AUTOMATED VIDEO SURVEILLANCE SYSTEM FOR HUMAN MOTION DETECTION

[1] HARDIK H JOSHI, [2] PROF. N. N. SHAH, [3] PROF. P. D. RAVAL

[1, 2] Department Of Electronics & Communication Engineering, R. K. University, Rajkot, Gujarat, India.
[3] Department Of Electrical Engineering, Government Engineering College, Rajkot, Gujarat, India.

hardikjoshi_hhj@yahoo.com, nishit.shah@rku.ac.in, pranavraval13@yahoo.co.in

ABSTRACT : *This paper presents a survey of various techniques related to video surveillance system. The aim of this paper is to review of various methods for motion detection as well as face detection. This paper focuses on moving object detection and then it will identify the human entity in the video sequence. The basic task of surveillance system is detecting and tracking moving object in the video sequence. Video surveillance is an important security asset to control theft, trespassing or traffic monitoring, banks, department stores, highways, crowded public places and borders.*

1. INTRODUCTION

Automated video surveillance system is most important in the field of security. The task of surveillance is to detect and track moving objects in the video sequence. Now-a-days, video surveillance system is an important security asset to control theft, trespassing or traffic monitoring, banks, department stores, highways, crowded public places, borders etc. In video surveillance, detection of moving objects from a video is necessary for object classification, target tracking, activity recognition as well as behaviour understanding.

In last one-two decades there are lots of work and research is doing in the area of video surveillance system with various applications. Prithviraj Banerjee and Somnath Sengupta [1] present the video surveillance for human motion detection and tracking. Their system also concentrates on tracking an object in motion as well as classifying it as a human or non-human entity. It is a novel combination of an Adaptive background modelling algorithm (which is based on the Gaussian Mixture Model) and a human detection for surveillance i.e. HDS system. The HDS system deals with a Histogram of Oriented Gradients based human detector which is famous for its performance in detecting humans in still images. The foreground is extracted from video sequence by learning a statistical model of the background and subtracting it from the original frame [1]. Background model deals with stationary part only and it ignore the moving foreground. This system deals with Gaussian mixture model for modelling the background adaptively. So motion regions are identified in the video frame, which consist the region of interest (ROI) for HDS system. The Histogram of Oriented Gradients algorithm is applied on the ROI to detect which category of object is present in the ROI. There are many categories present as ROI like human figure, a vehicle or an animal.

As we know the original image is a combination of stationary part and motion part. The static background subtraction algorithm obtain statistical background image and this background image is subtracted from the current frame of image and then threshold. After performing appropriate morphological operations on thresholded image we are getting foreground regions of interest. For HDS system, they are following Histogram of Oriented Gradient for human detection which is created by Navneet Dalal and Bill Triggs. Image window divide into small spatial regions which is called cells. In that each cell accumulated a local 1-D Histogram of Oriented Gradient directions or edge orientations of the pixel values in the cell [1]**.** To check out whether the ROI contains human figure or not SVM (Support Vector Machine) is useful. Linear SVM's are useful for baseline classifier because their performance is good relative to other linear classifiers and running fast. The HDS system makes a decision at each frame for a particular object being tracked [1]. After certain number of frames, object being tracked. The HDS systems compute majority decision taken by analysing 100 frames and then declare it as a human or non-human entity.

2. DISCUSSIONS OF VARIOUS METHODS

In A survey on Moving object Detection and Tracking in video surveillance system [2] by Kinjal A Joshi and Darshak G Thakore presented in his paper that how moving object detected and tracked through various method. Heikkila and Silven [3] present background subtraction with Alpha method. In that,

when system starts at that time after passing first few frames from video sequence reference background initialized and then it is updated to adapt dynamic change. The threshold and reference background are updated through foreground pixel information. It attempts to detect moving regions by subtracting the current image pixel by pixel from a reference background image that is created by averaging images over time in an initialized period [3]. The differences of pixels higher than threshold are known as foreground. After that with the use of some morphological operations like erosion, dilation are performed for reducing noise effect as well as extracting the detected regions.

