AJ Theft Prevention Alarm Based Video Summarization Algorithm

Ali Javed and Sidra Noman

Abstract—**The last two decades have witnessed a tremendous growth in the field of computer vision. Video Summarization is one of the hot research area in the field of Computer Vision. Since the quality of videos is increasing day by day and so the requirements to save storage space and transmission bandwidth Researchers have been proposing efficient and intelligent solutions for video summarization. The proposed theft prevention algorithm is an extension of the earlier effort we put towards the objective of proposing an intelligent and efficient algorithm designed for video summarization. Security cameras are widely used for security purposes and to detect any illegal movement inside or outside the campus. These cameras are backbone of modern security infrastructure in any critical or security sensitive organization e.g defense sector, education sector, hospitals etc. The system being developed actually captures each frame from the video, then it processes the frame, if the frame is of its interest, it retains the frames otherwise it discards the frame, hence the resultant video is very short and it only contains those frames in which a human being is carrying an object of our interest, i-e a Laptop, Projector and a PC in our case. Apart from summarization the system also generate an alarm in the form of a beep whenever it captures a frame in which there is a person carrying a specific object which is basically an alert to prompt security officials about an illegal movement in the surroundings.**

Index Terms—**Video Skims, Background Estimation, Bounding Box, Normalization, Convolution**

I. Introduction

The utilization of video cameras is increasing tremendously due to massive reduction in their prices in last decade. Researchers are showing keen interest to develop economical solutions for various problems using video cameras. With the increasing security risks and activities of theft researchers are working to utilize the security cameras also commonly known as CCTV (Close Circuit Television Camera) in the best possible way to hamper the illegal activities to maximum extent.

The primary objective of this project work is to develop an algorithm which basically extract key frames from the video and on the basis of these key frames, compile a new video which only consists of the key frames. Another task accomplished by the system is the generation of an alarm beep whenever it detects an event of a person leaves from the department carrying an object.

One of the main objectives of the system was to keep check on the persons who try to carry important objects like laptop, projector, pc etc, out of the building or campus. The objective was also to generate an alarm so that the concerned security authorities take suitable action immediately.

The proposed video summarization system is very economical to implement and deploy as it only requires a static digital video camera, which is available at very low price. It does not require any additional hardware to implement the system as the remaining task is done by the software component of the system.

Another major benefit of the system is that we don't have to waste time watching complete video; instead video only contains key frames in which there is any illegal movement. Hence lot of time can be saved using the proposed system.

Proposed system is designed to make a summary of the video based on principle of key frame extraction. System utilizes the benefit of decreasing prices of digital cameras in addition it provides a mechanism which saves a person to watch complete video instead of looking at each every part of the video.

The proposed AJ theft prevention algorithm is an extension to earlier work which summarizes the video further by retaining only those frames in which a person leaves the department while holding object like PC, Laptop or Projector.

II. Literature Survey and Existing Systems

Digital video have introduced a new technology competition among it companies and researchers are working to propose low price yet efficient systems in order to put these advancement into practical. Among different research areas of digital video processing, video summarization is considered the most important as it enables to browse and save large videos into smaller ones, saving time as well as space. Video summarization deals with making a small summary of the images which can be either series of frames or a moving video. Different solutions have been proposed by different researchers in this domain of video summarization. One solution is that a video is first divided into segments using segmentation techniques, then from these segments, called frame, the segments of our interests can be detected. Two passes are required by such techniques and they are generally profound. Another technique is classifying the video into clusters and then extracting the key frames from reducing the computational time to a substantial amount as compared to other computation techniques based on clustering.

To apply existing data mining techniques on video data, one of the most important steps is to transform video from

Manuscript received January 1, 2012; revised February 12, 2012.

The authors are with the Department of Software Engineering, Faculty of Telecommunication and Information Engineering, University of Engineering & Technology, Taxila (e-mail: ali.javed@uettaxila.edu.pk).

non-relational data into a relational data set. To facilitate this goal, we adopt a series of algorithms to explore shot detection, key frame extraction, and story segmentation.

