Detection and Tracking of People based on Computer Vision

B.VENBA[1]

[1]Assistant Professor, PG-Department of Computer Science, Saradha Gangadharan College, Puducherry, India.

S.SATHYA[2]

[2]Assistant Professor, Department of Computer Applications, Saradha Gangadharan College, Puducherry, India.

Abstract:

Computer vision is allied with the automatic extraction, psychoanalysis and comprehending useful information from a specific image or from the image series. It includes the growth of a theoretical and algorithmic base so as to attain automatic visual understanding. Identifying humans and evaluating their behavior by vision is vital for the computer systems nowadays so as to intermingle intelligently in a few human colonized environments. The visual surveillance systems are employed for the real-time scrutiny of targets like persons or vehicles, which leads to the depiction of activity of the object in that environment. Visual surveillance has been exercised for security scrutinizing, anomaly detection, intruder detection, traffic flow measuring, accident recognition on highways and routine continuance in nuclear amenities. Detecting and tracking of human beings from a video is the primary phase towards psychiatry and prediction of their actions and intention. The hitch of multiple object tracking is more knotty and exigent than the single object tracking owing to the abrupt appearance and disappearance of object. Though human detection has been investigated thoroughly in the computer vision, many conventional techniques are inappropriate for detecting targets of great variance in their exterior look. As a result, the robust human detection still remains a challenging problem because of the vastly articulated body postures, occlusion, and background clutter and viewpoint variations. In this paper, a 360-degree view is presented on the assessment of the diverse people tracking techniques.

Keywords: people tracking, human detection, multiple object detection, moving object, background subtraction.

1. Introduction:

The entire information the human observe, greater than 80% can be attained visually. The computer includes the capability like the human being is the aim of populace effort. While the computer does not have the capability to function proficiently, flexibly and skillfully like the humans, the people attain their aim gradually by means of the hard work of pollsters at abode and overseas. The computer vision, expert system and natural language are the main lively queries in the technology of artificial intelligence. By the evolution of modern science, technology and the growth of economic level, technologies in accordance with the computer vision merge every corner of our life steadily.

The computer vision technology is employed in several occasions like the intelligent monitoring scheme, perceptual interface, motion analysis and so on (Li, 2015). A vast open confront in the computer vision is to construct a system which dynamically tracks the people. A consistent elucidation exposes incredible probabilities from human computer interfaces to the video data mining to computerized observation. This chore is tricky as people can shift rapidly and randomly, can emerge in a range of poses and costume which repeatedly bounded by clutter. Consequently, several realistic computer vision systems presume a permanent camera setting that builds the object detection method more simple.

The increase in the video camera recorders of the normal investigation systems crowded communally the human operatives and the storage devices by huge data volumes and equipped, it impossible to guarantee appropriate observation of sensitive regions for elongated times. To sieve out the redundant information created by an array of cameras and amplify the response time to forensic events, supporting the human machinist with the recognition of main events in the video with the help of smart video surveillance schemes has turn out to be a significant obligation. The building of video surveillance schemes smart entails quick, dependable and robust algorithms for shifting object detection, classification, tracking and action psychiatry (Rajkumar et.al, 2015).

Various human detection techniques and tracking techniques have been built up nowadays. Person recognition and trailing paves the way for broad range of computer vision relevance fields. Few of them includes: video scrutiny and security conspires, biometric, rule enforcement, human computer interaction, video indexing and recovery, medical imaging, robotics and augmented practicality.

Generally there includes three phases of video analysis namely:

- Object detection,
- Object classification and
- Object tracking.

