A Survey of Multi Object Tracking and Detecting Algorithm in Real Scene use in video surveillance systems

Abouzar Ghasemi [#1] , C.N Ravi Kumar [*2]

Department of Computer Science & Egg Sri Jayachamarajendra College of Engineering, Mysore – 570 006, India

Abstract-*There are now large networks of CCTV cameras collecting great amounts of image data, many of which deploy Pan-Tilt-Zoom (PTZ) controllable cameras. A multi-camera and multi-sensor system has potential both for gaining improved imaging quality and for capturing more relevant details .Such a system can also cause overflow of information and confusion if data content is not analyzed in real-time. Video Analytics is the emerging technology where Computer Vision and Pattern Recognition techniques are used to filter and manage real time CCTV videos for security and intelligent monitoring. Background subtraction, object detection, object tracking, re-identification, and behavior analysis are the most important components for a Video Analytics system. The scientist has some of the cutting edge technologies in this area, which exploit recent statistical and differential geometric theories and adapt them to challenging tasks for example, individuating eye directions, tracking groups of people, re-identifying individuals in different days that take place in real case scenarios. Detecting and tracking human beings in a given scene represents one of the most important and challenging tasks in computer vision. We are interested to consider some of powerful methods in these issues for their implications in video-surveillance and driver assistance systems.*

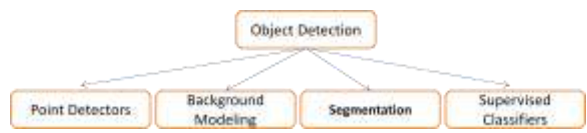
Keywords: *Multi object tracking, Multi object detection, video surveillance system.*

1. Introduction:

Before starting the description of the algorithm, we need to introduce some terminology.

1.2 What is Object Detection? Object Detection deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos. Every tracking method requires an object detection mechanism either in every frame or when the object first appears in the video. A common approach for object detection is to use information in a single frame. However, some object detection methods make use of the temporal information computed from a sequence of frames to reduce the number of false detections.

1.3 Taxonomy of Object Detection:



1.3.1 Point detectors :Point detectore are used to find interest points in images which have an expressive texture in their respective localities. Interest points have been long used in the context of motion, stereo, and tracking problems. A desirable quality of an interest point is its invariance to changes in illumination and camera viewpoint.

1.3.2 Background Subtraction: Object detection can be achieved by building a representation of the scene called the background model and then finding deviations from the model for each incoming frame. Any significant change in an image region from the background model signifies a moving object. The pixels constituting the regions undergoing change are marked for further processing. Usually, a connected component algorithm is applied to obtain connected regions corresponding to the objects. This process is referred to as the background subtraction.

1.3.3 Segmentation: The aim of image segmentation algorithms is to partition the image into perceptually similar regions. Every segmentation algorithm addresses two problems, the criteria for a good partition and the method for achieving efficient partitioning.

1.3.4 Supervised Learning: Object detection can be performed by learning different object views automatically from a set of examples by means of a supervised learning mechanism. Learning of different object views waives the requirement of storing a complete set of templates. Given a set of learning examples, supervised learning methods generate a function that maps inputs to desired outputs. A standard formulation of supervised learning is the classification problem where the learner approximates the behavior of a function by generating an output in the form of either a continuous value, which is called regression, or a class label, which is called classification. In context of object detection, the learning examples are composed of pairs of object features and an associated object class where both of these quantities are manually defined.

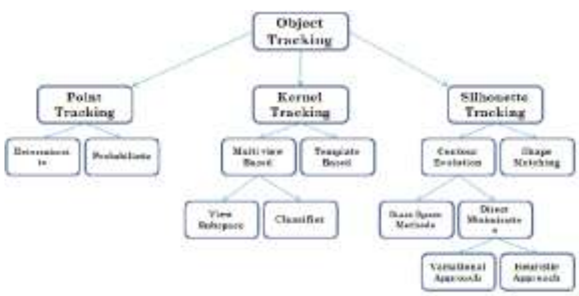
1.4 DETECTION PROBLEMS:

People detection is challenging because of the high variability of the human body appearance (different poses, clothes, viewpoints, etc.). Detecting a person in crowded scenarios (which are common in video-surveillance applications) is even harder, due to the severe occlusions that may occur.

