A Survey of Automatic Surveillance Systems for Detection of Person in Static Background and Dynamic Foreground

Mr. Sagar V. Bakale[1], Dr. Prashant V. Ingole[2]
Department of Electronics & Telecommunication,
GHRCEAM, Amravati
sagarbakale@gmail.com[1], prashant.ingole@raisoni.net[2]

Abstract- In computer vision technology, visual surveillance especially for humans and vehicles is currently one of the most active research topics. It has a promising application in human identification and anomalous behaviors. Generalized framework of visual surveillance in dynamic scenes includes the following stages: environment and its modeling, detection of motion, classification of moving objects, tracking, understanding and description of behaviors, human identification. In this paper the review recent developments and general strategies of all these stages are clarified. Finally, it analysis over possible research directions, e.g. the person detection is the first step to person identification and accordingly the work has been proceeded.

Index Terms-Motion Detection; Personal Identification; Tracking and Visual Surveillance.

1. INTRODUCTION

Visual surveillance in dynamic scenes attempts to detect, recognize and track certain objects from image sequences, and most probably to understand and describe object behaviors. The aim is to develop intelligent visual surveillance to replace the traditional video surveillance that is providing ineffective as the number of cameras exceeds the capability of human operators to monitor them. In short, the goal of visual surveillance is not only to put cameras in the place of human eyes, but also to accomplish the entire surveillance system fully automatic. It has wide applications, such as traffic surveillance, security guard for important buildings, etc. In this paper, applications involving the surveillance of peoples or vehicles which include the full range of surveillance methods are induced.

2. MOTION DETECTION

Mostly every visual surveillance system starts with motion detection. In Motion detection main aim is region segmenting corresponding to moving objects. Tracking and behavior recognition are greatly dependent on motion detection. The process of motion detection usually involves environment modeling, motion segmentation, and object classification.

A. Environment Modeling

The active construction and updating of environmental models are indispensable to visual surveillance. Fuzzy Extreme Learning Machine for Single Hidden layer Feed- forward neural networks (SLFNs) automatically find the threshold value for the given video sequence to detect the moving object, which is appropriate for Dynamic Environment [33]. For fixed cameras, the key problem is to automatically recover and update background images from a dynamic sequence. Unfavorable conditions, such as illumination variance, shadows and shaking branches, bring many difficulties for acquiring and updating the background images.

B. Motion Segmentation

Motion segmentation in frame sequences aims at detecting regions corresponding to moving objects such as vehicles and humans. Detecting moving regions provides a focus of attention for later processes such as tracking and behavior analysis because only these regions are consider in the later process. At present, most segmentation methods use either temporal or spatial information in the image sequence. The dynamic neural-fuzzy approach is used for segmentation of moving objects in dynamic backgrounds [1]. Segmentation is carried out by fuzzy c-means algorithm which allows the feature vector to have multiple membership grades to multiple clusters [4]. K means clustering is a method of cluster analysis which partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean [7]. Learning Algorithm Segmentation in this weights and biases are updated using the data obtained in the pre-processing. This process is repeated as many times as necessary in order to have a trained network [9]. The recursive flooding method detects clusters by "flooding" the "landscape" given by the Clusot surface [20]. Detaching the moving objects using fuzzy logic inference system which fuses multiple sources of information together for decision making and create a fuzzy silhouette by α-level cut [23]. Several conventional approaches for motion segmentation are outlined in the following.

1) ***Background subtraction:***Background subtraction is a popular method for motion segmentation, relatively in static background. It detects moving

regions in an image by taking the difference between the current image and the reference background image in a pixel-by-pixel fashion.The stopped foreground subtraction (SFS) algorithm is implemented for an incoming pixel in sequence frame, adopting the stopped layering mechanism with *L* layers which is used for constructing the moving foreground model [5]. Multi-scale image feature maps of color, intensity, and orientation are extracted, and local spatial contrast is estimated for each feature at each location, providing a separate conspicuity map for each feature. Such maps are combined to a single topographical saliency map that guided the attention focus in a bottom-up manner [6]. The Least Median of Squares method is used to construct the background which is insensitive to changes of color and texture of clothes therefore use binary silhouette [18]. The background subtraction and update procedureis done by SOM which find the Best Match to Current Sample and employ the Euclidean distance of vectors in the HSV color hex cone that gives the distance between two pixels [12]. The background model initialization presented in this study is depending on stationary pixel intensity value is brightness value which has the highest redundancy ratio on intensity values taken from a training sequence. Each intensity values taken by the pixel location *x* in the training sequence is compared with the intensity values taken hitherto [16]. An NN architecture used as a Bayesian classifier, based on the Parzen estimation to store the model of the background within its weights [14].

