International Journal of Computer Applications in Engineering Sciences
[VOL III, ISSUE I, MARCH 2013] [ISSN: 2231-4946]

A Critical Survey of Moving Object Detection Techniques and Related Proposed Research

Jeba Veera Singh.J[1] and Nancy Emymal.S[2]

[1]*Dept of CSE,SCAD College of Engineering and Technology,Tirunelveli,India.*
[2]*M.E. Graduate, Tirunelveli Dist., Tamil Nadu, India*

jebaveerasingh.j@gmail.com
s.nancyemymal@gmail.com

Abstract-- **Detection of moving objects in a video sequence is a difficult task and robust moving object detection in video frames for video surveillance applications is a challenging problem. This paper presents a critical survey of various techniques related to video surveillance system that can lead to improved object detection. In this critical survey we discuss various techniques such as, background subtraction, shape analysis, pixel level motion and others that are employed to detect moving objects in a video sequence.**

***Keywords*— Moving object recognition, Video sequences, Background subtraction, Shape Analysis, Motion Detection.**

I. Introduction

Detection of objects from static images is common and relatively simple, when compared to the detection of objects from video sequences. The moving object detection has many applications in video surveillance systems as employed by law enforcement agencies and the military, for example. The detection and tracking of moving objects draw great attention from the researchers in the area of computer vision. The video sequences can be captured from a single camera or multiple cameras placed at many places. Many video surveillance systems use stable and still camera for continuous surveillance, deals with captured sequence of images and frame, in order to detect a moving object correctly. In this paper we will discuss the various techniques used, their relative performance characteristics and a critique of each of them. The rest of the paper is organized as follows: section 2 provides a background to object detection and three sub-categories therein, while section 3 provides our proposed research work. The paper is summarized in the conclusions, which also provides pointers for further research in this area.

II. Object Detection

Object detection is a process that deals with detecting instances of semantic objects of a certain class, such as, humans, cars or buildings. Object detection mechanism can be classified into three categories, namely, Background Subtraction, Shape Analysis and Pixel level motion. These are further described in the following sub-sections.

A. Background Subtraction

1) Color and Edge Information:

Jabri, et al. [1] proposed an approach in which background modeling and subtraction approach are used to detect a human in the video images. This approach is used to segment the person from the background by computing the mean image for all video sequences. The incoming frame is subtracted from the mean image to identify the pixels which have changed the color. However, the problem with this approach is both the color and edge channel are subtracted separately before finding the result and, as a consequence, the computational time increases.

2) Standard Subtraction:

The method developed by Davis and Taylor [3] is a motion-based method for differentiating normal walking movements at multiple speeds when atypical or non-walking locomotion is involved. Human walking movements are detected using low level regularities and constraints. The person's shape in each video frame is extracted with standard background subtraction. This approach locates the head, waist and feet using the W4 approach [12]. Standard subtraction techniques, which use RGB pixel differences, dilations and removal of small pixel region, are employed. The centroid of the outline pixel is called the head pixel, while the mean value of silhouette pixels in the torso region is called the waistline. The waistline is divided into two halves in order to locate information relating to the feet. Dynamic regularity features are calculated using cycle time, stance/swing ratio and double support time. Dynamic regularity features are independent of the camera position, but this approach uses view-based constraint of extension angle, which is suitable for non-walking locomotions and not for other regular locomotions.

3) Background Subtraction and Temporal Differencing:

An approach for moving object detection has been introduced by Li, et al. [5] in which background subtraction and temporal differencing is used to segment the foreground from background. This approach segments the moving object with low false alarm. The color images will be transformed into grey scale image and the image background will be updated to analyze the changes of pixel intensity. The moving object is recognized by subtracting the incoming frames from background frame. During the temporal analysis phase, an estimate of the similarity between the values that a pixel assumes in two consecutive frames is performed. The limitation of this approach is the need for the regular update of background images.