Transformation of non relational dataset into a relational dataset is required, in order to apply current data mining techniques on the data that is in the form of a video. Different algorithms are devised in order to perform shot detection, key frame extraction and story detection. In order to make it sure that our video is summarized and concise, we have to make it sure that the summarized video is not having any redundancy or a redundant frame. And all the content is given equal priority and equal part in the summarized video. Key frames don't contain any information regarding the content, so summarization based on the selection of key frame is not suitable and it may not represent the video content uniformly. Users don't care about what technology is adopted for key frame selection neither does he care about individual frames, also not every frame is important, these frames become redundant in the video causing the video summary less efficient or useful or sometime useless. In order to deal with this rendering problem, a clustering algorithm has been devised which actually groups the similar frames. Clustering algorithm is now widely used for video summarization and segmentations. Video retrieval techniques are used in this clustering algorithm which considers the matches between different scenes and contents to generate a summary of key shot frames.

Video summarization approach is now widely being implemented in different organization for diverse purposes. Many vendors also provide different video summarization tools, which can be later on adapted to organizations own needs.

Video summarization approach is now widely being implemented in different organization for diverse purposes. Many vendors also provide different video summarization tools, which can be later on adapted to organizations own needs.

Brief Cam's is company which provides video summarization solution to its customer to allow effective access to recorded surveillance video and enable the end user to find and display any event in only a few minutes.

Brief Cam's product is based on its propriety, patent pending Video Synopsis (VS) technology. Video Synopsis tracks and analyzes moving objects, and converts video streams into a database of objects and activities. When a video summary is needed, all objects from the target period are collected, and are shifted in time to create a much shorter synopsis video showing maximum activity. A synopsis video clip contains the activities and objects executed at diverse points of time are displayed in chorus at real time. (Fig.1). The Brief Cam system comprises a hardware unit that connects to a LAN consisting of cameras and DVR/NVR. The end user can access the Brief Cam summary and index using a software client running on a PC. (Fig. 2). Using the Brief Cam client, the end user can create and view video summaries, as well as index the original video data. The Brief Cam server and client access the DVR/NVR just like any other client.

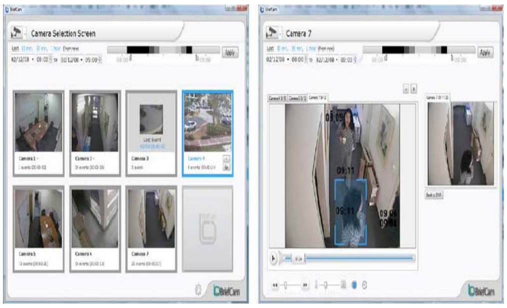


Fig. 1. Brief Cam Screen Capture (Left: Camera Selection Screen, Right: Summary Screen)

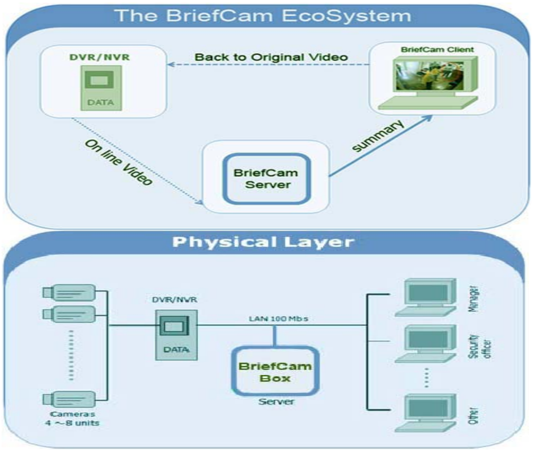


Fig. 2. Physical And Virtual Connection with the network

Many researchers have previously done tremendous work in the domain of human and object detection. Some of the work is reviewed here.