2) ***Temporal differencing:***Temporal differencing makes use of the pixel-wise differences between two or three consecutive frames in an image sequence to extract moving regions. Temporal differencing is very adaptive to dynamic environments, but generally does a poor job of extracting all the relevant pixels, e.g., there may be holes left inside moving entities. Temporal Median Filter compiles the temporal median value of the pixels using the long-term timer and the short-term timer [8].
3) ***Optical flow:***Optical-flow-based motion segmentation uses characteristics of flow vectors of moving objects over time to detect moving regions in an image sequence. The displacement vector field to initialize a contour based tracking algorithm, called active rays, for the extraction of articulated objects. Image Segmentation is done by FCM Clustering Algorithm. Successful results have been reported for image segmentation using such algorithm [32].

C. Object Classification

Different moving regions may correspond to different moving targets in natural scenes. For instance, the image sequences captured by surveillance cameras mounted in road traffic scenes probably include humans, vehicles and other moving objects such as animals and birds etc. At present, there are two main categories of approaches for classifying moving objects.

1) ***Shape-based classification:***Different descriptions of shape information of motion regions such as points, boxes, silhouettes and blobs are available for classifying moving objects. A pixel is classified according to bin size in the histogram and its Intensity channel of the model [3]. The approach for moving cast shadow detection is proved to be quite accurate and suitable for moving object detection [12]. The probabilistic segmentation algorithm is to classify the pixels such as intensity or red, green, or blue (RGB) color components can be used as basis for segmentation [14]. Human motion analysis is concerned with detecting periodic motion signifying a human gait and acquiring descriptions of human body pose over time [25].
2) ***Motion-based classification:***In general, non-rigid articulated human motion shows a periodic property, which can be used for classification of moving objects. The Multilayer Perceptron NN to perform both object classification and scene understanding [29]. Data points lying near each other in the input space are mapped onto nearby map units. Distances between and all the prototype vectors are computed which best matching unit (BMU) [21]. An FFT-based implementation of cross-correlation is used that computes the correlation at all pixel displacements between the two templates [22].

3. OBJECT TRACKING

After motion detection, surveillance systems generally track moving objects from one frame to another in an image sequence. During processing the tracking algorithms usually have considerable intersection with motion detection. Tracking involves matching objects in consecutive frames using features such as points, lines or blobs. Tracking methods are divided into four major categories: region-based tracking, active-contour-based tracking, feature based tracking, and model-based tracking.

A. Region-Based Tracking

Region-based tracking algorithms track objects based on variations of the image regions corresponding to the moving objects. The background image is maintained dynamically and motion regions are detected by subtracting the background from the current image. The use of small blob features to track a single human in an indoor environment. In their work, a human body is considered as a combination of some blobs respectively representing various body parts such as head, torso and the four limbs. The proposed Semantic Modelling technique is creating the road model which is used for obstacle detection. E.g. histogram has been used for creating the road model [3]. The obtained self-organizing neural network is organized as a 3-D grid of neurons, producing a representation of training samples with lower dimensionality, with the same time preserving topological neighborhood relations of the input patterns [5].

B. Active Contour-Based Tracking

Active contour-based tracking algorithms track objects by representing their outlines as bounding contours and updating these contours dynamically in successive frames. The TNNs to make navigation safer by detecting the contour of different possible objects [9].The contour projection analysis with shape analysis is used to remove the shadow effect [35].

C. Feature-Based Tracking

Feature-based tracking algorithms perform recognition and tracking of objects by extracting elements, clustering them into higher level features and then matching the features between images [13]. The Spatio-temporal interest points (Mo-SIFT) is used to select the object of interest (OBI), as horizontal stride and vertical distance are extracted from the binary images by tracking its motion [2]. Semantic concept of motion describes a class of behaviors for object tracking [17]. The winner neuron and all the neurons in its neighborhood may adapt their codebook vectors for object tracking [20]. Feature representation by 2D silhouette image can be converted to a 1D vector with the same dimension in a row-scan manner for trajectory comparing [28]. With the external points and angle between them this tracking algorithm works. Thick and thin white lines separately represent their trajectories produced by tracking algorithm [16]. A pixel surrounded by foreground labels should receive a foreground label than a pixel with background neighbors which can be accomplished by Markov random field [11].The features used in global feature-based algorithms include centroids, perimeters, areas, some orders of quadrature's and colors. A person is bounded with a rectangular box whose centroid is selected as the feature for tracking. The features used in local feature-based algorithms include line segments, curve segments, and corner vertices, etc. Position table for each index control points are present which have six different points for lip, 6 control points for right eye, 6 control points for left eye Bezier curve, left eye height and width, lip height and width and also right eye height and width which examines the action of the people [27].