1.5 What is Tracking Definition?

In its simplest form, tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around a scene. In other words, a tracker assigns consistent labels to the tracked objects in different frames of a video. Additionally , depending on the tracking domain, a tracker can also provide object-centric information, such as orientation, area, or shape of an object.

1.6 Taxonomy of Tracking Methods:



1.6.1Point Tracking: Objects detected in consecutive frames are represented by points, and the association of the points is based on the previous object state which can include object position and motion. This approach requires an external mechanism to detect the objects in every frame.

1.6.2 Kernel Tracking: Kernel refers to the object shape and appearance. For example, the kernel can be a rectangular template or an elliptical shape with an associated histogram. Objects are tracked by computing the motion of the kernel in consecutive frames .This motion is usually in the form of a parametric transformation such as translation, rotation, and affine.

1.6.3 Silhouette Tracking: Tracking is performed by estimating the object region in each frame. Silhouette tracking methods use the information encoded inside the object region. This information can be in the form of appearance density and shape models which are usually in the form of edge maps. Given the object models, silhouettes are tracked by either shape matching or contour evolution .Both of these methods can essentially be considered as object segmentation applied in the temporal domain using the priors generated from the previous frames.

1.7 The Object Tracking Problems:

The object tracking problem is deceptively simple to formulate: given a video sequence containing one or more moving objects, the desired result is the set of the trajectories of these objects. Unfortunately, in real world scenarios, there are several issues that make this result far from being easy to achieve:

1. Loss of information caused by projection of the 3D word on a 2D images.
2. Noise in images.
3. Partial and full object occlusions.
4. Nonrigid or articulated nature of objects.
5. Scene illumination changes.
6. Complex object motion.
7. Complex object shapes.
8. Real time processing requirement.

Because of these difficulties, many tracking algorithms have been proposed in the last years, but the problem is still considered open.

2. Tracking and Detection Algorithm Categories:

The algorithms present in the literature can be roughly divided into two categories:

- **First category**, tracking is performed after an object detection phase.
- **Second category**, detection and tracking are performed at once.

2.1 FIRST CATEGORY:

In the first one category, tracking is performed after an object detection phase. objects are detected in each frame using either some form of change detection (e.g. differences from a background model) or an a priori model of the objects. Algorithms in this category are usually faster, but they have to consider also the errors of the detection phase as spurious and missing objects, objects split into pieces, multiple objects merged into a single detected blob.

2.1.1 Sethi and Jain [1] solve the correspondence by a greedy approach based on the proximity and rigidity constraints. Their algorithm considers two consecutive frames and is initialized by the nearest neighbor criterion. The correspondences are exchanged iteratively to minimize the cost. A modified version of the same algorithm which computes the correspondences in the backward direction (from the last frame to the first frame) in addition to the forward direction is also analyzed.The demerit of this method is, it cannot handle occlusions, entries, or exits.

2.1.2 Occlusion and poor feature point detection are two of the main difficulties in the use of multiple frames for establishing correspondence of feature points. Salari and Sethi [2] handle these problems, by first establishing correspondence for the detected points and then extending the tracking of the missing objects by adding a number of hypothetical points.The demerit is the method does not address entry and exit of objects.

2.1.3 Rangarajan and Shah [3] propose a greedy approach, which is constrained by proximal uniformity. Initial correspondences are obtained by computing optical flow in the first two frames. If the number of detected points decrease, occlusion or misdetection is assumed. Occlusion is handled by establishing the correspondence for the detected objects in the current frame. For the remaining objects, position is predicted based on a constant velocity assumption. The demerit is this method does not address entry and exit of objects also.

2.1.4 Inthe work by Intille et al. [4], which uses a slightly modified version of Rangarajan and Shah (1991) for matching object centroids, the objects are detected by using background subtraction. The authors explicitly handle the change in the number of objects by examining specific regions in the image, for example, a door, to detect entries/exits before computing the correspondence. The demert is this method cannot handle occlusions, entries, or exits.

2.1.5 Veenman et al. [5] extend the work of Sethi and Jain (1987), and Rangarajan and Shah (1991) by introducing the common motion constraint for correspondence. The common motion constraint provides a strong constraint for coherent tracking of points that lie on the same object; however, it is not suitable for points lying on isolated objects moving in different directions. The algorithm is initialized by generating the initial tracks using a two-pass algorithm, and the cost function is minimized by Hungarian assignment algorithm in two consecutive frames. Merit & Demerit is this approach can handle occlusion and misdetection errors, however, it is assumed that the number of objects are the same throughout the sequence, that is, no object entries or exits .

2.1.6 Shafique and Shah [6] propose a multi frame approach to preserve temporal coherency of the speed and position. They formulate the correspondence as a graph theoretic problem. Multiple frame correspondence relates to finding the best unique path Pi = {x0, . . . , xk} for each point (the superscript represents the frame number). For misdetected or occluded objects, the path will consist of missing positions in corresponding frames. The directed graph, which is generated using the points in k frames, is converted to a bipartite graph by splitting each node (object) into two (+ and −) nodes and representing directed edges as undirected edges from + to − nodes. The correspondence is then established by a greedy algorithm. They use a window of frames during point correspondence to handle occlusions whose durations are shorter than the temporal window used to perform matching. The Merit is the proposed algorithm deals with the problems of occlusion, missed detections, and false positives by using a single noniterative greedy optimization scheme and, hence, reduces the complexity of the overall algorithm as compared to most existing approaches where multiple heuristics are used for the same purpose.

2.1.7 The w^4 system by Haritaoglu et al.[7] uses the overlap of the areas as a criterion to find a correspondence between the objects at the current and at the previous frame .When this criterion selects multiple objects, the algorithm considers split or merge hypotheses to deal with detection errors or with occlusions. After an occlusion, an appearance model of the objects is used to reassign the original object identities. Also, when an object is seen for the first time, the algorithm waits for a fixed number of frames before assigning it an object identifier, in order to filter out spurious objects due to detection errors. The demerit is the use of overlap works well with only high frame rates and objects that are not very fast.

2.1.8 The method proposed by Chen et al.[8] formulates the tracking problem as a bipartite graph matching, solving it with the well-known Hungarian algorithm. They describe a tracking algorithm to address the interactions among objects, and to track them individually and confidently via a static camera. It is achieved by constructing an invariant bipartite graph to model the dynamics of the tracking process, of which the nodes are classified into objects and profiles.The demerit is It recognizes an occlusion, but is able to preserve the object identities only if the horizontal projection of the detected blob shows a separate mode.

2.1.9 Dai et al.[9] have proposed a method able to track pedestrians by using shape and appearance information extracted from infra-red imagery. They present 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. The demerit is the method have some problems when objects quickly change their appearance or during occlusions.

2.1.10 The method by Pellegrini et al.[10] tries to predict the trajectories on the scene using a set of behavior models learned using a training video sequence. The model is trained with videos recorded from birds-eye view at busy locations, and applied as a motion model for multi-people tracking from a vehicle-mounted camera. The merit is the experiments on real sequences show that accounting for social interactions and scene knowledge improves tracking performance, especially during occlusions. The demerit is the method is very effective for repetitive behaviors, but may have some problems for behaviors that do not occur frequently.

2.1.11 The method by Ess et al.[11] uses stereo vision, coupled with a motion dynamic model and an object appearance model to perform the tracking.They propose such an approach, which jointly estimates camera position, stereo depth, object detection, and tracking. The interplay between those components is represented by a graphical model. Since the model has to incorporate object-object interactions and temporal links to past frames, direct inference is intractable. They, therefore, propose a two stage procedure: for each frame, they first solve a simplified version of the model to estimate the scene geometry and an over complete set of object detections. Conditioned on these results, they then address object interactions, tracking, and prediction in a second step. Our results show that the proposed integration makes it possible to deliver robust tracking performance in scenes of realistic complexity. The demerit is the method is not applicable where a stereo camera is not available; furthermore its computational cost is significant, requiring 0.3 s per frame only for the tracking part.

