

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

Survey on Human Detection Techniques in Real Time Video

Paresh M Tank¹, Hitesh A Patel²

Lecturer, Department of Computer Engineering, B & B Institute of Technology, V.V.Nagar, Gujarat, India¹

Lecturer, Department of Computer Engineering, B & B Institute of Technology, V.V.Nagar, Gujarat, India²

ABSTRACT: Video surveillance has been in use to monitor security sensitive areas such as banks, department stores, highways, crowded public places and borders. The advance in computing power, availability of large-capacity storage devices and high speed network infrastructure paved the way for cheaper, multi sensor video surveillance systems. Automated surveillance systems are of critical importance for the field of security. The task of reliably detecting and tracking moving objects in surveillance video, which forms a basis for higher level intelligence applications, has not fully solved and has many open questions. Segmentation, detection, and tracking of multiple objects of a known class are a fundamental problem in computer vision. This Survey focuses on detection and tracking of human in real time monocular video captured for the purpose of visual surveillance. Contemporary research work for the last decade has been thoroughly reviewed in this paper. This review has been intended for the monocular video captured online from still camera having moderate to complex background. Camera may be located in indoor as well as outdoor environment.

KEYWORDS: Video Surveillance, Human Detection, Human Tracking, Computer Vision.

I. INTRODUCTION

Understanding activities of objects moving in a scene by the use of video is both a challenging scientific problem and a very interesting domain with many promising applications. Thus, it draws attentions of several researchers, institutions and commercial companies [1]. Traditionally, the video outputs are processed online by human operators and are usually saved to tapes for later use only after a forensic event. The increase in the number of cameras in ordinary surveillance systems overloaded both the human operators and the storage devices with high volumes of data and made it infeasible to ensure proper monitoring of sensitive areas for long times. In order to filter out redundant information generated by an array of cameras, and increase the response time to forensic events, assisting the human operators with identification of important events in video by the use of “smart” video surveillance systems has become a critical requirement. The making of video surveillance systems “smart” requires fast, reliable and robust algorithms for moving object detection, classification, tracking and activity analysis.

Below are list of the some applications that smart surveillance systems and algorithms might handle.

Monitoring of banks, department stores, airports, museums, stations, private properties and parking lots, for crime prevention and detection. Patrolling of highways and railways for accident detection, Surveillance of properties and forests for fire detection. Observation of the activities of elderly and infirm people for early alarms and measuring effectiveness of medical treatments. Access control, Measuring traffic flow, pedestrian congestion and athletic performance, compiling consumer demographics in shopping centers and amusement parks. Extracting statistics from sport activities. Counting endangered species, Logging routine, maintenance tasks at nuclear and industrial facilities, Artistic performance evaluation and self learning, Measuring speed of vehicles, Detecting red light crossings and unnecessary lane occupation, Patrolling national borders, Measuring flow of refugees, Monitoring peace treaties, Assisting battlefield command and control.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

The use of smart object detection, tracking and classification algorithms are not limited to video surveillance only. Other application domains also benefit from the advances in the research on these algorithms. Some examples are virtual reality, video compression, human machine interface, augmented reality, video editing and multimedia databases. Our motivation in studying this problem is to create a visual surveillance system with real-time moving human detection, classification, and tracking and activity analysis capabilities.

Person detection and tracking has many applications in field of computer vision. Some of these are listed below.

- Visual security and surveillance
- Driving assistance system
- Human computer interaction
- Scene analysis and activity recognition
- Event detection
- Interpretation of video and logical inference
- Video annotations

In video surveillance system, complete algorithm can be divided into hierarchy of the four stages. In first stage input video is segmented and all moving object presents in video are detected. This first stage may be preceded by some fine ground level noise removal process. In second stage, all detected objects are classified into various classes as per requirement of the system. In third stage, all object of interest, which has been detected in second stage are tracked and its trajectories are followed by the system. Last part of the algorithm may utilize information received from third stage and forms higher level process to make logical inference.

II. CHALLENGES/ ISSUES

The most difficulty in building a robust system for human detection and tracking is the amount of variation in the videos and several other factors contribute to this as given below [2].

- Device creates dependencies on viewpoint such that even a small change in the object's position or orientation with respect to the camera centre may change its appearance considerably. An object detector system must **handle** the issues of viewpoint and scale changes.
- Most natural object classes have large within-class variations. For example, for humans both appearance and pose change considerably between images and differences in clothing create further changes
- Background clutter is common and varies from image to image. Examples are images taken in natural settings, outdoor scenes in cities and indoor environments. The system must be capable of distinguishing object class from complex background regions.
- Object color and general illumination varies considerably, for example direct sunlight and shadows during the day to artificial or dim lighting at night.
- Partial occlusions create further difficulties because only part of the object is visible for processing.

