IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)
Volume 6, Issue 5, Ver. I (Sep. - Oct. 2016), PP 10-16
e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197
www.iosrjournals.org

Multiclass SVM and HoG based object recognition of AGMM detected and KF tracked moving objects from single camera input video

Bob P. George[1], Anoop K. Johnson[2]
[1] *Dept. of Electronics and Communication Engineering Mar Baselios College of Engineering & Technology Kerala, India*
[2]*Asst. Professor, Dept. of Electronics and Communication Engineering Mar Baselios College of Engineering & Technology Kerala, India*

Abstract: *Object detection and tracking are two fundamental tasks in video camera surveillance. In the case of moving object detection and tracking, an integrated Kalman Filter based system can be used. Automatic object detection is usually the first task in a camera-based surveillance system and background modelling (BM) is commonly used to extract predefined information such as object's shape, geometry and etc., for further processing. But occlusion handling is required to perform advanced recognition and segmentation processes. Training based could be employed for recognition purposes on the frame outputs of a single surveillance camera video. A HoG feature based SVM trained classifier can be utilized for the recognition. The proposed framework configuration is not quite the same as existing multi-camera observation frameworks which use basic image data extricated from comparable field of vision (FOVs) to enhance the object discovery and tracking execution.*

Index Terms: *Object detection, tracking, recognition, background modelling, HoG, Kalman filter, SVM*

I. Introduction

Camera based surveillance is now an important security procedure present in almost all public spaces. With the advent of less expensive digital hardware, the use of camera has become a common occurrence. But manual presence is still necessary to monitor the surveillance videos to detect anomalies. Object detection and tracking along with recognition [5] can be used for the automatic analysis of object of importance in a particular scene. It can be inferred that moving objects tend to be the objects of importance in almost all surveillance videos or camera feeds. Surveillance is a very broad term which includes monitoring human activities, crowd activities, traffic, etc. The approach to surveillance thus depends on the application for which it is being used. There are many existing techniques for road traffic monitoring like the use of sensors based on radar and LIDAR. Other techniques employ the use of multiple cameras [3] [7] to monitor the scene or the use of stereoscopic cameras [1].

The proposed system employs the use of Gaussian Mixture Models (GMM) for the object detection process. Tracking is achieved using the Kalman filter which is able to predict the future state from the present state and the past states of the system. The recognition phase utilizes features based on HoG (Histogram of Gradients) and SVM (Support Vector Machine) based classification using a database of objects.

II. Proposed system

The proposed system is aimed to detect, track and recognize moving objects which are treated as objects of importance in a scene which is generally assumed to be a surveillance camera input with a static background. The overall system consists of three processing blocks performing the three main functions: object detection, object tracking and object recognition.

The object detection block primarily focuses on the task of segregating the background and foreground pixels to correctly identify the object of importance. This is the primary and critical step necessary for the later blocks like the tracking and recognition blocks of the system. The objects of importance are generally assumed to be the moving objects in the traffic monitoring or surveillance video. Object tracking block primarily deals with the successful tracking of a detected object over the subsequent frames until the object exits the scene. The presence of multiple object also must be acknowledged as well. Each different object must be tracked differently and a distinction must be made between the objects that are moving in the video. The tracking block makes use of the Kalman filter for this purpose. The property of the Kalman filter to work under cases of occlusion and to give satisfactory results in such cases is also utilized. The object recognition phase deals with the task of identifying the object with respect to an offline database which contain description of different objects. The

feature used for this process is the HoG [13] feature. Feature matching is performed and the classification is done based on an SVM classifier [15]. Each of these blocks are explained in detail below.

A. Object detection

Automatic object detection is usually the first task in a multi-camera surveillance system and background modelling (BM) is commonly used to extract predefined information such as object's shape, geometry and etc., for further processing. Pixel-based adaptive Gaussian mixture modelling (AGMM) [10] is one of the most popular algorithms for BM where object detection is formulated as an independent pixel detection problem. It is invariant to gradually light change, slightly moving background and fluttering objects.

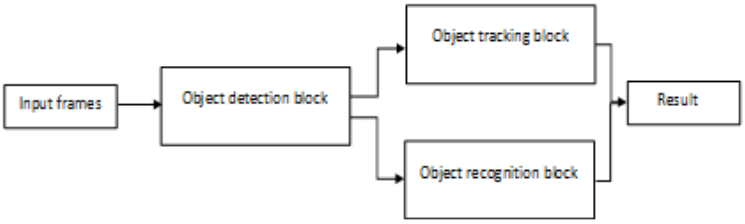


Fig. 1. Proposed system architecture

For extracting objects from the background, a wide variety of methods can be utilized. A normal background subtraction can be performed if a statistical model of the scene in available. An incoming object can be detected by recognizing the pixels of the image that do not fit the model. The result of a background subtraction is a binary thresholded image which highlights the incoming objects which are separated from the background. The simplest process of creating a reference background image is by using a time averaged background image where the training period comprises of background images devoid of any foreground objects. This method suffers from some problems like the false background problem where the foreground object which was motionless during training will be treated like background. Another problem is the motion of background objects after training period which will lead to their detection as foreground objects. