CHAPTER- 2

LITERATURE SURVEY

Multitude of Foreground Segmentation and Tracking algorithms has been proposed in past decades. A fairly comprehensive literature survey is provided here. Section 2.1 of this Chapter introduces the popular techniques in the area of foreground segmentation. Section 2.2 discusses the recent methods proposed in the area of object tracking. Section 2.3 concludes by showing the comparisons of the latest techniques with our proposed method.

2.1 In the Area of Foreground Segmentation

The basic component of many computer vision and other video analysis applications is background modeling and subtraction, which determines the performance of the various processes like real time tracking, activity analysis, visual surveillance etc. In a straightforward manner background modeling is a process of automatically obtaining a static image that contains only background from a frame sequence. More precisely a background subtraction method focuses on extracting and updating the background. It is accomplished by segmentation of foreground regions in image sequences. Due to it wide purview in different areas, an effective modeling of the background is the foremost requirement of any system. Foreground Segmentation is achieved by matching the learned background model with each new incoming video frame. However, it is difficult for most existing background models to detect foreground objects in dynamic scenes such as changing illumination, shadow movement, and trees swaying. Consequently, effective modeling scheme is crucial for foreground segmentation.

The research to update background models by building up Gaussian Model for each pixel was initially proposed by C. R. Wren [7] . In recent years, lot of work has been done in foreground segmentation. Stauffer et al. [8] proposed a robust, adaptive and flexible tracking system by modeling every pixel of the image as a mixture of Gaussians and using an on-line approximation for the updation of the generated background model. The work also proposed a method to decide about the number of Gaussians to best describe the background process.

Sheikh et al. [9] built a non parametric model by combining both spatial and temporal information. Gong et al. [10] proposed an effective stroke based algorithm to process live videos containing fuzzy boundaries where at each pixel two one-class SVM (Support Vector Machines) is trained and maintained that model local color distribution for foreground and background. It associates novel acceleration techniques to speed up the algorithm.

Guang Ye-Peng [11] proposed a novel foreground segmentation algorithm in context of real time face recognition. Background segmentation is performed by multi-scale wavelet transform. A simple automatic optimum threshold selector is employed. The main motive of real time face recognition is achieved by combining the projection matrices without iteration and Kernel Fischer discriminant analysis (KFDA).

Patrick Dickinson et al. [12] proposed a foreground segmentation framework wherein homogeneous regions of scene pixels are modeled as an adaptive Mixture of Gaussians (MOG) on color and spatial information. A probabilistic classification in order to handle significant distortion is used.

Jong-Eun Ha et al. [13] proposed a robust foreground segmentation technique using multiple difference images that require adjustment of a single parameter only. Ki Tae Park et al. (2011) discuss an automatic salient object detection technique from still image using Backward Elimination based on multi-scale segmentation. A region adjacency graph utilizes a trimap comprising of foreground, Background and undefined regions. A new graph cut based technique finally extracts foreground regions from candidates.

Rui Caseiro et al. [14] proposed a novel method of foreground segmentation using structure tensor based on data modeling by employing GMM (Gaussian Mixture Model) in a tensor domain keeping. An inline K-means approximation of Expectation Maximization (EM) algorithm is used to efficiently estimate the mixture properties.

2.2 In the Area of Object Tracking

Persistently ascertaining the motion of an intended target from the given frame sequences is the crux of a Visual Object Tracking problem. Visual object tracking deals with consistently inferring the motion of a desired target from image sequences. It is an active and fruitful research topic in the computer vision community for decades due to both its versatile applications in practice, e.g. in security surveillance, human-computer, robotics, interaction, medical imaging and multimedia applications, and diverse impacts in theory, e.g. Bayesian inference on graphical models, particle filtering, kernel density estimation, and machine learning algorithms. However, Object Tracking is a very challenging task for long-term robust tracking in unconstrained environment. An object tracking algorithm has to address the two core challenges, the computational efficiency constraint and the enormous unpredictable variations in targets due to lighting changes, deformations, partial occlusions, camouflage, quick motion and imperfect image qualities, etc.

Y. Li [15] developed a novel incremental PCA (Principal Component Analysis) algorithm. However, the conventional batch-mode PCA technique suffers from the limitations of being computationally intensive and susceptible to outlying measurement. Therefore, the authors extended the PCA to a robust PCA, which is a computationally more efficient and easy to implement technique. Its distinguishing nature is that, this updated model at each step of an incremental PCA algorithm can be used for outlier detection, i.e. given this “prototype” model, one does not need to go through the expensive iterative process, and the robust analysis can be preformed “in one go”. This method has been applied in dynamic background modeling and multi-view face modeling.

