A Comparative study on object detection and Tracking in video

[1]Praveen Kumar, [2]Bindu N.S
[1]Sahyadri College of Engineering, Dept of E&C, Mangalore, India

[2]Vidyavardhaka college of Engineering, Dept of E&C, Mysore, India

Abstract— **This paper presents a survey of various techniques related to video surveillance system improving the security. The goal of this paper is to review of various moving object detection and object tracking methods. This paper focuses on detection of moving objects in video surveillance system then tracking the detected objects in the scene. Moving object detection is first low level important task for any video surveillance application. Detection of moving object is a challenging task. Tracking is required in higher level applications that require the location and shape of object in every frame. This survey provides various techniques or methods that are used to recognize, detect and track objects in the shadowed region, crowded area, multi modality background, occluded object, and deformable based objects**

Index Terms- **Object detection, background subtraction, Temporal frame differencing, object tracking, video surveillance.**

I. INTRODUCTION

Surveillance for multiple objects is an active research area in computer vision. Accurate detection and tracking of objects are two essential components required by a variety of applications. The areas of application include Ambient Intelligence (AmI), automated surveillance, pedestrian detection in Surveillance, image compression and content-based multimedia storage and retrieval [1]. Due to a large number of potential applications, object detection and tracking has become an extremely active computer vision research. The result of this has been a significant amount of prior art proposing object segmentation techniques [2, 3] which includes the application of traditional 2D computer vision techniques and the use of other sensor modalities such as infrared, laser scanners, sonar or radar. Many of the proposed approaches produce good results when presented with constrained scenarios [3] that allow specific assumptions to be made. These constraining assumptions are generally introduced by techniques to reduce the number of complicating factors that are inherent in object detection, thereby making the problem tractable. They include assumptions about the environmental conditions, object appearance, flow density, background color intensity information, duration of time for which an object exists within the scene, object enters the scene un-occluded and even the number of objects within the scene. Unfortunately, due to these assumptions few approaches produce reliable results for long periods of time in unconstrained environments [4]. Figure.1 shows the basic block diagram of object detection and tracking. Pre-processing is the initial step in object detection and tracking where image size and camera parameters are taken into account. Moving object detection is the basic step for further analysis of the detection process.
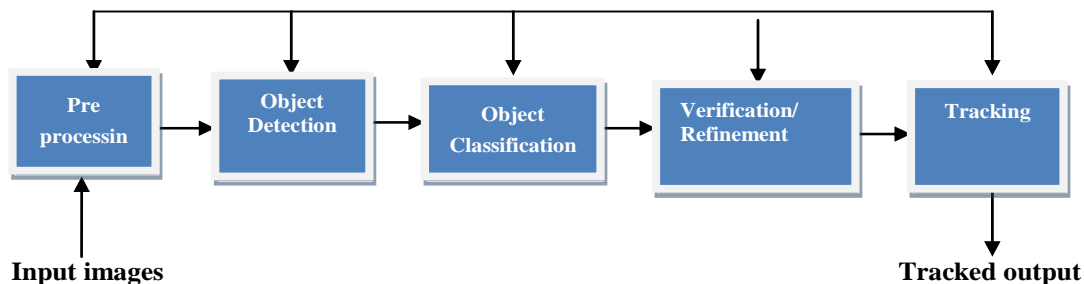

Figure 1: Basic block diagram of object detection and tracking system

www.ijert.org

It involves segmentation of moving objects from stationary background. This not only creates a focus of attention for higher level processing but also decreases the computation time. Commonly used techniques for object detection are foreground segmentation, statistical models, temporal differencing and optical flow. Due to dynamic environmental conditions such as illumination changes, shadows and occlusion, object detection is a difficult and significant problem that needs to be handled well for a robust visual surveillance system. Recently, 3D stereovision has been studied seriously as a technique to overcome some of the issues inherent in robustly detecting objects. The use of stereo information carries some distinct advantages over conventional 2D techniques. Stereo vision is a reliable tool to obtain image and depth data of a scenery simultaneously. The accuracy of the results depends on the choice of the stereo camera system and stereo correspondence algorithm. Stereo correspondence has remained in the focus of the machine vision community from last few decades. It is dictated by the biological observation that, two slightly displaced images of the same scenery provide enough information to perceive the depth of the objects depicted and third dimension of the scenery is retrieved from it. Therefore the importance of stereo correspondence is obvious in the fields of machine vision, object detection, robot navigation, depth measurements and 3D environment reconstruction, as well as in many other aspects of production, security, defence, exploration, and entertainment.

