VIDEO SURVEILLANCE SYSTEM: A REVIEW

Garima Sharma *

ABSTRACT

Video surveillance is gaining its important in almost every field of day to day life. Surveillance is being done not only in military areas or airports but also in offices, schools, shopping areas, old age home and many more areas. The primary purpose of this paper is to provide a general review on the overall process of a surveillance system used in the present time. The processing framework of the video surveillance system includes the following stages: moving object detection, object segmentation, representation, classification, tracking of objects, activity recognition and prediction.

Keywords: *Video Surveillance, Background Subtraction, Objects Classification, Blobs, Tracking.*

* UIET, Kurukshetra University, Kurukshetra, Haryana.

I. INTRODUCTION TO VIDEO SURVEILLANCE

There has been an accelerated expansion of CCTV surveillance in recent years mainly due to rising crime and its threat to security and safety. Video surveillance is defined as the task of analyzing video sequences to detect abnormal and unusual activities. Unusual events are defined as abnormal behavior patterns that are not represented by sufficient samples in a training data set. An unusual activity is defined as one which is seldom or rarely occurs or has not been observed before [10]. The goal of automatic behavior profiling is to define a model that is capable of distinguishing between normal and abnormal behavior patterns [1].

The effectiveness of a behavior profiling algorithm is determined by how effectively and accurately an anomaly is detected [1]. The set up for an automated visual surveillance system includes the following steps: moving object detection, object classification, object tracking, activity analysis, and anomaly detection.

II. NEED OF VIDEO SURVEILLANCE

Some of the areas where video surveillance system is deployed are:

- Patrolling of highways and railways for accident detection.
- Surveillance of forests for fire detection
- Extracting statistics for sports activities
- For military security
- Patrolling of country borders
- Keeping security check in shopping malls, naval bases
- Monitoring of old age home, nursing homes, schools.
- Tracking of activities in hospitals, parking areas.

A video surveillance system works as follows [7]:

- Detects moving objects
- Classifies the detected objects
- Tracks them through the sequence of images
- Analyze their behaviors

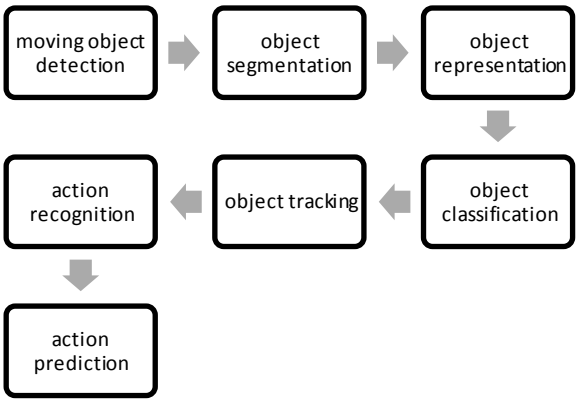


Fig 1: Video Surveillance System Block Diagram

III. MOVING OBJECT DETECTION

Almost all the video surveillance system starts with the moving object detection step. A number of surveillance systems have been proposed in the recent years. Different surveillance techniques are used in every system for e.g. Pfinder [18] tracks the whole body of person but it considers only single person in the scene. Pfinder uses unimodal background model to locate the objects. Stauffer and Grimson [15] use multimodal background subtraction method to locate the objects. This method is capable of dealing with slow changes in the background like illumination changes and changes in the scene.

In any scene considered there can be many objects present in it so it is necessary to classify them according to the priority and need. It handles segmentation of moving objects from stationary background objects. Segmentation of the video is first done to a sequence of image frames, which aims at segmenting regions corresponding to moving objects as compared with the rest of the image.

Fig 2: Background model without and with a foreground object

Motion detection step generally involves two main parts background modeling and motion segmentation. Background modeling involves the differentiation between the background and the foreground objects. Foreground means the part of the image that is nearer to and in front of the viewer. Background modeling creates a reference image representing the background. Motion segmentation involves the extraction of moving objects from stationary background.

Different techniques are used for the motion segmentation such as background subtraction, temporal differencing and optical flow.

A. Background subtraction in this method the current frame is compared with a reference frame [9]. Pixels in the current frame that deviate from the background are considered to be moving objects or 'foreground'. Background subtraction is basically done to determine the background state of each pixel.

In this method the current frame is subtracted from a reference frame, if the difference between the two is greater than a set threshold value (Th), the pixel is then considered as a part of foreground else it is taken into the background category.