Another method is used, to overcome the shortcoming of the basic background methods is called Statistical method [2]. This method is inspired by background subtraction method. The statistical methods' example is an adaptive background mixture modelled by a mixture of Gaussians which was described by Stauffer and Grimson [4]. To extract moving regions, take the pixel wise differences between two or more consecutive frames, this method is known as Temporal differencing. It is highly adaptive approach to dynamic scene changes however, it fails to extract all relevant pixels of a foreground object especially when the object has uniform texture or moves slowly [5] . The limitations of temporal differencing method is when foreground object does not move at that time the change between consecutive frames is fail. Eigen background subtraction [6] method is proposed by Oliver et al. In that, Eigen space model presented for segmentation of moving object with the help of Principle Component Analysis (PCA), the dimensionality of the space which is generated by sample images is reduced. It is proposed by that the reduced space after PCA should present only the static part of the scene, remaining moving objects, if an image is projected on the space [2]. Chris Stauffer and W. E. L. Grimson [7] described the adaptive background mixture model in his paper. Their goal is to develop a robust, adaptive tracking system which is flexible when there is moving scene clutter, variations in lighting, multiple moving objects etc. changes during the scene observed.

In online mixture model, assume the values of a particular pixel over time as a "pixel process". Consider a particular pixel, at time t, {x0, y0}, is its history

$$\{X1,.....,Xt\} = \{I(x0,y0,i) ; 1 \leq i \leq t \}$$

Where I is the image sequence [7]. If pixel process could be considered a stationary process, a standard method for maximizing the likelihood of the observed data is expectation maximization [8]. But unfortunately, each pixel process varies with time so approximate method is needful which is consider sample size 1 for each new observation and to integrate new data uses standard learning rules. The great advantages of this method are when anything becomes a part of background then it does not destroy existing model of background. The method uses an approximation to Gaussian mixture modeling to describe the recent history of color and depth scene observation at each pixel, is described by Michael Harville, Gaile Gordon and John Woodfill in paper [9]. The input to the algorithm is a time series of spatially registered, time-synchronized pairs of color and depth images obtained by static cameras [9]. Each pair of corresponding pixels in the two images for a given time step provides one scene observation in the combined input space of color and depth [9]. Nahum Kiryati, Shay Rochel et al. present a novel real-time abnormal motion detection scheme [10]. This algorithm uses the macro block motion vectors that are generated anyway as part of standard video compression method [11]. Improbable motions indicate abnormal motion at the time of real time operation.

A method for motion detection in the given video sequence is Foreground extraction using background subtraction. As we know, image is a combination of stationary part and motion (moving) part. The part which is stable during the video sequence is called stationary part. For instance: background, while the moving part is not stationary. With the use of Static background subtraction algorithm, we get original image (which contain motion as well as stationary part) and estimated background image (which is having only stationary part) and after that subtracted original image from estimated background image operation gives motion part only. In other words, we can say that foreground regions of interest are extracted through various morphological operations like Thresholding, Erosion, Dilation etc.

Prithviraj Banerjee and Somnath Sengupta [1] proposed the overview of surveillance system as shown in figure 1. In that, the image is acquired from camera which is consisting stationary and motion part. With the use of Adaptive background modelling algorithm background subtracted and foreground extracted. Histograms of Oriented Gradients for human detection detect the human motion and then it can classify as a human or non-human entity through SVM (Support Vector Machine).

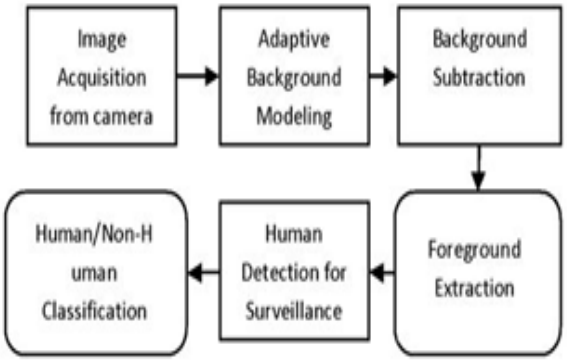


Figure 1: Surveillance System Overview [1]