Wren *et al.* [1]-[7]. in his work presented realtimepfinder system for detecting and tracking humans. Agaussian distribution is used by the background model at each pixel in the YUV space, after that the background model is continually updated. Beleznai *et al.* [8].Proposed a multi model probability distribution technique in which the difference of intensities between an input frame and reference image is calculated. Mean shift computation is used to perform mode detection. Haga *et al.* [9].In his paper presented Image motion's spatial uniqueness, human motion's temporal uniqueness and the temporal motion continuity is used to classify a moving object as human Eng *et al.* Reference [10] presented a combination of background subtraction in the form of bottom up approach and a human shape model incorporated by top down approach is presented for multiple overlapping humans and partially occluded human, in this paper. Elzein e*t al.* [11] in his work used Frame differencing perform optic flow which is used to detect moving human. Fixed reference point in the image is to compute a time to collision, by using optic flow capacity in this paper. Toth and Aach [12] Illumination invariant background subtraction is first performed according to method proposed in this paper using an adaptive threshold, frame differencing, window based sum of absolute differences called (SAD) aggregation. Zhou and Hoang [13] proposed a method in this paper which is used to detect and track the body of a human. Firstly, foreground object is first detected using a background subtraction method using consecutive frame temporal differencing. Yoon and Kim [14] describes a human detection approach in a composite form, in which foreground of the candidate is detected using skin color and motion information of the human for human detection, and then objects are classified using more sophisticated techniques.Xu and Fujimura [15] presented in this paper an approach which is supposed to work well in indoor environments, which is basically used to detect the pedestrians. Image information and depth information is simultaneously given using a new sensing device. The part of the image between Dmin and Dmax which is the specified depth values is selected. Walls and other similar background areas are removed using preprocessing. These background objects are large texture less figures making the image present in between Dmin and Dmax of the selected areas.

III. Proposed System

Proposed system is designed to make a summary of the video based on principle of key frame extraction. System utilizes the benefit of decreasing prices of digital cameras in addition it provides a mechanism which saves a person to watch complete video instead of looking at each every part of the video.

A. *System Flow Chart*

The System Flow Chart has been presented in Figure 3.