The robust detection of objects under unconstrained conditions introduces a variety of complicating factors that makes it one of the most challenging problems in computer vision. These complicating factors have to be acknowledged and addressed by computer vision systems if robust object detection is needed in real-world scenarios. Some of the object detection challenges are object Appearances such as a large variability in objects local and global appearance can be caused by various types and styles [5]. Therefore, an object detection technique of a particular feature, such as the appearance of shapes, may not be applicable for all classes of object. Another difficulty in object detection is its pose, because object is a rigid body and an object's global shape can undergo a large range of transformations [6, 8] due to the variety of possible poses. This leads to miss detection of the object. An object can be viewed at a variety of possible [6] orientations with respect to the camera image, this is also another cause for detection process. Position of an object in a scene will have various distances to the camera and appearance of objects close to the camera [7] can differ significantly from those at a greater distance. Self occlusion of object's silhouette may also be perturbed by a variety of occluding accessories such as neighbouring objects and shadow. Object classification step categorizes detected objects into predefined classes such as human, vehicle, animal, clutter, etc. It is necessary to distinguish objects from each other in order to track and analyze their actions reliably. Currently there are two major approaches towards moving object classification. Which are shape-based and motion-based methods. Object and non object verifications are done in one special block that is verification and refinement. The next step in this process is object tracking, which can be simply defined as the creation of temporal correspondence among detected objects from frame to frame. This procedure provides temporal identification of the segmented regions and generates cohesive information about the objects in the monitored area such as trajectory, speed and direction. The output produced by tracking step is generally used to support and enhance motion segmentation, object classification and higher level activity analysis. Feedback approach is used to update the object position during this tracking process. The goal of many computer vision applications such as security, which needs object flow information. Other allocation such as smart rooms, require information not only about where the objects are at a given instant of time, but also what these objects are doing. More useful information about the activities of object within the scene may be obtained if the object is reliably tracked through time. Therefore Robust object tracking introduces further challenges. Object Movement such as an object may travel in a non-linear and unpredictable fashion, resulting in difficult tracking. For example in real-world scenarios a given [9] object can move, stop or turn around unexpectedly. This creates serious problem in tracking. Occlusion is another problem in tracking. Depending on the tracking technique, when two or more tracked objects occlude [8], it is possible that one track can be temporarily lost. The tracking algorithm must be able to take this into account so that after occlusion, each track should maintain the appropriate object (i.e. the one which was tracked before occlusion) [10].Splitting is also affecting the performance of the tracking system. A track may split into two or more pieces. This can be due to poor segmentation in the current frame, poor segmentation in the previous frames, or possibly due to an object depositing another object in the scene. If the reason is due to poor segmentation the system should be able to recognize this and remedy the situation. New objects, which have not been previously tracked, must be recognized [9] as new and not confused with previously tracked objects and so on.

II RELATED WORK

Object detection and tracking using stereo vision is not new, and has been introduced to computer vision from several years. Recently, David Gero nimo, Antonio M. Lo´pez, Angel D. Sappa, and Thorsten Graf [11] has reviewed by surveying many papers and compared the performance of both monocular and binocular vision based pedestrian detection and tracking systems. According to the above authors, binocular vision based detection and tracking systems outperform the best rather than the monocular vision based system especially in unconstrained environment. In order to detect and track any moving object in an unconstrained environment, stereo matching is the basic technique for foreground segmentation in stereo vision based object detection system. Ling Cai, LeiHe, YirenXu, YumingZhao, XinYang [12] used Normalized Cross Correlation (NCC) and Kernel Density Estimation (KDE) algorithm for stereo matching and object detection respectively. An iterative position updating method has been used for tracking purpose and he has also tried to address the problem of radiometric changes in images by introducing NCC. NCC itself is sensitive to the changes in radio metric conditions which results in poor performance. Dong-Jin Seo, Ju-Kyong Jin, Sang-il Na, Dong-Seok Jeong [13] used Sum of Absolute Difference (SAD) and Sum of Squared Difference (SSD) methods for stereo matching technique. Object regions are used for the detection and classification. A region updating algorithm has been used for object tracking and Change in illumination has not been addressed by the above authors. After object detection, the tracking algorithm is used to associate the object positions in consecutive frames and provide a trajectory of each object. Eduardo parilla, juan-r.torregrosa, Jaime riera, jose-l.hueso [14] developed a tracking system based on Optical flow and stereo vision, combined with adaptive filters in order to predict the expected direction and movement of the objects. Fuzzy control system has been used by the above authors in order to couple the tracking and predictive algorithm without considering the effects of radiometric changes. Young-chul lim, minho lee, chung-hee lee [15] used Zero-Mean NCC for stereo matching and edge features for detecting occluded objects. The authors have used Extended Kalman Filter for tracking. Like NCC, ZNCC is also sensitive to the radiometric variations which results in poor performance.