Steps involved in HDS System [12]:

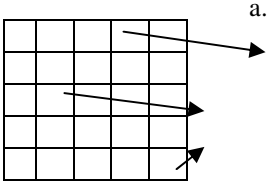
Histogram for Oriented Gradients (HOG) for human detection algorithm described by Navneet Dalal and Bill Triggs. According to that, the image window is divided into small spatial regions which is called "cell".

a. With the use of Adaptive background modelling bounding box generate and then bounding box resized to a dimension 128x64 pixel size.
b. Gradients in the image computed using simple 1-D mask. There are three types of gradients taken: Horizontal Gradients, Vertical Gradients and L1 norm Gradients.
c. Image is divided into cell which is a multiple regions of fixed rectangular size as shown in figure 2.

Histogram of Gradient directions is computed. [A to I Histogram bins with angular difference of 20°.] [12].

d. 9 component feature vectors are generated for the entire selected regions.
e. 128x64 pixel image consisting of 7 blocks width wise and 15 blocks height wise and its generating 3780 dimension feature vector.
f. SVM (Support Vector Machine) classifier decides that the regions of interest having a human figure or not.

The system architecture which is described below is understood by Farhan S. Khan and Salman A. Baset in their paper.



a. The end-user performs the system initialization using System Initialization Agent (SIA) at the beginning. SIA stores as well as identify the current position, band colors and also status of hands-legs in the history. SIA divides initial frame into square blocks of size. With the use of Range based mode filter each blocks is consider as a moving object or background.
b. Object Detection Agent (ODA) is locate the object in the region given by Proximity Calculation Agent (PCA). With the use of Range mode filter each block in region and labels it as a part of object or background.
c. If object is detected successfully then their coordinates of unified blocks are stored in history and classification agent will classify new position. If the object is not identifying then Self-Occlusion Detection Agent (SODA) initiates the unidentified object's position.
d. When Motion Detection Agent fails to identify the object, self-occlusion agent is initialized.
e. Another agent is Classification Agent (CA), which is identifying the current position of a hand/leg. Moreover it also identifies whether the change in state compared to previous frame has occurred or not.

Neural network is also useful in the field of Video surveillance system. Alberto Amato et al [13] focused their paper in this area and with this the proposed system is detected mobile objects as well as identifies their movements. With the use of Alarm detection whenever unacceptable movements are detected, alarm will raise. The process consisting the steps that first of all cameras sends a frame then signature extracted from frame. The neural network store the knowledge about it as "known" or "unknown" entity and according to that it classify and if "unknown" entity classify then system raised an alarm. The current frame is compared with the centroids, which is having background knowledge. If the current frame signature is having minimum distance from centroid then the frame is considered as a similar. But the distance is more than experimentally-defined threshold, and then frame is not similar compared to centroid so alarm is raised and with that we can know that something suspicious may occurred.

Brajesh Patel and Neelam Patel [14] discuss various algorithm related to frame based motion detection in surveillance system. The action of sensing a physical movement in the video sequence/area is called motion detection. With the use of it we can also identifies the movement of an object within two or more successive frames. Motion detection can be achieved by Mechanical form or Electronics form. In mechanical form of motion detection, a tripwire, this is a simple form of motion detection. If a moving object steps into the tripwire's field of view (i.e. trips the wire), then a simple sound device (e.g. bells) may alert the user [14]. While in electronic form of motion detection, motion detection have sensors which detect the movement of an object and according to that it send signals to sound device which produce an alarm or switch on to the image recording device. In that method, they approach three stages for video frame which are: Tracking Step, Detection Step and Validation Step. In tracking step, the objects which have been previously identified are tracked to find their position and shape within the current video frame [14]. Same time motion of these objects is

estimated. In detection Step, the new moving objects are detected and their shape as well as motion is estimated. This step also consisting creation of new hypotheses regarding new moving objects. In final step i.e. validation step, if any of the hypotheses are deemed valid, then we have identified a new moving object at this frame and this will now be tracked through subsequent frames [14].