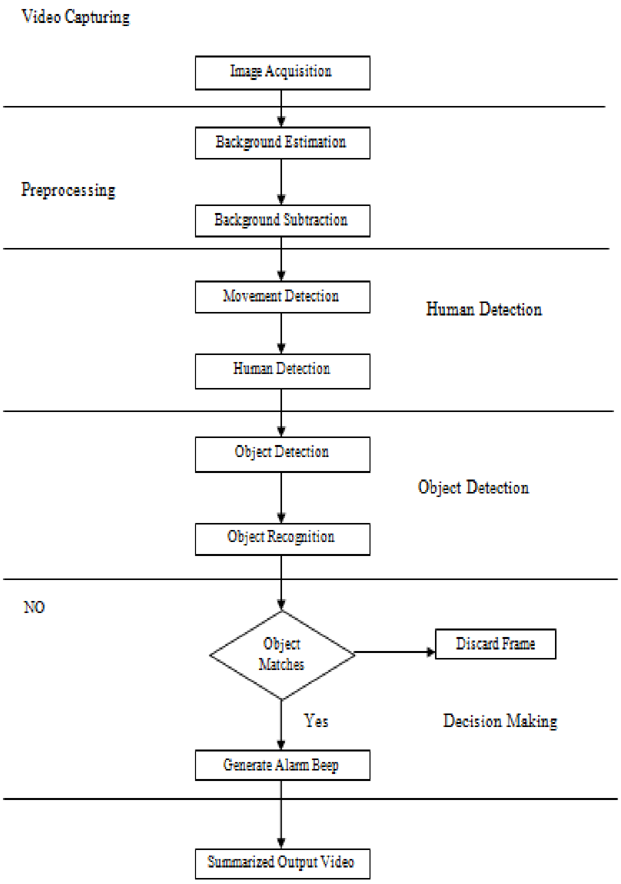


Fig. 3. System Flow Chart

Video recording of every moment inside the department is captured through digital camera to preprocess it before applying algorithm. Preprocessing phase includes two interlinked stages, including background estimation followed by the background subtraction. As a result of preprocessing an image is obtained which only contain objects which are not the part of background making human detection easy. Human detection is the next phase of the system. In this phase dilationtechnique is used for the detection of moving component from the image. As result of dilation, moving human is detected from the image. As the thesis is mainly concerned with the detection of the specific object, on the basis of which the video is summarized, object detection is the next phase of the proposed system. The techniques used for object detection are convolution and normalization. Image is matched with a set of already stored images in the database to find the best match. A bounding box is formed around the detected object, if this is the object of our interest. The last phase of the flow chart is the decision making. Decision making is the simplest phase as it only has to make a decision that if the object is of our interest i-e Laptop, Pc or a projector, it generates the beep and add the frame to summarized video, else it discards it.

The proposed video summarization system capture image of a moving object from the scenario using a static video camera. Proposed video summarization algorithm takes this video as input and applies processing on it. Processing starts with background estimation, background of the scenario is estimated in order to detect moving objects. Background subtraction is used to detect objects which are not the part of background .The technique here is used to detect human from the video. Background subtraction is most popular and effective method for human detection in which frames are subtracted and moving object is detected as foreground. Human tracking is required to capture contents from the region occluded by the human in the previous frames, which have been achieved through bounding box using the technique of dilation. Dilation is applied at 0 and 90 degrees in order to detect human movement and corresponding bounding box is formed. Within the bounding box the techniques of convolution and normalization are applied to generate a graph of the image which serves as the basis of object detection

B. *Proposed Methodology*

1) *Image Acquisition*

The proposed system acquires the input video by using CCTV camera which is capturing video at a resolution of 280 × 180 pixels. The CCTV camera is fixed on the wall and capturing video at a frame rate of 18 frames/sec. To detect moving objectsbecomes quite complicated in the presence of noise, reflections, shadows, illumination conditions. The proposed algorithm is designed to detect objects and human despite all these factors. Two different images in which a person is carrying a PC, acquired from the camera are shown in Figure 4(a) and Figure 4(b)

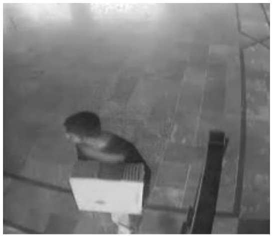


Fig. 4(a). Person carrying an object (First Scenario)

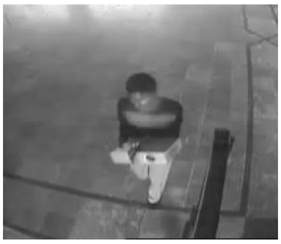


Fig. 4(b). Person carrying an object(Second Scenario)

2) Preprocessing

Preprocessing phase includes two interlinked activities of background estimation and background subtraction.

a) Background Estimation.

Background estimation deals with estimating background of an image. For the purpose of background estimation a reasonable amount of captured frames are processed in order to estimate the background. Background estimation involves two approaches. First approach deals with capturing the background only, this is the simplest case for background estimation, as there is no other object in the image and whole frame is considered as the background. A single frame is sufficient for back ground estimation in such case. The other approach deals with the detection of the static objects in the image. A reasonable amount of frames are required for this approach as there may be a movement later in the video, so large number of frames are captured to make it sure that image contains only the background and all the moving objects are discarded. Proposed system estimates the background after capturing every 63 frames and takes the median value of pixels to form the background image. Estimated background from original image is shown in figure 5.

