

Multiple View Surveillance using Image Registration

G. Sandhya Devi
CS&SE Department
Andhra University
A.P, India

P.V.G.D. Prasad
Reddy, Ph.D
CS&SE Department
Andhra University
A.P, India

G. Suvarna Kumar,
Ph.D
CSE Department
M.V.G.R.C.E
A.P, India

Vijaya Chitnaya
CSE Department
M.V.G.R.C.E
A.P, India

ABSTRACT

In the area of activity monitoring, video surveillance was mainly used for monitoring the scene by security personnel. However, due to high increase of unethical activities it is being strongly considered as a must for vigilance authorities, even sometimes for some unique purposes such as exam surveillance. The main theme is to ensure that students take their examinations the right way instead of indulging in unfair methods or malpractices. The proposed system aims to detect suspicious activities which occur in the classroom, thereby ruling out copying. We aim to develop a system that will identify the students from the top view camera using the image registration technique and check to see if he/she indulges in any form of malpractices or suspicious activities during the entire course of an examination. To get the aimed results the whole system has been divided into two modules through which the suspicious activities are classified from the normal activities. The first one includes the image registration technique which is used as a novel approach to determine the position of the student from the top view camera. The second module proposes the usage of object detection and skin detection. Finally we compare the normal frames and the suspicious frames based on their respective threshold values which are generated using the threshold technique. Also the accuracy of the system is checked using SVM classification.

1. INTRODUCTION

Surveillance is the monitoring of behavior, activities, or other changing information, usually of people for the purpose of influencing, managing, directing, or protecting them. This can include observation from a distance by means of electronic equipment (such as CCTV cameras), or interception of electronically transmitted information (such as Internet traffic or phone calls); and it can include simple, relatively no- or low-technology methods such as human intelligence agents and postal interception. Surveillance is very useful to governments and law enforcement to maintain social control, recognize and monitor threats, and prevent/investigate criminal activity.

Surveillance cameras are video cameras used for the purpose of observing an area. They are often connected to a recording device or IP network, and may be watched by a security guard or law enforcement officer. Cameras and recording equipment used to be relatively expensive and required human personnel to monitor camera footage, but analysis of footage has been made easier by automated software that organizes digital video footage into a searchable database, and by video analysis software (such as VIRAT and Human ID). With cheaper production techniques, surveillance cameras are

simple and inexpensive enough to be used in home security systems, and for everyday surveillance.

Video surveillance is found in different areas. Traditionally multiple security cameras are positioned throughout the area, linked to computer screens. Unfortunately humans are unable to monitor a huge number of screens and may lose their focus over time. In this we describe an automated multi-camera video surveillance system which is able to track people and detect suspicious situations

Automation is used in complex situations to reduce the workload put on the human operator and consequently to increase the productivity and enhance the performance. Workload assessment of operators while using an automatic surveillance system can provide system developers with information about the systems' effectiveness. Furthermore, depending on automation is not only related to the accuracy of the system but also to the human operators' reliance on automatic system.

The design of automatic surveillance systems can determine the level of reliability and increase the effectiveness of the surveillance. Automatic surveillance systems are interactive systems and human operators play a central part. Therefore, applying principles of Interaction design should theoretically enhance the performance.

We can break down the general problem of an automated surveillance system into a series of sub problems. In general, a surveillance system must be able to detect the presence of objects moving in its field of view, track these objects over time, classify them into various categories, and detect some of their activities. It should also be capable of generating a description of the events happening within its field of view (FOV). Each of these tasks poses its own challenges and hurdles for the system designers. We intend to automatically track students in an examination hall and analyze their behavior. The environment is monitored by a multi camera surveillance system where top view camera mounted sufficiently high above to provide a top-view of the scene. This is required since all the mal practices of students cannot be identified by the front camera surveillance. As a part of this, we implemented skin and object detection methods. With the help of these we aim to detect any such suspicious behavior in the examination hall.