Mathematically defining background subtraction as:

$$|frame_i - frame_{i-1}| > Th \qquad (1)$$

Background subtraction method has a main advantage compared to the other methods that it easily adapts to the changing background. This method is very sensitive to the threshold value so it must be set accordingly.

Fig 3: Result of background subtraction

B. Temporal differencing makes use of pixel wise differences between two or more consecutive frames in an image sequence to extract the foreground objects [17]. Temporal differencing has an advantage that it changes with the changing environment but the disadvantage is that it does not recognize moving objects. Temporal differencing mathematically is defined as [17] :

$$\Delta_n = |I_n - I_{n-1}| \qquad (2)$$

C. Optical flow turns (bends) one image into another image and describes the motion of a point or feature between images. Mostly background subtraction method is used to extract the objects from the scene.

IV. OBJECT SEGMENTATION

Subtraction operation classifies the type of the pixel i.e. the pixel is the part of background or foreground. Foreground pixels are then segmented into regions using connected component

method. Connected component analysis (CCA) is done to cluster and label foreground pixels into object blobs from which some appearances and motion features can be extracted.

According to [4], connectivity means that pixels having same intensity values and adjacent to each other are grouped together. Commonly 4-pixel connectivity (four pixels touching to form a cross) or 8-connectivity (eight pixels touching to form a square around the pixel in question) is used.

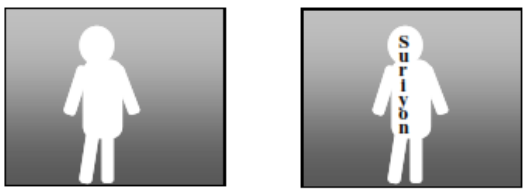
The connected regions are classified into object categories; human, human group and vehicle. The process of grouping similar regions into homogeneous regions called as blobs is called as object segmentation method.

V. OBJECT REPRESENTATION

The blobs created by CCA method prove quite useful in the classification and tracking steps. But the blobs occupy large pixel area and the memory to store and display them due to their formation, so it is necessary to compact the blobs to represent the objects.

For object representation mostly two representations are used quiet often: silhouette based method and bounding boxes.

***A. Silhouette*:** it is defined as the image of the person consisting of an outline. Object silhouettes are used when the shape or posture of the object needs to be identified. Generally edge detection techniques like Sobel or Canny or Kirch edge detection methods are applied to foreground region to extract the silhouette of the region. According to [11] silhouettes are used to hide a person's appearance to protect person's privacy.



(a) Silhouette display(SD) (b) Silhouette display with name (SN)

Fig 4: Silhouette display for protecting person's privacy [11].

B. Bounding box: it is a box with smallest measure (area, volume) within which all the points lie. A bounding box gives the system measures for the height; width and center of blob which can be used in object classification and tracking.

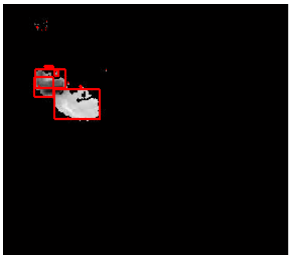


Fig 5: Bounding box on the selected object

VI. OBJECT CLASSIFICATION

Different regions of the image correspond to different moving objects in an image. It is necessary to detect, track and analyze their behavior, so the need of object classification arises. Object classification distinguishes between the different objects present in the image into predefined classes such as human, vehicle, animal, clutter, etc.

Two approaches are employed for the object classification:

A. Shape based classification: mostly used feature in this technique are bounding rectangle, points, area, silhouette and gradient of detected object region. The moving object blobs are classified into four classes' namely human, human groups, vehicles and clutter using three layer neural network classifier [14]. The input to the neural network are the dispersedness, area and aspect ratio of the object region and camera zoom magnification.

According to [17] a classification metric based on knowledge is done according to which humans are smaller than the vehicles and they have more complex shapes. Dispersedness [17] is used as a classification metric which is defined as

$$\text{Dispersedness} = \frac{perimeter^2}{area} \qquad (3)$$

B. Motion based classification: this is done to classify the moving objects as rigid (humans) or non rigid (vehicles) based on their appearances in multiple frames [16]. The method proposed in [13] is based on temporal self similarity of moving object. The moving object is characterized according to its periodic motion.

Classification Metrics

Object classification tries to identify the types of objects that are present in the environment. Different classification measures are:

A. Size metric: used when the scene has objects of different sizes for e.g. to distinguish between a bag and a human [3] [11].

