International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 4, Issue 4, April 2015

A Survey on Human Motion Detection and Surveillance

Miss. Sapana K .Mishra
Dept. of Electronics and Telecommunication Engg.
J.T.M.C.O.E., Faizpur , Maharashtra, India

Prof. K .S .Bhagat
Dept. of Electronics and Telecommunication Engg.
J.T.M.C.O.E., Faizpur, Maharashtra, India

Abstract - **Over years detecting human beings in a video scene of a surveillance system is one of the most active research topics in computer vision. This interest is driven by wide applications in many areas such as virtual reality, smart surveillance and perceptual interface, human gait characterization person counting in a dense crowd, person identification, gender classification, and fall detection for elderly people. Video surveillance system mainly deals with tracking and classification of moving objects. The general processing steps of human motion detection for video surveillance includes modeling of environments, detection of motion, object detection and classification human detection, activity recognition and behavior understanding. The aim of this paper is to review recent developments and analyze future open directions in visual surveillance systems.**

Index Terms— **Behavior understanding, background subtraction, motion detection, statistical methods, temporal frame differencing. Object detection, object tracking, video surveillance,**

I. INTRODUCTION

Video surveillance has received a great attention as active application-oriented research areas in image processing computer vision, artificial intelligence. The process of video surveillance aims at analyzing video sequences. Video surveillance activities can be manual, semi-autonomous or fully-autonomous. Manual video surveillance involves analysis of the video content by a human. . Semi-autonomous video surveillance involves some form of video processing but with significant human intervention. In a fully-autonomous system only input is the video sequence taken at the scene where surveillance is performed. In such system there is intervention of human and the system does both tasks, like motion detection and tracking, and also decision making tasks like abnormal event detection and gesture recognition. The surveillance system starts with motion and object detection. Motion detection aims at segmenting regions containing moving objects from rest of the image. This is followed by object tracking and behavior analysis and recognition. The process of motion detection and object detection usually involves background modeling and motion segmentation. Motion segmentation in image sequence aims at detecting regions corresponding to moving objects such as humans, vehicles, animals, birds etc. Once the region involving motion is detected then for subsequent processes such as tracking and behavior analysis only these regions need to be investigated. After motion and object detection the surveillance system usually tracks moving objects from one frame to another in an image sequence. Behavior understanding involves analysis and recognition of motion patterns and description of actions and interactions among objects. Fig. 1 shows the general framework of visual surveillance in dynamic scenes. The primary purpose of this paper is to give a general review on the overall process of a visual surveillance system. The paper is organized as follows. Section II discusses motion detection, modeling of environment, segmentation of motion, and classification of moving objects. Section III reviews tracking of objects and Section IV discusses understanding and description of behaviors.

II. MOTION DETECTION

The process of motion detection includes environment modeling, motion segmentation, and object classification, which intersect each other during processing