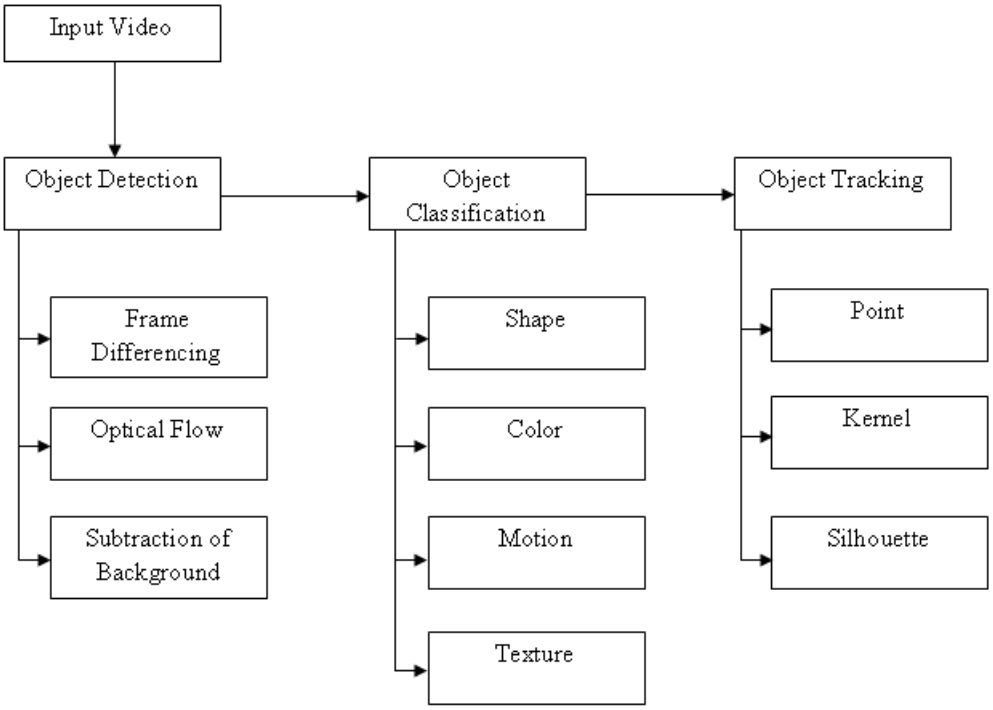

Figure.1: Classification of Techniques in Moving Object Tracking

2. Review of Literature

2.1 Object Detection Techniques

The purpose of object detection is to distinguish the objects in the sequence of videos and to huddle the pixels of those objects. Human detection is one of the main exigent troubles in computer vision owing to the huge variations of articulation, viewpoint and manifestation (Schiele et.al, 2009), video camera position, broad range of poses espoused by human beings, alterations in brightness, illumination, contrast levels or backgrounds and human occlusions. Few noteworthy investigations has been dedicated for placing and pursuing humans in static and dynamic images, as several applications includes position and movement of the human beings.

2.1.1 Frame Differencing

An automatic multiple person detection and tracking scheme for still camera video succession has been intended by (Tudor Barbu, 2014). Initially, the movie has to be pre-processed prior to carrying out the needed video psychiatry on it. Few PDE-based noise exclusion and restoration approaches have been employed which assists the object detection process. The multiple moving person detection method includes two main stages. The primary phase epitomizes a foreground segmentation of the sifted video sequence, whereas the detection of the objects symbolizes the persons is moved to the subsequent stage. The video frames of the analyzed sequence are altered into grayscale and the video movement of this frame sequence then undergoes evaluation via temporal differencing approaches. Subsequent to the detection stage, the identified video objects signifies the pedestrians and found out by assessing various circumstances regarding the human parts and detect the regions of skin from the movie frames. An excellent human tracking approach via the Histogram of Oriented Gradient based template matching process is then employed. This human detection and tracking scheme is then subjected to testing on several video datasets including moving people. The object detection rate, attained by the projected temporal differencing based technique holds high rate of around 90%.

An enhanced technique is introduced by (Ruiyue et.al, 2015) to identify and track multiple heads by deliberating them as stiff body parts. The exterior form of human heads is rationalized in accordance with the fusion of color histogram and oriented gradients. An associative mechanism of detection and tracking has been built to recuperate the transient missed detections and repress the transient false detections. The object identity can be set as invariant at the time of tracking although when inevitable occlusion takes place. Moreover, the intended approach is speedy to identify and track multiple human in a dynamic scene devoid of any hypothesis for the situation contents earlier.

To detect a moving object inside the current frame (Matteucci et.al, 2006) devised a hybrid method that utilizes the frame by frame difference as well as the background subtraction for video surveillance applications. They reveal that a mutual utilization of frame by frame difference with a background subtraction technique facilitates to have a sturdy and fast pixel foreground classification with no need of the previous background learning. The joint difference algorithm employs frame difference information to rectify pixels classification done by means of a background subtraction technique whilst selectively modernizing the background model in accordance with such classification.

Lee et al in the year 2015 have lectured a proficient moving object segmentation approach via motion orientation histogram in adaptively divided blocks. In chase of effectiveness and consistency, each fundamental block motion vector was quantized into one of several signifying alignments. The initial shape of a moving object was extracted using frame difference and LBP algorithm. They partitioned the variable-size blocks in accordance with the initial shape for removing outlier motions. They analyzed the direction of one or more moving objects using the motion orientation histogram (MOH). The proposed moving object segmentation method is a reasonable alternative to existing state-of-the-art background modeling-based methods for high-definition video surveillance systems.

Technique	Method used	Merits	Demerits
Chun-Ming Tsai et.al, 2013	Adaptive frame differencing	High performance outcomes, less computation	Pass over frames to enhance the speed of computation which paves the way for negotiating vital frames
Bahadir Karasuluand Serdar Korukoglu, 2011	Frame differencing for subtraction, entropy based simulated annealing	• Accurate segmentation of frames, • Frame differencing rate to enhance the speed of segmentation • Superior performance based upon precision and recall	Cost function in simulated annealing is classy
Gopal Thapa, 2014	Differencing and summing technique	• Simple and low in calculation complexity, • Proficiently recognizes the moving objects and segments them from the static background	Does not consider the shadows of the moving objects, as a result they are also segmented as moving objects if they are large in size as related to the threshold value
Bing Leng and Qionghai Dai, 2007	Accumulate based frame differencing, thresholding	Low error rate in segmentation	Post processing occasionally requires more time
Spagnolo et.al, 2006	Classify the points in each image as moving or static using temporal analysis, background updating	• Threshold selection is better by means of temporal analysis • Appropriate for both indoor and outdoor environments	Fixed threshold in background updating Endures from the trouble of segmenting shadows by deliberating as an object
Tudor Barbu, 2014	Temporal differencing, HOG-based template matching technique	High performance rate in object detection and appreciably less computation time	Performs fairly inferior than the methods utilizing HOGs in the identification stage for front view pedestrian detection Rear view pedestrian detection and partly occluded pedestrian detection
Yifeng Liu, 2016	Weighted aggregation, binarized normed gradients, bootstrapped SVM	High accuracy although of slight deformation, partial occlusion, and complex background	Not real-time owing to the computational complexity of HASP descriptor

Mustafa Oral and Umut Deniz, 2007	Moment and centre of mass	• The model is resistant to for noise and sudden illumination variations owing to the effect of ‘‘noise’’ on the location of CoMs of the image regions • Simple to implement • Processing speeds are remarkably less	The exploit of standard deviation of the preceding positions as threshold value in decision mechanism outcome with the insensitivity to noise and intensity variations in regions. Hence alteration mechanism is employed in changing weather circumstances. This leaves holes within large moving objects. In order to evade this, post processing is performed which takes more time

2.1.2 Optical Flow

It gives the uprooting of picture elements in two back to back pictures of one camera, and relies on upon the movement of spectator and moreover the movement of the concerning 3D notion. Direct optical flow examination gives quick identification result.