D. Model-Based Tracking

Model-based tracking algorithms track objects by matching projected object models, produced with prior knowledge, to image data. The basic ICA model describes how the observed mixture signals are generated by a process that uses the mixing matrix to mix the latent source signals and get matched object [10]. The basic idea in model-based approach consists of keeping a model of foreground objects and classifying as stopped objects those whose model holds the same features for several consecutive frames; remaining foreground objects are consequently classified as moving objects [5].

4. PERSONAL IDENTIFICATION FOR VISUAL SURVEILLANCE

The problem of "who is now entering the area under surveillance" is of increasing importance for visual surveillance. Such personal identification can be treated as a special behavior-understanding problem. Human face and gait are now regarded as the main biometric features that can be used for personal identification in visual surveillance systems.

A. Model-Based Methods

In model-based methods, parameters are to be measured such as joint trajectories, limb lengths, and angular speeds.

B. Statistical Methods

Statistical recognition techniques usually characterize the statistical description of motion image set which will be developed in automatic gait recognition. The purpose of PCA training is to obtain several principal components to represent the original gait features from a high-dimensional measurement space to a low-dimensional Eigen space [18].

C. Physical-Parameter-Based Methods

Physical-parameter-based methods make use of geometric structural properties of a human body to characterize a person's gait pattern. The parameters used include height, weight, stride cadence and length, etc. The two-point Gait segmentation statistics of optical flow, achieve a highly discriminative, body-shape robust representation of human gait [26]. Region is denoted by the 2D coordinates of the centroid, P, a ratio between the total number of foreground pixels (T) and the size of the bounding box (B), $R = T{=}B$, and the color/gray level characteristic, D. The position plus predicted velocity of the region exit/enter from/to scene are easily used for determining to have exited/entered the scene [16].

D. Fusion of Gait with Other Biometrics

The fusion of gait information with other biometrics can further increase recognition robustness and reliability. For optimal face recognition, they set a virtual camera to capture the frontal face. Gait cycle analysis serves two important functions. First, it determines the frequency and phase of each observed gait sequence. Secondly, it provides data reduction by summarizing the sequence with a small number of prototypical key frames [22]. Adaboost machine learning to choose useful features from the HOG.Adaboost is composed of simple weak classifiers and decides the weights and get responses [34].

E. Spatiotemporal motion

An extension of 2D image correlation to 3D correlation in the space and time domain to better capture its spatial structural and temporal transitional characteristics [18].

5. FUTURE DEVELOPMENTS

The state-of-the-art of visual surveillance for humans and vehicles sorted by a general framework of visual surveillance systems. Although a large amount of work has been done in visual surveillance for humans and vehicles, many issues are still open and deserve further research, especially in the following areas.

A. Occlusion Handling

Occlusion handing is a major problem in visual surveillance. During occlusion, only portions of each object are visible and often at very low resolution. To reduce ambiguities due to occlusion, better models need be developed to cope with the correspondence between features and body parts, and thus eliminate correspondence errors that occur during tracking multiple objects.

B. Anomaly Detection and Behavior Prediction

Anomaly detection and behavior prediction are significant in practice. In applications of visual surveillance, not only should visual surveillance systems detect anomalies such as traffic accidents and car theft etc, according to requirements of functions, but also predict what will happen according to the current situation and raise an alarm for a predicted abnormal behavior.

6. CONCLUSION REMARKS

Visual surveillance in dynamic scenes is an active and important research area, person-specific identification in certain scenes, crowd flux statistics and anomaly detection, etc. are some of the wide applications. We have presented an overview of recent developments in visual surveillance within a general processing framework for visual surveillance systems. As for the detection of moving objects, it involves environmental modeling, motion segmentation and object classification. Three techniques for motion segmentation are addressed: background subtraction, temporal differencing, and optical flow. We have discussed four intensively studied approaches to tracking: region based, active-contour based, feature based, and model based. As to personal identification at a distance, we have divided gait recognition methods into four classes: mode based, statistics, physical-parameter based, and gait based. At the end of this survey, a research work is person identification with the help of gait and biometric method.

REFERENCES

[1] M. I. Chacon-Murguia and S. G. Duarte, “An Adaptive Neural-Fuzzy Approach for Object Detection in Dynamic Backgrounds for Surveillance Systems,” *IEEE Trans. on Industrial Electronics*, vol. 59, no. 8, Aug. 2012.

[2] Deepak. N. A andDr. U. N. Sinha, “Silhouette Based Human Motion Detection and Recognizing their Actions from the Captured Video Streams,” *Int. J. Advanced Networking and Applications,* vol. 02, issue: 05, 2011.

[3] S. Santhanam, V. Balisavira, and V. K. Pandey, “Real-Time Obstacle Detection by Road Plane Segmentation,” *IEEE 9th International Colloquium on Signal Process and its Applications,* Mar. 2013.

[4] Dr. K Perumal and V. G. Karthikram, “A Fusion of SOM and Fuzzy C-Means for Image Processing,” *IJCTA*, vol.1 July-Aug. 2013.

[5] L. Maddalena and A. Petrosino, “Stopped Object Fetection by Learning Foreground Model in Videos,” *IEEE Trans. on Neural Networks and Learning systems,* vol.24, no. 5, May 2007.

[6] Dongxiang Chi, “Self-Organizing Map-Based Color Image Segmentation with k -Means Clustering and Saliency Map,”*ISRN Signal Proc.*, vol. 2011.

[7] K. Jalali, M. Heydari, and A. Tanavar, “Image Segmentation with Improved Distance Measure in SOM and K Means Algorithms,”*Journal of Mathematics and Computer Science*, vol. 8, 2014.

[8] F. C. Cheng, S. C. Huang, and S. J. Ruan, “Scene Analysis for Object Detection in Advanced Surveillance Systems Using Laplacian Distribution Model,” *IEEE Trans. on Systems, Man, And Cybernetics Applications and Reviews,* vol. 41, no.5, Sep. 2011.

[9] F. Moreno, J. Alarcón, R. Salvador, and T. Riesgo, “Reconfigurable Hardware Architecture of a Shape Recognition System Based on Specialized Tiny Neural Networks with Online Training,” *IEEE Trans. on Industrial Electronics,* vol. 56, no. 8, August 2009.

[10] D. Tsai and S. C. Lai, “Independent Component Analysis-Based Background Subtraction for Indoor Surveillance,” *IEEE Trans. on Image Process.*, vol. 18, no. 1, Jan. 2009.

[11] J. M. McHugh, J. Konrad, and V.Saligrama, and P. M. Jodoin, “Foreground Adaptive Background