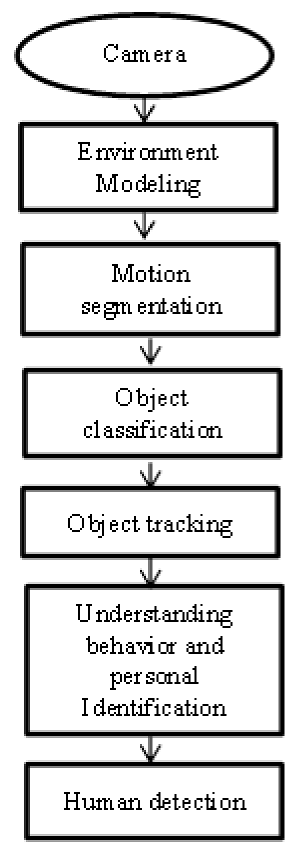

Fig 1 General framework of visual surveillance

A. Environment Modeling

ISSN: 2278 – 909X

Environment modelling is also known as background modelling. It is currently used to detect moving objects in video acquired from static cameras. Many statistical methods have been developed over the recent years. Murshed, M proposed an edge segment based statistical background modelling [1] algorithm and a moving edge detection framework for the detection of moving objects. This paper actually focused about various methods of background modelling like traditional pixel based, edge pixel based and edge segment based approaches. Yun Chu Zhang analyzed the background mechanism using GMM [2] model. Here this model updates new strategy which weighs the model adaptability and motion segmentation accuracy. Wei Zhou [3] proposed the dynamic background subtraction using spatial color binary patterns. ViBe: A Universal Background Subtraction Algorithm [4] for Video Sequences paper present a technique for motion detection which stores a set of value taken in the past in the same location or in the neighborhood. It then compares this set to current pixel value in order to determine whether the pixel belongs to the background and to adopt the model which substitutes from the background model. M.Hedayati [5] suggested Gaussian-based approach is the best approach for real-time applications. He evaluated and compared five well known background technique like Median filtering, Approximate Median, Running Gaussian Average (RGA), Gaussian Mixture Modal (GMM) and Kernel Density Estimation (KDE). Among these methods Gaussian based approached (RGA, GMM) gives better results in issues of speed, accuracy and memory usage for real time application.

B. Motion Segmentation

Motion segmentation in image sequences corresponds at detecting regions corresponding to moving objects such as vehicles and humans. Motion segmentation separates foreground images from background images and it is followed by object classification and Tracking. Several conventional approaches for motion segmentation are outlined as follows:

a. Background subtraction:

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 reference frame is commonly known as ‘background image’ or ‘environment model’. A good background model should be adaptive to the changes in dynamic scenes. It is simple, but extremely sensitive to changes in dynamic scenes derived from lighting and other similar events. The most common ones are adaptive Gaussian mixture, non-parametric background, temporal differencing, warping background and hierarchical background models.

b. Temporal differencing:

In temporal differencing the reference image is the previous images. Hence the previous frame is subtracted from the current image and the subtraction value must be greater than a threshold value which gives a difference image. To extract moving regions temporal differencing makes use of pixel-wise difference between two or three consecutive frames in an image sequence. The advantage of temporal differencing is that it is adaptive to dynamic environments but on other hand it has limitation of extracting the entire relevant feature pixels. e.g. possibly generating holes inside moving entities. Also temporal differencing method loses the object when a foreground object stops moving, temporal differencing method fails in detecting a change between consecutive frames. As an example of this method, Lipton et al. [6] used temporal differencing and detected moving targets in real video streams. After finding the absolute difference between the current and the previous frame, a threshold function was used to determine change. By using a connected component analysis, the extracted moving sections were grouped in motion regions. An improved version is to use three-frame differencing instead of two-frame differencing. For instance, VSAM [7] has developed a algorithm for motion segmentation by combining a three-frame differencing technique with an adaptive background subtraction algorithm. This hybrid algorithm is very fast and effective for detecting moving objects in image sequences.

c. Optical flow:

Optical flow is a vector-based approach that estimates motion in video by matching points on objects over image frame(s). Under the assumption of brightness constancy and spatial smoothness, optical flow is used to describe coherent motion of points or features between image frames. Optical flow based motion segmentation uses characteristics of flow vectors of moving objects over time to detect moving regions in an image sequence. One key benefit of using optical flow is that it is robust to multiple and simultaneous cameras and object motions, making it ideal for crowd analysis and conditions that contain dense motion. In Bregler’s work [8], each pixel was represented by its optical flow. These flow vectors were grouped into blobs having coherent motion and characterized by a mixture of multivariate Gaussians. More detailed discussion of optical flow can be found in Barron’s work [9].

B. *Object Classification*:

Object classification is a standard pattern recognition task. To track objects and analyze the behavior, it is essential to correctly classify moving objects. There are two different categories of approaches for classifying moving objects like, shape based and motion based classification.

a. Shape-based classification:

Different descriptions of shape information of motion regions such as points, boxes, silhouettes and blobs are available for classifying moving objects. VASM [3] takes image blob dispersedness, image blob area, apparent aspect ratio of the blob bounding box, etc., as key features, and classifies moving-object blobs into four classes: single human, vehicles, human groups, and clutter, using a viewpoint-specific three-layer neural network classifier.

b. Motion-based classification.

