2018 4th International Conference on Computing Communication and Automation (ICCCA)

Object Detection, Classification and Tracking Methods for Video Surveillance: A Review

Rakesh Chandra Joshi
Deptt. of ECE
GBPUA&T
Pantnagar, India
rakeshchandraindia@gmail.com

Mayank Joshi
Deptt. of ECE
GBPUA&T
Pantnagar, India
mayankjoshi63@gmail.com

Adithya Gaurav Singh
Deptt. of ECE
GBPUA&T
Pantnagar, India
agaurav51@gmail.com

Sanjay Mathur
Deptt. of ECE
GBPUA&T
Pantnagar, India
mathur_sanjay@rediffmail.com

Abstract— **Video surveillance is going through an enormous change in the last decade and lots of research is undergoing in this area. As object tracking is an integral part of such systems, it becomes necessary to review all state-of-the-art methods and approaches which belongs to object detection, classification, and tracking. This paper consists of all such methods in a classified manner. After detecting object, classification is done in order to track that later. Object tracking is the method to locate the target in subsequent frames of video. There are also some challenges in video surveillance which make it complex. Object tracking is characterized on some parameters like Illumination variation, co-ordinates matching, environmental issues, tracking object's variety, pose variation, occlusion, motion blur etc. A vast classification is considered for review and complete analysis of all possible techniques so that each aspect is thus covered along with present research undergoing in that area.**

Keywords- Object Tracking, Object Detection, Object Classification, Video Surveillance

I. INTRODUCTION

Nowadays security requirements are increasing day by day and human based surveillance systems are not enough to deal with it. Thus, quest for development of intelligent video surveillance systems become necessary to have good security. Video surveillance systems which include object detection and tracking to implement security measures. Surveillance is the analysis and monitoring of the activities and behavior of suspicious objects for safety of the public. Automatic video surveillance such as in [1], detects abnormal behavior of objects and make financial benefit by saving the money on human-based security systems. As discussed in [2-4], many challenges are involved in developing a video surveillance system such as illumination, co-ordinates matching in case of multiple camera systems, environmental issues, tracking object's variety, pose variation of the object, occlusion, motion blur etc. and to perform tracking with these challenges in real time make tracking tedious.

Illumination gradually changes throughout the day because of the position of the sun, clouds in the sky and other weather conditions etc. Low illumination induces noise and sometimes, lighting conditions cause shadows or white-out effect. Most of the background subtraction (BS) based tracking techniques suffered from it. It is difficult to handle shade and shadow caused by illumination change [5]. But there are some algorithms which are able to handle these illumination conditions but needs time of the order of several frames for estimation and training of the background model [6].

For large area surveillance, multiple cameras are needed and all cameras should be coordinated for taking multi-view analysis of the single object or association of field of view (FoV) of different cameras as done in [7]. Environmental conditions like fog, smog, rain, fire, snow, smoke etc. can create the false object detection or system may lose its robustness or may be unable to track correctly. Variety of object is also an important parameter e.g. for a car tracking system, there is huge variety of cars available with respect of size, color, logo and shape. Thus, the system should be trained with different physical features of the car to detect and track accurately. Camera should also place in such a way that pose variation will not affect tracking and it can get the maximum features of tracked object.

Motion blur is an issue arises in case of an object in motion where only foreground is suffered from the blur, not background. It is due to insufficient shutter speed or frame rate [8]. The resolution requirement of the system varies, as if the system made for human detection it needs only 10% of the view, for recognition it needs 50% and for further identification it can need 100% and more. Number of cameras reduces with camera having better resolution. Small lens magnification on the edges compared to center of Field of View (FoV) results into barrel distortion (short focal length) problem in which object of the same size contains less pixels at the edge compared to center of FoV. [9, 10] worked on barrel correction. Occlusion is a challenging problem while tracking. At the instant of the occlusion, feature based tracking get failed. Hence, to continue tracking, prediction is essential. Classically, Kalman filtering [11] is used but simple it is not enough. Shape, size, texture, geometry, color and 3-D approaches like depth estimation, azimuth angle calibration taken as features for better precision. Merging happens when two or more objects grouped together and splitting is the separation from the merged object. Split happens after the merge and these are detected by ratio of height and length.

The remainder of the paper is organized as follows. Section II describes object detection methods. Classification methods are given in section III. Section IV is dedicated to object tracking methods. Conclusion is given in section V.

II. DETECTION OF THE OBJECT

Object detection is classified as- Motion-based and Appearance-based object detection. Motion-based approaches include the velocity, acceleration, direction and trajectory of the object whereas appearance-based methods include the features like color, edge, shape, size and any other static features.