This paper implements a technology in which real time videos are analyzed and are used for student activity analysis in an online examination, thus helping to classify whether the particular person's activity is suspicious or not. The goal of the project is to implement the system having following specifications:

- Automatically identify abnormal head motions, thereby prohibiting copying i.e. one student looking into another student's paper.

- Analyze the activities of students in an examination and alert the educational institute's administration on account of any malpractices/suspicious activities.

1.1 Object detection

The first step toward automated activity monitoring is detecting interesting objects in the camera's FOV. While the definition of an interesting object is context dependent, for a general automated system any independently moving object—such as a person, vehicle, or animal—is deemed interesting. We can achieve object detection by building a representation of the scene called the background model and then finding deviations from the model for each incoming frame.

Once the system detects the interesting objects, it's useful to have a record of their movement over time. We can define tracking as the problem of estimating the trajectory of an object as the object moves around a scene. Simply stated, we want to know where the object is in the image at each instant in time. If the object is continuously observable and its shape, size, or motion doesn't vary over time, then tracking isn't difficult.

1.2 Skin detection

In general the human skin is characterized by a combination of red and melanin (yellow, brown) and there is somewhat a range of hue for skin and saturation that represent skin-like pixels. The main goal of skin detection and classification is to build a decision rule that discriminate between skin and non-skin pixels. Identifying skin color pixels involves finding a range of values for which most skin pixels would fall in a given color space. In our proposed method we use skin detection for identifying the suspicious activity of a student in an examination. As all the suspicious activity of a student cannot be detected from front view so we use another camera which records from top view. We take the video feed as input from the top view camera and acquire the frames from the video. Skin detection is applied to the frame. If the skin is detected at places like –in between two students or at the immediate back of a student, we consider as some suspicious activity is taking place at that specified position.

1.3 Tracking across cameras

In general, surveillance systems are required to observe large areas like airport perimeters, naval ports, or shopping malls. In these scenarios, it isn't possible for a single camera to observe the complete area of interest because sensor resolution is finite and structures in the scene limit the visible areas. Therefore, surveillance of wide areas requires a system with the ability to track objects while observing them through multiple cameras. It's better if the tracking approach doesn't require camera calibration or complete site modeling, since the luxury of calibrated cameras or site models isn't available in most situations.

The main reason we use multiple camera surveillance is because there are a few malpractices that are not detectable from the front view camera hence we need a camera mounted at the top of the room to monitor the scenario.

Examination malpractice shall be defined as all forms of cheating which directly or indirectly falsify the ability of the students, outside an examination hall and any involvement in all illegal examination-related offences.

1.4 Types of malpractices in an Examination hall

Cheating within an examination Hall/Room

1. Copying from one-another/exchanging questions/answers sheets.
2. Bringing in prepared answers, copying from textbooks, notebooks, laboratory specimens or other instructional aids smuggled into the examination hall.
3. Collaboration with an invigilator/lecturer where it involves the lecturer proving written/oral answers to a student in the examination hall.
4. Oral/written communication between/amongst students.
5. Bringing in prepared answers written or oral from any person(s) outside an examination hall.
6. Receiving information whether written or oral from any person(s) outside an examination hall.
7. Impersonation

2. RELATED WORKS

A huge amount of research has been made in the field of computer vision. The main areas include human tracking [3,15,7], automatic human detection[14,9,13], counting [8,12]. It is expensive to hire staff to constantly monitor the feeds, and it is only a very small percentage of the video that contains useful information [5]. Multiple camera tracking and mixture models prove to be an effective joint solution in terms of robustness, even in presence of varying conditions as commonly happens in surveillance scenarios[1]. [6] proposed an efficient solution to people tracking in distributed video surveillance is requested to monitor crowded and large environments. This use of the Entry Edges of Field of View (E2oFoV) is used to solve the consistent labeling problem between partially overlapped views. Also, the estimation of visual focus of attention as detecting the orientation of body and following the head movements to find where the person is looking at is described by[2]. Integrating multiple computer vision based approaches to surveillance video analysis to support user retrieval of video segments showing human activities is given in [11]. The issues faced in developing both generic and specific analysis tools and how they were integrated for use in the new TRECVID interactive surveillance event detection task are described in[10]. Glas *et al.* [4] studied the problem of orientation of both body and head in a closed environment by combining video data and laser scanner data. Their work extracts the position of the arms and the head from a top-view appearance and finds the orientation of the human body in the scene.