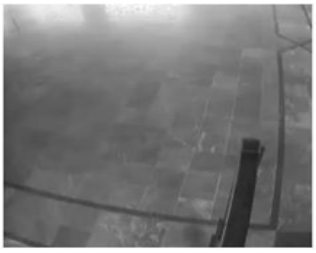


Fig. 5. Result of background estimation

b) Background Subtraction

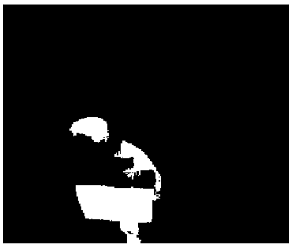


Fig. 6(a). Background Subtraction (First Scenario)

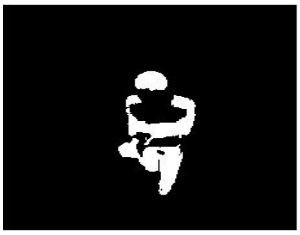


Fig. 6(b). Background Subtraction (Second Scenario)

Background estimation is followed by background subtraction in every image processing algorithm. Background subtraction is the most commonly used technique for object detection. It deals with subtracting background from the image in order to detect object components from the image. Background subtraction is done in pixels domain, where process is applied pixel by pixel. In proposed system pixel by pixel background subtraction is done with a tolerance 35. By subtracting Figure 5 from Figure 4(a) and Figure 4(b) yields Figure 6(a) and Figure 6(b).

c) Motion and Human Detection

Image obtained as a result of background subtraction only contains moving object which is human in this case. Technique of dilation is further applied in order to enhance the object so that human can be easily detected, as we can see as a result of background subtraction, there are certain areas which are not completely filled in the subtracted image. Dilation is applied in order to enhance or enlarge the edges in to fill these areas and make the object detectable as a human. Proposed system applies dilation in 45 and 90 degrees.

Images obtained as a result of dilation are shown in figure 7(a) and figure 7(b).

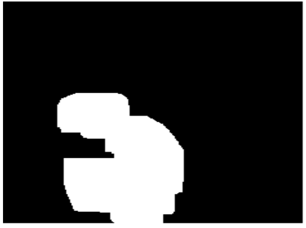


Fig. 7(a). Dilated image (First Scenario)

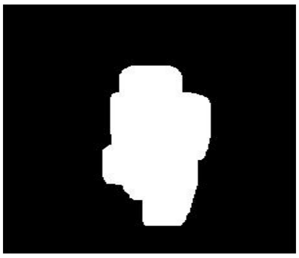


Fig. 7(b). Dilated Image(Second Scenario)

After dilated image contains complete object including theareas which were missing after background subtraction. As the object is human, and human must be carrying an object with him, so instead of further processing the complete frame, we are only concerned with the object. A bounding box is formed around human and further processing is done only inside the bounding box, making algorithm fast and efficient. An example of how bounding box is formed around a human is shown in Figure 8.

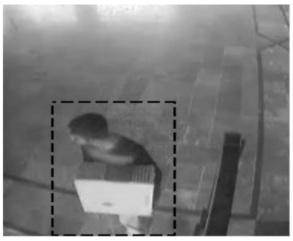


Fig. 8. Bonding box after dilation

d) Object Detection and Recognition

After the detection of human and formation of bounding box, next step is to detect the presence of an object. Whether human is carrying an object with him or not. For the object detection purpose the techniques of convolution and normalization are used.

e) Convolution and Normalization

Proposed system generates a graph as a result of applying convolution and normalization to the bounding box region. This graph serves as basis for object detection. Graph shows different variation between different intensity value as in Figure9 Double mean of the graph values is calculated. If there is a certain difference between the maximum graph value and the double mean it shows that there is some sort of object present within the bounding box, due to which there is an abrupt variation in the graph. In the proposed system the value of double mean should be between 0.95 to 1.00 of the maximumgraph value. If the value of double mean is less than 0.95, there is a very strong possibility of the presence of object.