A. Motion-based object detection

Mainly, the motion-based methods are classified as- Background Subtraction, Frame Differencing method and Optical Flow method.

1. Background Subtraction (BS)

BS is the widely used method to detect an object in motion by taking a difference of current and background frames as in [12], shown mathematically in Eq. 1 and 2. Thus, the changes are characterized as moving body. One threshold is set and where the difference is above the threshold, termed as foreground, while other pixels are termed as the background pixels.

$$\text{Foreground Pixel: } \left| F_C - F_B \right| \geq \text{Threshold} \quad (1)$$
$$\text{Background Pixel: } \left| F_C - F_B \right| < \text{Threshold} \quad (2)$$

where, F_C and F_B are current and background frame respectively.

Before thresholding, colored image is converted into the grey-scale or binary image. BS is good to detect movement in the places where small movement is significant such as banks after the office time, places where luxury materials or sensitive documents are placed or the places related to defense and explosives.

The background should be dynamic (ability to update time to time) to provide good detection. Flow graph for BS is shown in Fig. 1.

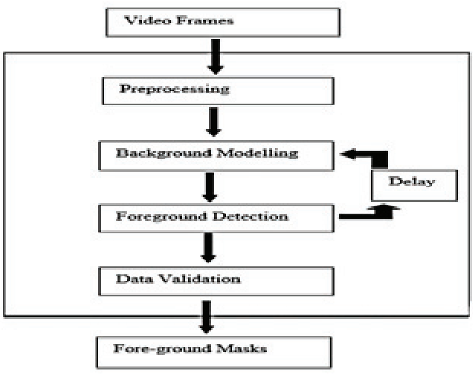

Fig. 1: Generalized BS flow graph [13]

There are some widely used BS methodologies such as-

i) Gaussian Mixture Model Method [14]
ii) Approximate Median Filter method
iii) Eigen Background method

But BS is not suitable to detect static objects and affects very much on sudden illumination change, shadows etc.

2. Frame Differencing (FD)

FD is also an image subtraction method like BS. But in this type of tracking such as [15, 16], difference between current and any other previous frame is taken, shown mathematically in Eq. 3 and 4. Hence, two successive frames or set frame interval e.g. 5 frames is taken and thresholding is done to avoid noise due to leaves of tree, rippling water etc.

$$\text{Foreground Pixel: } \left| F_C - F_P \right| \geq \text{Threshold} \quad (3)$$
$$\text{Background Pixel: } \left| F_C - F_P \right| < \text{Threshold} \quad (4)$$

where, F_C is the current frame and F_P is the previous frame.

For better accuracy, threshold should be dynamic and also adaptive to the background model. Combination of BS and FD can also be utilized to detect the abnormal event e.g. unattended objects in a crowd places [17].

3. Optical Flow method

Optical flow shows pattern of the movement in the image of two consecutive frames due to displacement of object or imagining device and represented in the form of the displacement vector field and to do that successive image matching is done. These algorithms mostly use the frequency information, gradient and correlation. As it considers all the pixels, thus optical flow is not a good choice for the real time application without any modification. RBCs in blood and vehicles are also tracked with this [18, 19].

Most methods which use optical flow and tried to optimize it are based on the assumption that the grey values of the pixel do not change while having the pixel displacement from one point to other in consecutive frames. Mathematically, it is shown as-

$$I(x + a, y + b, t+1) = I(x, y, t) \quad (5)$$

where, I is an image in two dimensions, x and y at time t and a, b are the displacement in those coordinates.

After converting the image into grey scale image and applying optical flow method with suitable threshold value and some morphological operation gives better results. Although it is accurate, yet it is slow due to large number of the calculations. Convolution of optical flow with magnitude of gradient information results in the improvement in tracking [20].

B. Appearance-based object detection

Primarily, the appearance-based methods can be classified as- Template Matching and Feature Extraction.

1. Template matching

It is simplest detection procedure and capable to detect static object in set of frames. It takes the image part known as template which is to be search or detect. Single template or set of templates (for better accuracy) can be taken. Then, next step is matching of that template in the whole image or frame of the video. After setting the threshold, matching criteria is made. If matched then, the requisite object is found. But it becomes slow in case of large ROI. To make it robust as [21], system is to be trained with huge dataset of certain object e.g. vehicles for the detection of vehicles, such as in convolutional neural network. These datasets are now available free of cost on the internet.

