International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 12, pp 171-176, December-2023

https://doi.org/10.47001/IRJIET/2023.712024

A Method of Detecting an Object Using the Latest Technology

[1]Dr. Ramesh Palanisamy, [2]Mr. Mohamed Osman Akarma Al-Tigani Mohamed, [3]Dr. Mathivanan Viruthachalam, [4]Dr. Kumar Kaliyamoorthy, [5]Senthil Jayapal

[1,2,3,4,5]Department of Information Technology, University of Technology and Applied Sciences – Ibra, Sultanate of Oman

Authors E-mail: [1]ramesh.palanisamy@utas.edu.om, [2]osman.mohamed@utas.edu.om, [3]mathivanan.viruthachalam@utas.edu.om, [4]kumar.kaliyamoorthy@utas.edu.om, [5]senthil.jayapal@utas.edu.om

***Abstract* - Multiple object detection and tracking are the essential components required by a variety of intelligent applications. Object detection identifies the location of the object in a scene whereas object tracking associates the detected object over a sequence of frames. A variety of techniques has been developed in the past few decades, which can be broadly classified into 2D and stereo based 3D techniques. Majority of these techniques produce reliable results under specific assumptions in constrained scenarios [1]. These constraining assumptions are introduced to reduce the number of complicating factors, which are inherent in object detection and tracking. The most common assumptions are about environmental conditions, object appearance, flow density, background color intensity information, duration of time for which an object exists in the scene, objects occlusion, limitation regarding number of objects within the scene, etc. The reliability of these techniques is not guaranteed in real-time applications. The robust object detection and tracking in an unconstrained environment is the key requirement of state-of-the-art surveillance system [2].**

Keywords: Object Detection, Tracking, Location, Kalman Filter.

1. Introduction

Temporal differencing method is similar to background subtraction, where a background model is set in correspondence with the previous frames. Optical flow based motion segmentation determines the 2D projection of the 3D velocity map on the camera plane [10]. This process creates a velocity field in the image, which will be transformed from one image into the next image in a sequence [23]. This technique has many advantages over background subtraction and temporal differencing. For example, discontinuity in the optical flow helps in segmenting background from foreground region and also for segmenting multiple object regions. Even though this technique is efficient than the previous methods, it is computationally expensive and sensitive to noise. It is observed that the reliability of aforementioned techniques mainly depends on the initial assumptions and specific application of the algorithm. A generalized block diagram of the video surveillance system is shown in Figure1.1. Input video sequence is pre-processed in terms of resizing, denoising, camera calibration, etc. Identifying the moving object is one of the fundamental and critical steps in object detection, which is usually carried out by segmenting motion in the scene.

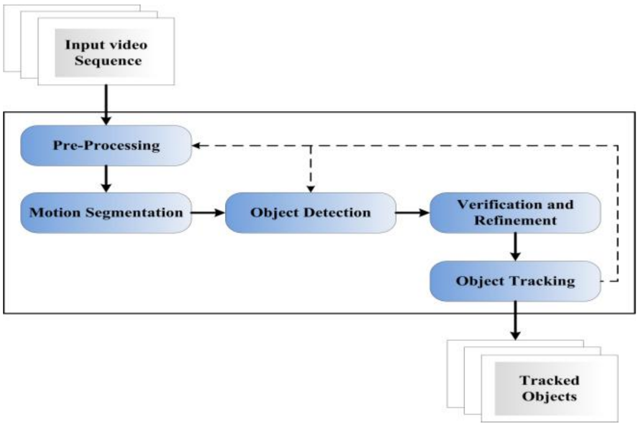

Figure 1.1: A generalized block diagram of video surveillance system

Generally, there are three basic techniques used for motion segmentation viz., background subtraction [3, 4, 5, 6], temporal differencing [7, 8, 9] and optical flow based techniques [10, 11]. Background subtraction is the commonly used technique, which models the background with respect to foreground image features. Commonly used background models are mean or median based model [12, 11, 13, 14], Gaussian distribution model [15, 16, 17], Mixture of Gaussian (MOG) model [18, 19, 20], Stereo vision based disparity model [21, 22], etc. Stereo vision based background subtraction is one of the efficient methods, which integrates 3D information with 2D disparity points for background subtraction. Initially, it generates a disparity map by matching two video sequences of the same scene with different viewpoints. Obtained disparity map along with 3D world co-ordinate points are used for identifying foreground objects [21, 22].