As we have seen from the above literature, correlation based stereo matching algorithms namely NCC, SAD, SSD, ZNCC are the very common techniques for the background subtraction. These stereo based correlation algorithms are sensitive to serve radiometric variations such as change in illumination, camera setting variation, image noise, exposure variation etc. This variation will affect the performance of stereo matching algorithms which results in poor disparity output. Even though NCC is insensitive to global illumination variation it still sensitive to local variation. Selection of correlation window size is another parameter which will affect the output disparity. Large sized correlation window gives better disparity as well as it blurs the object boundaries near depth discontinuities. Hence there is a need to address the above mentioned issues in stereo matching algorithms so that the object detection system will be more robust for these variations. Andreas Ess, Bastian Leibe [16] used depth map for object detection process and Bayesian network along with the depth information for classification purpose. The developed system fails to detect the object which are nearer to the camera and also which are entering with the occlusion. Eduardo parilla, juanr.torregrosa, Jaime riera, josel.hueso [17] proposed adaptive filters in order to predict the expected direction and movement of the objects. The authors used neural network for the classification. The experiment was carried out by considering very less number of objects and they did not address the problem of multiple objects with occlusion. Rafael Munoz-Salinas,EugenioAguirre,Miguel Garcia-Silvente, Antonio Gonzalez [18] proposes color and depth information based algorithm for detection process. An adaptive color based particle filter and Kalman filter techniques are applied to check the robustness of the algorithm without considering the occlusion problem. Kunsoo Huh, Jaehak Park, Junyeon Hwang, Daegun Hong [19] used Harris corner detector and NCC for feature extraction and matching respectively. Tracking has been done using Kalman filter and the authors did not address the Object occlusion problem. In the above papers, varieties of detection and tracking algorithms have been proposed. Even though some of the algorithms are very efficient in object detection, detection within the occluded region is not been addressed properly. Hence there is a necessary to develop a system that can handle the detection of various occluded object followed by tracking.

III. PARAMETERS AND METHODOLOGIES IN VIDEO TRACKING AND DETECTION

Parameters play a major role in calculating the efficiency of detection and tracking. There are several other parameters available in the field of vehicle detection and training. Some of the parameters are, color, texture, qualitative and quantitative analysis, false positive and false negative etc.,

TABLE 1: Analysis of Object detection and tacking

SI. No	Year	Proposed Method	Dataset Used	Parameter Used For Evaluation	Results
1.	Feb 2012	vehicle retrieval based on fine-grained attributes	urban environments	1. Quantitative analysis 2. Minimizing Transaction/Inquiry Response Time	Object retrieval using in challenging environment
2.	June 2012	Vehicle detection in traffic jams and complex weather conditions such as sunny days, rainy days, cloudy days, sunrise time, sunset time, or night time	Video to be tracked	Color, Height., weight, length of the car	Detected vehicle from traffic jam complex environment
3.	Sept 2012	Robust Detection and tracking an object Particle filter- Kalman filter	Video from traffic	False Negative and False positive	Detection and tracking of lane markings using visual inputs from a camera
4.	March 2013	Video-based traffic surveillance using a fuzzy hybrid information inference mechanism (FHIIM).	Video traffic	Recall and Precision	A new approach for a traffic monitor system that overcame the problems of congested conditions by using FHIIM
5.	Oct 2013	An extended Markov chain Monte Carlo (MCMC) method for tracking and an extended hidden Markov model (HMM) method	Traffic surveillance video.	Recall and precision	An extended MCMC method, i.e., EMCMC, for robust tracking and an HMM-based trajectory recognition method with an interactive system for efficient learning. Our method is an enhancement of the previous MCMC

The following inferences are made. Auto-allocation of symbols from images and a hand-drawing function for learning trajectories; locations and regions can be freely set, named, and recognized. Analysis results showed that the segmentation algorithm is memory efficient, and that the memory reduction rate is at least 66%. The method (MTTF) is useful for the pitch extraction of moving targets but each time only one pitch is extracted from the acoustic signal in which multiple pitches may coexist. The algorithm overcomes the disadvantage of traditional TMD method and it is well applied in the application of tracking vehicles in occlusion case. The proposed method can achieve an average vehicle detection rate of 97% and an average vehicle-tracking rate of 86%. A limited number of data collection points enable vehicle tracking without full network coverage. The proposed system can easily be set up without being given any environment information in advance.