ISSN: 2278 – 909X

In general, nonrigid articulated human motion shows a periodic property, so this has been used as a strong cue for classification of moving objects.

The two common approaches can also be combined for classification of moving objects. It is expected that more accurate results can be obtained by features such as color and velocity

III. TRACKING:

The objective of video tracking is to associate target objects in consecutive video frames. The association can be especially difficult when the objects are moving fast relative to the frame rate. Tracking is a particularly important issue in human motion analysis since it serves as a means to prepare data for pose estimation and action recognition. After motion detection, surveillance systems generally track moving objects from one frame to another in an image sequence. Tracking over time typically 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. Useful mathematical tools for tracking include Kalman filter, the Condensation algorithm, Dynamic Bayesian Network, etc. Kalman Altering is a state estimation method based on Gaussian distribution. Unfortunately, it is restricted to situations where the probability distribution of the state parameters is unimodal. That is, it is inadequate in dealing with simultaneous multi-modal distributions with the presence of occlusion, cluttered background resembling the tracked objects, etc. The Condensation algorithm has shown to be a powerful alternative. It is a kind of conditional density propagation method for visual tracking. Based upon sampling the posterior distribution estimated in the previous frame, it is extended to propagate these samples iteratively to successive images. Tracking methods are divided into four major categories: region-based tracking, active-contour-based tracking, color cues in motion segmentation. Feature based tracking, and model-based tracking.

A. Region-Based Tracking

Region-based tracking algorithms track objects according to variations of the image regions corresponding to the moving objects.

For these algorithms, the background image is maintained dynamically and motion regions are usually detected by subtracting the background from the current image. Wren *et al.* [10] explores the use of small blob features to track a single human in an indoor environment. Work of McKenna et al. [11] proposed an adaptive background subtraction method that combined color and gradient information to effectively cope with shadows and unreliable color cues in motion segmentation. The region-based tracking approach works reasonably well. However, difficulties arise in two important situations. The first is that of long shadows, and it may result in connecting up blobs that should have been associated with separate people. The more serious, and so far intractable, problem for video tracking has been that of congested situations. Hence these algorithms cannot satisfy the requirements for surveillance against a cluttered background or with multiple moving objects.

B. Active Contour-Based Tracking

Tracking based on active contour aims at directly extracting the shape of the subjects. The idea is to have a representation of the bounding contour of the object and keep dynamically updating it over a time. Paragios *et al.* [12] detect and track multiple moving objects in image sequences using a geodesic active contour objective function and a level set formulation scheme. Peterfreund [13] explores a new active contour model based on a Kalman filter for tracking nonrigid moving targets such as people in spatio-velocity space. Isard *et al.* [14] adopt stochastic differential equations to describe complex motion models, and combine this approach with deformable templates to cope with people tracking. In contrast to the region-based tracking approach, the advantage of having an active contour-based representation is the reduction of computational complexity. However, it requires a good initial fit.

C. Feature-based tracking

Abandoning the idea of tracking objects as a whole, this tracking method uses sub-features such as distinguishable points or lines on the object to realize the tracking task. Its benefit is that even in the presence of partial occlusion, some of the sub-features of the tracked objects remain visible. Feature-based tracking includes feature extraction and feature matching. Low-level features such as points are easier to extract. It is relatively more difficult to track higher-level features such as lines and blobs. Polana and Nelson’s work [15] is a good example of point-feature tracking. In their work, a person was bounded by a rectangular box, whose centroid was selected as the feature point for tracking. In recent work of Jang and Choi [16], an active template that characterized regional and structural features of an object was built dynamically based on the information of shape, texture, color, and edge feature of the region. Feature-based tracking algorithms can handle partial occlusion by using information on object motion, local features and dependence graphs. However, there are several serious deficiencies in feature-based tracking algorithms.

a. The recognition rate of objects based on 2-D image features is low, because of the nonlinear distortion during perspective projection and the image variations with the viewpoint’s movement.

b. These algorithms are generally unable to recover 3-D pose of objects.

c. The stability of dealing effectively with occlusion, overlapping and interference of unrelated structures is generally poor.