Ross et al. [16] introduced a novel method of tracking objects that does not employ fixed appearance models of the target. Such models are trained using appearance data only that is available before tracking begins in short durations and in well controlled environments. It incrementally learns a low-dimensional subspace representation by efficiently adapting online the changes in the appearance of the target using two important features, the sample mean and a forgetting factor. The estimation of locations of the target objects in consecutive frames is achieved by using a sampling algorithm with likelihood estimates. There is no constraints for the camera motion.

Allan D. Jepson et al. [17] proposed a robust, adaptive appearance model for motion-based tracking of complex natural objects. The model has better adaptability to slow varying appearance and at the same time maintains a natural measure of the stability of the observed image structure

during tracking. Weighting of the properties are done as per the importance. The online version of EM (Expectation Maximization) algorithm is used to learn efficiently the parameters of the mixture model.

The wavelet-based model allows for stability at different scales. It has proved its robustness with respect to partial occlusions, significant image deformations, and natural appearance changes.

S Gu et al.[18] proposed a novel method based on the idea of tracking by detection by using neural network as the substratal non-parametric classifier, ESS (Efficient Subwindow Search).It speculates the target locations by handling scale changes and a feature updating and pruning method is used to achieve proper balance between flexibility and stability. It relies on the fact that no training is required in neural network classifier and works well where data size is small i.e in tracking by detection scenarios.

David S. Bolme et al. [19] critically analyzed the tracking by detection methodology by pointing out that performing tracking is easier than detection, as it require few computations than running an object detector on every new frame. They then proposed a tracking method based on MOSSE (Minimum Output Sum of Squared Error) filters. the proposed method is robust to illumination variations, scale, pose, and non-rigid deformations and operate at 669 frames per second. It detects occlusion depending on peak-to side lobe ratio which then enables the tracker to stop and resume where it left off when the object reappears. Its adaptability relies on the characteristics that correlation filters are capable of tracking complex objects through rotations, occlusions and other distractions with around 20 times the rate of current state-of the art techniques.

P. Barrera et al. [20] proposed a tracking algorithm that utilizes a particle filter based on condensation algorithm in contour tracking inside an image by determining the current 3D location of a moving object and robustly tracking it which requires only color information of the object to be tracked and not any other stringent motion model. Calibrated cameras are required which uses projective geometry to forward project particles into all the frames. Closeness of the particles to their respective real 3D location of the colored object is determined by a feedback about color of such projection and the feedback from its neighbors about the closeness of the particles to the real 3D and the corresponding location of colored object. A motion model of the particles is constructed from the Gaussian random noise.

Z. Han et al. [21] proposed a new approach to obtain an effective object representation for tracking to deal with unpredictable appearance variation of the object and its background, partial occlusion and deterioration in object images. The method is based on Sample-Based Adaptive Sparse Representation (AdaSR) to achieve proficiently visual object tracking goals. In this method the object to be tracked is adaptively and compactly expressed with predefined samples.

S. Datla [22] proposed a novel methodology which aims at attaining a much lesser complex tracking system by reducing undesirable and redundant computations. The video frames are passed through a pre-processing stage which inputs to the tracking algorithm only motion detected blocks. Moreover, the preprocessing stage itself senses the frames containing little motion in the search area of the target and thus barring from further processing.

Zulfiqar Hasan Khan et. al. [23] proposed a method using a joint point feature correspondence and object appearance similarity. In point feature based tracking a candidate tracker utilizes the two set of separate points features correspondences in the foreground and in the background. The foreground features give the information of the presence of the object and the background features provide indications of the occlusions. In the objects appearance based tracking the tracker is based upon the enhanced anisotropic mean shift with fully tunable bounding box. The accumulated tracking error propagation in frames is prevented by the re-initialization process. the mean shift tracker process is learned online.

Feng Tang et. al [24] made efforts in order to solve the problem of tracking objects with changing appearance by employing novel sparse, feature based object representation named as the attributed relational feature graph. SIFT(Scale Invariant Feature Transform) features and their geometric relations are encoded in form of dynamic graph by adding new stable features while removing inactive features.

Qiang chan et. al [25] proposed a two stage object tracking method that combines a region based method and a contour based method . The object region is located using a kernel based method which is initialized using a Kalman filter and Battacharya distance respectively. The tracking precision is improved by object contour evolution using diffusion snakes.

Takashi et. al [26] presented a real time cooperative multi target tracking using group of Active Vision Agent(AVAs), which is formed by a logical model of network connected computer with an active camera. It has the capability of actively track multiple objects under complex dynamic backgrounds.

Aishy et. al [27] proposed an automatic object tracking method for the real time content oriented video applications using both object segmentation and motion estimation and reliably work in presence of noise shadows, occlusions etc.

Karthik Hariharakrishan et. al [28] proposed a fast object tracking algorithm that works with predicting the object contour using the block motion vector information and avoiding the segmentation step. In motion estimation the gap between frames is controlled by an efficient modulation scheme, The disocclusions are detected by first estimating the uncovered region from the displaced frame difference and the uncovered region thus obtained are the classified into actual disocclusions and false alarms.