2. Feature Extraction

Feature extraction is basically the method to take characteristic information from the image and then, represent it in low dimensionality space. In other words, it is an art to find the information pattern in an image in the form of the feature vectors. These types of detectors are also known as point detector. Then, next step is the feature selection which means to retain only those features which has strong or most relevant information out of all extracted features to reduce calculation complexity and to increase the detection speed. There are several feature extractions methods such as Gabor features, contour profiles, deformable templates, Zernike moments, SIFT (Scale Invariant Feature Transform) features, Haar-like features, Gradient features, Projection histograms, unitary image transforms, Speeded Up Robust Features (SURF) etc. Template matching become effective when feature extracted from that template and then tracked as extraction of features from the whole template make it less complex and detection fast. Local Binary Pattern (LBP) features are used in the local feature extraction based on texture. Ahonen *et al.* [22] utilized LBP for recognition of face. It is grey scale invariant method and measures grey values.

III. CLASSIFICATION OF THE OBJECT

Object Classification is the next step after the object detection. An object is differentiated from other with respect to four characteristics - Shape and size, color, texture and motion.

A. Shape and size based classification

This method takes very less computational time and accuracy is good enough. The information regarding the shape is collected from the box around the ROI, blob and point. Hence, it checks the respective box, blob or point in the image frames. It is not suitable for the dynamic situations and to detect internal movements.

B. Color-based classification

As it doesn't depend upon size, orientation and scale of the image, it gives more accurate results than other methods but with increase in computational time. It gives wide range of classification of the objects. Color remains constant that why it is easy to identify the object with the help of color. To check distribution of color component among different image frames Gaussian distribution model is created and accordingly, image is differentiated among foreground and background. Thus, color based histogram can be utilized in real-time applications such as for forest delineation in [23] where texture information is also utilised.

C. Texture-based classification

Object can also characterized by its texture. Texture has information about the surface characteristics. There are different texture characteristics on basis of which object is classified such as directionality, contrast, smoothness, brightness change, roughness, coarseness etc. Steerable Pyramids, Wavelet and Gabor wavelet transforms are also used for the analysis of texture. This type of classification also requires lots of computations to give accurate results.

D. Motion-based classification

This classification categorizes the object on its acceleration, velocity and sometimes changes in the position. The accuracy and the computational time is good enough but is not suitable for the static objects.

IV. OBJECT TRACKING

Object tracking is defined as the generation of the path and trajectory of object in the image plane by finding and locating the position of object in subsequent frames. Simply, it gives the region where the object lies in every frame. Then, the relationship is obtained in between the positions of the object in different frames. The object is tracked in two ways- Forward and back tracking.

In forward tracking, object position is estimated from earlier image frames analysis after their segmentation as done in [24]. It finds the location of object in each frame by means of any detection method. If the detection is in the form of the point, translation is obtained as trajectory. In back tracking, information about location of object and associated regions obtained in the previous frame, is updated in every frame. After segmenting the current frame into foreground and background, correspondence is established (with the help of template in majority cases) with earlier frames, used in [25] for handheld objects.

Tracking is generally classified into three parts- Point Tracking, Kernel Tracking and Silhouette-based object tracking.

A. Point Tracking

When the object location is given as point (position) or set of points (in case of feature extraction) after the object detection then to locate those points in the subsequent frames, point tracking is used. It is difficult to use it in case of occlusion because once the object is completely occluded then it may result in the false detection or miss-detection. Point tracking is classified into two types- Deterministic and Statistical or Probabilistic Method.

1. Deterministic Method -

This method considers some parameters to establish relation between the previous frame and current frame. These constraints or parameters [26] are -

a) Maximum velocity- Limit on maximum velocity and correspondences to the locality of the object in circular manner.
b) Smooth motion- No abrupt change in speed and direction.
c) Common motion- If there is multi-point representation of the object, velocity is assumed similar in small area.
d) Proximity- There should not be a huge difference in the position of the object in subsequent frames.
e) Proximal uniformity- It is combination of two parameters, smooth motion and proximity.
f) Rigidity-The distance between any of the two points on the object should remain the same always.

These parameters are not restricted to only deterministic method but can also be used with statistical methods. After establishing the parameter correspondence in between two frames, it can also be extended to the multiple frames. Multiple parameters can be taken in multiple frames to establish the relationship so that tracking give good results.