4) Object Extraction:

The algorithm proposed by Yoginee, et al. [7] has moving object segmentation, blob analysis and tracking. Blob analysis is used to count the vehicle from which the speed and flow are calculated. Boundary Block Detection (BBD) algorithm is used for moving object detection by identifying the blocks which contain the moving objects boundaries. The system requires the model background with no moving objects and scene which contain moving objects. The system finds the boundary of the moving objects and the number of moving objects from a given video scene. Aviread function [7], is used to extract all frames in the video. Background subtraction extracts the object, while the pixels of the background model image are used as threshold. All images are divided into two parts, viz., background and foreground in binarization. The new video frame was subtracted from those background images, if the pixel difference is higher than the threshold, that images are foreground or object. If the pixel significantly differs from the background image, then the pixel is marked as a moving object. Each image frame must update the threshold level. To count the moving object flow, the algorithm tracks each vehicle within successive image frames. This algorithm works only for the videos obtained from fixed cameras and which has the normal background and stable videos. The algorithm can be modified to work on complex background and videos that are not stable. In addition, the performance can be improved by using optimizing algorithm such as fuzzy logic and neural network.

B. Shape Analysis

1) Hypothesize and Verify:

Zhao, et al.[2] proposed a method in which 3D model based approach was used to detect and track the human in complex situations. At first the foreground blobs are extracted by background subtraction method and human hypothesis are generated by shape analysis of the foreground blobs using human shape model. Human hypothesis are computed by boundary analysis to find the top candidate and shape model. In boundary analysis top most (peak) point represents a head of the top candidate. Additionally, flat peaks are also considered and midpoint of the flat peak is known as head for other than top candidate. The peaks which do not have enough foreground pixels are abolished. Once the head of the top candidate is selected, the potential height has to be found. Once the height is set, width of the ellipse will be computed by average height. In this process, the shadow pixels get eliminated. These detected human hypotheses are tracked in 3D in every subsequent frame using Kalman Filter [13]. Using the static image from one frame does not verify the correctness of the hypothesis; therefore, the hypotheses depict a human walking pattern. Walking humans are recognized using human articulated motion model and the motion of legs used to verify the hypotheses. But the perspective effect of the camera position may cause a difference in the noise measurement in 3D and, as a consequence, newer techniques are needed to verify humans walking along the camera view angle.

2) Region of Interest:

Sidla, et al.[4] introduced an approach in which pedestrian motion and shape are computed using region of interest. All the foreground pixels are extracted using KLT algorithm [11], which forms the region of interest. Histograms are used to select the background, which has the highest grey value. When the pedestrian makes movements, the change in grey value leads to the detection of the pedestrian. The histograms of every new frame will be compared in order to find the shape of the objects in the foreground. Using Active Shape Model, the pedestrian's shape is identified based on human head shoulder regions [10]. The movement was tracked using KLT algorithm or Kalman filter, however this detection technique is dependent on the position of camera. An interesting aspect of this algorithm is that the speed of this technique can be improved to be on par with real time.

3) Joint 3D Motion:

Harguess, et al. [6] used multiple cameras to detect the full motion of the object, which can be recovered robustly and accurately. This approach fuses multiple camera information into a joint 3D motion calculation. The faces in multiple cameras are then tracked by using this approach. The motion calculation is realized to utilize the 3D cylinder model. Full motion recovery is applied from multiple cameras model to face recognition. This method extends the full motion recovery in multiple cameras which track the face independently. To improve the robustness of motion estimation of the face, the motion information from multiple cameras combined simultaneously. But the limitation of the algorithm is that the loss of tracked faces in independent cameras cannot be recovered.

C. Pixel Level Motion

1) Space Time Interest Points:

Lejeune [8] used the Space Time Interest Points (STIP) approach, which uses simple and low level features that efficiently characterizes moving objects in a video and changes in the movements of the noted objects. STIP focuses on thousands of pixels which contain information, but the detection rate decreases as the detection of STIPs increases.

2) Background Modeling:

Pollard and Antone [9] proposed the Background Modeling approach, which provides a simple, fast and robust motion detection, which handles parallax. The pixel level motions are detected using a Gaussian Mixture Model (GMM) to handle misalignment and parallax. The spatial temporal filtering is used for quickly discarding spurious detections. The pixels are grouped into discrete blobs in motion segmentation. The shape of each object is learned adaptively over multiple frames based on the detection blobs associated with that object. This method incorporates multi-cue fusion for state evolution that enables tracking through confusion, occlusion, and stops. The sporadic false alarms are rejected using temporal consistency. To improve efficiency the geometric attributes (position, size, and shape) are considered. But the problem is that the shape update has to be disabled, if the object velocity falls below a threshold level.

III. PROPOSED RESEARCH WORK

To improve the accuracy of moving object detection, our proposed method will employ a four steps mechanism as noted below: See Fig.1.