3. PROPOSED METHOD

Top view camera surveillance is needed in cases where the position of the student in the frame does not match with any case that can be detected by front view surveillance camera.

But to examine the student position from top view, initially the exact coordinates of him/her needs to be evaluated. This is done by registering of images of both front-view and top-view grids. We have used control point registration to overlay both the grids. From the result of this registration, the top view coordinates of the student is calculated. At that particular position, object and skin detection methods are applied to

identify the suspicious activity performed by him/her. This is shown in figure 3.1

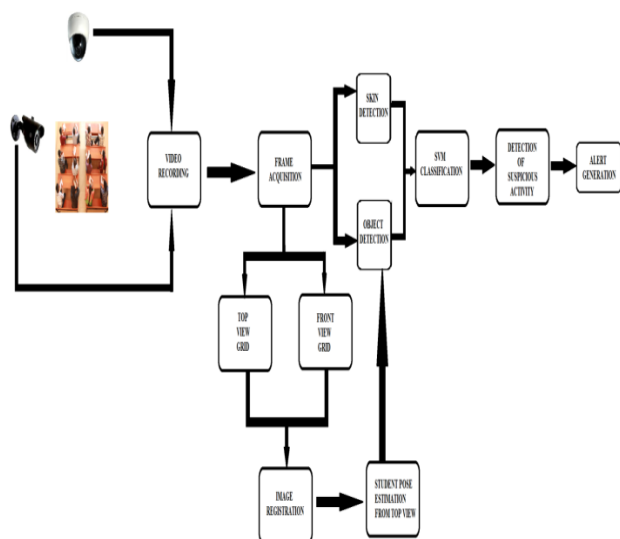


Figure 3.1: Flow diagram

Two-dimensional registration problem:

Let us assume that $F(x)$ and $G(x)$ be two functions which gives the respective pixel values at each location x in two images, where x is a vector. Then for any one-dimensional registration problem, the horizontal disparity is given by the following equation.

$$h = \sum_x \frac{w(x)[G(x) - F(x)]}{F'(x)} / \sum_k w(x)$$

Where

$$w(x) = \frac{1}{|G(x) - F(x)|}$$

Then the iteration of the disparity can be given by:

$$h_0 = 0,$$

$$h_{k+1} = h_k + \sum_x \frac{w(x)[G(x) - F(x + h_k)]}{F'(x + h_k)} / \sum_k w(x)$$

Where

$$\begin{aligned} 0 &= \frac{\partial E}{\partial h} \\ &= \frac{\partial}{\partial h} \sum_x [F(x) + hF'(x) - G(x)] \\ &= \sum_x 2F'(x)[F(x) + hF'(x) - G(x)], \end{aligned}$$

But, the two dimensional linear approximation occurs in a different form where the weight function is :

$$w(x) = F'(x)^2$$

and the corresponding iterative form with weighting is given by:

$$h_0 = 0,$$

$$h_{k+1} = h_k + \frac{\sum_x w(x)F'(x + h_k)[G(x) - F(x + h_k)]}{\sum_x w(x)F'(x + h_k)^2}$$

The following figures show a front view frame and a top view frame(Fig 3.3) for which grids are drawn.