2. Statistical or Probabilistic Method

These methods take uncertainties into the consideration during the tracking. State space estimation is done to calculate or predict next state on the basis of the position, velocity and acceleration of the object. This method can be classified further on basis of number of object is to be tracked - Single and Multiple object Tracking.

a. Single Object Tracking

This type of tracking can be done with any of these methods-

i. Kalman Filter (KF)

KF is used to estimate unknown variable in the system with much more accuracy by analyzing the series of the measurements with noise and inaccuracies in the system parameters over the period of time. As Rudolf E. Kalman is the main contributors to this theory, it is named after him. Joint probability distribution is calculated from variables for each time-duration. It is practical tool of estimation which works on real-time. But now adaptive Kalman gain is used in many processes instead of traditional constant gain KF. The performance parameter is the difference of actual and estimated values. This is a recursive algorithm where estimation of the value is done after prediction and at the end, value (correction) is updated to reduce mean square error iteratively. It depends only upon previous calculated and current state, not on any other past value. It assumes Gaussian distribution of the variables to estimate the state variables. It is suitable only for the linear motion. Unscented KF [27] and extended KF [28, 29] are used in non-linear cases but the estimation is affected when there is very high non-linearity. KF can also track the multiple objects by treating them as a separate i.e. taking multiple objects as many single objects [30, 31]. But it is difficult to establish good correspondence among different points and become more tedious when objects are approaching towards each other.

ii. Particle Filter

As KF is only limited to the Gaussian or normally distributed state variables and estimate them very poorly. This problem is overcome by the particle filtering. Particle filtering as used in [29, 32], can estimate the internal states from partial observations. This method is good for non-Gaussian and multi model probability density function. Del Moral originated the term particle filtering in 1996 [33]. In this method, without need of the assumptions it generates the samples from any type of probability density function (PDF), known as Importance sampling. Then, these set of samples with discrete values are represented as particles. Each particle contains one set of value for each state variable [34]. Instead of giving the exact representation (as in case of Gaussians) it gives the approximate representation. Then, weight is assigned for each particle which represents probability of that particle to be sampled from the PDF. Dark particle gets more weight [35]. But there is problem of weight collapse due to weight inequality. It can be solved through resampling or adaptive sampling. In resampling, low or negligible weights are replaced by the higher ones in that neighborhood. Resampling is done to remove old particles which are of no use now and if these unlike particles kept on, these can create many more unlikely states and condition of particle depletion where one particle has highly likelihood and other particles have almost zero-probability.

There are many resampling techniques available such as Multinomial, Residual, Stratified and Systematic. This method consists of three steps [36]-

- Prediction - To obtain prior probability where each particle is added with random sample taken from motion model
- Update – Assign the weight to each particle according to probability of measurement and then normalize.
- Resample – Insertion of new set of particles to avoid unlike particles to form proper PDF.

b. Multiple Object Tracking

Multiple objects tracking using probabilistic or statistical approach requires multiple correspondences of the different objects and state estimation as shown in fig. 2.

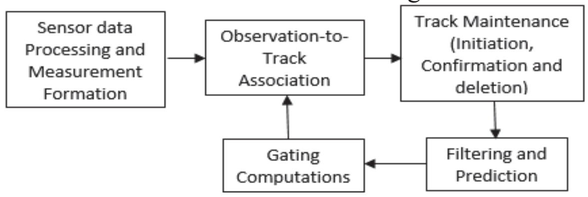

Fig. 2: Basic elements of the conventional multiple object tracking

Traditional methods are based on the Global Nearest Neighbour (GNN) approach [37, 38] which finds the best allocation of the input observations from previous tracks. It is good where observations are produced by the single target but there are chances of false detection when multiple objects are close to each other.

There are two methods for multiple object tracking-

i. Joint Probability Data Association Filter (JPDAF)

A good correspondence of the points in between the multiple objects can be established when all are separate but it is difficult to explain the correspondence between objects which are near to each other or in neighborhood. Joint Probability Data Association (JPDA) has a problem of coalescence. At near points, chances of incorrect correspondence are more and even a single incorrect correspondence can lead to a system failure. Thus, there should good data association between the objects and state estimation should be predicted correctly [39]. JPDAF works on the framework of the KF and it based on likelihood estimate. As JPDA method facilitates updating of track by the weighted sum of all observations, uncertain conditions of association are avoided. But it results into increase in the size of covariance matrix of the KF and can result into more false observation in track gate.

ii. Multiple Hypothesis Tracking (MHT)

As most of the surveillance systems require multiple object tracking, to establish multiple correspondences in between the objects and their state to solve data association problem. For this purpose, MHT works as an alternative, used in [39, 40]. Here, in case of conflict, alternate data association hypothesis formed as shown in fig. 3. It is most preferred data association method in present-day systems.

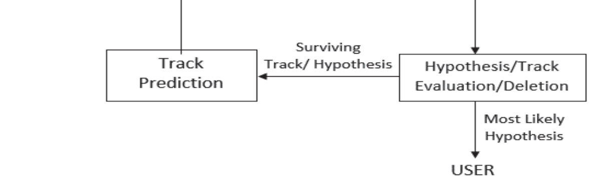

Fig. 3: Multiple Hypothesis Tracking [41]

In MHT tedious data association decisions are postponed until more data is not received. MHT is classified into two approaches-Hypothesis oriented and Track oriented MHT (TOMHT) [42].