III. MOVING OBJECT DETECTION TECHNIQUES

Introduction

Each application with smart video processing has different needs but have something in common that is moving objects. Thus, detecting regions that correspond to moving objects such as people and vehicles in video is the first basic step of almost every vision system since it provides a focus of attention and simplifies the processing on subsequent analysis steps. Due to dynamic changes in natural scenes such as sudden illumination and weather changes, repetitive motions that cause clutter (tree leaves moving in blowing wind), motion detection is a difficult problem to process. Most widely used techniques for moving object detection are background modeling, statistical methods, temporal differencing and direct segmentation on each frame, whose descriptions are given below.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

Temporal Differencing

Temporal differencing attempts to detect moving regions by making use of the pixel-by-pixel difference of consecutive frames (two or three) in a video sequence. This method is highly adaptive to dynamic scene changes; however it generally fails in detecting whole relevant pixels of some types of moving objects. For the mono colored region of the object the temporal differencing algorithm fail in extracting all pixels of the object's moving region. Also, this method fails to detect stopped objects in the scene. Additional methods need to be adopted in order to detect stopped objects for the success of higher level processing. Lipton et al. presented a two-frame differencing scheme where the pixels that satisfy the are marked as foreground.

In order to overcome shortcomings of two frame differencing in some cases, three frame differencing can be used. For instance, Collins et al. developed a hybrid method that combines three-frame differencing with an adaptive background subtraction model [6]. The hybrid algorithm successfully segments moving regions in video without the defects of temporal differencing and background subtraction.

Background Subtraction

Background subtraction is particularly a commonly used technique for motion segmentation in static scenes. 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 initialization period. The pixels where the difference is above a threshold are classified as foreground. After creating a foreground pixel map, some morphological post processing operations such as erosion, dilation and closing are performed to reduce the effects of noise and enhance the detected regions. The reference background is updated with new images over time to adapt to dynamic scene changes. There are different approaches to this basic scheme of background subtraction in terms of foreground region detection, background maintenance and post processing.

In [7] Heikkila and Silven uses the simple version of this scheme where a pixel at location (x, y) in the current image It is marked as foreground if it is satisfied condition where T is a predefined threshold.

The foreground pixel map creation is followed by morphological closing and the elimination of small-sized regions. Although background subtraction techniques perform well at extracting most of the relevant pixels of moving regions even they stop. They are usually sensitive to dynamic changes when, for instance, stationary objects uncover the background or sudden illumination changes occur.

Background Modeling (Subtraction)

Background subtraction is a very popular conventional method. Various researchers have used this method for detection purposes. This conventional method is widely used in tracking systems and especially in surveillance applications. The name "background subtraction" comes from the simple technique that subtracts all information of an image and finally thresholds the information to generate the object of interest [2]. This method is typically used to segment moving regions in image sequences taken from a static camera. The changes or moving region can be detected by simply taking the absolute difference by comparing each new frame to a model of the scene background.

Mixture of Gaussian

Up until the late 90s background subtraction was known as a powerful preprocessing step but only in controlled indoor environments. In 1998, Stauffer and Grimson [8] presented the idea of representing each pixel by a mixture of Gaussians (MoG) and updating each pixel with new Gaussians during run-time. This allows background subtraction to be used in outdoor environments. Normally the updating was done recursively, which can model slow changes in a scene, but not rapid changes like clouds.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

In this technique, the values of a particular pixel are calculated as a mixture of Gaussians. Based on persistence and the variance of each of the Gaussians of the mixture, the background and foreground are determined. Each pixel values that do not fit to the background distributions are considered as foreground. This technique is an interesting technique and has been used in the background subtraction approach. In a real time indoor and outdoor tracker, Stauffer and Grimson have employed an adaptive Gaussian mixture model in order to evaluate and determine the background of an image [8]. Each pixel is classified based on a Gaussian distribution to consider a part of the background model. The result has proven that this technique is reliable in dealing with lighting changes, repetitive motions from clutter.

Eigen Background

Oliver, et al. [10] has proposed an eigenspace model for moving object segmentation. In this method, dimensionality of the space constructed from sample images is reduced by using Principal Component Analysis (PCA). Their claim is that, after the application of PCA, the reduced space will represent only the static parts of the scene, yielding moving objects, if an image is projected on this space. Although the method has some success in certain applications, it cannot model dynamic scenes completely. Hence, it is not very suitable especially for outdoor surveillance tasks. The main steps of the algorithm can be summarized as follows [11]:

A sample of N images of the scene is obtained, mean background image, μ_b , is calculated and mean normalized images are arranged as the columns of a matrix, A .