V. CONCLUSION

This paper presents an extensive survey on object recognition, detection and tracking techniques that are able to detect an object in multimodality background, detect a deformable object, detect and track in abrupt change in the trajectories, environmental condition, shadow region, Occlusion etc., Many research issues have been highlighted and direction for future work have been suggested. Many new ideas have been

suggested by the researchers as future work. Many open issues have been highlighted by the researchers such as dealing with occlusion, Handling Multimodality background, Background modelling, Handling abrupt change in the trajectory and environmental conditions are some of the challenging work for junior researchers in the field of computer vision and computer graphics

ACKNOWLEDGEMENT

REFERENCES

1. Yuan Li, C. Huang, R. Nevatia, "Learning to associate: hybrid boosted multi- target tracker for crowded scene", In *Proc of IEEE Conference on Computer Vision and Pattern Recognition*, pp.2953–2960, 2009.

2. M. Harville, "Stereo person tracking with adaptive plan-view templates of height and occupancy statistics". *International Journal of Computer Vision*, vol.22, pp.127–142, 2004.

3. Bastian Leibe, Edgar Seemann, and Bernt Schiele, "Pedestrian Detection in Crowded Scenes", In *Proc of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1063-1072, 2005

4. Jurij leskovek, "detection of human bodies using computer analysis of sequence of stereo images", In *Proc of IEEE Conference on Computer Vision and Pattern Recognition,* pp 01-19, 2004

5. M. Bertozzi, E. Binelli, A. Broggi, and M. Del Rose, " Stereo vision-based approaches for pedestrian detection", In *Proc of IEEE International Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum*, 2005.

6. J. Garcia, N. Da Vitoria Lobo, M. Shah, and J. Feinstein, "Automatic detection of heads in colored images". In *Proc of Canadian Conference on Computer and Robot Vision*, pp. 276–281, 2005.

7. A.Broggi, M.Bertozzi, A. Fascioli, and M. Sechi, "Shape-based pedestrian Detection", In *Proc of IEEE Intelligent Vehicles Symposium*, pp 215–220, 2000.

8. T. Zhao and R. Nevatia, "Bayesian human segmentation in crowded situations", In *Proc of IEEE Conference on Computer Vision and Pattern Recognition*, vol.2,pp. 459–466, 2003.

9. A. Baumberg, "Learning Deformable Models for Tracking Human Motion". PhD Thesis, University of Leeds, 1995.

10. S. Dockstader, T.A. Murat, "Multiple camera tracking of interacting and occluded human motion", In *Proc of IEEE Conference on Computer Vision and Pattern Recognition*, pp.1441–1455, 2001.

11. David Gero nimo, Antonio M. Lopez, Angel D. Sappa, and Thorsten Graf, "Survey of Pedestrian Detection for Advanced Driver Assistance Systems", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 7, July 2010.

12. Ling Cai, LeiHe, YirenXu, YumingZhao, XinYang, "Multi-object detection and tracking by stereo vision", *Journal of Pattern Recognition,* vol.43, pp.4028–4041, 2010.

13. Dong-Jin Seo , Ju-Kyong Jin, Sang-il Na, Dong-Seok Jeong," A region- based fast stereo matching and tracking", *IJCSNS International Journal of Computer Science and Network Security*, Vol.10, No.4, 2010.

14. Eduardo parilla, juanrtorregrosa, Jaime riera, jose.hueso, "Fuzzy control for obstacle detection in stereo video sequences", *Journal of mathematical and computer modelling*, pp. 1-5, 2010.

15. Young-chul lim,minho lee,chung-hee lee, "Improvement of stereo vision-based position and velocity estimation and tracking using a stripe-based disparity estimation and inverse perspective map-based extended Kalman filter", *Journal of Optics And Laser Engineering,*Vol.48, pp 859-868, 2010.

16. Andreas Ess, Bastian Leibe, "Robust Multiperson Tracking from a Mobile Platform", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 10, 2009.

17. Eduardo parilla, juanrtorregrosa, Jaime riera, josel.hueso, "handling occlusion in object tracking in stereoscopic video sequences", *Journal of mathematical and computer modelling*, Vol.50, pp 823-830, 2009.

18. Rafael Munoz-Salinas , Eugenio Aguirre , Miguel Garcia-Silvente , Antonio Gonzalez , "A multiple object tracking approach that combines color and depth information using a confidence measure", *Journal of pattern   recognition letters*,Vol.29, pp 823-830, 2008.

19. Kunso based obstacle detection system in vehicles", *Journal of optics and laser engineering, Vol.46,pp* 168-178, 2008.

20. Rafael Munoz-Salinas, Eugenio Aguirre, Miguel Garcia-Silvente, "People detection and tracking using stereo vision and color", *Journal of Image And Vision Computing*,Vol.25, pp. 995–1007, 2007

21. www.**vision**.ee.**ethz**.ch