D. Model-Based Tracking

Model-based tracking algorithms track objects by matching projected object models, produced with prior knowledge, to image data. The models are usually constructed off-line with a manual measurement, CAD tools or computer vision techniques. As model-based rigid object tracking and model-based non rigid object tracking are quite different, we review separately model-based human body tracking (nonrigid object tracking) and model-based vehicle tracking (rigid object tracking)

.

a. Model-Based Human Body Tracking

ISSN: 2278 – 909X

The general approach for model-based human body tracing is known as analysis-by-synthesis, and it is used in a predict-match-update style. Firstly, the pose of the model for the next frame is predicted according to prior knowledge and tracking history. Then, the predicted model is synthesized and projected into the image plane for comparison with the image data. Generally, model-based human body tracking involves three main issues:

1. Construction of human body models
2. Representation of prior knowledge of motion models and motion constraints
3. Prediction and search strategies

The above three issues are discussed as follows

b. Human body models

Traditionally, the geometric structure of human body can be represented in the following four styles.

1. Stick figure.
2. 2-D contour
3. Volumetric models.
4. Hierarchical mode

1. **Stick figure**

The essence of human motion is typically addressed by the movements of the torso, head and four limbs, so the stick-figure representation can be used to approximate a human body as a combination of line segments linked by joints. The stick figure is obtained in various ways, e.g., by means of median axis transform or distance transform.

2. **2-D contour**

This kind of human body model is directly relevant to human body projections in an image plane. The human body segments are modeled by 2-D ribbons or blobs

3. Volumetric models

The disadvantage of 2-D models is its restriction to the camera's angle; so many researchers are trying to depict the geometric structure of human body in more detail using some 3-D models such as elliptical cylinders, cones, spheres, etc. The more complex 3-D volumetric models, the better results may be expected but they require more parameters and lead to more expensive computation during the matching process.

c. Motion models

Motion models of human limbs and joints are widely used in tracking. They are effective because the movements of the limbs are strongly constrained. These motion models serve as prior knowledge to predict motion parameters to interpret and recognize human behaviors or to constrain the estimation of low-level image measurements

d. Search strategies:

Pose estimation in a high-dimensional body configuration space is intrinsically difficult, so, search strategies are often carefully designed to reduce the solution space. Generally, there are four main classes of search strategies: dynamics, Taylor models, Kalman filtering, and stochastic sampling.

IV. BEHAVIOR UNDERSTANDING

After successfully tracking the moving humans from one frame to another in an image sequence, the problem of understanding human behaviors from image sequences follows naturally. Behavior understanding involves action recognition and description. As a Anal or long-time goal, human behavior understanding can guide the development of many human motion analysis systems. In our opinion, it will be the most important area of future research in human motion analysis. Behavior understanding is to analyze and recognize human motion patterns, and to produce high-level description of actions and interactions. It may be simply considered as a classification problem of time varying feature data, i.e., matching an unknown test sequence with a group of labeled reference sequences representing typical human actions. The general analytical methods for matching time-varying data are outlined in the following.

A. Dynamic time warping

Dynamic time warping (DTW) used widely for speech recognition in the early days, is a template-based dynamic programming matching technique

a. Hidden Markov models

Hidden Markov models (HMMs) a kind of stochastic state machine, is a more sophisticated technique for analyzing time-varying data with spatio-temporal variability.

b. Neural network

Neural network (NN) is also an interesting approach for analyzing time-varying data. As larger data sets become available, more emphasis is being placed on neural networks for representing temporal information.

