SCHEMATIC MODEL FOR ANALYZING MOBILITY AND DETECTION OF MULTIPLE OBJECT ON TRAFFIC SCENE

Pushpa D.¹, Dr. H.S. Sheshadri²

¹Asst. Prof: Dept of Information Science & Engg, Maharaja Institute of Technology, Mysore, India
²Prof.: Dept. of Electronics and communication Engg, PES College of Engg, Mandya, India

ABSTRACT

Detection, counting as well as gathering features to perform analysis of behavior of natural scene is one of the complex processes to be design. My previous work has introduced a simple model for identifying and counting the moving objects using simple low level features. The current work is an extension of the previous work where the focus is not only to detect and count the multiple moving objects but also to understand the crowd behavior as well as exponentially reduce the issues of inter-object occlusion. The image frame sequence is considered as input for the proposed schematic model. Unscented Kalman filter is used for understanding the behavior of the scene as well as for increasing the detection accuracy and reducing the false positives. Designed on Matlab environment, the result shows highly accurate detection rate.

Keywords: Background Image, Foreground Image, Kalman Filter, Multiple Object Detection, Object tracking, Object counting.

I. INTRODUCTION

The area of computer vision is still in its infancy. It has many real-world applications, and many breakthroughs are yet to be made. Most of the companies in existence today that have products based on computer vision can be divided into three main categories: auto manufacturing, computer circuit manufacturing, and face
There are other smaller categories of this field that are beginning to be developed in industry such as pharmaceutical manufacturing applications and traffic control. Auto manufacturing employs computer vision through the use of robots that put the cars together. Computer circuit manufacturers use computer vision to visually check circuits in a production line against a working template of that circuit. Computer vision is used as quality control in this case. The third most common application of computer vision is in face recognition. This field has become popular in the last few years with the advent of more sophisticated and accurate methods of facial recognition. Traffic control is also of interest because computer vision software systems can be applied to already existing hardware in this field. By traffic control, we mean the regulation or overview of motor traffic by means of the already existing and functioning array of police monitoring equipment. Cameras are already present at busy intersections, highways, and other junctions for the purposes of regulating traffic, spotting problems, and enforcing laws such as running red lights. Computer vision could be used to make all of these tasks automatic. In the last few years, object detection and localization have become very popular areas of research in computer vision. Although most of the categorization algorithms tend to use local features, there is much more variety on the classification methods. In some cases, the generative models show significant robustness with respect to partial occlusion and viewpoint changes and can tolerate considerable intra-class variation of object appearance [1, 2, 3]. However, if object classes share a high visual similarity then the generative models tend to produce a significant number of false positives. On the other hand, the discriminative models permit us to construct flexible decision boundaries, resulting in classification performance often superior to those obtained by only generative models [4, 5]. However, they contain no localization component and require accurate localization in positions and scale. In the literature, the standard solution to this problem is to perform an exhaustive search over all position and scales. However, this exhaustive search imposes two main constraints. One of them is the detector’s computational complexity. It requires large computational time for relatively large number of objects. The second is the detector’s discriminance, since a large number of potential false positives need to be excluded.

Problems like inter-object occlusion, different object moving at all different speed, walking in arbitrary direction etc. will pose a huge challenge in the development as well as design of multiple object detection as well as tracking from the crowded scene. Majority of the video monitoring techniques are normally experimented with permanent image capturing device which is comparatively unproblematic in terms of real time application in order to accomplish image capturing parameters. Therefore, the current research work has been experimented with monocular adjusted image capturing device, which facilitates to design a distinctive connection between the subjects walking in real time and image plane. The phenomenon is considered for isometric view of the subjects walking. In Section 2, we review recent work on multi-object detection. In Section 3, problem description is discussed in brief followed by proposed system in section 4. Section 5 discusses about algorithm design that is implemented in the proposed system which is followed by Section 6 that discusses about result discussion. Finally concludes by summarizing the entire research work in section 7.
II. RELATED WORK