2.1.12 The method by J. Berclaz et al.[12] use the information from different cameras with overlapping fields of view in order to perform the occlusion resolution. The merit of this method is using the k-shortest paths algorithm, which is very fast. This new approach is far simpler formally and algorithmically than existing techniques and lets us demonstrate excellent performance in two very different contexts. The demerit of this method is this kind of technique is limited to the cases where multiple cameras are available; furthermore, most of the methods adopting this approach require a full calibration of each camera, which could make the deployment of the system more complicated.

2.1.12 Rosario Di Lascio [13] present a real-time tracking algorithm that is able to deal with complex occlusions involving a plurality of moving objects simultaneously. The rationale is grounded on a suitable representation and exploitation of the recent history of each single moving object being tracked. The object history is encoded using a state, and the transitions among the states are described through a Finite State Automata. In presence of complex situations the tracking is properly solved by making the FSA's of the involved objects interact with each other. The merit of this method is the way for basing the tracking decisions not only on the information present in the current frame, but also on conditions that have been observed more stably over a longer time span. The object history can be used to reliably discern the occurrence of the most common problems affecting object detection, making this method particularly robust in complex scenarios.

2.2 Second Category:

In the second category, detection and tracking are performed at once, usually on the basis of an object model that is dynamically updated during the tracking. These methods are computationally more expensive, and often have problems with the initial definition of the object models, that in some cases has to be provided by hand.

2.2.1 The paper by Comaniciu et al.[14] proposes the use of Mean Shift, a fast, iterative algorithm for finding the centroid of a probability distribution, for determining the most probable position of the tracking target. The demerit of this method is, It requires a manual selection of the objects being tracked in the initial frame, and deals only with partial occlusions.

2.2.2 Tao et al.[15] have proposed a method based on a layered representation of the scene, that is created and updated using a probabilistic framework. This paper introduces a complete dynamic motion layer representation in which spatial and temporal constraints on shape, motion and layer appearance are modeled and estimated in a maximum a-posteriori (MAP) framework using the generalized expectation-maximization (EM) algorithm. The Merit of this method performance is compared with that of a correlation-based tracker and a motion change-based tracker to demonstrate the advantages of the new method. The Demerit of their method is able to deal with occlusions, but is extremely computational expensive, requiring up to 30–40 s per frame.

2.2.3 The method by Wu and Nevatia [16] tracks people in a crowded environment. This paper proposes a method for human detection in crowded scene from static images. An individual human is modeled as an assembly of natural body parts. We introduce edgelet features, which are a new type of silhouette oriented features. Part detectors, based on these features, are learned by a boosting method. Responses of part detectors are combined to form a joint likelihood model that includes cases of multiple, possibly inter-occluded humans. The human detection problem is formulated as maximum a posteriori (MAP) estimation. The merit of this method is, method combined detection method results in better performance for individual human detection and furthermore can deal with crowded scenes. The demerit of method is, It uses an a priori model of a person, that is not extendable to other kind of objects.

2.2.4 K. Bhuvaneswari, H. and Abdul Rauf [17] proposed method in 2009. In this paper human detection is based on a silhouette oriented feature called edgelet feature. The system automatically detects and tracks possibly partially occluded humans from a single camera, which is stationary. The discriminative classifiers of objects of a known class are learnt and applied to the video sequence frame by frame. The output of the detection module is a soft decision which consists of a set of detection responses of different confidence levels. The combined detection responses provide the observations used for tracking. The responses of a multiple view detection system are taken as the observation of the human hypotheses. Trajectory initialization and termination rely on the confidences computed from the detection responses. Finally the human is tracked by mean shift style tracker. The system tracks human with inter object and scene occlusions with static or non-static backgrounds. The Merit of method is the experimentation shows better performance to detect and track human in gray level or color based images and demerit of proposed method is it does not handle total occlusions, and, because of the Kalman filter, it works better if the people are moving with uniform direction and speed.

2.2.5 The method by Yogameena et al.[18] uses a skin color model to detect and then track the faces in the scene. The task of correctly identifying and tracking people in a shadow environment for understanding the group dynamics is of paramount importance in many vision systems. This work presents a real time system for detecting and tracking people, in an environment where, people have similar attire. The proposed frame work contains shadow removal in HSV color space, detection through occlusion, person identification by developing skin color model and tracking by extracting image features. Merit & Demerit of the method is, it is able to deal with crowded scenes where the persons are dressed with very similar attire, but it works only as long as the face of each person remains clearly visible.