V. CONCLUSIONS

Computer-vision-based human motion analysis has become an active research area. It is strongly driven by many promising applications such as smart surveillance, virtual reality, advanced user interface, etc. Recent technical developments have strongly demonstrated that visual systems can successfully deal with complex human movements. It is exciting to see many researchers gradually spreading their achievements into more intelligent practical applications. Bearing in mind a general processing framework of human motion analysis systems, we have presented an overview of recent developments in human motion analysis in this paper. The state of the art of existing methods in each key issue is described and the focus is on three major tasks: detection, tracking and behavior understanding. As for human detection, it involves motion segmentation and object classification. Four types of techniques for motion segmentation are addressed: background subtraction, statistical methods, temporal differencing and optical flow. The statistical methods may be a better choice in more unconstrained situations. Tracking objects is equivalent to establish correspondence of image features between frames. We have discussed four

ISSN: 2278 – 909X

approaches studied intensively in past works: model-based, Active-contour-based, region-based and feature-based.

REFERENCES

[1] M. Murshed, A.Ramirez, O.Chae, (2010) "Statistical Background Modelling: An Edge Segment Based Moving Object Detection Approach" IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS).
[2] Yun Chu Zhang, Ru Min Zhang, Shi Jun Song, (2011) "Research on GMM Background Modelling and its Covariance Estimation" Journal on Advanced Materials Research (Volumes 383 - 390).
[3] Wei Zhou; Yu Liu; Weiming Zhang; Liansheng Zhuang; Nenghai Yu, (2011) "Dynamic Background Subtraction Using Spatial-Color Binary Patterns" , International Conference on Image and Graphics (ICIG).
[4] Barnich.O, Van Droogenbroeck.M,(2011) " ViBe: A Universal Background Subtraction Algorithm for Video Sequences" IEEE Transactions on Image Processing, Vol 20,No.6.
[5] M.Hedayati, Wan Mimi Diyana Wan Zaki, Aini Hussain (2010), "Real-Time Background Subtraction for Video Surveillance: From Research to Reality", (CSPA).
[6] J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video," in Proc. IEEE Workshop Applications of Computer Vision, 1998, pp. 8–14.
[7] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, and L.Wixson, "A system for video surveillance and monitoring," Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep., CMU-RI-TR-00-12, 2000.
[8] Bregler, "Learning and recognizing human dynamics in video sequences", Proceedings of the IEEE CS Conference on Computer Vision and Pattern Recognition, 1997, pp. 568–574.
[9] J. Barron, D. Fleet, S. Beauchemin, Performance of optical Kow techniques, Int. J. Comput. Vision 12 (1) (1994) 42–77.
[10] R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: real-time tracking of the human body," IEEE Trans. Pattern Anal. Machine Intell., vol. 19, pp. 780–785, July 1997.
[11] S. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking groups of people," Comput. Vis. Image Understanding, vol. 80, no. 1, pp. 42–56, 2000.
[12] N. Paragios and R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects," IEEE Trans. Pattern Anal. Machine Intell., vol. 22, pp. 266–280, Mar. 2000.
[13] N. Peterfreund, "Robust tracking of position and velocity with Kalman snakes," IEEE Trans. Pattern Anal. Machine Intell., vol. 22, pp. 564–569, June 2000.
[14] M. Isard and A. Blake, "Contour tracking by stochastic propagation of conditional density," in Proc. European Conf. Computer Vision, 1996, pp. 343–356.
[15] R. Polana and R. Nelson, "Low level recognition of human motion," in *Proc. IEEE Workshop Motion of Non-Rigid and Articulated Objects*,Austin, TX, 1994, pp. 77–82.
[16] D.-S. Jang and H.-I. Choi, "Active models for tracking moving objects," *Pattern Recognit.*, vol. 33, no. 7, pp. 1135–1146, 2000.

ISSN: 2278 – 909X