Moving object detection is the basic step for further analysis of video. Every tracking method requires an object detection mechanism either in every frame or when the object first appears in the video. It handles segmentation of moving objects from stationary background objects. This focuses on higher level processing. It also decreases computation time. Due to environmental conditions like illumination changes, shadow object segmentation becomes difficult and significant problem. 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. This temporal information is usually in the form of frame differencing, which highlights regions that changes dynamically in consecutive frames. A brief survey on the past research work is described here. One of the significance work carried out by Koller et al. [6][7] is design of an algorithm which uses offline calibrated camera step to aid the recovery of the 3D images, and it is also passed through Kalman Filter to update estimates like location and position of the object. The concept of Bayesian technique for image segmentation based on feature distribution is also deployed where a statistical mixture model for probabilistic grouping of distributed data is [8] adopted. It is mainly used for unsupervised segmentation of textured images based on local distributions of Gabor coefficients. Toufiq P. et al., in [9] describes background subtraction as the widely used paradigm for detection of moving objects in videos taken from static camera which has a very wide range of applications.

Polana and Nelson [10] has adopted spatial-temporal discretization which is allocated to every region of image, accomplished normally by overlapping a grid on the image and its mean visual stream. The episodic movement prototype is estimated by performing Fourier analysis on the resulting attribute vector. The use of movement field by Shio and Sklansky [11] has shown the accomplishment by correlation technique over consecutive image frames, which results in identification of movement regions with uniform direction when smoothening is done along with quantization of image regions.

Deployment of clusters of pixels has been done by Heisele et al. [12] to be attained by different grouping methods as fundamental constituents for tracking. The grouping analysis is performed by both spatial data and color information. The concept behind this is that appending spatial data generates grouping more steady than deploying only color data. There has been also abundant use of k-means algorithm while working on clustering techniques as evident in Cutler and Davis [13] for designing a change detection algorithm. The prototype also facilitates a map of the image elements signifying moving subjects along with precise thresholding.

Use of wavelets transform has been seen in the work of Papageorgiou and Poggio [14] for characterizing a moving object shape and then examines the transform technique of the image to evaluate the movement patterns. Implementation of two dimensional contour shape analyses is seen in the work of Wren et al. [15] which tends to recognize different parts of body of the moving subjects using heuristics techniques. Cai and Aggarwal [16] proposed a model with a uncomplicated head-trunk replica to trail humans across multiple image capturing devices. Beymer and Konolige [17] have worked for fitting a trouble-free shape on the candidate image.
Dimitrios Makris [18] with the non-trivial problem of performance evaluation of motion tracking. We propose a rich set of metrics to assess different aspects of performance of motion tracking. We use six different video sequences that represent a variety of challenges to illustrate the practical value of the proposed metrics by evaluating and comparing two motion tracking algorithms. Max et al [19] describes performance results from a real-time system for detecting, localizing, and tracking pedestrians from a moving vehicle. It achieves results comparable with alternative approaches with other sensors, but offers the potential for long-term scalability to higher spatial resolution, smaller size, and lower cost than other sensors. But issue of this work is it cannot segment people or objects in close contact. Jifeng [20] has presented a novel object tracking algorithm where the author has used the joint color texture histogram to represent a target and then applying it to the mean shift framework.

To analyze images and extract high level information, image enhancement, motion detection, object tracking and behavior understanding researches have been studied. In this section, we have studied and presented different methods of moving object detection from prior research work. Various methods like background subtraction, temporal differencing, statistical methods were explored. Detection techniques into various categories, here, we also discuss the related issues, to the moving object detection technique. The drawback of temporal differencing is that it fails to extract all relevant pixels of a foreground object especially when the object has uniform texture or moves slowly. When a foreground object stops moving, temporal differencing method fails in detecting a change between consecutive frames and loses the track of the object. The survey also identified background subtraction method in deep because of its computational effectiveness and accuracy. This section gives valuable insight into this important research topic and encourages the new research in the area of moving object detection as well as in the field of computer vision. It was also found that research on object tracking can be classified as point tracking, kernel tracking and contour tracking according to the representation method of a target object. In point tracking approach, statistical filtering method has been used to estimating the state of target object. Kalman filter and particle filter are the most popular filtering method. In kernel tracking approach, various estimating methods are used to find corresponding region to target object. Now a day, the most preferred and popular kernel tracking techniques are based on mean-shift tracking and particle filter. Contour tracking can be divided into state space method and energy function minimization method according to the way of evolving of contours. The survey has witnessed various publications that are focused on addressing issues on object detection. Finally, Table 1 will show the evidence of some recent publications related to the issue of Object detection. Majority of the Journals based on moving objects are selected for the review purpose as shown below:
Table 1: Existing Approaches used in Moving Object Detection