B. Speed metric: used when the objects in the environment have different speed for e.g. pedestrians and vehicle speed vary [5].

C. Dispersedness: this metric is also used to differentiate between different size objects [10].

Fig 6: Classification Metric based on the size of the object

VII. OBJECT TRACKING

Object tracking is defined as segmenting an object of interest from the image and keeping track of its location, motion, and occlusion. Tracking creates time limitations or boundaries on the objects detected from frame to frame. The output of the tracking step is object trajectory information which is used to calculate direction and speed of the objects in the scene.

The purpose of tracking a foreground object or blob in an environment is to determine what kind of action a blob is performing and from the present action it becomes easy to predict the next action.

The targets are represented in 2D space and 3D space by some features [7]. Some of these features are:

- Size: this is determined by the minor and major axis of the ellipse
- Position: this is determined by the centroid of the target
- Velocity: the target centroid velocity is determined with the help of kalman filtering.
- Target type: this feature depends on the size, shape of the target. It is determined with the help of Bayesian network.
- Trajectory: trajectory of the target is determined with the help of comparing the position of target in the previous frames.

Object matching

The aim of the object tracking is to establish a relation between the object or object parts in consecutive frames and to extract time based information about the objects such as trajectory, position, speed and direction.

The most important step in tracking of object is to recognize that the same blob is being tracked in the consecutive frames. Then the blob location and the distance moved from the

last frame are calculated. Data record is generated for each object and labels are created. Kalman filter can be used for the labeling of objects [5]. For each new frame positions are compared with previous frames. If matching object is found then the object is considered successfully tracked and if not matched then the object is considered as new.

Different methods of object matching are used:

A. Proximity based: this method keeps track of the blob position in each frame and calculates the displacement measured by the blob. A bounding box centroid is calculated for each blob to measure the displacement. The displacement of the bounding box centroid is measured by Euclidian distance between the consecutive frames [2]. If the distance is below a certain threshold value then it is assumed that the same blob has been matched and tracking is continued.

B. Predictive based: in this technique some characteristics of the blob are predicted like speed, direction, location. Kalman filter can be used to predict an object's location or speed in the next frame [5].

W^4 [12] is also a predictive technique to estimate where the centroid of the blob would be located in subsequent frames in order to track and match the blobs.

C. Using blob characteristics: several blob characteristics like size, speed or direction of movement of blob can also be used for matching and tracking of the objects. For e.g. if a person is moving away then the size of the blob will decrease continuously in the successive frames.

Many systems for target tracking are based on other methods. [17] Uses a method which combines temporal differencing and template matching for tracking of objects. This combination method provides a good tracking performance even in presence of clutter. In temporal differencing frames are compared to find out the changes occurred over the course of time. In template matching image frame is compared with the image template.

VIII. ACTION/ BEHAVIOR RECOGNITION

The final step of an automated surveillance system is to recognize the behavior of the objects and create a high level semantic description of their activities. Behavior is defined as the order of events, with or without the time constraint. With the increasing video surveillance it is necessary to find out what all is happening in the environment not only in the security areas but also in hospitals, schools, nursing homes, offices etc. The action recognition system is based on the behavior of the blobs according to which the action can be classified as simple

or complex. A good tracking system is necessary to recognize the actions correctly and accurately which helps to determine what kind of action the blob is performing.

The action recognition is divided into two groups:

A. Simple actions: these are the actions that can be directly analyzed by blob tracking data.

B. Complex actions: these cannot be determined by simple analysis of tracking data. Complex actions are made up of many simple actions. Detecting interaction between different blobs is a complex task. Detecting an abandoned luggage in a crowded area is also an example of complex action recognition task [3] [6]. This is done by detecting the size of the blob of the left luggage.

IX. ACTION PREDICTION

Acc to [8] normal behavior is recognized by regular movement, speed, stopping etc whereas abnormal behavior comprise of opposite movement in the crowd, fighting, deviation from the normal motion. The crowd behavior can be predicted depending on the crowd motion information like direction, velocity. Prediction has been used by several researchers including [3] [12] for predicting the location or speed of an object in the next frame or time step for tracking purposes. Action prediction is a difficult thing to do. For e.g. the weather predictions made by the meteorological department are incorrect many times.