- With the help of a well-defined background model, the foreground objects will be identified using background subtraction.
- After feature extraction, tile function will be used to detect a moving object.
- As noted before, camera position may cause a difference in the noise measurement in 3D. We will therefore be examining this problem carefully so that verification of human walking along the camera view angle can be better achieved.
- We would also be improving upon the Kalman filter algorithm so that its performance in real-time scenarios will be acceptable.

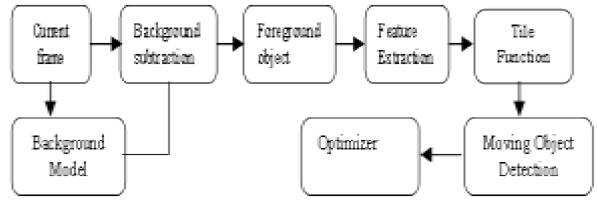

Fig.1. Framework of Proposed Research Work

IV. CONCLUSIONS

We discussed a variety of techniques to detect moving object in video frames and their advantages and limitations. In the background subtraction method, the subtraction of color and edge channels is performed separately before finding out the result. The background model should be updated at frequent intervals, while in shape analysis the threshold level of the velocity should be different for different type of objects. The detection of STIPs is inversely proportional to the detection rate of objects and hence the detection of STIPs increases as the detection rate decreases. The techniques can be improved in order to improve the accuracy and time consumption. Furthermore, in this paper we have proposed a combined method of feature extraction and tile function, such that the time and accuracy attributes can be improved.

REFERENCES

[1] Sumer Jabri, Zoran Duric, Harry Wechsler, Azriel Rosenfeld, "Detection and Location of People in Video Images Using Adaptive Fusion of Color and Edge Information," In Proc. 15th Int'l Conf. on Pattern Reg, 2000,vol. 4,pp. 627 – 630.

[2] Tao Zhao,Ram Nevatia,Fengjun Lv, "Segmentation and Tracking of Multiple Humans in Complex Situations," In Proc. IEEE Computer Society Conf. on Computer Vision and pattern Recognition, 2001, Vol.2, pp. 194 – 201.

[3] James W. Davis, Stephanie R. Taylor, "Analysis and Recognition of Walking Movements," In Proc.16th Int'l Conf. on pattern Recognition, 2002, vol.1, pp. 315 – 318.

[4] Oliver Sidla, Yuriy Lypetskyy, Norbert Brandle, Stefan Seer, "Pedestrian Detection And Tracking For Counting Applications In Crowded Situations," In Proc. IEEE Int'l conf. on advanced video and signal based surveillance, 2006,pp.70.

[5] Guanglun Li,Yanling Wang, Weiqun Shu, "Real-time moving object detection for video monitoring systems," 2nd Int'l Sym. on Intelligent Info. Tech. Application (IITA),2008,vol.3,pp. 163 – 166.

[6] Josh Harguess, Changbo Hu, J.K. Aggarwal, "Full Motion Recovery From Multiple Video Cameras Applied To Face Tracking And Recognition," IEEE

Int'l Conf.on Computer Vision Workshops (ICCVW),2011,pp. 1889 – 1896.

[7] Yoginee B. Bramhe(Pethe), P.S. Kulkarni, "An implementation of Moving Object Detection, Tracking and Counting Objects for Traffic Surveillance System," Int'l Conf. on Computational Intelligence and Comm. Networks (CICN), 2011, pp. 143 – 148.

[8] Alain Simac Lejeune, "Moving Object Analysis in Video Sequences Using Space Time Interest Point," 7th Int.Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications,2012, pp.201-204.

[9] Thomas Pollard, Matthew Antone, "Detection and Tracking All Moving Objects Wide Area Aerial Video," IEEE Computer Society Conf. on Computer Vision and Pattern Recog. Workshops (CVPRW),2012, pp.15 – 22.

[10] T.Zhao and R.Nevatia, "Bayesian human segmentation in cowded situations," In Proc IEEE conf.computer vision and pattern Recog.II,2003,pp.459-466.

[11] C. Tomasi and T. Kanade, "Detection and Tracking of Point Features", Carnegie Mellon University Technical Report CMUCS- 91-132, April 1991.

[12] I. Haritaoglu, D. Harwood, and L. Davis, "W4: Who? When? Where? What? A real time system for detecting and tracking people". In Proc. Int. Conf. Auto. Face and Gesture Recog., 1998, pages 222–227.

[13] G. Welch, G. Bishop, "An Introduction to the Kalman Filter," TR-95-041, Dept. of Computer Science, University of North Carolina at Chapel Hill.