<table>
<thead>
<tr>
<th>Year</th>
<th>Authors</th>
<th>Problem Focused</th>
<th>Techniques Used</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>Djamel, Nicolas [21]</td>
<td>Counting people</td>
<td>Skeleton graph, Background subtraction</td>
<td>Objects other than human are not considered</td>
</tr>
<tr>
<td></td>
<td>Conte e.t. al [22]</td>
<td>Counting people</td>
<td>Motion vector, Support vector Regressor</td>
<td>Objects other than human are not considered</td>
</tr>
<tr>
<td></td>
<td>Burkert e.t. al [23]</td>
<td>Monitoring people based on aerial images</td>
<td>Microscopic &amp; Mesoscopic parameter, trajectory</td>
<td>Inter-object Occlusion not focused on</td>
</tr>
<tr>
<td></td>
<td>Conte e.t. al [24]</td>
<td>Counting Moving People</td>
<td>Detection using interesting points</td>
<td>Objects other than human are not considered</td>
</tr>
<tr>
<td></td>
<td>Widhalm [25]</td>
<td>Flow of Pedestrian</td>
<td>Growing Neural Gas, Dynamic Time Warping</td>
<td>Objects other than human are not considered, Inter-object Occlusion not focused on</td>
</tr>
<tr>
<td></td>
<td>Dehghan e.t. al [26]</td>
<td>Automatic Detection &amp; Tracking of Pedestrians</td>
<td>Static Floor Field, Boundary Floor Field, Dynamic Floor Field</td>
<td>Addressed Occlusion, good results obtained, not focused on multiple objects moving</td>
</tr>
<tr>
<td></td>
<td>Spampinato [27]</td>
<td>Event Detection in crowd</td>
<td>Lagrangian Particle Dynamics</td>
<td>Results found with Issue in noises in frames affecting the performance of flow map estimation</td>
</tr>
<tr>
<td></td>
<td>Jae Kyu Suhr [28]</td>
<td>Moving Object Detection</td>
<td>Background compensation method using 1-D feature matching &amp; outlier rejection</td>
<td>Not focused on inter-object occlusion, Multiple objects are not considered</td>
</tr>
<tr>
<td></td>
<td>Lei, Su, Peng [29]</td>
<td>Maritime Surveillance</td>
<td>Trajectory Pattern Mining</td>
<td>Better result but cannot be used in crowded scenes</td>
</tr>
<tr>
<td></td>
<td>Sirmacek [30]</td>
<td>Tracking people from airborne images</td>
<td>Kalman filter</td>
<td>Not focused on inter-object occlusion, Multiple objects are not considered</td>
</tr>
<tr>
<td>2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cui, Zheng [31]</td>
<td>Moving Object Detection</td>
<td>Graph cuts</td>
<td>Not focused on inter-object occlusion, Multiple objects are not considered</td>
</tr>
<tr>
<td></td>
<td>Chen e.t. al [32]</td>
<td>Counting People in crowd</td>
<td>2-Stage Segmentation</td>
<td>Better result, but impact of occlusion &amp; multiple objects are not considered</td>
</tr>
<tr>
<td></td>
<td>Gowsikhaa [33]</td>
<td>Human Activity Detection</td>
<td>Artificial Neural Network</td>
<td>Topic is narrowed down to face recognition system, Occlusion issue is not addressed</td>
</tr>
<tr>
<td></td>
<td>Rosswog [34]</td>
<td>Object detection in crowd</td>
<td>SVM</td>
<td>Occlusion issue is not addressed</td>
</tr>
<tr>
<td></td>
<td>Arif [35]</td>
<td>Counting Moving Object</td>
<td>Neural Network</td>
<td>Illumination state &amp; Occlusion issues are not addressed</td>
</tr>
</tbody>
</table>