2.2.6 Z. Han, Q. Ye, J. Jiao[19] published paper in 2011.In this paper, proposed a combined feature evaluation approach in filter frameworks for adaptive object tracking. First, a feature set is constructed by combining color histogram (HC) and gradient orientation histogram (HOG), which gives a representation of both color and contour. Then, to adapt to the appearance changes of the object and its background, these features are assigned with different confidences adaptively to make the features with higher discriminative ability play more important roles in the instantaneous tracking. To keep the temporal consistency, the feature confidences are evaluated based on Kalman and Particle filters. The merit of method is that experiments and comparisons demonstrate that object tracking with evaluated features have good performance even when objects go across complex backgrounds. Demerit of the method is, it is not able to handle large scale changes of the target objects.

2.2.7 Yizheng Cai, Nando de Freitas, and James J. Little [20] in 2006 proposed method based on particle filter. The new particle filter framework is more suitable for tracking a variable number of targets. The rectification technique compensates for the camera motion and make the motion of targets easier to predict by the second order autoregression model. The linear optimization algorithm achieves the global optimal solution to correctly assign boosting detections to the existing tracks. Finally, the mean-shift embedded particle filter is able to stabilize the trajectory of the targets and improve the dynamics model prediction. It biases particles to new locations with high likelihood so that the variance of particle sets decreases significantly. The advantages is the experimental results show that our system is able to automatically and robustly track a variable number of targets and correctly maintain their identities regardless of background clutter, camera motion and frequent mutual occlusion between targets. The disadvantages is the computational cost is still too high for real-time applications, especially with multiple occluding targets, with a processing time ranging from 0.5 to 45 s per frame.

2.2.8Shimin Yin, Jin Hee Na, Jin Young Choi, Songhwai Oh [21] in 2011 presented a new tracking method with improved efficiency and accuracy based on the subspace representation and particle filter. The subspace representation has been successfully adopted in tracking, e.g. the Eigen-tracking algorithm, and it has shown considerable robustness for tracking an object with changing appearance. Their tracking algorithm requires a significantly small number of particles while maintaining robustness and accuracy. They proposed two methods in our tracking algorithm: first, we analyze object motion in acoarse-to-fine way and use hierarchical strategy to estimate it, in which the Kalman filter estimates global linear motion and the particle filter handles the local nonlinear motion, second, we give a more physically meaningful proposal distribution of the particle filter with consideration of the nature of motion. The merit is the experiments demonstrate the effectiveness of our tracking algorithm in real video sequences in which the target objects undergo rapid and abrupt motion. Furthermore, we provide quantitative comparisons between the existing tracking algorithm and the proposed tracking algorithm and the demerit is the computational cost is too high for real-time applications, especially with multiple occluding targets, with a processing time ranging from 0.5 to 45 s per frame.

2.2.9 Henry Medeiros et al.[22] presented a parallel implementation of a histogram-based particle filter for object tracking on smart cameras based on SIMD processors. They specifically focus on parallel computation of the particle weights and parallel construction of the feature histograms since these are the major bottlenecks in standard implementations of histogram-based particle filters. The proposed algorithm can be applied with any histogram-based feature sets they show in detail how the parallel particle filter can employ simple color histograms as well as more complex histograms of oriented gradients (HOG). The merit is, the algorithm was successfully implemented on an SIMD processor and performs robust object tracking at up to 30 frames per second a performance difficult to achieve even on a modern desktop computer and the demerit is that one of the major limitations of current approach is that the target models are not updated. Therefore, the algorithms are not robust to large variations in the appearance of the target.

2.2.10 X. Song, J. Cui, H. Zha, H. Zhao[23] presented a paper in 2008.They presented an on-line supervised learning based method for tracking multiple interacting targets. When the targets do not interact with each other, multiple independent trackers are employed for training a classier for each target. When the targets are in close proximity or present occlusions, the learned classiers are used to assist in tracking. The tracking and learning supplement each other in the proposed method, which not only deals with tough problems encountered in multi-target tracking, but also ensures all the process to be completely on-line. The advantages of this method performs better than previous methods when the interactions occur, and can maintain the correct tracking under various complex tracking situations, including crossovers, collisions and occlusions and the disadvantages of the method assume that each object enters the scene unoccluded; furthermore, it is based on the Particle Filters framework, and so it is computationally expensive.