B. Kernel Tracking

Kernel is the central part of anything. In context of object tracking, kernel is the central component of the non-rigid object. Isotropic kernel is masked with the target spatially to generate similarity function and according to that, object is localized in image frame. Mostly, Bhattacharya coefficient is used for finding the correlation score which gives the similarity measure between the target models and target candidate and object is located fast [43].

The methods under the Kernel tracking differs from each other in the terms of object representation and classified as-Template Tracking, Mean Shift Method, Support Vector Machine and Layering based Matching

1. Template Tracking

It is basically a brute force technique to find and locate the object template (reference image) in the larger image or a frame of the video. Pixel by pixel matching is done with the given template by sliding the pixel matrix at each possible position in image frame. Then, same procedure is repeated for subsequent frame and respective location of object is located. Higher value of the cross correlation is achieved at the places where pixel value of template matched with image frame. Brightest location represent the highest matching parts. Occlusion robust template matching is presented in [44].

Template tracking is mostly done for the single object and image is converted into binary image, as a result only grey values are compared. To make matching effective, template image is divided into many small sub-images. It helps in the case of partial occlusion e.g. to track the playing card occluded by the fingers. Frequency domain filtering is used to speed up the matching procedure with help of convolution theorem. To make effective and accurate matching template matching is combined with feature extraction [45]. Correlation can fail if there is variation of intensity with position i.e. when template pixel is correlate with the bright pixel of image and give higher value than others.

2. Mean Shift Method

This non-parametric robust statistical method, was first presented by Fukunaga and Hostetler in 1975. It can locate the local maxima or modes of any PDF. At first, image frames of video are represented as probability distribution i.e. every pixel has some probability value according to color which makes it good choice for color-based object tracking. After placing search window over one part of probability distribution, window moved to the position of maxima and recursively mean is calculated until it finds the local maxima and mean converges. But it becomes difficult to calculate in case of high dimensional space. As search window has the fixed size, this method is good only for single image and statistic distributions.

There are various improvements in the mean-shift algorithms. In Continuously Adaptive Mean Shift algorithm (CAMShift) method [46], there is a facility of dynamic search window which change its size by invariant moments according to object size in every frame. Performance is increased by the quantization of the 1-D histogram which is generated by the probability values of each pixel of image.

3. Support Vector Machines (SVM)

SVM is basically a classifier based upon statistical learning theory which separates two lineally separable pattern in an optimized way at decision boundary. Support vectors are those data points which are at the separating boundary and are difficult to classify [47]. It is also used for the regression and outlier detection. This classifies data in terms of the positive values and negative values. Positive values corresponds points of an object or foreground and negative values corresponds to background or the points which are similar to the foreground but not to be considered while tracking. Then, tracking system is trained by these positive and negative values in single or multiple stages, known as supervised learning [48]. Subset of the training samples are the support vectors. Thus, negative values become very important to perfectly track the object and to avoid false detection and tracking. SVM works perfectly for outliers as it ignores those points after choosing optimal hyperplane. If patterns are not separable linearly then, these points are mapped to a new space with the help of kernel function. Only that hyperplane will be selected which separates two patterns very well and has higher margin from both set of points and has no classification error. Then, with help of kernel function (decision function) those points which are not separable, converted into the higher dimensional space instead of a single line. Therefore, SVM is memory efficient and is mainly used for the higher dimensional spaces even when dimensions are more than samples.

4. Layering Based- Matching

Motion layer has the coherent information about two dimensional motion and shows homogeneous motion in a sequence of images. As in [49, 50], layer-based method of tracking computes the motion. Motion models with supporting layers give the compact representation of a scene. For which, segmentation and motion computation is done iteratively and whole image is modelled as set of layers. Each layer has motion (translation or rotation), shape, layer appearance, intensity information etc. There is one layer for the background and one layer for each object. Various motions and complex interactions i.e. pass (objects are moving in opposite directions) and stop are handled with help of dynamic layer tracking which is more robust to track the object because it includes appearance information, global segmentation prior and complete temporal consistency constraints with motion model and its segmentation [51]. Shape of two layers is maintained during passing because of global shape prior function.

C. Silhouette-Based Object Tracking

When object is represented by the outlines with only single solid color in between the outlines made up of edges is known as silhouette. It is used where object cannot be represented by the simple geometric shapes or by set of points. Thus, it is mostly used for tracking of human [52]. Human can be differentiated from others with help of silhouette. Silhouette is feature-less, therefore object model is created with help of contour, edge information, color histogram etc. Then, object is located in video frames with help of these object models. The silhouette tracking is also used for human action recognition (HAR) and common procedure is shown in fig. 4.