Technique	Methods	Merits	Demerits
Jeongho, 2005	Non-prior training (NPT) active feature model (AFM),Principal component analysis	Real-time, Robust tracking system	The performance of tracking is based upon the object’s motion. In case of dealing with multiple objects with comparable motion, the tracking algorithms will not succeed.
Xingzhi Luo And Suchendra, 2007	elastic matching algorithm, Kalman filter	Very proficient in treating occlusions and variations in scale and illumination	Appropriate foreground or background model required to build apposite for long term tracking, a single object hypothesis is not adequate to treat all probable tracking scenarios.
Marek Schikora, 2011	Particle filter	Capable to identify the human beings in occlusion case and dealt with false siren	Less processing time for the tracking could be attained via parallel execution of particle filter on GPU.
Ningning, 2015	Mixture Gaussian model, Lucas-Kanade optical flows	It can offer real-time and robust multi-object tracking in the occluded areas	This method cannot attain accuracy in target tracking when the objects are moving in the same direction and sheltered.

2.1.3 Background Subtraction:

It is a widespread approach employed for moving object segmentation owing to its computational simplicity; however it often paves the way for erroneously segmented regions due to noise and flicker dynamic illumination. Tao et.al in the year 2008 presented a model based technique to construe the image observations via multiple, partially occluded person hypotheses inthe frame work of Bayesian. They describedthe joint image likelihoodfor multiple humans depending upon the appearance of humans,the visibility of body acquired by means of occlusion reasoning, and foreground or backgroundseparation. The optimal resolution is achieved with the help of a competent sampling technique, data-driven Markov chainMonte Carlo (DDMCMC) that employs image observations forproposal probabilities. The knowledge of several facets together withhuman shape, camera model and image cues are incorporatedin one hypothetically sound structure.

The mainconfront faced by the algorithms intended for multipleobjects tracking is the identity switches that happenamid tracked objects owing to occlusions andinteractions involving these identical objects. (Merad et.al, 2016) devised an approach to exploit a re-identification approachso as toavertthese identity switches. This re-recognition method depends upon the segmentation of the detected individuals intohead, chest and legs in addition to the categorization of their appearances into the frontage and backside positions. Thisre-recognitioncomponent is incorporatedinside our tracking system to fuse tracklets acquired from a particlefilter based tracking structurein a mono-camera tracking scheme. The grouping of those trackingand re-recognition components lets the revival of worldwide routes for the tracked individuals.

2.2 Object Classification Techniques

Object characterization endeavors to distinguish the sort of object(s) being watched and consequently it can add urgent data to video investigation. Object characterization hence takes into account more nitty gritty depiction of occasions or practices. Besides it expands identification exactness and it makes abnormal state metadata that explain the video grouping. So far numerous object characterization methods have been found for characterization of individuals in video.

2.2.1 Appearance based classification:

These methods involve the extraction of visual features which bridge the gap between low level appearance features and higher level semantic features. This model can be further divided into two main sub-models depending on the types of appearance features used, these are: Appearance Models with Local Features, and Appearance Models with Global Features. One other model that mostly combines between the concepts of global and local features and that is Appearance models with textural features. The Features describe the change in face texture when particular action is performed such as wrinkles, bulges, forefront, regions surrounding the mouth and eyes. Image filters are used, applied to either the whole face or specific regions in a face image to extract a feature vector. As per the study appearance based algorithms are of wide-range. These include Principal Component Analysis (PCA), Independent Component Analysis (ICA), Locality Preserving Projections (LPP), Linear Discriminate Analysis (LDA), Gabor wavelets, Local Binary Pattern (LBP).
In Alvaro García-Martin *et.al* (2012) they cover two different aspects: people detection and tracking. A whole detection/ tracking system that integrates appearance, motion and tracking information is presented. This system uses the information provided by each of the independent tasks to improve the final result of the system. The tracking information is integrated in the detection task improving the detection results and vice versa. In Julian *et.al* (2015) have proposed a novel method for keeping track of multiple objects in provided regions of interest, i.e. object detections, specifically in cases where a single object results in multiple co-occurring detections. Their method identifies a minimal set of objects to explain the observed features, which are extracted from the regions of interest in a set of frames. Focusing on appearance rather than temporal cues, we treat video as an unordered collection of frames, and ‘‘unmix’’ object appearances from inaccurate detections within a Latent Dirichlet Allocation (LDA) framework, for which we propose an efficient Variational Bayes inference method. After the objects have been localized and their appearances have been learned, we can use the posterior distributions to ‘‘back-project’’ the assigned object features to the image and obtain segmentation at pixel level. Yixiao *et.al* (2016) e main novelties of this paper include: (1) representing the dynamic appearance, shape and motion of a target person each being points moving on a different Riemannian manifold; (2) characterizing the dynamics of different features by computing velocity statistics of their corresponding manifold points, based on geodesic distances; (3) employing a feature weighting approach, where each statistical feature is weighted according to the mutual information; (4) fusing statistical features learned from different manifolds with a two-stage ensemble learning strategy under a boosting framework. Costantino *et.al* (2011) they described the AD-HOC tracking system, a complete approach for multiple people tracking in video surveillance applications. In particular, our effort was focused on overcoming large and long-lasting occlusions by using an appearance driven tracking model. The main novelty of the system was the classification of non-visible regions into three classes, which aims at distinguishing between actual occlusions, occlusions with an object belonging to the background, and shape changes. Therefore, based on classification results, a different behavior can be adopted to keep memory of the occluded parts of each object and to recover them once they appear again. Min yang *et.al* (2016) have proposed tracking was very robust and fast; it has been adopted in several projects of indoor and outdoor people surveillance, with many people and real operating conditions. propose a feature selection algorithm to describe the appearance variations with mid-level semantic features, and demonstrate its usefulness in terms of temporal dynamic appearance modeling. Moreover, the appearance model was learned incrementally by alternatively evaluating newly-observed appearances and adjusting the model parameters to be suitable for online tracking. Reliable tracking of multiple persons in complex scenes was achieved by incorporating the learned model into an online tracking-by-detection framework. Dai congxia *et.al* they presented an approach toward pedestrian detection and tracking from infrared imagery using joint shape and appearance cues. A layered representation is first introduced and a generalized expectation-maximization (EM) algorithm is developed to separate infrared images into