© 2023-2017 IRJIET All Rights Reserved

Object detection process will be initiated after segmenting moving regions from the background. Commonly used techniques for object detection are foreground blobs [24, 25], template matching [26, 27], explicit 3D shape model [28, 29, 30], statistical shape model [27, 31], low level features [32, 33, 34], multi-cue based approach [35, 36, 9], etc. Foreground blob based techniques use a set of blobs, which are obtained during background subtraction and are examined using analytical techniques. It works based on the assumption that objects are segmented individually in the foreground region and the object occlusion will never occur in the scene. This assumption may not hold true in real-time applications. Template based techniques use a single or multiple 2D object models or templates for object detection. The key idea is to search the developed template within the surrounding regions of an image. The major challenge in this technique is to make the template flexible enough to handle different scale and orientations. Unlike previous methods, explicit 3D shape based techniques use a 3D object model to determine the pose and orientation and the process of creation of the model is simple than the previous methods. In addition, it adopts hypothesis-and- verify approach to enhance the accuracy of the object detection process. Statistical shape model based techniques determine variations in object appearance using a single 2D statistical model. The developed model is trained with initial samples andthe testing is carried out on a very large feature space for determining similar shape. Low level feature based techniques use multiple shifting window of various sizes over the image at different resolutions and a pattern classification algorithms are used for classifying object and non-object regions. The detection of object with varying scale, pose, occlusion and orientation are the critical problems in this technique. Multi-cue based approaches integrate multiple features such as skin color, face and multiple body part detectors in order to enhance the detection accuracy with additional computation load.

2. Research Objectives

This research work developed robust object detection and tracking system using stereo vision with the following key objectives.

1) Developing illumination invariant stereo matching algorithm for disparity estimation.
2) Enhancing the efficiency of the object detection algorithm in an occluded region.
3) Detection and tracking multiple object using illumination invariant stereo matching algorithm.

3. Research Contributions

The major research contributions of this thesis are in three modules namely, disparity estimation, multiple object detection and multiple object tracking. However, in each module there are several sub contributions as given below.

Disparity estimation module: Following disparity estimation techniques are developed to handle the key challenges such as depth discontinuities and radiometric variations in the stereo image.

- A shape adaptive local support region for disparity estimation.
- Hybrid Correlation based stereo data cost for illumination invariant disparity estimation.
- Correlation Fusion based stereo data cost for robust disparity estimation.
- Multiple object detection module: An efficient multiple object detection algorithm is developed with the following sub modules.
- Efficient ground plane estimation and extraction using RANSAC and Mode filtering.
- 3D Region of Interest (ROI) generation.
- 3D ROI refinement techniques.
- Plan-view Significance map generation for occluded object detection on the ground plane.
- Multiple object detection algorithm using connected component based labeling and depth layering technique.
- Multiple object tracking module: A robust multiple object tracking technique is developed with the following sub modules.
- Formulating Kalman filter for multiple object tracking.
- Multiple object tracking algorithm using Kalman filter and object detection rollback loop.

4. Literature Survey

Multiple object detection and tracking in an unconstrained environment encounters variety of real-time challenges. These challenges were broadly tackled using two categories of systems viz., monocular and stereo/multi camera based techniques. The monocular camera systems have addressed majority of the problems, on the other hand some of the critical problems such as multiple object occlusion, object pose, change in illumination condition, etc., are addressed by stereo or multi-camera systems.

Stereo or multi camera system records the scene from more than one different viewpoints and the sparse disparity information is extracted from successive views. The obtained disparity is used to extract depth information, which can also be used for further processing such as object detection and tracking [46, 47]. Multi camera systems are broadly classified into two types based on the distance between the cameras viz., wide baseline system and short baseline system. Wide baseline systems provide flexible viewing angle than short baseline system without prior calibration. Hence, it is an obvious

ISSN (online): 2581-3048
Volume 7, Issue 12, pp 171-176, December-2023
https://doi.org/10.47001/IRJIET/2023.712024

choice by most of the state-of-the-art stereo vision based tracking algorithms. It uses sparse set of feature points for disparity estimation, which reduces the effect of noise in the generated disparity map. Region based stereo algorithm and M2 tracker [47] are designed based on this wide baseline framework, which use 16 cameras to derive object position in 3D space. Thesesystems can solve object occlusion problem and produce global optimal results for multiple object detection and tracking. The major drawback of these techniques is the computation time to develop multiple correspondences among different views. In addition, change in environmental condition will largely affect corresponding point estimation process. Another wide baseline system can be seen in Khan and Shah [48], where people feet on the ground plane is corresponded in different views using planar holography. It allows determining the people location on the ground plane so that it can avoid partial occlusion. This system uses color as a key feature in order to build correspondence among various cameras and this system faces difficulty when different objects have same appearances. In order to overcome this problem, Berdaz et al. [49] developed a motion model using human movement on the ground plane. Radiometric variation between different views has influence on deriving this motion model, which results in inaccurate disparity points.