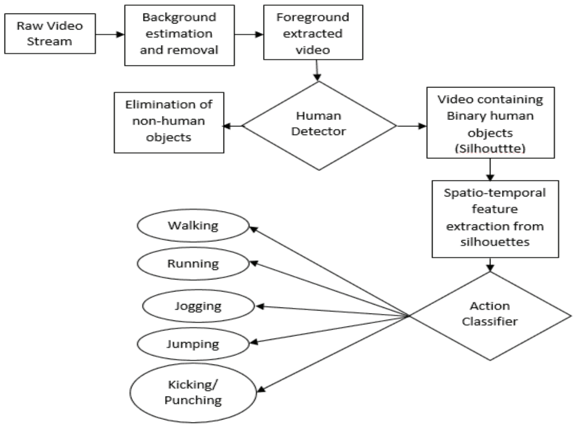

Fig. 4: General procedure for human action recognition

There are primarily two method of silhouette tracking on the basis of the object model-Shape Matching and Contour matching.

1. Shape matching

In this procedure, silhouette of object and its object model is searched in current frame. Also, it is assumed that there is only a translation of silhouette in between the frames. Hence, shape matching based silhouette tracking is not suitable for non-rigid motion. Information is captured in terms of edges, density function or silhouette boundary and mapping of those is done to track that object [53]. Huttenlocher ***et al.*** [54] used the edge based information for shape matching using the Hausdorff distance. Hausdorff distance is measure of mismatch in between the shapes. Thus, this approach focuses on those edges which do not change much while tracking the object such as object length or height. Edge information and color information can also be combined to make tracking robust [55].

2. Contour matching

This technique discussed in [56, 57, 58]. Points are selected on the boundary of object in each frame. Then, a contour is developed gradually from the previous frame to its new position in ongoing frames. According to the contour position in previous and current frame, object is tracked but it requires some overlapping of object region in between those frames.

Tracking using the contour matching can be done by two methods i.e. using the state space model or using direct minimization of contour energy function. Object motion and shape are the parameters to define the state of object.

V. CONCLUSION

This paper provides an overview on all techniques on object detection, classification and tracking. The paper mainly focuses on the working principle of all these methods. As the object detection and tracking deals with many challenges. Every method has its own advantage but one particular method is unable to deal all the problems alone.

Hence, it is necessary to develop a method to track the object which has the capability to track the object accurately and efficiently in terms of the computation. As video surveillance is on demand, this paper will help to understand all the basics to design a tracking algorithm which adapt to various conditions.

REFERENCES

[1] G. L. Foresti, L. Marcenaro and C. S. Regazzoni, "Automatic detection and indexing of video-event shots for surveillance applications," in *IEEE Transactions on Multimedia*, vol. 4, no. 4, pp. 459-471, Dec 2002

[2] S. Abdurrahman, "Smart video-based surveillance: Opportunities and challenges from image processing perspectives," *2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, Semarang, 2016, pp. 10-10.

[3] S. Ezatzadeh and M. R. Keyvanpour, "Fall detection for elderly in assisted environments: Video surveillance systems and challenges," *2017 9th International Conference on Information and Knowledge Technology (IKT)*, Tehran, 2017, pp. 93-98.

[4] S. Little, K. Clawson, A. Mereu and A. Rodriguez, "Identifying and addressing challenges for search and analysis of disparate surveillance video archives," *5th International Conference on Imaging for Crime Detection and Prevention (ICDP 2013)*, London, 2013, pp. 1-6.

[5] W. Kim and C. Jung, "Illumination-Invariant Background Subtraction: Comparative Review, Models, and Prospects," in *IEEE Access*, vol. 5, pp. 8369-8384, 2017.

[6] M. Shakeri and H. Zhang, "Object detection using a moving camera under sudden illumination change," *Proceedings of the 32nd Chinese Control Conference*, Xi'an, 2013, pp. 4001-4006.

[7] L. Havasi and Z. Szlávik, "Using location and motion statistics for the localization of moving objects in multiple camera surveillance videos," *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, Kyoto, 2009, pp. 1275-1281.

[8] B. Ma, L. Huang, J. Shen, L. Shao, M. H. Yang and F. Porikli, "Visual Tracking Under Motion Blur," in *IEEE Transactions on Image Processing*, vol. 25, no. 12, pp. 5867-5876, Dec. 2016.

[9] H. T. Ngo and V. K. Asari, "A pipelined architecture for real-time correction of barrel distortion in wide-angle camera images," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 3, pp. 436-444, March 2005.

[10] S. L. Chen, H. Y. Huang and C. H. Luo, "Time Multiplexed VLSI Architecture for Real-Time Barrel Distortion Correction in Video-Endoscopic Images," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 11, pp. 1612-1621, Nov. 2011.