2.2.11 M. Wang, H. Qiao, B. Zhang [24] in 2011published a paper. In this paper, a new kind of manifold subspace is introduced, in which the intrinsic features of the target's motion can be best preserved, and the dimensionality of feature is very low. Manifold learning has been a popular method in many areas such as classification and recognition. In the proposed subspace, variations of continuous pedestrian postures can be represented well by these intrinsic features. This also validates our conjecture that the movement of pedestrians can be described by some intrinsic and low-dimensional features, which are significant for tracking. Although intrinsic features are useful for tracking, algorithms that directly apply intrinsic features could not guarantee stable performance due to the influence from a complicated background. To address this issue, a foreground extraction method is introduced to enhance tracking stability by selecting the most discriminative color features to automatically distinguish the foreground from the candidate image. This preprocessing stage is proven to promote the accuracy of low-dimensional feature representation in pedestrian tracking. The whole tracking procedure, particularly dimensionality reduction, is linear and fast without complicated calculations.

2.2.12 Abdul-Lateef Yussiff et al.[25] proposed a novel approach to robust and flexible person tracking using an algorithm that integrates state of the arts techniques; an Enhanced Person Detector (EPD) and Kalman filtering algorithm. This proposed algorithm employs multiple instances of Kalman Filter with complex assignment constraints using Graphics Processing Unit (GPU-NVDIA CUDA) as a parallel computing environment for tracking multiple persons even in the presence of occlusion. A Kalman filter is a recursive algorithm which predict the state variables and further uses the observed data to correct the predicted value. Data association in different frames are solved using Hungarian technique to link data in previous frame to the current frame. The benefit of this research is an adoption of standard Kalman Filter for multiple target tracking of humans in real time. This can further be used in all applications where human tracking is needed and the demerit of method is, the parallel implementation has increased the frame processing speed by 20- 30 percent over the CPU implementation.

2.2.13 Michael D. Breitenstein et al.[26] in 2011 published paper.They propose a novel approach for multiperson tracking-bydetection in a particle filtering framework. In addition to final high-confidence detections, our algorithm uses the continuous confidence of pedestrian detectors and online-trained, instance-specific classifiers as a graded observation model. Thus, generic object category knowledge is complemented by instance-specific information. The main contribution of this paper is to explore how these unreliable information sources can be used for robust multiperson tracking. The algorithm detects and tracks a large number of dynamically moving people in

complex scenes with occlusions, does not rely on background modeling, requires no camera or ground plane calibration, and only makes use of information from the past. Hence, it imposes very few restrictions and is suitable for online applications. Our experiments show that the method yields good tracking performance in a large variety of highly dynamic scenarios,
such as typical surveillance videos, webcam footage, or sports sequences. We demonstrate that our algorithm outperforms other methods that rely on additional information. Furthermore, we analyze the influence of different algorithm components on the robustness.

2.2.14 Andreas Opelt et al. The objective of this work is the detection of object classes, such as airplanes or horses. Instead of using a model based on salient im-
age fragments, they show that object class detection is also possible using only the object's boundary. To this end, they develop a novel learning tech- nique to extract class-discriminative boundary fragments. In addition to their shape, these \codebook" entries also determine the object's centroid (in the manner of Leibe *et al.*). Boosting is used to select discrim- inative combinations of boundary fragments (weak detectors) to form a strong \Boundary-Fragment-Model" (BFM) detector. The generative aspect of the model is used to determine an approximate segmentation.We demonstrate the following results: (i) the BFM detector is able to
represent and detect object classes principally de¯ned by their shape, rather than their appearance; and (ii) in comparison with other published results on several object classes (airplanes, cars-rear, cows) the BFM detector is able to exceed previous performances, and to achieve this with less supervision (such as the number of training images).
2.2.15 Krystian Mikolajczyk et al.published a paper in 2010. In this paper they propose an approach capable of si-multaneous recognition and localization of multiple object classes using a generative model. A novel hierarchical rep-resentation allows to represent individual images as well as various objects classes in a single, scale and rotation invari-ant model. The recognition method is based on a codebook representation where appearance clusters built from edge based features are shared among several object classes. A probabilistic model allows for reliable detection of various objects in the same image. The approach is highly effi-cient due to fast clustering and matching methods capable of dealing with millions of high dimensional features. The system shows excellent performance on several object cate-gories over a wide range of scales, in-plane rotations, back- ground clutter, and partial occlusions. The performance of the proposed multi-object class detection approach is com-petitive to state of the art approaches dedicated to a single object class recognition problem.