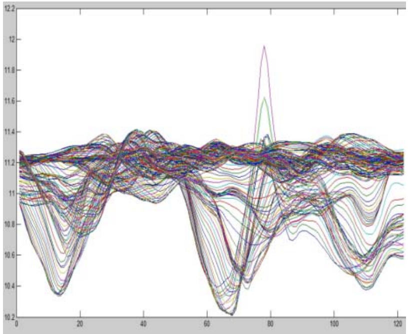


Fig. 9. Convolution Graph

A bounding box is formed around the object using these convolution values, leaving us with two bounding boxes, one around the human and the other bounding box around the object as shown in Figure 10(a) and 10(b). Now the scope of processing is further limited as we have to only deal with the inner box now, making the system even more efficient.

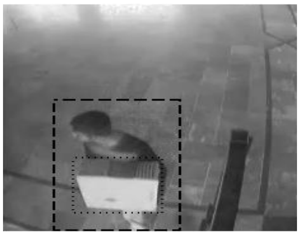


Fig. 10(a). Object Detection from convolution (1st Scene)

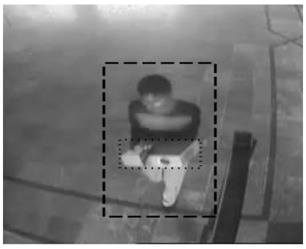


Fig. 10(b). Object Detection from convolution (2ndScene)

f) Decision Making

Proposed system has a very simple decision making algorithm, based on feature matching. A set of images of PC's, Laptops and Projectors (Object of interests) from different angles are stored in a database. Object detected as a result of convolution in the bounding box is matched with the database values using the technique of feature matching. If there is a match between the bounding box object and the defined objects like PC, laptops and projectors after feature matching with objects features stored in the dataset, proposed system generates an alarm beep as well as it stored the frame in the resultant summarized video leaving us with a very short video as compared to the large original video. The decision making algorithm is displayed in figure 11.

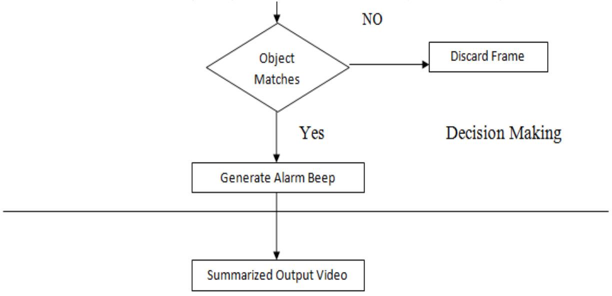


Fig. 11. Decision Making Algorithm

IV. Experimental Setup and Final Result

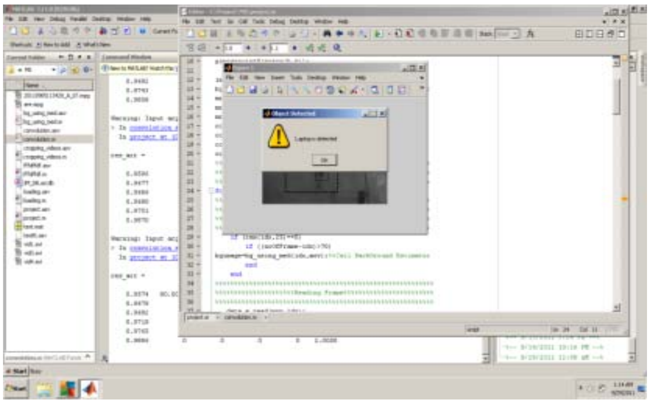
Fig. 12(a). Left Side view of CCTV Camera Setup

Fig. 12 (b). Right Side view of CCTV Camera Setup

Fig. 12 (c). Front view of CCTV Camera Setup

The proposed system is designed to operate with a CCTV camera which is located above the main entry/exit gate of the department. Video sequence is captured from the camera located above the main gate. Camera can capture image within the range of 10 to 12 meters in the area in front of the gate. The proposed system setup is shown in figure 12. The camera is placed right above the main entry/exit gate with the help of a hanger, setup is very similar to close circuit TV camera (CCTV). Figure 12 demonstrate different views of the proposed system setup.