background (still) and foreground (moving) layers regardless of camera panning. In the two-pass scheme of detecting pedestrians from the foreground layer: shape cue is first used to eliminate non-pedestrian moving objects and then appearance cue helps to locate the exact position of pedestrians. Templates with varying sizes are sequentially applied to detect pedestrians at multiple scales to accommodate different camera distances. To facilitate the task of pedestrian tracking, we formulate the problem of shot segmentation and present a graph matching-based tracking algorithm that jointly exploits the shape, appearance and distance information. The merits and demerits of the methods are given below.

Technique	Methods	Merits	Demerits
Alvaro García-Martin and Jose M. Martinez, 2012	Implicit Shape Model	High detection rate	Takes more time for computation
Jianguo Zhang and Shaogang Gong, 2009	Integral gradient orientation histogram map	Better detection rate	Requires adaboost training to increase the computation speed.
Julian et.al, 2015	Variational Bayes inference	Enhanced identification outcomes based upon tracked and lost identities, and of identity switches, related to both global batch-mode and on-line multi-view state-of-the-art trackers	Temporal motion restraints with dynamics for the appearance model, eliminating the prior target upper-limit K, and by support for multiple object classes should be enhanced
Yixiao Yun and Irene Yu-HuaGu, 2016	The shape descriptor is based on the histogram of oriented gradient, histogram of optical flow	High detection rate	Degraded specificity, processing speed ought to be enhanced
Costantino, 2011	Appearance driven tracking	Very robust and fast, appropriate for indoor and outdoor people surveillance	Splitting group of people should be rectified, identity detection needs to be enhanced
Min Yang and Yunde Jia, 2016	Temporal dynamic appearance model (TDAM) by means of Hidden Markov Model	Improved performance when person is in closer appearance	Performance should be enhanced by improving feature selection approaches
Congxia, 2007	expectation-maximization,	Better performance rate	Ought to be enhanced for work under dissimilar environments

2.2.2 Color Based classification:

Unlike many other image features (e.g. shape) color is constant under viewpoint changes and also easy to be acquired. Though color is not always appropriate as the sole means of detecting objects and tracking these objects, but the low computational cost of the algorithms proposed makes color a desirable feature to be used when necessary. To detect and track objects or pedestrians in real-time color histogram based technique is used. Mixture Model is created to describe the color distribution within the sequence of images and to segment the image into background. Object occlusion was handled using an occlusion buffer. In Rafael *et.al* (2007) they presented a system able to visually detect and track multiple people using a stereo camera placed at an under-head position. That camera position was especially appropriated for human–machine applications that require interacting with people or to analyze human facial gestures. The system models the background as height map that was employed to easily extract foreground objects among which people are found using a face detector. Once a person has been spotted, the system is capable of tracking him while is still looking for more people. their system tracks people combining color and position information (using the Kalman filter). Tracking based exclusively on position information is unreliable when people establish close interactions. Thus, we also include color information about the people clothes in order to increase the tracking robustness. Daniel Rowe *et.al* (2007) improves tracking by means of particle filtering, where occlusions are handled considering the target's predicted trajectories. Model drift is tackled by careful updating, based on the history of likelihood measures. A colour-based likelihood, computed from histogram similarity, was used. Chandrajit M *et.al* (2016) they presented a feature based method to track the multiple moving objects in surveillance video sequence is proposed. Object tracking was done by extracting the color and Hu moment's features from the motion segmented object blob and establishing the

association of objects in the successive frames of the video sequence based on Chi-Square dissimilarity measure and nearest neighbor classifier. Pashalis *et.al* (2013) they collected and updated a representation of the color appearance of the persons in the environment. The combination of volumetric and color information reinforces tracking robustness, even when a person is not visible by any of the cameras for extended time intervals.