Face recognition in video surveillance is described by Christophe Pagano et al [15] in their paper. A generic system for video-based face recognition works like first it takes images from camera and then after segmentation Region Of Interest (ROI) identifies. The features vectors for extraction and tracking is completed and the biometric data base is decided the face recognition. This is an example of a system which is a combination of spatial and temporal computations. They also proposed a modular multi-classifier system (MCS) for accurate recognition of individuals in video-to-video surveillance applications. It is composed of a long-term memory (LTM), an ensemble of binary 2-class classifiers or detectors (EoDs) per individual and a dynamic multi-objective optimization module.

4. CONCLUSION

This paper gives the basic idea about various techniques used for motion detection in the field of video surveillance system. Video surveillance system is very useful in the field of security. With the use of motion detection we are tracking moving object and identify its activity. Extraction of information due to surveillance provides us object classification, define the difference between vehicles and humans, or between animals and humans, human identification. Some useful applications like to define human activities in video sequence which also includes animals or vehicles. Another application like track the vehicles as well as human simultaneously. The main feature of all of above is to ability to track the particular objects.

5. REFERENCES

[1] Prithviraj Banerjee and Somnath Sengupta. Human motion detection and tracking for video surveillance. Available at http://www scf.usc.edu/~pbanerje/Homepage/HMD. pdf.

[2] Kinjal A Joshi and Darshak G Thakore, "A survey on moving object detectionand tracking in video surveillance system", IJSCE, ISSN: 2231-2307, volume-2, issue-3,july 2012.

[3] Elgammal, A. Duraiswami, R., Hairwood, D., Anddavis, L. 2002. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. Proceedings of IEEE 90, 7, 1151–1163.

[4] S. Zhu, and A. Yuille. Region competition: unifying snakes, region growing, and bayes/mdl for multiband image segmentation. IEEE Trans. Patt. Analy. Mach. Intell. 18, 9, 884–900, 1996.

[5] N. Paragios, and R. Deriche.. Geodesic active contours and level sets for the detection and tracking of moving objects. IEEE Trans. Patt. Analy. Mach. Intell. 22, 3, 266–280, 2000.

[6] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In IEEE International Conference on Computer Vision . 694–699, 1995.

[7] C. Stauffer and W. Grimson. Adaptive background mixture models for real time tracking. In Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, page 246-252, 1999.

[8] A Dempster, N. Laird, and D. Rubin. "Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal Statistical Society, 39 (Series B):1-38, 1977.

[9] Michael Harville, Gaile Gordon and John Woodfill, "Foreground segmentation using adaptive mixture models in color and depth", IEEE workshop on Detection and Recognition of Events in Video, 3-11, 2001.

[10] Kiryati, N.; Raviv, T.R.; Ivanchenko, Y.; Rochel, S., "Real-time Abnormal Motion Detection in Surveillance Video", International conference on Pattern Recognition, 1-4, 2008.

[11] V. Bhaskaran and K. Konstantinides, Image and Video Compression Standards: Algorithms and Architectures, Kluwer, 1997.

[12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", Proceedings of the Conference on Computer Vision and Pattern Recognition, San Diego, California, USA, pp. 886–893, 2005.

[13] Alberto Amato, Vincenzo Di Lecce, Vincenzo Piuri, "Neural network based video surveillance system", IEEE conference on Computational Intelligence for Homeland Security and Personal Safety, Orlando, FL,USA, page 85-89, 2005.

[14] Brajesh Patel and Neelam Patel, "Frame based motion detection for real-time surveillance", International Journal Computer Technology and Applications, ISSN 2229-6093, Volume-3, Issue-2, page 650-655, 2012.

[15] Christophe Pagano, Eric Granger, Robert Sabourin and Dmitry O. Gorodnichy, "Detector Ensembles for Face Recognition in Video Surveillance", International Joint Conference on Neural Networks (IJCNN), Brisbane, QLD, 2012.