2.2.16 Bastian Leibe published paper in 2006. In this paper, we address the problem of detecting pedestrians in crowded real-world scenes with severe overlaps. Our basic premise is that this problem is too difficult for any type of model or feature alone. Instead, we present a novel algorithm that integrates evidence in multiple iterations and from different sources. The core part of our method is the combination of local and global cues via a probabilistic top-down segmentation. Altogether, this approach allows to examine and compare object hypotheses with high precision down to the pixel level. Qualitative and quantitative results on a large data set confirm that our method is able to reliably detect pedestrians in crowded scenes, even when they overlap and partially occlude each other. In addition, the flexible nature of our approach allows it to operate on very small training sets.

2.2.17 Amir Akramin Shafie et al. published paper in 2014. Video surveillance is an active research topic in computer vision. In this paper, humans and cars identification technique suitable for real time video surveillance systems is presented. The technique we proposed includes background subtraction, foreground segmentation, shadow removal, feature extraction and classification. The feature extraction of the extracted foreground objects is done via a new set of affine moment invariants based on statistics method and these were used to identify human or car. When the partial occlusion occurs, although features of full body cannot be extracted, our proposed technique extracts the features of head
shoulder. Our proposed technique can identify human by extracting the human head-shoulder up to 60%–70% occlusion. Thus, it has a better classification to solve the issue of the loss of property arising from human occluded easily in practical applications. The whole system works at approximately 16−29 fps and thus it is suitable for real-time applications. The accuracy for our proposed technique in identifying human is very good, which is 98.33%, while for cars_ identification, the accuracy is also good, which is 94.41%. The overall accuracy for our proposed technique in identifying human and car is at 98.04%. The experiment results show that this method is effective and has strong robustness.

3. Conclusion:

The multiple target tracking has become a major application area in computer vision. This development has been driven by the many interesting applications that lie ahead in this area and the recent technological advances involving the real-time capture, transfer, and processing of images on widely available low-cost hardware platforms. A number of promising application scenarios were discussed. In real time scenarios, there are several difficulties, Because of these difficulties, many tracking algorithms have been proposed in the last years, but the problem is still considered open.

4. References:

[1]. SETHI, I. AND JAIN, R. 1987. Finding trajectories of feature points in a monocular image sequence. IEEE Trans. Patt. Analy. Mach. Intell. 9, 1, 56–73.

[2]. Feature point correspondence in the presence of occlusion SALARI, V. AND SETHI, I. K. 1990. IEEE

[3]. Establishing motion correspondence RANGARAJAN, K. AND SHAH,M. 1991. Conference Vision Graphies Image Process

[4]. Real-time closed-world tracking INTILLE, S.,DAVIS, J., ANDBOBICK,A. 1997. IEEE

[5]. Resolving motion correspondence for densely moving points VEENMAN, C., REINDERS, M., AND BACKER, E. 2001.. IEEE

[6]. A non-iterative greedy algorithm for multi-frame point correspondence .SHAFIQUE, K. AND SHAH, M. 2003 In IEEE

[7]. I. Haritaoglu, D. Harwood, L.S. David, W4: Real-time surveillance of people and their activities, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 809–830.

[8]. Hwann-Tzong Chen, Horng-Horng Lin, T.-L. Liu, Multi-object tracking using dynamical graph matching, in: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, vol. 2, 2001, pp. 210–217.

[9]. C. Dai, Y. Zheng, X. Li, Pedestrian detection and tracking in infrared imagery using shape and appearance, Comput. Vis. Image Understand. 106 (2007) 288–299.