The covariance matrix, $C=AAT$, is computed.

Using the covariance matrix C , the diagonal matrix of its eigenvalue, L , and the eigenvector matrix, Φ , is computed.

The M eigenvectors, having the largest eigenvalues (eigenbackgrounds), is retained and these vectors form the background model for the scene.

If a new frame, I , arrives it is first projected onto the space spanned by M eigenvectors and the reconstructed frame I' is obtained by using the projection coefficients and the eigenvectors.

The difference $I - I'$ is computed. Since the subspace formed by the eigenvectors well represents only the static parts of the scene, outcome of the difference will be the desired change mask including the moving objects.

Non Parametric Kernel Density Estimation

Kernel density estimation is another important technique in background subtraction. This technique is applied to solve the non parametric problem for tracking systems. Generally, kernel density estimation is a nonparametric technique for density estimation in which a known density function (the kernel) is averaged across the observed data points to create a smooth approximation. In [9], Elgammal et al has applied kernel density estimation to handle situations where the scene background is not completely static, but contains small motions such as moving tree branches and bushes or illumination changes. On the other hand, it is able to suppress false detections that arise due to small camera displacements.

Direct Segmentation on Frame

In addition to various techniques explained above, we can directly utilize some versatile image segmentation techniques available in literature like Mean shift and Normalized cut.

Mean Shift is a powerful and versatile non parametric iterative algorithm that can be used for lot of purposes like finding modes, clustering, analysis of multi modal feature space and to delineate arbitrary shaped contour in it etc. Mean Shift was introduced in Fukunaga and Hostetler [13] and has been extended to be applicable in other fields like Computer Vision. If the input is a set of points then Mean shift considers them as sampled from the underlying probability density function. If dense regions (or clusters) are present in the feature space, then they correspond to the

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

mode (or local maxima) of the probability density function [12]. We can also identify clusters associated with the given mode using Mean Shift.

Mean shift can be used to segment images or here frames of video. Mean shift segments image in two steps [12]. In first step discontinuities preserving smoothing are done in feature space using mean shift mode seeking procedure. This first step smooths the image without perturbing edges in image. In second step, all the clusters residing close to each another in some known parameters are grouped together and form the segment in image.

Normalized cut is a versatile technique, which is suitable for large class of problem related to image segmentation. In this technique, image segmentation problem treated as graph partitioning problem. To partition a graph various techniques are available with varying demerits. Shi and Malik [14] propose a modified cost function, normalized cut, to partition a graph into two parts. They compute the cut cost as a fraction of the total edge connections to all the nodes in the graph and called this disassociation measure the normalized cut (Ncut):

IV. HUMAN DETECTION TECHNIQUES

Introduction

After video segmentation, we have to identify the class of object in which we are interested. The object classification stage receives a list and size of the blobs, which are likely to contain a human. In this stage, they are classified as human or nonhuman aiming, with the goal of minimizing the number of false positives and false negatives. Detecting a human in video is current area of the research. In last decade many researcher has developed techniques to detect a human. These techniques are summarized below.

For human detection, some widely used features include Haar wavelet, HOG, shapelet, edge orientation histogram (EOH), Edgelet, region covariance, and LBP [15]. Researchers also explore various combinations of these features to achieve good results.

Histogram of Oriented Gradient

Dalal et al [4] presented a novel system for recognizing humans in images using Histograms of Oriented Gradients. The algorithm is based on evaluating normalized local histograms of image gradient orientation in a dense grid. The method tries to characterize local object appearance and shape by a general idea of the distribution of local intensity gradients or edge directions.

It is well known that HOG (Histogram of Oriented Gradients) is quite a popular method of detecting people in static image. However, it can't be used in video due to its slow computing of features. It needs to be improved if we use it in video. The HOG feature describes the distribution of gradient magnitude and gradient orientation in the local area of the image. First of all, it extracts the histograms of oriented gradient feature from local image, through training sample, and then classifies the object by classifiers in pattern recognition in order to achieve human detection. It is unique that HOG features describe the distribution of gradient strength and gradient direction in local area, which is a good characterization of appearance and shape. It is proved that it can be well used in human detection, and it has become a mainstream approach.

Edge Oriented Histogram and Oriented LBP

Yingdong Ma at al [15] used a fast method to detect human in video based on the edge oriented histogram and a novel oriented local binary patterns features. They have concluded combination of these features captures salient features of humans and together with rejection cascade achieves an efficient and accurate human detection system.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

They have integrated the EOH feature with an oriented LBP feature. The EOH feature extracts gradient orientation based shape information whereas oriented LBP features records local intensity based information. They have demonstrated their method outperform some up-to-date approaches.