In addition, it is very difficult to adopt this system to the real-time applications because of the usage of particular techniques or specific features. Salinas et al. [50] made an attempt to reduce the number of cameras and developed a plan-view map known as confidence map using three stereo camera. Even though the method can detect occlusion up to some extent, it demands separate particle filter for tracking each object. Hence, the computation time of the algorithm increases with the increase in number of objects. Short baseline system constructs dense disparity map using corresponding point matching between successive views. Estimated disparities are used for extracting depth information and also to determine the objects of interest in the scene. However, estimated disparity map is sensitive to radiometric variations and majority of the techniques fail to estimate accurate disparity within homogeneous and depth discontinuity regions [51, 52, 53]. In addition, proposed detection algorithms are also sensitive to object occlusion, pose and orientations. Following section gives the details of the techniques, which are developed with an objective to address the aforementioned problems.

5. Object Occlusion

Object detection during close interaction or occlusion is one of the key challenges, which needs to be addressed for real-time application. Recent techniques have shown that, it can be addressed using depth image extracted by a stereo camera[59]. Depth image comprises 3D information, which separates objects during close interaction or occlusion. It enhances the object model by combining scale, shape and appearance information. The first attempt was made in this direction by Darrell et al. [51], where objects were detected and tracked directly using depth image. Again this work has been extended using ground plane estimation and 3D projection [52]. Initially, 3D points along with the calibration parameters are used for identifying the planar region and the direction normal to the planar region is determined. 3D points are projected onto the estimated planar or ground plane region. This method combines multiple stereo views and uses a plan-view statistics known as Occupancy or Density map for detection and tracking. Depth noise is the major cause for this technique as it uses conventional stereo matching algorithm integrated in the stereo camera for disparity estimation.

Harville et al. [22] extended this work by introducing another plan-view statistic known as Height map. It contains highest point of each vertical bin of planar histogram, which preserves the object shape information compared to Occupancy map. The main drawback of Height map is identifying the objects, which are having lesser height than the pre-defined threshold and also due to noise in the depth map. Harville et al. [60] introduced one combination approach, which combines two plan-view statistics such as Height map and Occupancy map for integrated tracking of multiple object on the ground plane. A refinement technique is introduced on the raw Occupancy and Height map, which can remove some of the non-significant objects from the raw map. These refined maps are combined together to track the multiple object. Recently, Tang et al. [61] proposed a fusion based technique, which combines plan-view statistics with the object appearance on the ground plane. Developed technique is efficient than all the previous methods, but lack of appearance features during close interaction restricted it from accurate tracking. It is observed that noise in the depth map and the type of stereo matching algorithm used for disparity estimation are the major cause for stereo vision based tracking techniques. Better efficiency can be achieved by reducing noise in the depth map or by refining the disparity map prior to depth map extraction.

6. Effect of Radiometric Variation on Disparity Estimation

Majority of algorithms proposed in the last few decades work based on the assumption that corresponding pixels will have similar color value. As a consequence, these algorithm utilize a data cost which includes intensity or color difference of raw image for corresponding point estimation. This approach may not hold good for the stereo image, which does not have similar corresponding color value. Few attempts have

ISSN (online): 2581-3048
Volume 7, Issue 12, pp 171-176, December-2023
https://doi.org/10.47001/IRJIET/2023.712024

been made in this direction to address this problem. Illumination variation more specifically radiometric difference between stereo image is one of the important factors, which might be due to variation in camera settings, exposure variations, image noise, non-Lambertian surface, etc. Under this variation, majority of the existing stereo algorithms fail to identify corresponding point which leads to poor disparity estimation [62].

Hirschmuller et al. [62] carried out evaluation of various data cost for disparity estimation with radiometrically different stereo image caused by various factors such as light configuration, change in exposure, variations in gamma correction, noise, etc. Evaluation compares Birchfield and Tomasi (BT) data cost [63], BT with Laplacian of Gaussian (LOG) [64], BT with mean filtering, BT with Rank Transform (RT) [65], Normalized Cross Correlation (NCC) and also some of the local, global and semi global methods. BT is a popular data cost, which is insensitive to camera sampling but sensitive to radiometric variations as it employees linearly interpolated function of intensity values. Although LOG filter is insensitive to outliers due to the usage of the second order derivative, it is sensitive to radiometric variation. BT with Rank Transform [65] shows the robustness to global radiometric variations due to the principle of rank ordering based on pixel intensities but weak at local radiometric variation. Ogale et al. [66] presented a contrast invariant stereo matching algorithm, which can compensate only global variations. Normalized Cross Correlation (NCC) [67] is one of the popular similarity measures, which can compensate global variation but suffers from fattening effect across the object boundaries. Heo et al. [68] developed a new correlation measure known as Adaptive Normalized Cross Correlation (ANCC) using log chromaticity color. This correlation measure can compensate local variation but is weakened at camera exposure variation. The extension of this work using mutual information and SIFT descriptor are efficient in terms of disparity error with additional computation load, which prevent them from being used in real-time applications. Recently, Jung et al.[69] proposed color adaptation and pseudo-disparity vector for stereo matching under radiometric variations. Method can compensate global variations and be weakened at local variations.