X. CONCLUSION

Video surveillance system can be classified in a number of areas considering the environment in which they operate: indoor, outdoor or airborne. There are a number of challenges associated with the chosen scenario like repetitive background, illumination changes, occlusions, shadows, highlights etc. the existing methodology employs detection of the moving objects, segmentation, object representation, classification, tracking, behavior recognition and prediction. Although a large amount of work has already been done in the video surveillance field but many areas are still open. Video surveillance is a vast and important area of research in the fields like anomaly detection and alarming, person to person identification etc.

REFERENCES

1. Tao Xiang and Shaogang Gong, “Video Behavior Profiling for Anomaly Detection”, IEEE transactions on pattern analysis and machine intelligence, vol. 30, no. 5, may 2008
2. Euclidean distance Retrieved July 1, 2007, from Wikipedia, the free encyclopedia: http://en.wikipedia.org/wiki/Euclidean_distance

3. Grabner, H., Roth, P. M., Grabner, M. and Bischof, H. 2006. Autonomous Learning of a Robust Background Model for Change Detection. In Proceeding of the 9th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (New York, USA, June 18, 2006). PETS 2006. IEEE Computer Society, Los Alamitos, CA.
4. Fisher, R., Perkins, S., Walker, A. and Wolfart, E. Connected Components Labeling. Retrieved June 8, 2006, From Image Analysis- Connected Components: http://homepages.inf.ed.ac.uk/rbf/HIPR2/label.htm.
5. Bodor, R., Bennett, J. and Papanikolopoulos, N. 2006. Vision-Based Human Tracking and Activity Recognition. Retrieved April 4, 2006, from Artifical Intelligence, Robotics and Vision Laboratory, Department of Computer Science and Engineering, University of Minnesota: http://mha.cs.umn.edu/Papers/Vision_Tracking_Recognition.pdf.
6. Smith, K., P. Quelhas, P. and Gatica-Perez, D. 2006. Detecting Abandoned Luggage Items in a public Space. In Proceeding of the 9th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (New York, USA, June 18, 2006). PETS 2006. IEEE Computer Society, Los Alamitos, CA, 75-82.
7. Pankaj Kumar, "Framework for Real-Time Behavior Interpretation From Traffic Video", IEEE transactions on intelligent transportation systems, vol. 6, no. 1, March 2005.
8. Shobhit Saxena and Fran¸cois Br´emond, Monnique Thonnat, "Crowd Behavior Recognition for Video Surveillance", Department of Computer Sc. and Engg, Indian Institute of Technology, Hauz Khas, New Delhi, India, 2004
9. Background subtraction techniques: a review Massimo Piccardi, Computer Vision Group, Faculty of Information Technology, University of Technology, Sydney (UTS), Australia, 2004 IEEE International Conference on Systems, Man and Cybernetics.
10. Hua Zhong, Jianbo Shi and Mirk´o Visontai "Detecting Unusual Activity in Video", Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)
11. Tansuriyavong, S. and Hanaki, S. 2001. Privacy protection by concealing persons in circumstantial video image. In Proceedings of the 2001 Workshop on Perceptive User interfaces (Orlando, Florida, November 15 - 16, 2001). PUI '01, vol. 15. ACM Press, New York, NY, 1-4.

12. Haritaoglu, I., Harwood, D., and David, L. S. 2000. W4: Real-Time Surveillance of People and Their Activities. IEEE Trans. Pattern Analysis Machine Intelligence, vol. 22, no. 8 (Aug. 2000), 809-830.
13. R. Cutler and L.S. Davis. Robust real-time periodic motion detection, analysis and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8:781– 796, 2000.
14. R. T. Collins et al. A system for video surveillance and monitoring: VSAM final report. Technical report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May 2000.
15. 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, pages 246-252, 1999.
16. L. Wixson and A. Selinger. Classifying moving objects as rigid or non-rigid. In Proc. Of DARPA Image Understanding Workshop, pages 341–358, 1998.
17. Lipton, A. J., Fujiyoshi, H. and Patil, R. S. "Moving target classification and tracking from real-time video", In Proc. IEEE Workshop on Applications of Computer Vision, (Princeton, NJ, USA, October 1998). IEEE Computer Society, Los Alamitos, CA.
18. Wren, C. R., Azarbayejani, A., Darrell, T., and Pentland, A. P. 1997. Pfinder: Real-Time Tracking of the Human Body. IEEE Trans. Pattern Anal. Mach. Intell. vol. 19, no. 7, (July, 1997). IEEE Computer Society, Los Alamitos, CA, 780-785.