Although, we can’t conclude that the previous work has got any demerits or disadvantages, but it can be seen that with each discoveries of research, there is a new idea which keeps the research on ongoing pace. But, in nutshell, it can be said that majority of the work has not much integrated the process of Multiple (different) objects detection along with solution for inter-object occlusion and reduction of false positives. Hence a research gap can be explored in this viewpoint.
III. PROBLEM DESCRIPTION

Searching for a known object in a specified scene and locating a given object are inherently different problems. Object recognition is a computationally expensive process. Fast algorithms are essential at all stages of the recognition process. Object recognition is tricky because a combination of factors must be considered to identify objects. These factors may include limitations on allowable shapes, the semantics of the scene context, and the information present in the image itself. Given a video clip, the initial problem is segregating it into number of frames. Each frame is then considered as an independent image, which is in RGB format and is converted into Gray scale image. Next the difference between the frames at certain intervals is computed. This interval can be decided based on the motion of moving object in a video sequence. If the object is moving quite fast, then the difference between every successive frame has to be considered, posing a difficulty in designing smart object identification system.

The key issue of multiple targets tracking in the image sequences is the occlusion problem. During the occlusion time, unpredictable events can be activated. Thus, the tracking failure and missed tracking often happened. To cope with this problem, a behavior analysis of multiple targets is required the first time. A specific target enters into the specific field of view, and then it is moving, stopping, interacting with other targets. And finally, it leaves the field of view. We can classify the behaviors of multiple targets accordingly as

1. A specific target enters into the scene.
2. Multiple targets enter into the scene.
3. A specific target is moving and forms a group with other targets, or just moves beside other targets or obstacles.
4. A specific target within the group leaves a group.
5. A specific target continues to move alone, or stops moving and then starts to move again.
6. Multiple targets in a group continue to move and interact between them, or stop interacting and then start to move again.
7. A specific target leaves a scene.
8. A specific group leaves a scene

The events of (1), (4), (5), and (7) can be tracked using general tracking algorithms. However, the events of (2), (3), (6) and (8) cannot be tracked reliably. Thus, to resolve this problem, we propose the occlusion reasoning method that detects occlusion activity status using Kalman Filter.

IV. PROPOSED SYSTEM

The prime aim of the proposed work is to design a robust framework that is capable of identifying mobile objects using sequences of frames captured from video. The research work is designed not only for its utility in crowd detection that was presented in my previous work [36] but now the proposed framework will be used for extensive analysis of the video dataset. Various knowledge and data about the coordinates and characteristics of the mobile object at various instance of time as the fundamental of identifying abnormal mobile object detection pertaining to its movement. The flow of the proposed system is as highlighted in Fig.1.
The proposed study is designed to accept the sequential frames of the video, which is converted to blob image or binary images. Image plane is defined for the purpose of frame representation as well as for the purpose of estimating the centroid location. (we are using the relative distance from centroid in image plane from the position of foreground object). A Gaussian distribution is applied to understand the pixel density, so that distinction of every pixel changes is easy to find/estimate. Another purpose of Gaussian distribution is to find the peak in pixel intensity distribution which will be used for threshold estimation. The white
pixels are compared with the threshold in order to distinguish the foreground and background object. However, after estimating the relative position of newly detected foreground object, background is also found and finally state transition matrix (STm) is compared with the threshold value (Tm). The condition is checked to predict distinctly about the foreground and background object. The Kalman filter is the best part of the contribution.

The proposed system uses the frames captured from the still camera, where the visibility is restricted to a defined area of focus of the aperture of the camera. This method is adopted to analyze the preciseness of the mobility of the object, which could not be done on mobile video sensors. The system uses distinctive technique to filter out the foreground objects from the background for better accuracy. Basically, the background model is considered as prime module as the experimental framework designs for every pixels of the captured frame as a distribution of the typical values considered by background module. The main concept behind this is the most frequent values of a pixel that are corresponding to the background metaphors. The system is designed in such a way that the iterative mobility in the background is also captured. That will mean that even a slightest displacement on any object on the considered visibility screen will also be captured.