3.1 Grid formation

Initially, a linear best fit curve is drawn connecting all faces of the students whose seating positions are in a column. This is repeated for all rows and columns covering every learner in the classroom which forms a grid on the image. The below equations are used to find linear best fit curve which passes through all the learners' faces in rows and columns. Referring to x and y axes, learners' face coordinates are $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. The equation 3.11 determines the linear best fit curve by using all coordinates of faces in rows and columns.

$$\tilde{y} = a_0 + a_1 \quad (3.11)$$

Where \tilde{y}_i for each x_i is the predicted value of y for x_i , based on the straight line that is found to best represent the data. a_0 and a_1 are two constants to be determined from the obtained data. Using the above linear best fit curve equation, a grid is formed for all learners in a class room which is shown in the figure 3.2



Figure 3.2: Grid formation

Similarly, grid drawn for a frame (Figure:3.3) in examination hall is figure:3.6 and the corresponding front view grid for the frame (Figure:3.4) is shown in the figure:3.5. Here, the grid sizes are minimized for simplicity in presentation.

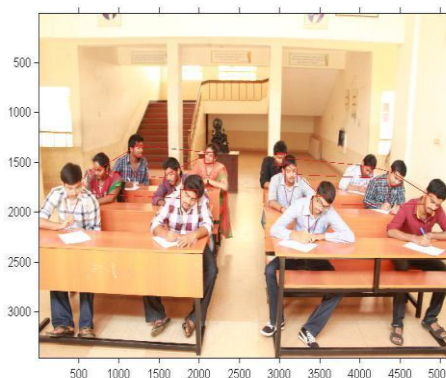


Figure 3.3: front view frame



Figure 3.4: Top view frame

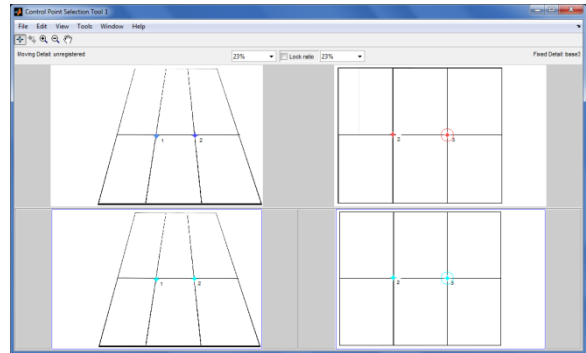


Figure 4.1: Control Point Registration

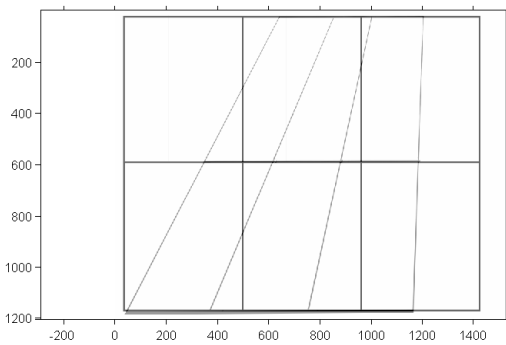


Figure 3.5: top-view grid

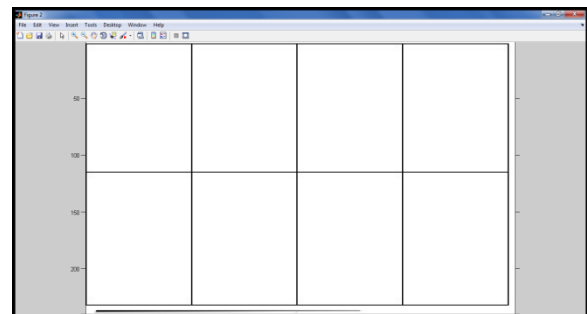


Figure 4.2: Registered Image

Finally, the grids are registered as shown in the figure 4.2. From the registered image, positions of students are calculated and tabulated as below:

Table 4.3: Positions of Students

Student Number	Front View	Top View
1	(35,1178)	(35,1178)
2	(374,1178)	(494,1178)
3	(755,1178)	(953,1178)
4	(1163,1178)	(1418,1178)
5	(347,596)	(35,596)
6	(617,596)	(494,596)

At a particular position where it is needed, object and skin detection is applied as shown in below figures(4.4, 4.5, 4.6, 4.7):