The other way to deal with illumination variation is by extracting illumination invariant image from the raw color image and finding correspondence between them [70]. Color constancy and color invariant approaches are the two techniques used for finding the illumination invariant image [71]. Color constancy algorithms [72, 73, 74, 75, 76] attempt to separate the illumination and the reflectance component from an image. The Gamut-mapping algorithm [73] and color-by-correlation algorithm [73] are the first attempt in this direction, which can estimate the illuminant of a given image. Retinex algorithm is another technique, which calculates lightness sensations rather than the physical reflectance of a given image and effectively compensates non-uniform lighting [72, 74, 75]. Color invariant approach [77, 78] tries to find lighting and device independent function of an image.

Among these approaches, chromaticity normalization and gray-world assumption are the commonly used methods [79]. Chromaticity normalization is used to remove lighting geometry effects, while gray-world assumption is used to remove illuminant color effects. However, neither chromaticity normalization nor gray-world assumption can remove both lighting geometry and illuminant color dependencies simultaneously.

7. Effect of Depth Discontinuity on Disparity Estimation

Local stereo matching algorithms are commonly used for real-time applications. It uses a local neighborhood of similar depth for disparity estimation. Depth discontinuity in the extracted local neighborhood poses critical challenges and varieties of algorithms have been proposed in the literature for disparity estimation within the depth discontinuity regions. These algorithms can be broadly classified into two different types. In the first type, there are two different techniques; single/multiple window techniques and shape adaptive techniques. First technique selects optimal support region based on the pre-defined single/multiple windows [80, 81] and the second technique adapts window shape/size based on the pixel information [82, 83]. These methods use the rectangular window even at depth discontinuities. Lu et al. [84] proposed anisotropic polygon based shape adaptive support region and the generated polygon is not flexible enough for approximating different scene structure as it is built on the varying scale sector. The second type concentrates on the adaptation of weights associated with the selected support region with fixed shape and size. Xu et al. [85] proposed radial computation based adaptive weight using initial disparity estimation. It uses three parameters to determine the support weight viz., certainty, color and disparity distribution correlation and generates certainty map. The obtained certainty map of initial disparity estimation is used to decide the size of the support window. It is observed that the obtained initial disparity is sensitive at depth discontinuities and texture less regions due to the variations in the assigned weights. The assigned weights increase whenever the difference between neighboring pixels decreases or disparity distribution correlation increases. A pixel color similarity based adaptive weight technique is proposed in [86, 87], which consumes huge memory due to storage of central pixel dependent support weight. Zhang et al. [88] proposed cross based local stereo matching, which uses horizontal and vertical segments of the anchor pixel for constructing local support region.

It can be observed that majority of state-of-the-art algorithms use vertical or horizontal segments for constructing shape adaptive local support region. These algorithms are computationally expensive and none of them is flexible enough to capture different scene structures.

8. Conclusion

Variety of stereo vision based object detection and tracking techniques are outlined in this chapter. It also includes detailed review of the state-of-the-art techniques, which are developed to address object occlusion and disparity estimation problems. These techniques use stereo information in a variety of approaches to increase the robustness of the algorithm for handling object occlusion and disparity estimation. However, it is observed that quality of the disparity map decides the accuracy of the detection and tracking methods. In addition, it is also observed that only few attempts have been made to develop higher quality disparity map. Instead, it uses conventional stereo matching algorithm followed by expensive refinement techniques. For example, method proposed in [22] uses conventional stereo matching algorithm integrated in the Digiclops stereo vision camera for disparity estimation [89, 90]. Obtained disparities are used for occluded object detection using plan-view statistics [22, 60]. It is known that conventional stereo matching algorithms do not produce reliable disparities in homogenous and depth discontinuity regions. It is also sensitive to small variation in radiometric condition between stereo image, which results in creation of unknown regions in the disparity map. Moreover, state-of-the-art plan-view statistics based occluded object detection algorithms are sensitive to depth noise. Hence, it demands an efficient stereo matching algorithm, which can handle depth discontinuity and radiometric variations during disparity estimation. This disparity estimation technique avoids expensive refinement technique and automatically improves robustness in handling object occlusion in the plan-view.