![Figure 2 showing Original Scene, Mask](image1.png)

![Figure 3 showing Original Scene, Mask](image2.png)
The categorization of the pixel into background or foreground is designed depending on the collected module of background using outlier-detection. According to this design, the pixels with the intensity values sufficiently distinctive from the background distribution are indexed as foreground. The framework will also use gradual alteration in the illumination factor by spontaneously upgrading the background model of every pixel in the frames. The same phenomenon will also be applicable to the static objects in the considered scene. An example input image and the corresponding pixel labeling computed from a background model can be seen on the left where foreground pixels were labeled as white. The person crossing the street can be clearly seen in Fig.2 and Fig.3. From Fig.2 and Fig.3, by lumping large connected foreground regions into blobs, foreground objects, such as the person seen crossing the street is identified. However, the left images also show that an additional foreground object was erroneously detected in the top right corner due to violent motion of the trees in the background. Such occasional misdetections are taken into account at later stages of the object detection algorithm.

The consecutive stage will include a sequence of the object blobs together across various frames of the scene. In order to accomplish this task, various attributes like position, velocity, and size are used to represent every object block. Unscented Kalman filter is specifically used in this work for this purpose for determining the actual coordinates of the mobile object identified on the scene-frames. The main purpose of adopting this step is to truncate the possibility of noise from blob estimations which could possibly lead to generation of the false-positives values.

V. ALGORITHM DESIGN

The proposed work is performed on 32-bit Windows OS. In order to smoothly run the heavy computation from large datasets of frames, it is highly essential to use high configuration system with min 2.20 GHz processor with Intel core-i3 along with 4 GB RAM size. The programming is done on Matlab environment due to its faster computation of complex problems. The image sequence datasets has been taken from [36].

By deploying the computed indexing, foreground objects are identified, however, direct use of these blobs for tracking is sensitive to noise and fails to determine object identities at different times in a sequence. In order to couple a sequence of state measurements, that is, relevant blob properties such as the centroid, centroid velocity between consecutive frames, or size into a denoised state trajectory, The filter operates as a two-stage process switching between absorbing evidence from the most recent foreground blob and forming a predictive distribution for the next. In the context of tracking multiple objects simultaneously, the key difficulty lies in determining which Unscented Kalman filter should absorb which new state measurement. If this association task is not solved adequately then Unscented Kalman filters are corrupted with wrong information.
Algorithm: Detection of object mobility
Input: Image sequence datasets
Output: Appropriate detection of object movement in visualization
START
1 Load set of image sequence
2 Returns data matrix
3 Reading Indexed individual frames
4 Reading RGB components of the scene
5 Reading frame number
6 Estimate the log of the Gaussian distribution for each point in x using μ and σ.
7 x=matrix of column-wise data points.
8 μ =column vector
9 σ=Covariance matrix of respective size
10 Estimate the log probability for each column in x-centered.
11 Deploy other parameters:
12 weight=matrix of mixing proportion
13 σ =Covariance
14 Active_Gaussian=components that were either matched or replaced.
15 Background_Threshold=Threshold standard deviation for outlier detection on each Gaussian.
16 K= Quantity of components used in the mixtures.
17 Alpha=How fast for adapting the component weights
18 Rho=How fast to adapt the component means and covariances
19 Threshold_Deviation=Threshold used for finding matching versus unmatched components
20 INIT_Var= Initial variance for newly placed components
21 INIT_MixProp= Initial prior for newly placed components
22 COMPONENT_THRESH: Filter out connected components of smaller size
23 BACKGROUND_THRESH: Percentage of weight that must be accounted for by background models
24 Design a helper function for structure of Kalman filtering
25 Define Kalman State type
26 Create best background image
27 Computes a β probability density on image coordinates pos
28 Set the state transition matrix
29 Update using position only
30 Update using position and velocity
31 Compute a cost matrix
32 Computes the squared distance between the centroid predictions of a list of objects and a list of observed blob centroids.
33 Compute a cost matrix
34 Compute the square distance between Kalman filter predictions and observations
35 Create foreground_map
36 Binary matrix
37 Define chromaticity coordinates
38 Show visualization of object detection
END
VI. RESULT DISCUSSION

In order to accomplish the research work, the whole detection protocol was assessed on a number of different sequences in order to analyze its accurateness and weakness. The previous work [36] done is able to detect objects from crowded scene, but false positive are quite moderate in number. Hence, this paper will use the technique of foreground detection that can be used for extracting useful information under various distinctive conditions, particularly when the data stream is recoded in different color space. The cumulative results obtained are as below in Fig.3

![Figure 4](image-url) Detection window showing object detection

The above Fig. 4 shows frame rates and movement of each object, along with its count. The interesting thing is that it also shows every minute movement even for shaking leaves of tree. Manual inspection of various detection results suggested that multiple object tracking using the devised association technique performs adequately in many situations.