[11] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan and S. S. Sastry, "Kalman filtering with intermittent observations," in *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453-1464, Sept. 2004.

[12] M. Piccardi, "Background subtraction techniques: a review," *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, 2004, pp. 3099-3104 vol.4

[13] Aseema Mohanty and Sanjivani Shantaiya. Article: A Survey on Moving Object Detection using Background Subtraction Methods in Video. *IJCA Proceedings on National Conference on Knowledge, Innovation in Technology and Engineering (NCKITE 2015)*NCKITE 2015(2):5-10, July 2015

[14] Dar-Shyang Lee, "Effective Gaussian mixture learning for video background subtraction," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 827-832, May 2005.

[15] J. Kang, D. V. Anderson and M. H. Hayes, "Face recognition for vehicle personalization with near infrared frame differencing," in *IEEE Transactions on Consumer Electronics*, vol. 62, no. 3, pp. 316-324, August 2016.

[16] M. Zhu and H. Wang, "Fast detection of moving object based on improved frame-difference method," *2017 6th International Conference on Computer Science and Network Technology (ICCSNT)*, Dalian, 2017, pp. 299-303.

[17] N. Srivastav, S. L. Agrwal, S. K. Gupta, S. R. Srivastava, B. Chacko and H. Sharma, "Hybrid object detection using improved three frame differencing and background subtraction," *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, Noida, 2017, pp. 613-617.

[18] P. K. Bhaskar, S. P. Yong and L. T. Jung, "Enhanced and effective parallel optical flow method for vehicle detection and tracking," *2015 International Symposium on Mathematical Sciences and Computing Research (iSMSC)*, Ipon, 2015, pp. 138-143.

[19] D. Guo, A. L. van de Ven and X. Zhou, "Red Blood Cell Tracking Using Optical Flow Methods," in *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 3, pp. 991-998, May 2014.