Technique	Methods	Merits	Demerits
Rafael et.al, 2007	Hungarian Method	Lack of registration of false detections in the performed tests.	Architecture should be enhanced for real time purposes
Daniel et.al, 2007	particle filter	sampling impoverishment and mutual occlusions are dealt by redefining the weight normalization and envisaging and handling occlusions	Disambiguation of targets from clutter should be enhanced
Chandrajitet.al, 2016	Chi-Square dissimilarity measure	Robust and efficient	Occlusion handling strategy and shadow purging ought to be taken into consideration.
Pashaliset.al, 2013	volumetric representation	Robust and efficient	Should be enhanced for dynamically varying background

2.2.3 Texture Based classification:

Texture is an innate property of virtually all surfaces, the grain of wood, the weave of fabric, the pattern of crop in fields, etc. It contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment. Since the textural properties of images appear to carry useful information, for discriminating purpose features have always been calculated for textures although it is quite easy for a human observer to recognize and describe in empirical terms, texture has been extremely adverse to precise definition and analysis by computer. Texture is represented by means of texture descriptors. Xuan *et.al* (2016) have proposed new methods based on fusion of textures, angle histograms and color moments to find a specific person. Texture features of three parts are extracted by means of this network. Back propagation neural network, multi-class SVM and KNN are used as classifiers. For improving recognition rate, different fusion methods have been studied such as the fusion of texture features and other features in three body parts, and decision fusion using voting mechanism, probability combination etc. Xuewei Shen *et.al* represented the target with perceptual hashing (PH) and color self-similarity at patch level, and design an adaptive patch-based similarity measurement which simultaneously considers the spatial and temporal appearance information of the target. Hajer *et.al* (2015) they presented a method to enhance human detection and tracking in crowded scenes. It was based on introducing additional information about crowds and integrating it into the state-of-the-art detector. Ahmed *et.al* (2014) they proposed a novel tracking algorithm based on PSO method. Over 30 video sequences and having over 20,000 frames are used to test the developed PSO-based object tracking algorithm and compare it to classical object tracking algorithms as well as previously published PSO-based tracking algorithms. GustavoFühr *et.al* (2014) they presented a new approach for tracking multiple people in monocular calibrated cameras combining patch matching and pedestrian detection. Initially, background removal and pedestrian detection are used in conjunction with the vertical standing hypothesis to initialize the targets with multiples patches.

Technique	Method	Merits	Demerits
Xuan et.al, 2016	Neural Network, multi-class SVMand KNN, Spiking Neural Network	Better performance by using serial featurefusion and probability fusion	Consume more time
Xuewei, 2016	perceptual hashing, Particle Filter	Robust, better in shadow removal	Background subtraction should be improved
Hajer, 2015	local crowd densityand geometrical correctionfilters	Training is not needed, effective detection rate	contextual information should be considered
Ahmed, 2014	PSO,Similarity (SIM) and Bhattacharyya coefficient	Efficient in occlusion handling	Time should be minimized

Gustavo, 2014	Patch matching, weighted vector median filter	Handle short term occlusions	Cost should be considered

2.3 Object Tracking Techniques:

It is a significant and challenging chore in the field of computer vision. The rise of high-poweredcomputers, the availability of high quality and cheap cameras, and the growing requirement for the automated videoanalysis has sourcedan immensecuriosity in the object tracking algorithms. The aim of Object tracking is to create the route for an object over time by finding its position in video sequences.Object tracking can be categorized as a) pointtracking, b) kernel based tracking and c) silhouette based tracking.

Human detection and tracking in movie sequences turn out to be a greatinterest in the field of computer vision inearlier decades which signifies the chiefsignificant subclass ofthe moving object detection and the tracking research field.The Human detection chore includes recognition of the existence of human in video streams and discriminates humans from thenon-human video objects. Human tracking traces the occurrence of each detected person in frames of the assessed movie (Tudor, 2012).

Babaee and Rigoli, (2016) introduced a method that traces superpixels in place of detection boxes in multi-view video sequences. Particularly, we initially haul out the superpixels from detectionboxes and then relate them inside each detection box, over various views and time steps thatinitiate a collective segmentation, reconstruction and trailing the superpixels. We build a flow graphand incorporate both the visual and geometric cues in a global optimization structure to lessen its cost

A hybrid technique is devised by Nagaraju et.al for the segmentation of gradually moving objects in video sequence intending to attain perceptually reliable outcomes. Initially, the temporal information of the differences amid multiple frames is used to identify the initial moving regions. After that, the Gaussian mixture modelis used and an enhanced expectationmaximization technique is instituted to segment a spatial image into homogeneous regions. Lastly, the outcomes of motiondetection and spatial segmentation are combined to extort the final moving objects.

Tracking the humans from thevideos is significant in the field of surveillance. Xue et.al in the year of 2016 have deliberated the hitch of human tracking in RGBD videos filmed by sensors like MS Kinect and Primesense. Their objective was to track the humans where the mass of people was known previously or every human in the video have emerged in the very beginning. Consequently they trained a classifier to aid in categorizing and tracking the persons in the video. A deeplearning form the trained with huge data has been confirmed to be a competent classifier for variousobjects. They intended to train a deep convolutional neural network that enhances the performance of tracking and to categorize people. Moreover, a motion model in accordance with the spatial and kinetic clues is pooled with the networkto track people in the scene. They revealed the efficiency of our approach by assessing it on various datasets and comparedit with that of the conventional approaches like SVM.