![Figure 5](image-url) Result showing successful detection rate
The assessment is done considering three case scenarios e.g. i) Successful detection rate of object: Fig.5 shows very effectively both truck and human along with the person riding bicycle. ii) Successful detection of multiple-object and counting: Fig.6 shows very distinctively the human as well as truck as a moving objects. iii) Preciseness in multiple-object detection counting: Fig.7 shows an accurate detection of a person riding bicycle, a person walking, and a person walking along with pram. All the three objects are accurately shown and counted too. The above implemented result shows better detection rate for multiple object with precise bounding box over the object. It successfully counts the objects too. The previous model has some detection issue due to inter-occlusion. However, 95% of the inter-occlusion issue are overcome in this model with highly precise detection rate.

VII. PERFORMANCE DISCUSSION

For evaluating the performance of the application, the selected video clip in chosen to understand the various frames as mentioned in previous section. The classifications of the images are done and consecutively Machine is trained where the total number of classes has to be defined with class name. Finally object recognition module is executed. This video segmentation method was applied on three different video sequences one of which are depicted in Fig 8 and Fig 9, which represents original pictures, mask, and detected Object.
Figure 9 showing Original Scene, Mask, Detected Object (Second Sequence)

For the first video sequence Figure 2 & 3 shows the original image, the background registered image, and image obtained after background subtracted, and finally shows the count of the detected objects. The same is repeated for the next video sequence.

Figure 10 Object counting for moving objects

The system is able to track and count most vehicles successfully. Although the accuracy of vehicle detection was 100%, the average accuracy of counting vehicles was 99% (Fig 10). Each video sequences are tested with various types of object and also in complex background too. The application is successfully able to identify as well as track various pedestrians separately along with intersection. The application also successfully conducts object recognition (Fig 11). This is due to noise which causes detected objects to become too large or too small to be considered as a vehicle.

Figure 11 Moving Object recognition for separated objects moving
However, two vehicles will persist to exist as a single vehicle if relative motion between them is small and in such cases the count of vehicles becomes incorrect.

Figure 12 Different Moving object recognition in one scene.

Figure 13 Performance Analysis for Accuracy
Figure 14 Performance Analysis for Object Identification

Figure 15 Performance Analysis for Object Recognition.
Fig 13 represents the performance analysis of accuracy level considering all the three videos for evaluation. The accuracy rate, identification and recognition rate is better approximated to 100% respectively. Fig 14 shows the identification analysis which represents the better identification rate in all the AVI files. Fig 15 shows better and enhanced performance in recognition for all types of moving object.

VIII. CONCLUSION

This paper has presented a novel and highly precise model of detecting multiple objects in real time street scene with highly uncertainty of the types of objects, their movements, and their count rates. The previous model presented has been addressed for object detection from crowded scene with better detection rate, but even some moderate false positives were also raised due to inter-occlusion and variances in illumination. However, using unscented Kalman filter such issues were overcome in this model. The scope of the application for this model is now much higher. If the previous model could be used to detect multiple object than this model can be used to analyzed the behavior of object in the real time scene. The simulation result shows highly precise detection rate. However, these will eventuality limits the specification of the model in terms of segregating the targeted object with not so important object. For example, due to proposed algorithm that has used both foreground and background model, the framework detects even the minor movement of leaves of trees. This property might be little unwanted when we want to design an application specific to analyze pedestrian or car or some major moving object visually that keeps highest importance in application. Hence, the next further work will be enhanced the model using homographic transformation

REFERENCES


[34] James Rosswog, Kanad Ghose, Detecting and Tracking Coordinated Groups in Dense, Systematically Moving, Crowds, SIAM-2012