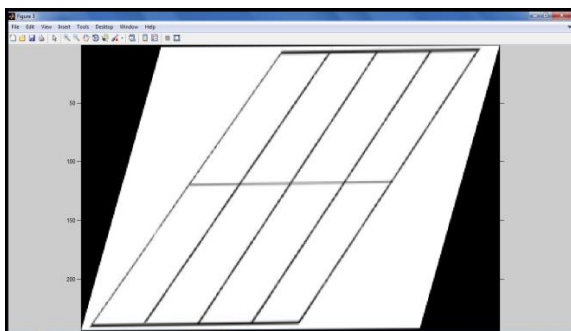


Figure 3.6: front-view grid

Control point registration is applied to these grids which results in a registered formation. The result is shown in results section. Now, all the positions of students from top view are calculated with respect to their front-view positions.

So at a particular student position, if the activity is not identified by the front-view camera, his/her top-view coordinates are extracted and object, skin detection techniques are applied.

4. RESULTS AND CONCLUSION

Initially, to obtain the position of student from top view, control point registration is done. The result is shown in the figure:4.1



Figure 4.4: Frame before Mobile is detected

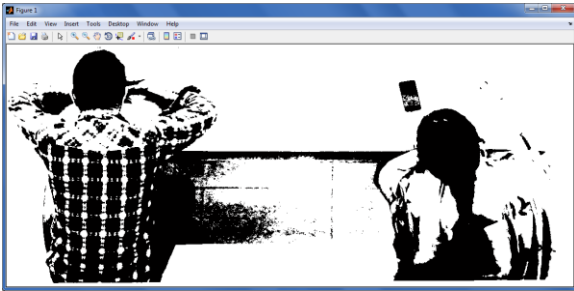


Figure 4.5: Frame after Mobile is detected

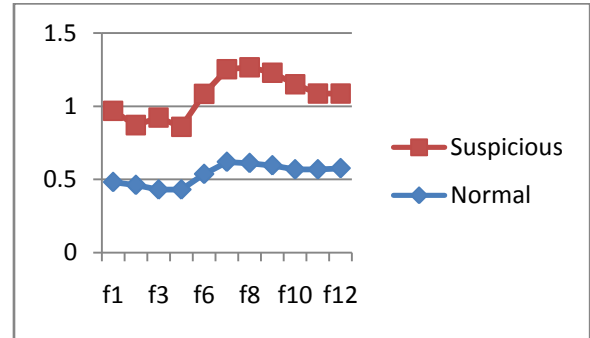


Figure 4.6: Frame of Suspicious activity



Figure 4.7: Frame after skin detection is applied

Threshold of suspicious and normal frames are calculated (figure 4.8) and the differentiation is shown in the form of graph (figure 4.9) as shown below:



Graph 4.1: Differentiation graph

To check the accuracy of the proposed approach SVM classification is done and the corresponding results are shown below .

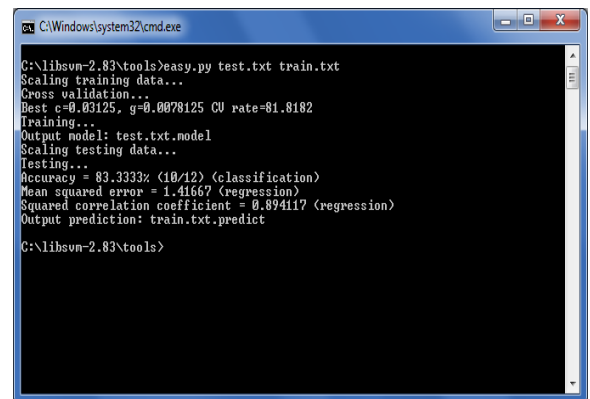


Figure 4.10: Screenshot showing accuracy of results of Object detection.