Geometry feature based detection

Pablo JP at al [17] has proposed a new approach to detect a human in video sequences based on geometry features and AdaBoost learning. In that work they calculate geometrical features on detected silhouette part by part. so they have claimed their approach efficiently applicable for partially occluded humans and groups of peoples. They have utilized change detection techniques to boost the algorithm.

To achieve a fast, yet efficient, change detection they proposed an approach that focuses on the regions with the highest occupancy, aiming to process only the regions that interest the application (i.e., where the presumed pedestrians are likely to be located). They achieve that by concentrating in the areas of the image where the change density is higher, i.e., where a large number of pixels has changed. This aims to exclude regions with relatively small changes due to non-stationary backgrounds (i.e., swaying trees, etc.). Therefore, it allows filtering most noise and image artifacts while focusing on potential humans.

Edgelet

Bo Wu at al [18] proposed part based detection approach. Their detection method consists of two stages. In first stage they detect a part and then their combination. Detector learns from novel set of silhouette oriented features that they call Edgelet features. They claimed these features are suitable for human detection because it is relatively invariant to clothing differences unlike color features.

In the second stage of detection, they combines the result of various part detectors. Performance of the combined detector is better than that of any individual part detector. They mentioned advantages of these methods are: handles partial occlusion, less false alarms, more tolerant to point changes and pose variations of articulated objects.

V. CONCLUSION

In this paper, ongoing contemporary work on human detection in stationary video has been thoroughly surveyed. All methods have their own pros and cons which have been discussed.

REFERENCES

1. L.Wang, W. Hu, and T. Tan. Recent developments in human motion analysis. *Pattern Recognition*, 36(3):585–601, March 2003.
2. Junzo Watada, Zalili Musa, Lakhmi C. Jain, and John Fulcher. Human Tracking: A State-of-Art Survey. *KES 2010, Part II, LNAI 6277*, pp. 454–463, 2010.
3. David Gero´ nimo, Antonio M. Lo´ pez, Angel D. Sappa. Survey of Pedestrian Detection for Advanced Driver Assistance System. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, VOL. 32, NO. 7, JULY 2010
4. Piotr Dollár, Christian Wojek, Bernt Schiele, and Pietro Perona. Pedestrian Detection: An Evaluation of the State of the Art. *SUBMISSION TO IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*. 2010
5. J. Lipton, H. Fujiyoshi, and R.S. Patil. Moving target classification and tracking from real-time video. In *Proc. of Workshop Applications of Computer Vision*, pages 129–136, 1998.
6. 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.
7. J. Heikkila and O. Silven. A real-time system for monitoring of cyclists and pedestrians. In *Proc. of Second IEEE Workshop on Visual Surveillance*, pages 74–81, Fort Collins, Colorado, June 1999.
8. Stauffer, W.E.L. Grimson, Adaptive background mixture models for real-time tracking, in: *Computer Vision and Pattern Recognition*, Santa Barbara, CA, June 1998
9. Elgammal, A.M., Harwood, D., Davis, L.S.: Non-parametric Model for Background Subtraction. In: Vernon, D. (ed.) *ECCV 2000*. LNCS, vol. 1843, pp. 751–767. Springer, Heidelberg (2000)

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

10. Nuria M. Oliver, Barbara Rosario, and Alex P. Pentland, A Bayesian Computer Vision System for Modeling Human Interactions. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 22, NO. 8, AUGUST 2000
11. Piccardi, M. "Background subtraction techniques: a review." Systems, Man and Cybernetics, 2004 IEEE International Conference, Vol 4, pp.3099, 2004,
12. Dorin Comaniciu and Peter Meer, "Mean Shift : A Robust approach towards feature space analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 24, No 5, May 2002.
13. Fukunaga and Hostetler, "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition", IEEE Transactions on Information Theory vol 21 , pp 32-40 ,1975
14. J Shi and J Malik, "Normalized cut and image segmentation", vol.22, NO. 8, IEEE Transactions on Pattern Analysis and Machine Intelligence,, IEEE, 2000.
15. Yingdong Ma, Xiankai Chen, Liu Jin, and George Chen. A Monocular Human Detection System Based on EOH and Oriented LBP Features, ISVC 2011, Part I, LNCS 6938, pp. 551–562, 2011
16. N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," In IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 886 – 893, 2005.
17. Pablo Julian Pedrocca and Mohand Saïd Allili, Real-Time People Detection in Videos Using Geometrical Features and Adaptive Boosting. ICIAR 2011, Part I, LNCS 6753, pp. 314–324, 2011
18. BO WU AND RAM NEVATIA, Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet based Part Detectors, International Journal of Computer Vision,2007