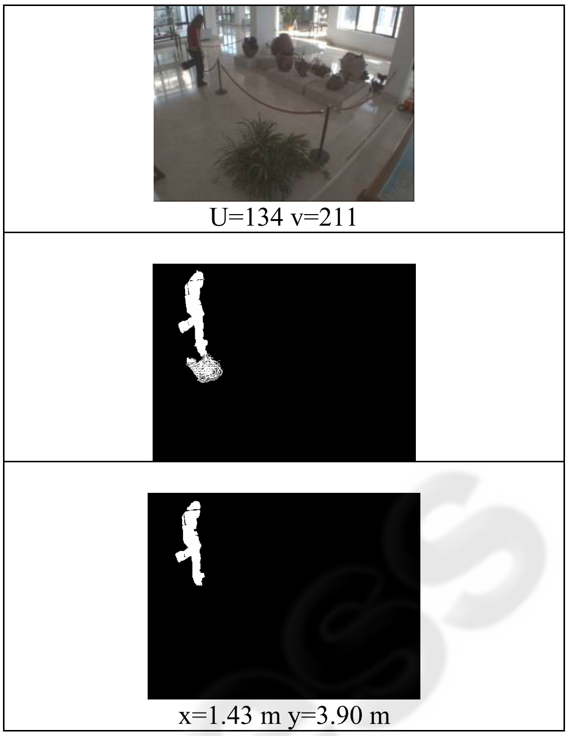

Figure 1: A frame extracted while an actor stepped over the cord and the corresponding segmented images before and after shadow removing. The relative position of the moving person on the image plane and onto the ground plane are also reported.

ACKNOWLEDGEMENTS

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

Wu,Y., Huang,T.S., (1999). Vision-Based Gesture Recognition: A Review. *Lecture Notes in Computer Science* 1739, 103-114.
Cedras, C., Shah, M., (1995). Motion based recognition: a survey. *Image and Vision Computing*, 13(2), 129-155.
Gravila, D.M., (1999). The visual analysis of human movement: a survey. *Computer Vision Image Understanding*, 73(1),82-98.
Dee,H., Hogg, D., (2004). Detecting inexplicable behaviour. *British Machine Vision Conference*.
Bobick, A. F., Davis,J. W. ,(2001).The recognition of human movement using temporal templates". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3), 257–267.

Collins, R.T., Lipton, A.J., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N., Hasegawa, O., 2000.A System for Video Surveillance and Monitoring. *Technical Report CMU-RI-TR-00-12*, Carnagie Mellon University.

Wren, C., Azarbayejani, A., Darrell, T., Pentland, A.,1997. Pfinder: Real-time tracking of the human body. *IEEE Transactions on PAMI*, 19(7), 780–785.

Cucchiara, R., Grana, C. , Tardini, G., 2004. Track-based and object-based occlusion for people tracking refinement in indoor surveillance. *ACM 2nd International Workshop on Video Surveillance & Sensor Networks.* 81-87.

Connell, J., 2004. Detection and Tracking in the IBM People Vision System. *IEEE ICME.*

Spengler, M. and Schiele, B., 2003. Automatic Detection and Tracking of Abandoned Objects. *IEEE Int. Work. on VS-PETS*

Smith., S.M., 1992. A new class of corner finder. *Proc. 3rd British Machine Vision Conference,* 139-148

Aggarwal, J.K., Cai Q., 1999. Human motion analysis: a review, *Computer Vision Image Understanding* 73 (3), 428-440.

Hu,W., Tan, T., Wang, L., Maybank, S. 2004. A Survey on Visual Surveillance of Object Motion and Behaviours. *IEEE Transactions on Systems, man and Cybernetics part. C Applications and review*, 34(3).

Haritaoglu,I., Harwood, D., Davis,L.S. 2000. W4: Real-Time Surveillance of People and Their Activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 809 – 830.

Mittal, A., Davis, L.S., 2003. M2Tracker: A Multi-View Approach to Segmenting and Tracking People in a Cluttered Scene. *International Journal of Computer Vision* 51(3), 189–203.

Remagnino, P., Shihab, A.I., Jones, G.A, 2004. Distributed intelligence for multi-camera visual surveillance. *Pattern Recognition* 37(4), 675-689.

Spagnolo, P., D'Orazio,T., Leo, M., Distante , A, 2006. Moving Object Segmentation by Background Subtraction and Temporal Analysis. *Image and Vision Computing* 24, 411-423

Spagnolo, P., D'Orazio,T., Leo, M., Distante, A., 2005. Advances in Shadow Removing for Motion Detection Algorithms. *Second International Conference on Vision, Video and Graphics*, 69-75.

Hartley,R., Zisserman, A., 2003. Multiple view geometry in computer vision *Cambridge University Press*, Cambridge, 2nd edition.

Figure 2: An example of automatic detection of removed object.