Bastian et.al. [29] presented a multi-object tracking approach by object detection and segmentation as a coupled optimization problem. The successful trajectory hypothesis are fed back in order to guide object detection in future frames.

Ba Ngu et. al. [30] performed the detection of multiple objects by estimation of their states from the image observations as a joint problem. The collection of states are modeled as a random finite set in a Bayesian framework and their posterior distribution is derived from the various prior distributions of this random finite sets under the required assumptions.

Junlian et. al. [31] developed a two stage multi-object tracking method using online detection, where the two stages are the local stage which uses a particle filter to handle the particle occlusion, in order to generate a set of reliable tracklets. In global stage, temporal sliding window is used to deal with any ambiguity caused due to full object occlusion for generating a set of potential tracklets. The tracklets of both the stages are associated by Hungarian algorithm and applied on a modified pair wise tracklets association cost matrix to obtain the global optimal association.

2.3 Comparisons over the current techniques

An incessant work is going on in the fields of Foreground Segmentation and Object Tracking. A brief survey of some of the latest research done in the above fields is outlined below. Yan-Chen

et. al. [32] proposed a motion feedback based segmentation and tracking using a moving object detector kernel that has a background subtraction block with it. Background modeling is done using GMM

(Gaussian Mixture Model).An adaptive threshold depending on estimated object motion of the detected object is used for foreground segmentation. Morphology and connected component labeling is used for further improvisation. Tracking is performed by matching of color histograms and spatial location between successive frames. GMM modelling does not work well in the case *when there are fast moving objects*. And, the tracking method used is also prone to failure as the features chosen are not robust.

R. Kapoor et al. [33] proposed a real time tracking using a potential function approach. Priori knowledge of the object in the form of diffusion matrices and the current estimates of non-linear scaling and drift vector is used to make the current estimates. In a potential field the concept of attractors and repellers is used to identify the direction of motion. But, its object representation is not suitable which is a centroid and the complexity increases as the authors attempt to improve accuracy.

Junliang Xing et al. [34] proposed a multiplayer object tracking algorithm by developing a progressive observation modelling that gives strong tracking information. It uses a dual mode two-way Bayesian inference approach to perform tracking of single isolated object. And, then extends it to MOT (Multiple Object Tracking) integrally by forward filtering and backward smoothing.

Michael D. et.al [35] proposed an online multi person tracking technique by detection without using background modeling .Tracking is performed using particle filtering and greedy algorithm is used for Data association. This technique does not always provide the best optimal solution.

Jie Shao.et.al [36] proposed a real time foreground segmentation algorithm for MOT. It utilizes the spatial distribution of color and temporal distribution of pixels in consecutive frames, in order to make segmentation more effective. But the approach offers problems in the case when objects do not move for long time. Another drawback of the algorithm is that the features used are not so reliable.

Jing-Ming Guo et. al [37] used a Pixel-Bases Hierarchical Feature Adaboosting (PBHFA) , segmentation of skin color and background subtraction using codebook technique.

Michael Yu Wang et. al [38] proposed a modified Schoenemann’s method for foreground segmentation,that can handle shape and structure change of the object under consideration. But, it poses timing limitations and thus cannot be used for real time analysis.

Pan Pan . Dan Schonfeld [39] proposed an extension of first order Markov Chain Model to high order MonteCarlo Markov Chains. Posterior density is obtained by Sequential importance sampling. This approach has the drawback of higher memory requirements.

David A. Ross et.al [16] proposed an incremental vector subspace learning algorithm that utilizes the large volume of information that becomes available during tracking and thus, could model intrinsic as well as extrinsic appearance variations. But, it is not robust to drifts and also cannot handle non-linear aspects of the appearance variations.

Weiming Hu et.al [6] proposed an online incremental tensor subspace learning algorithm for foreground segmentation. The particle filter is used for tracking.

FuatCogun and A. Enis Cetin [40] proposed an amplitude invariant features i.e Cepstrum based co-difference tracker where, a minimum distance metric is used for tracking.

Min Wan et. al [41], and Bo Zhang proposed a robust pedestrian tracking algorithm based on manifold learning. As, the motion of the pedestrian can be defined by few intrinsic and low dimensional features and based on this the method has been developed. This method requires learning thus, can be used for specific applications only.

Vikas Reddy et. al. [42] proposed an undirected graph based method of foreground detection where the background initialization is carried out in Markov Random Field (MRF). Iterated Conditional modes (ICM) are employed for optimal labelling solutions. Spatial continuity as well as local temporal statistics is used in order to deal with the fast moving backgrounds as compared to the foregrounds.

Jianwei Dinget. Al [43] proposed a modified version of the above MRF based method. It is online and unsupervised and is better than existing methods as it provides good segmentation results.