The above made observations have motivated to develop stereo vision based object detection and tracking technique, which can handle several real-time difficulties within a single framework. Block diagram of the developed technique is given in Figure 2.1. First contribution of this thesis develops efficient disparity estimation techniques. Initially it developed a shape adaptive local support region based disparity estimation technique to address depth discontinuities followed by two robust correlation measures for compensating both local and global radiometric variations during disparity estimation. The second contribution of the thesis developed an occluded object detection technique using 3D ROI and a plan-view Significance map. Third contribution developed a multiple object tracking technique by integrating Kalman filter with detection rollback loop under continuous detect-and-track framework.

REFERENCES

[1] DucPhu Chau, Fran_cois Bremond, Monique Thonnat. Object Tracking in Videos: Approaches and Issues. The International Workshop "Rencontres UNS- UD" (RUNSUD, Danang, Vietnam. 2013.

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In The International Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, June 2005, pp. 886–893.

[3] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In The Proceeding of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), April 2003, pp. 511–518.

[4] F. Faradji, A.H. Rezaie, M. Ziaratban, A morphological-based license plate location, in: Proceedings of IEEE International Conference on Image Processing, 2007, pp. 57–60.

[5] K. Zheng, Y.X. Zhao, J. Gu, Q.M. Hu, License plate detection using Haar- like features and histogram of oriented gradients, in: Proceedings of IEEE International Symposium on Industrial Electronics, 2012, pp. 1502–1505.

[6] H.H.P. Wu, H.H. Chen, et al., License plate extraction in low resolution video, in: Proceedings of the 18th International Conference on Pattern Recognition, 2006, pp. 824–827.

[7] J. Kwon, K.M. Lee, F.C. Park, Visual tracking via geometric particle filtering on the affine group with optimal importance functions, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 991–998.

[8] N. Wang, D. Yeung, Learning a deep compact image representation for visual tracking, Proceedings of 27th Annual Conference on Neural Information Processing Systems (NIPS), 2013.

[9] J. Zhao, Z. Li, Particle filter based on Particle Swarm Optimization resampling for vision tracking, Elsevier Expert Syst. Appl. 37 (12), 2010, pp. 8910–8914.

[10] B. Zhang, W. Tian, Z. Jin, Robust appearance-guided particle filter for object tracking with occlusion analysis, Elsevier Int. J. Electron. Commun. 62 (1), 2008, pp. 24–32.

[11] R. Qin, S. Liao, Z. Lei, S.Z. Li, Moving cast shadow removal based on local descriptors, ICPR, 2010.

[12] N. Martel-Brisson, A. Zaccarin, Learning and removing cast shadows through a multi distribution approach, TPAMI 29 (7), 2007, pp. 1133–1146.

[13] [2.10] N. Martel-Brisson, A. Zaccarin, Kernel-based learning of cast shadows from a physical model of light sources and surfaces for low-level segmentation, CVPR, 2008.

[14] Lee, W, Lee, G, Ban, S.-W, Jung, I, & Lee, M. (2011). Intelligent video surveillance system using dynamic saliency map and boosted Gaussian mixture model. In International conference on neural information processing (pp. 557– 564). Springer.

[15] Wang, M. ,Qiao, H. , & Zhang, B, A new algorithm for robust pedestrian tracking based on manifold learning and feature selection. IEEE Transactions on Intelligent Transportation Systems, 12, 2011, pp. 1195–1208.

[16] Zhang, L, Li, Y &Nevatia, R, Global data association for multi-object tracking using network flows. IEEE conference on computer vision and pattern recognition (CVPR), 2008, pp. 1–8.

[17] Doucet, N. D. Freitas, and N. Gordon, Eds., Sequential Monte Carlo Methods in Practice. Berlin, Germany: Springer, 2001.

[18] H. Seo and P. Milanfar, ―Training-free, generic object detection using locally adaptive regression kernels,‖ IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, Sep. 2010, pp. 1688–1704.

Citation of this Article:

Dr. Ramesh Palanisamy, Mr. Mohamed Osman Akarma Al-Tigani Mohamed, Dr. Mathivanan Viruthachalam, Dr. Kumar Kaliyamoorthy, Senthil Jayapal, "A Method of Detecting an Object Using the Latest Technology" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 7, Issue 12, pp 171-176, December 2023. Article DOI https://doi.org/10.47001/IRJIET/2023.712024

****