2.3.1 Point Based Tracking:

The objects could be successfully depicted by a couple focuses which facilitate the way toward following in video. For reason for protest following is characterized by its position and by a descriptor. Position of points could be chosen by cursor point locator. Descriptor is highlight vector with depicts appearance of little neighborhood of point. The descriptor could contain two or three unique elements. In Mahdi Hashemzadeh and Nacer Farajzadeh, et.al 2016, they presented the estimate of crowd count, the combination of keypoint-based and segment-based (foreground) features. Various features are extracted from each foreground segment together with the corresponding keypoints which are highly correlated with the size, density, and occlusion level of the crowd. Finally, a combination of the segment-based and keypoint-based features is used to estimate the number of people in crowds. In Huiyu, *et.al* 2009 they presented a scale invariant feature transform (SIFT) based mean shift algorithm is presented for object tracking in real scenarios. SIFT features are used to correspond the region of interests across frames. Meanwhile, mean shift is applied to conduct similarity search via color histograms. The probability distributions from these two measurements are evaluated in an expectation–maximization scheme so as to achieve maximum likelihood estimation of similar regions. This mutual support mechanism can lead to consistent tracking performance if one of the two measurements becomes unstable

Technique	Methods	Merits	Demerits
Mahdi Hashemzadeh and Nacer Farajzadeh, 2016	Key point-based segmentation, neural network	Able to count the number of people in crowd	Time consuming, overestimating of crowd size
Huiyu, 2009	Expectation maximization, sift, mean-shift,	Better performance under various situations	In real scenarios, the performance should be improved

2.3.2 Kernel Based Tracking:

In this method the target is spatially masked with an isotropic kernel. A spatially even similarity function is described and the target localization trouble is then performed by means of a gradient based optimization approach, based on mean shift filter. This method has been demonstrated to successfully work. In Wei Chen 2016, they present a patch based tracker which adaptively integrates the kernel correlation filters with multiple effective features. To take full advantage of the useful information from different parts of the target, we train each template patch by kernel correlation filtering method, and adaptively set the weight of each patch for each particle in a particle filtering framework. In Hong, 2012 they proposed a constrained optimization approach to improving both the robustness and accuracy of kernel tracking which was appropriate for real-time video surveillance due to its low computational load. Typical tracking with histogram-wise matching provides robustness but has insufficient accuracy, because it does not involve spatial information. The merits and demerits are given below

Technique	Methods	Merits	Demerits
Wei Chen et.al, 2016	Patch based tracking,adaptive weight selection,HOG and CN	handles the occlusion troubleeffectively patch selection by means of adaptive weight selection	High calculation time.
Hong, 2014	Histogram wise weight matching, pixel wise weight matching, Lucas-Kande	Suitable under occlusion changes, robust	Need to work under poses

2.3.3 Silhouette based Tracking

The silhouettes are anything but difficult to separate giving important data about the position and state of the individual. At the point when the camera is static, background subtraction procedures can give high precision measures of human silhouettes by displaying and redesigning the background picture. In Troung et.al 2010, relied on the spectral classification of the appearance-based signatures extracted from the detected person in each sequence. they proposed a new feature called ‘‘color-position’’ histogram combined with several illumination invariant methods in order to characterize the silhouettes in static images. Then, they develop an algorithm based on spectral analysis and support vector machines (SVM) for the re-identification of people. In Nicholas R. Howe *et.al* 2007 they described a simple yet effective algorithm for tracking articulated pose, based upon looking up observations (such as body silhouettes) within a collection of known poses. The new algorithm runs quickly, can initialize itself without human intervention, and can automatically recover from critical tracking errors made while tracking previous frames in a video sequence. In Madrid-cuevas *et.al* 2012, they proposed a novel approach to solve Shape-from-Silhouette (SfS) with inconsistent silhouettes from an octree based perspective. The inconsistencies are dealt by means of the Dempster–Shafer (DS) theory and we employ a Butterworth function for adapting threshold values in each resolution level of the octree. In Rafael *et.al* 2012 they presented two main novelties. First, a three-dimensional reconstruction algorithm that is especially appropriate for people detection and tracking because of its robustness to errors in the background subtraction. Second, a new technique for creating plan-view maps from volumetric reconstructions that reduces the amount of false positives in the occluded regions. The merits and demerits are shown below

Technique	Method	Merits	Demerits
Truong et.al, 2010	Color-position histogram, Support Vector Machine	Produces Better Results	• Dimension of features should be reduced • Need to improve the performance under occlusion, partial detection
Nicholas R. Howe, 2007	Markov Chain model	Pose is considered	Performance should be improved
Madridet.al, 2012	Octree-based Shape-from-Silhouette (SfS) algorithm	Simple and posses high performance, did not require specifyingpriors or conditionals	Computation time should be minimized
Rafael et.al, 2012	Plan-view map,multiple particlefilter	Robust in segmentation, reliable in detection	Computationally more expensive, computation time is high

3. Discussion

From the above mentioned literature review we conclude that detection and tracking of moving objects from a video is a basic and risky task since isolating required object from a video scene is necessary for further process of surveillance. The process of differentiating the interested object from other background objects also became a typical problem. Hence it becomes a dynamic condition to understand videos and its components. We have analyzed the various merits and demerits of many techniques used by many researches mention in the review. There are many methods or algorithms those are aimed for the simple Object Detection and Tracking. Even though they are proposed for different purposes like to deal with the noise or to be worked on specific images (e.g. Grayscale images) or to reduce complexity, etc; there are still some limitations to the implementation of these various methods or algorithms. Many of the algorithms are concentrated to solve just the object detection problem or just the tracking problem independently. Though they are sufficient to solve their intended problems, all of them cannot solve all the problems. So it is better to solve maximum number of problems by implementing the combination of algorithms to obtain optimal results. Considering the problems rose in the field of people detection and tracking, the researches need to find and evaluate the appropriate solutions for the corresponding problems discussed above in this review.

4. Conclusion

Detecting the persons and analyzing their behavior by means of visualization is a prime factor for the computer systems now a days to interact cleverly in a few human populated areas. Visual surveillance systems are used for the real time surveillance of targets like persons or vehicles will lead to the description of objects' activities in that environment. Visual surveillance has been used for security observation, anomaly recognition, interloper detection, computing traffic flow, mishap detection on the highways and scheduled maintenance. We have provided a detail survey on people detection and tracking from 2006 to 2016. So far many techniques have been proposed. The proposed techniques have several advantages while detection and tracking but also have some disadvantages. So there is need of efficient technique to detect and track people with higher performance rate and less computation cost.