[10]. S. Pellegrini, A. Ess, K. Schindler, L. van Gool, You'll never walk alone: Modeling social behavior for multi-target tracking, in: IEEE 12th International Conference on Computer Vision, 2009, pp. 261–268.

[11]. Robust Multiperson Tracking from a Mobile Platform Andreas Ess, Student Member, IEEE, Bastian Leibe, Member, IEEE, Konrad Schindler, Member, IEEE, and Luc van Gool, Member, IEEE,2009

[12]. J. Berclaz, F. Fleuret, E. Turetken, P. Fua, Multiple object tracking using kshortest paths optimization, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011) 1806–1819.

[13]. A real time algorithm for people tracking using contextual reasoning Rosario Di Lascio a, Pasquale Foggia b, Gennaro Percannella b, Alessia Saggese b,⇑, Mario Vento b 2013

[14]. D. Comaniciu, V. Ramesh, P. Meer, Real-time tracking of non-rigid objects using mean shift, in: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, vol. 2, 2000, pp. 142–149

[15]. H. Tao, H. Sawhney, R. Kumar, Object tracking with bayesian estimation of dynamic layer representations, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 75–89

[16]. "B. Wu, R. Nevatia, Detection of multiple, partially occluded humans in a single image by Bayesian combination of Edgelet part detectors, in: Tenth IEEE Int. Conf. on Computer Vision, vol. 1, 2005, pp. 90–97."

[17]. K. Bhuvaneswari, H. Abdul Rauf, Edgelet based human detection and tracking by combined segmentation and soft decision, in: International Conference on Control, Automation, Communication and Energy Conservation, INCACEC

[18]. B. Yogameena, S. Roomi, S. Abhaikumar, Detecting and tracking people in a homogeneous environment using skin color model, in: Seventh Internationa Conference on Advances in Pattern Recognition, ICAPR '09, 2009, pp. 282–285

[19]. Z. Han, Q. Ye, J. Jiao, Combined feature evaluation for adaptive visual object tracking, Comput. Vis. Image Understand. 115 (2011) 69–80

[20]. Y. Cai, N. de Freitas, J. Little, Robust visual tracking for multiple targets, in: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision ECCV 2006, Lecture Notes in Computer Science, vol.

[21]. Hierarchical Kalman-particle filter with adaptation to motion changesfor object tracking Shimin Yin, Jin Hee Na, Jin Young Choi, Songhwai Oh 2011

[22]. A parallel histogram-based particle filter for object tracking on SIMD-based smart camerasHenry Medeiros *, Germán Holguín, Paul J. Shin, Johnny Park

[23]. X. Song, J. Cui, H. Zha, H. Zhao, Vision-based multiple interacting targets tracking via on-line supervised learning, in: Proceedings of the 10th European Conference on Computer Vision: Part III, ECCV '08, Springer-Verlag, Berlin

[24]. M. Wang, H. Qiao, B. Zhang, A new algorithm for robust pedestrian tracking based on manifold learning and feature selection, IEEE Trans. Pattern Anal. Mach. Intell. 12 (2011) 1195–1208

[25]. Yussiff, Abdul-Lateef, Suet-Peng Yong, and Baharum B. Baharudin. "Parallel Kalman filter-based multi-human tracking in surveillance video." Computer and Information Sciences (ICCOINS), 2014 International Conference on. IEEE, 2014.

[26]. Michael D. Breitenstein Online Multiperson Tracking-by-Detection from a Single, Uncalibrated Camera IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 33, NO. 9, SEPTEMBER 2011

[27]. Andreas Opelt1, Axel Pinz1, and Andrew Zisserman2 A Boundary-Fragment-Model for Object Detection.IEEE

[28]. Krystian Mikolajczyk Bastian Leibe Bernt Schiele Multiple Object Class Detection with a Generative Model. IEEE

[29]. Krystian Mikolajczyk et al Pedestrian Detection in Crowded Scenes.IEEE

[30]. Amir Akramin Shafie et al. Smart Objects Identification System for Robotic Surveillance 11(1),February 2014, 59-71 DOI: 10.1007/s11633-014-0766-9.IEEE