[20] T. Senst, V. Eiselein and T. Sikora, "Robust Local Optical Flow for Feature Tracking," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 9, pp. 1377-1387, Sept. 2012.
[21] V. John, K. Yoneda, Z. Liu and S. Mita, "Saliency Map Generation by the Convolutional Neural Network for Real-Time Traffic Light Detection Using Template Matching," in *IEEE Transactions on Computational Imaging*, vol. 1, no. 3, pp. 159-173, Sept. 2015.
[22] T. Ahonen, A. Hadid and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041, Dec. 2006.
[23] Z. Wang and R. Boesch, "Color- and Texture-Based Image Segmentation for Improved Forest Delineation," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3055-3062, Oct. 2007.
[24] E. Marchiori, M. Marchiori and J. N. Kok, "Forward-tracking: a technique for searching beyond failure," *Proceedings Eighth IEEE International Conference on Tools with Artificial Intelligence*, 1996, pp. 324-331.
[25] K. Chaudhary, Y. Mae, M. Kojima and T. Arai, "Autonomous acquisition of generic handheld objects in unstructured environments via sequential back-tracking for object recognition," *2014 IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, 2014, pp. 4953-4958.
[26] Alper Yilmaz, Omar Javed and Mubarak Shah, "Object Tracking: A Survey", ACM Computing Surveys, December 2006.
[27] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," in *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401-422, Mar 2004.
[28] A. Kamann, J. B. Bielmeier, S. Hasirlioglu, U. T. Schwarz and T. Brandmeier, "Object tracking based on an extended Kalman filter in high dynamic driving situations," *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, Yokohama, 2017, pp. 1-6.
[29] L. ChongYi, L. Cheng, F. LinYu and Y. JingTing, "Target tracking based on extended Kalman particle filter," *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, Chengdu, 2017, pp. 1715-1719.
[30] L. Marcenaro, M. Ferrari, L. Marchesotti and C. S. Regazzoni, "Multiple object tracking under heavy occlusions by using Kalman filters based on shape matching," *Proceedings. International Conference on Image Processing*, 2002, pp. III-341-III-344 vol.3.
[31] C. M. Bukey, S. V. Kulkarni and R. A. Chavan, "Multi-object tracking using Kalman filter and particle filter," *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, Chennai, 2017, pp. 1688-1692.
[32] L. Fu, "Particle Filter Pedestrian Tracking Algorithm Based on Selected Region RGB Histogram," *2018 International Conference on Robots & Intelligent System (ICRIS)*, Changsha, China, 2018, pp. 287-290.
[33] Del Moral, P.; Guionnet, A. Central limit theorem for nonlinear filtering and interacting particle systems. Ann. Appl. Probab. 9 (1999), no. 2, 275—297
[34] M. S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," in *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174-188, Feb 2002.
[35] J. H. Kotecha and P. M. Djuric, "Gaussian particle filtering," in *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2592-2601, Oct. 2003.
[36] Kaijen Hsiao, Henry de Plinval-Salgues and Jason Miller," Particle Filters and Their Applications", Cognitive Robotics April 11, 2005
[37] A. Sinha, Z. Ding, T. Kirubarajan and M. Farooq, "Track Quality Based Multitarget Tracking Approach for Global Nearest-Neighbor Association," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1179-1191, APRIL 2012.
[38] M. Shi, Q. Ling, Z. Yu and J. Zhu, "Association using modified Global Nearest Neighbor in the presence of bias," *Proceedings of the 32nd Chinese Control Conference*, Xi'an, 2013, pp. 4688-4691.
[39] X. Liu, K. Wang, D. Li, J. Xu and J. Pan, "A two-stage gating algorithm for joint probability data association filter," *IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS*, Beijing, 2010, pp. 381-384.
[40] K. Ahmadi and E. Salari, "A novel Multiple Hypothesis Testing (MHT) scheme for tracking of dim objects," *2015 IEEE International Conference on Electro/Information Technology (EIT)*, Dekalb, IL, 2015, pp. 365-369.
[41] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," in *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 5-18, Jan. 2004.
[42] Y. Kosuge, M. Kojima and S. Tsujimichi, "Multiple manoeuvre model track-oriented MHT (multiple hypothesis tracking)," *SICE Annual, 1999. 38th Annual Conference Proceedings of the*, Morioka, 1999, pp. 1129-1134.
[43] D. Comaniciu, V. Ramesh and P. Meer, "Kernel-based object tracking," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564-577, May 2003.
[44] H. T. Nguyen, M. Worring and R. van den Boomgaard, "Occlusion robust adaptive template tracking," *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, Vancouver, BC, 2001, pp. 678-683 vol.1.
[45] H. T. Nguyen and A. W. M. Smeulders, "Template tracking using color invariant pixel features," *Proceedings. International Conference on Image Processing*, 2002, pp. I-569-I-572 vol.1.
[46] O. D. Nouar, G. Ali and C. Raphael, "Improved Object Tracking With Camshift Algorithm," *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, Toulouse, 2006, pp. II-II.
[47] Chih-Wei Hsu and Chih-Jen Lin, "A comparison of methods for multiclass support vector machines," in *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415-425, Mar 2002.
[48] S. Avidan, "Support vector tracking," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1064-1072, Aug. 2004.
[49] Jiangjian Xiao and M. Shah, "Motion layer extraction in the presence of occlusion using graph cuts," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1644-1659, Oct. 2005.
[50] F. Xu and Q. Dai, "Occlusion-Aware Motion Layer Extraction Under Large Interframe Motions," in *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2615-2626, Sept. 2011.
[51] Hai Tao, H. S. Sawhney and R. Kumar, "Object tracking with Bayesian estimation of dynamic layer representations," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 75-89, Jan 2002.
[52] I. Haritaoglu, D. Harwood and L. S. Davis, "Hydra: multiple people detection and tracking using silhouettes," *Proceedings 10th International Conference on Image Analysis and Processing*, Venice, 1999, pp. 280-285.
[53] Satyabrata Maity, Debotosh Bhattacharjee and Amlan Chakrabarti, "A Novel Approach for Human Action Recognition from Silhouette Images", Computer Vision and Image Understanding 00 (2015) 1–18, ELSEVIER, 15 Oct, 2015.
[54] L. Wang and D. Suter, "Learning and Matching of Dynamic Shape Manifolds for Human Action Recognition," in *IEEE Transactions on Image Processing*, vol. 16, no. 6, pp. 1646-1661, June 2007.
[55] D. P. Huttenlocher, J. J. Noh and W. J. Rucklidge, "Tracking non-rigid objects in complex scenes," *1993 (4th) International Conference on Computer Vision*, Berlin, 1993, pp. 93-101.
[56] A. Mondal, S. Ghosh and A. Ghosh, "Efficient silhouette based contour tracking," *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Mysore, 2013, pp. 1781-1786.
[57] J. Schiel and R. Green, "Adaptive human silhouette extraction with chromatic distortion and contour tracking," *2013 28th International Conference on Image and Vision Computing New Zealand (IVCNZ 2013)*, Wellington, 2013, pp. 288-292.
[58] A. M. Baumberg and D. C. Hogg, "An efficient method for contour tracking using active shape models," *Proceedings of 1994 IEEE Workshop on Motion of Non-rigid and Articulated Objects*, Austin, TX, 1994, pp. 194-199.