References:

[1] Li, Jiude. "Research On Camera-Based Human Body Tracking Using Improved Cam-Shift Algorithm." International Journal on Smart Sensing & Intelligent Systems 8, no. 2 (2015).
[2] Rajkumar, Macha, and Eliyaz Mahammad. "Development of Image Processing based Human Tracking and Control Algorithm for a Service Robot." (2015).
[3] Schiele, Bernt, Mykhaylo Andriluka, Nikodem Majer, Stefan Roth, and Christian Wojek. "Visual people detection: Different models, comparison and discussion." In IEEE International Conference on Robotics and Automation. 2009.
[4] Barbu, Tudor. "Pedestrian detection and tracking using temporal differencing and HOG features." Computers & Electrical Engineering 40, no. 4 (2014): 1072-1079.
[5] Xu, Ruiyue, Yepeng Guan, and Yizhen Huang. "Multiple human detection and tracking based on head detection for real-time video surveillance." Multimedia Tools and Applications 74, no. 3 (2015): 729-742.
[6] Migliore, Matteucci and Naccari, M., 2006, October. A revaluation of frame difference in fast and robust motion detection. In Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks (pp. 215-218). ACM.
[7] Lee, Seungwon, Nahyun Kim, Kyungwon Jeong, Inho Paek, Hyunki Hong, and Joonki Paik. "Multiple moving object segmentation using motion orientation histogram in adaptively partitioned blocks for high-resolution video surveillance systems." Optik-International Journal for Light and Electron Optics 126, no. 19 (2015): 2063-2069.