Jist of the proposed system is to generate an alarm message as soon as a person is detected having a projector, laptop or a pc carrying with him. The alarm generation has been generated accompanied by a Dialog box displays the warning message as shown in the figure 13.


Object Detected

Laptop is detected

OK

Fig. 13. Alarm generation and Message

REFERENCES

[1] Y. T. Zhuang, Y. Rui, T. S. Huang, and S. Mehrotra,“Key frame extraction usingunsupervisedclustering,” in *Proceedings of the IEEE International Conference onImage Processing*, Chicago, US, pp. 866-870, 1998.

[2] D. Besiris, F. Fotopoulou, N. Laskaris, and G.Economou, “Key frame extraction in video sequences: advantage points approach,” *IEEE 9th Workshop onMultimedia Signal Processing (MMSP), Greece*, pp.434-437 [2007].

[3] G. Ciocca and R. Schettini, “Supervised andunsupervised classification post-processing for visualvideo summaries”, *IEEE Transactions on ConsumerElectronics*, Vol. 52, No. 2, pp. 630-638, 2006.

[4] X. D. Sun and M. S. Kankanhalli, “Videosummarization using R-sequences,” *Real-Time Imaging* vol. 6, no. 6, pp. 449-459, 2000.

[5] X. Zhu, X. Wu, and J. Fan,“*Exploring* video content structure for hierarchial summarization,” MultimediaSystem, *IEEE Inernational symposium on Electronics Commerce and security* vol. 10, no. 2, pp. 98-115, 2004.

[6] S. Pfeiffer, R. Lienart, S. Fisher, and W. Effelsberg, “Abstracting Digital MoviesAutomatically,” *Journal of Visual Communication and Image Representation*, vol.7, no. 4, pp. 345-353, 1996.

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

[8] Beleznai, B. Fruhstuck, and H. Bischof, “Human detection in groups using a fast mean shift procedure,” *International Conference on Image Processing*, 1:349–352, 2004.

[9] T. Haga, K. Sumi, and Y. Yagi, “Human detection in outdoor sceneusingspatio-temporal motion analysis,” *International Conference on Pattern Recognition*, 4:331–334, 2004.

[10] H. Eng, J. Wang, A. Kam, and W. Yau, “Abayesian framework for robust human detection and occlusion handling using a human shape model,” *International Conference on Pattern Recognition*, 2004.

[11] H. Elzein, S. Lakshmanan, and P. Watta. A motion and shapebasedpedestrian detection algorithm, *IEEE Intelligent Vehicles Symposium*, pages 500–504, 2003.

[12] D. Toth and T. Aach, “Detection and recognition of moving objects using statistical motion detection and fourier descriptors,” *International Conference on Image Analysis and Processing*, pages 430–435, 2003.

[13] J. Zhou and J. Hoang, “Real time robust human detection and tracking system,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3:149 – 149, 2005.

[14] S. M. Yoon and H. Kim, “Real-time multiple people detection using skin color, motion and appearance information,” *International Workshop on Robot and Human Interactive Communication,* pages 331–334, 2004.

[15] F. Xu and K. Fujimura, “Human detection using depthand gray images,” *IEEE Conference on Advanced Video and Signal Based Surveillance*, pages 115–121, 2003.

Engr. Ali Javed has been a lecturer since April 2008 in the Department of Software Engineering, University of Engineering and Technology Taxila, Pakistan. He accomplished his M.Sc in Computer engineering from university of Engineering and Technology Taxila, Pakistan in February, 2010. He graduated from University of Engineering and Technology taxila in Software Engineering in September, 2007. His areas of interest are Video Summarization, Digital Image Processing, Computer vision, Software Quality Assurance, Software testing andSoftware Requirements Analysis.

Engr. Sidra Noman is an MS Scholar in Software Engineering Department, at University of Engineering and Technology Taxila, Pakistan. She graduated from Fatima Jinnah Women University in Software Engineering in 2004. Her core areas of interest are Digital Image Processing, Computer vision and Software Engineering.