Subtraction," *IEEE Signal Process. Letters*, vol. 16, no. 5, May 2009.
[12] L.Maddalena and A.Petrosino, "A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications," *IEEE Trans. on Image Process.*, vol.17, no. 7, July 2008.
[13] C. BenedekandT. Szirányi, "Bayesian Foreground and Shadow Detection in Uncertain Frame Rate Surveillance Videos," *IEEE Trans. on Image Process.*,vol. 17, no. 4, April 2008.
[14] D. Culibrk, O. Marques, D. Socek, H. Kalva, and B. Furth, "Neural network approach to background modeling for video object segmentation," *IEEE Trans. On Neural Networks*, vol.18, no. 6, 2007.
[15] M. Valera and S. A. Velastin, "Intelligent Distributed Surveillance Systems: A Review," *IEE Proc.-Vis. Image Signal Process.*,Vol. 152, No. 2, April 2005.
[16] Murat EK_INC_I and Ey¨up GED_IKL_I "Silhouette Based Human Motion Detection and Analysis for Real-Time Automated Video Surveillance," *Turk J ElecEngin*, vol.13, no.2 2005.
[17] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 34, no. 3, 2004.
[18] Liang Wang, Tieniu Tan, HuazhongNing, and Weiming Hu, "Silhouette Analysis-Based Gait Recognition for Human Identification," *IEEE Trans. on Pattern Analysis and Machine Intell.*,vol. 25, no. 12, Dec. 2003.
[19] H. M. Dee, and S. A. Velastin, "How close are we to solving the problem of automated visual surveillance?," *Machine Vision Applications*, vol. 19, no. 5-6, Sep. 2008.
[20] DominikBrugger, Martin Bogdan, and Wolfgang Rosenstiel, "Automatic Cluster Detection in Kohonen's SOM," *IEEE Trans on Neural Networks*, vol. 19, no. 3, March 2008.
[21] JuhaVesanto and EsaAlhoniemi, "Clustering of the Self-Organizing Map," *IEEE Trans. on Neural Networks*, vol. 11, no. 3, May 2000.
[22] Robert T. Collins, Ralph Gross, and Jianbo Shi, "Silhouette-based Human Identification from Body Shape and Gait," *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2002.
[23] Xi Chen, Zhihai He, James M. Keller, Derek Anderson, and Marjorie Skubic, "Adaptive Silhouette Extraction In Dynamic Environments Using Fuzzy Logic,"*IEEE Int. Conf. on Fuzzy Systems*, July 2006.
[24] S. Nissi Paul and Y. Jayanta Singh, "Survey on Video Analysis of Human Walking Motion," *Int. Journal of Signal Process., Image Processing and Pattern Recognition*, vol.7, no.3, 2014.
[25] Robert T. Collins, Alan J. Lipton, and Takeo Kanade, "Introduction to the Special Section on Video Surveillance," *IEEE Trans. on Pattern Analysis and Machine Intell.*,vol. 22, no. 8, Aug 2000.
[26] Stephen Lombardi, Ko Nishino, Yasushi Makihara and Yasushi Yagi, "Two-Point Gait: Decoupling Gait from Body Shape" *Computer vision foundation*.
[27] K.Maithili, K. Rajeswari, R. Mohanapriya, and D. Krithika , "An Efficient Human Action Recognition System Using Single Camera and Feature Points,"*Int. Journal of Advanced Research in Computer. and Comm. Engg.*, vol. 2, Feb. 2013.
[28] Wei Huang and Q. M. Jonathan Wu, "Human Action Recognition Based On Self Organizing Map," *IEEE ICASSP*, 2010.
[29] Andrea Teschioni, Franco Oberti and Carlo Regazzoni "A Neural-Network Approach For Moving Objects Recognition In Color Image Sequences For Surveillance Applications," *Int. Journal of Research in Computer Applications and Robtics*, vol. 2, issue 8, 2011.
[30] E. Thomas Gilmore, Preston D. Frazier and M. F. Chouikha, "Improved Human Detection Using Image Fusion," *IEEE ICRA*, May 2009.
[31] Ananya R. Pathak and kandarpa K. Sarma, "Human Detection and Surveillance System Using Realtime Video Feed and ANN," *IRNet Trans. on Electrical and Electronics Engg.*, vol-1, 2012.
[32] Yong Yang, "Image Segmentation by Fuzzy C-Means Clustering Algorithm with A Novel Penalty Term," *Computing and Informatics*, vol. 26, 2007.
[33] N. Keerthana, K. S. Ravichandran, and B. Santhi,"Detecting the Moving Object in Dynamic Backgrounds by using Fuzzy-Extreme Learning Machine,"*Int. Journal of Engg. and Tech.*,vol. 5 no. 2 Apr-May 2013.
[34] YutaroTakayanagi and JiroKatto, "Human Body Detection using HOG with Additional Color Features,"*Fundamental Science and Engineering*.
[35] NehaDhavale andSeemaBiday, "Study on Self Organizing approach for moving object detection and tracking for visual surveillance," *Int. Journal of Emerging Trends & Tech. in Comput. Science*,vol. 2, Nov–Dec 2013.

Mr. Sagar V. Bakale received the Diploma in Department of Electronics Engineering from MSBTE in 2006 and B.E. in Electronics and Telecommunication Engineering form Prof. Ram Meghe Institute of Technology and Research, SGBAU Amravati University, Amravati in 2009. He is pursuing his M.E.

in Electronics and Telecommunication from G. H. Raisoni College of Engineering and Management, Amravati under SGBAU Amravati University, Amravati. His area of research is Video Processing and its Applications. He is a Life member of ISTE.

Dr. Prashant V. Ingole received his B.E degree in 1989 in Electronics Engineering and M.E. degree in Automatic Control and Robotics from College of Engineering Badnera, Amravati University, Amravati and Faculty of Engineering and Technology M. S. University, Baroda respectively. He received his PhD degree in 2010 from Visvesvaraya National Institute of Technology, Nagpur in the area of Medical Image Processing and Image Retrieval. He worked as Lecturer, Assistant Professor and Professor at P.R.M.I.T&R, Badnera (Formerly College of Engineering, Badnera). Currently, he is working as Prinipal of G H Raisoni College of Engineering and Management, Amravati. He has more than 30 research papers in his credit in National and International Conferences and Journals. His research interests include Digital Signal Processing, Soft Computing, Image Processing, Embedded system. He is having a Fellowship of IETE and IEI and a Life member of ISTE.