[8] Thapa, Gopal; Sharma, Kalpana and MK Ghose, Moving Object Detection and Segmentation using Frame differencing and Summing Technique, International Journal of Computer Applications, vol. 102, issue 7, pp. 20-25, 2014
[9] Tsai, Chun-Ming, and Zong-Mu Yeh. "Intelligent moving objects detection via adaptive frame differencing method." In Asian Conference on Intelligent Information and Database Systems, pp. 1-11. Springer Berlin Heidelberg, 2013.
[10] Karasulu, Bahadir, and Serdar Korukoglu. "A software for performance evaluation and comparison of people detection and tracking methods in video processing." Multimedia Tools and Applications 55, no. 3 (2011): 677-723.
[11] Leng, Bing, and Qionghai Dai. "Video object segmentation based on accumulative frame difference." (2007).
[12] Spagnolo, Paolo, M. Leo, and A. Distante. "Moving object segmentation by background subtraction and temporal analysis." Image and Vision Computing 24, no. 5 (2006): 411-423.
[13] Liu, Yifeng, Lian Zou, Jie Li, Jia Yan, Wenxuan Shi, and Dexiang Deng. "Segmentation by weighted aggregation and perceptual hash for pedestrian detection." Journal of Visual Communication and Image Representation 36 (2016): 80-89.
[14] Oral, Mustafa, and Umut Deniz. "Centre of mass model–A novel approach to background modelling for segmentation of moving objects." Image and Vision Computing 25, no. 8 (2007): 1365-1376.
[15] Shin, Jeongho, Sangjin Kim, Sangkyu Kang, Seong-Won Lee, Joonki Paik, Besma Abidi, and Mongi Abidi. "Optical flow-based real-time object tracking using non-prior training active feature model." Real-Time Imaging 11, no. 3 (2005): 204-218.
[16] Luo, Xingzhi, and Suchendra M. Bhandarkar. "Tracking of multiple objects using optical flow based multiscale elastic matching." In Dynamical Vision, pp. 203-217. Springer Berlin Heidelberg, 2007.
[17] Wang, Zhijie, Mohamed Ben Salah, Hong Zhang, and Nilanjan Ray. "Shape based appearance model for kernel tracking." Image and Vision Computing 30, no. 4 (2012): 332-344.
[18] Schikora, Marek, Wolfgang Koch, and Daniel Cremers. "Multi-object tracking via high accuracy optical flowand finite set statistics." In 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1409-1412. IEEE, 2011.
[19] Wang, Ningning, Mingli Luo, and Xiaohui Luo. "Multi-object Tracking In The Overlapping Area Based on Optical Flows." (2015).
[20] Zhao, Tao, Ram Nevatia, and Bo Wu. "Segmentation and tracking of multiple humans in crowded environments." IEEE transactions on pattern analysis and machine intelligence 30, no. 7 (2008): 1198-1211.
[21] Merad, Djamal, Kheir-Eddine Aziz, Rabah Iguernaissi, Bernard Fertil, and Pierre Drap. "Tracking multiple persons under partial and global occlusions: Application to customers' behavior analysis." Pattern Recognition Letters (2016).
[22] Garcia-Martin, Álvaro, and José M. Martínez. "On collaborative people detection and tracking in complex scenarios." Image and Vision Computing 30, no. 4 (2012): 345-354.
[23] Zhang, Jianguo, and Shaogang Gong. "People detection in low-resolution video with non-stationary background." Image and Vision Computing 27, no. 4 (2009): 437-443.
[24] Kooij, Julian FP, Gwenn Englebienne, and Dariu M. Gavrila. "Identifying multiple objects from their appearance in inaccurate detections." Computer Vision and Image Understanding 136 (2015): 103-116.
[25] Yun, Yixiao, and Irene Yu-HuaGu. "Human fall detection in videos via boosting and fusing statistical features of appearance, shape and motion dynamics on Riemannian manifolds with applications to assisted living." Computer Vision and Image Understanding 148 (2016): 111-122.
[26] Vezzani, Roberto, Costantino Grana, and Rita Cucchiara. "Probabilistic people tracking with appearance models and occlusion classification: The ad-hoc system." Pattern Recognition Letters 32, no. 6 (2011): 867-877.
[27] Yang, Min, and YundeJia. "Temporal Dynamic Appearance Modeling for Online Multi-Person Tracking." Computer Vision and Image Understanding (2016).
[28] Dai, Congxia, Yunfei Zheng, and Xin Li. "Pedestrian detection and tracking in infrared imagery using shape and appearance." Computer Vision and Image Understanding 106, no. 2 (2007): 288-299.
[29] Munoz-Salinas, Rafael, Eugenio Aguirre, and Miguel García-Silvente. "People detection and tracking using stereo vision and color." Image and Vision Computing 25, no. 6 (2007): 995-1007.
[30] Rowe, Daniel, Ivan Huerta, Jordi Gonzàlez, and Juan J. Villanueva. "Robust multiple-people tracking using colour-based particle filters." In Iberian Conference on Pattern Recognition and Image Analysis, pp. 113-120. Springer Berlin Heidelberg, 2007.
[31] Chandrajit, M., R. Girisha, and T. Vasudev. "Multiple Objects Tracking In Surveillance Video Using Color And Hu Moments."
[32] Padeleris, Pashalis, Xenophon Zabulis, and Antonis A. Argyros. "Multicamera tracking of multiple humans based on colored visual hulls." In 2013 IEEE 18th Conference on Emerging Technologies & Factory Automation (ETFA), pp. 1-8. IEEE, 2013.
[33] Wang, Xuan, Qingxiang Wu, Xiaojin Lin, Zhiqiang Zhuo, and Liuping Huang. "Pedestrian identification based on fusion of multiple features and multiple classifiers." Neurocomputing 188 (2016): 151-159.
[34] Shen, Xuewei, Xiubao Sui, Kechen Pan, and Yuanrong Tao. "Adaptive pedestrian tracking via patch-based features and spatial–temporal similarity measurement." Pattern Recognition 53 (2016): 163-173.
[35] Fradi, Hajer, Volker Eiselein, Jean-Luc Dugelay, Ivo Keller, and Thomas Sikora. "Spatio-temporal crowd density model in a human detection and tracking framework." Signal Processing: Image Communication 31 (2015): 100-111.
[36] Tawab, Ahmed M. Abdel, Mohamed B. Abdelhalim, and SE-D. Habib. "Efficient multi-feature PSO for fast gray level object-tracking." Applied Soft Computing 14 (2014): 317-337.
[37] Führ, Gustavo, and Cláudio Rosito Jung. "Combining patch matching and detection for robust pedestrian tracking in monocular calibrated cameras." Pattern Recognition Letters 39 (2014): 11-20.
[38] Meha Patel and Bhumika Bhat, A Comparative Study of Object Tracking Techniques, International Journal of Innovative Research in Science, Engineering and Technology,Vol.4, No.3, 2015
[39] Hashemzadeh, Mahdi, and Nacer Farajzadeh. "Combining keypoint-based and segment-based features for counting people in crowded scenes." Information Sciences 345 (2016): 199-216.
[40] Zhou, Huiyu, Yuan Yuan, and Chunmei Shi. "Object tracking using SIFT features and mean shift." Computer vision and image understanding 113, no. 3 (2009): 345-352.
[41] Chen, Wei, Kaihua Zhang, and Qingshan Liu. "Robust visual tracking via patch based kernel correlation filters with adaptive multiple feature ensemble." Neurocomputing (2016).
[42] Choi, Hong Seok, In Su Kim, and Jin Young Choi. "Combining histogram-wise and pixel-wise matchings for kernel tracking through constrained optimization." Computer Vision and Image Understanding 118 (2014): 61-70.
[43] Cong, D-N. Truong, Louahdi Khoudour, Catherine Achard, Cyril Meurie, and Olivier Lezoray. "People re-identification by spectral classification of silhouettes." Signal Processing 90, no. 8 (2010): 2362-2374.
[44] Howe, Nicholas R. "Silhouette lookup for monocular 3D pose tracking." Image and Vision Computing 25, no. 3 (2007): 331-341.
[45] Diaz-Mas, L., Francisco José Madrid-Cuevas, Rafael Muñoz-Salinas, A. Carmona-Poyato, and R. Medina-Carnicer. "An octree-based method for shape from inconsistent silhouettes." Pattern Recognition 45, no. 9 (2012): 3245-3255.
[46] Muñoz-Salinas, Rafael, Enrique Yeguas-Bolivar, L. Diaz-Mas, and Rafael Medina-Carnicer. "Shape from pairwise silhouettes for plan-view map generation." Image and Vision Computing 30, no. 2 (2012): 122-133.

[47] Babaee, M., Y. You, and G. Rigoll. "Combined segmentation, reconstruction, and tracking of multiple targets in multi-view video sequences." Computer Vision and Image Understanding (2016).
[48] Nagaraju, Sreedhar, Praneeth, Vamsi Krishna, and Gudlavalleru Krishna GEC. "A Novel Algorithm Based on Frame Difference and GMM for Segmentation of Slowly Moving Objects."
[49] Xue, Hongyang, Yao Liu, Deng Cai, and Xiaofei He. "Tracking people in RGBD videos using deep learning and motion clues." Neurocomputing 204 (2016): 70-76.