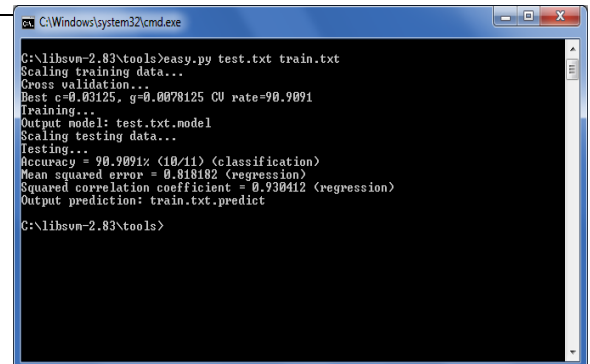


Figure 4.11: Screenshot showing accuracy of results of Skin detection.

	Normal Frame	Threshold	Suspicious Frame	Threshold
4		0.4314		0.4275
5		0.5373		0.5472

Figure 4.8: Threshold calculation

Hence from the above figures (4.10, 4.11), we conclude that our approach achieves 83.3% accuracy for object detection and 90% for skin detection from top view coordinates. This also demonstrates that the proposed approach higher accuracy than the older ones.

5. REFERENCES

- [1] Simone Calderara , Rita Cucchiara , Andrea Prati “Multimedia Surveillance: Content based Retrieval with Multi camera People Tracking”, 2004
- [2] Ovgu Ozturk, Toshihiko Yamasaki, Kiyoharu Aizawa” Tracking of Humans and Estimation of Body/Head Orientation from Top-view Single Camera for Visual Focus of Attention Analysis”, 2011
- [3] R. D. C. Yang and L. Davis. Fast multiple object tracking via a hierarchical particle filter. *Proc. IEEE Intl. Conf. on Computer Vision*, 1:212–219, 2005.
- [4] H. I. D. Glas, T. Miyashita and N. Hagita. Laser tracking of human body motion using adaptive shape modeling. *Proc. IEEE/RSJ Conf. on Intelligent Robots and Systems*, pages 602–608, 2007.
- [5] Jens Rosenkiaer Andersen, Paul Toft Duizer, Dennis Molholm Hansen, Bjarne Kondrup Mortensen “Automatic Annotation of Humans in Surveillance Video Recordings”, 2006
- [6] Simone Calderara, Roberto Vezzani, Andrea Prati, Rita Cucchiara “Entry Edge of Field of View for multi-camera tracking in distributed video surveillance “,2005
- [7] Q. Y. Z. Han and J. Jiao. Online feature evaluation for object tracking using kalman filter. *Proc. IEEE Intl. Conf. on Pattern Recognition*, 2008.
- [8] C. M. L. Snidaro and C. Chiavedale. Video security for ambient intelligence. *IEEE Trans. on Systems, Man, and Cybernetics-part A:Systems and Humans*, 35(1):133–144,2005.
- [9] B. W. L. Zhang and R. Nevatia. Detection and tracking of multiple humans with extensive pose articulation. *Proc. IEEE Intl. Conf. on Computer Vision*, pages 1–8, 2007
- [10] Iveel Jargalsaikhan, CemDirekoglu, Noel E. O’Connor, Alan F. Smeaton “An Information Retrieval Approach to Identifying Infrequent Events in Surveillance Video”,2013
- [11] Iveel Jargalsaikhan, CemDirekoglu, Noel E. O’Connor, Alan F. Smeaton “An Information Retrieval Approach to Identifying Infrequent Events in Surveillance Video”,2013
- [12] J. M. S. Sidla and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.
- [13] P. Sabzmeydani and G. Mori. Detecting pedestrians by learning shapelet features. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1–8, 2007.
- [14] B. Wu and R. Nevatia. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. *Proc. IEEE Intl. Conf. on Computer Vision*, 1:90–97, 2005.
- [15] T. Y. S. L. Y. Li, H. Ai and M. Kawade. Tracking in low frame rate video: a cascaded particle filter with discriminative observers of different life spans. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30(10):1728–1740, 2008.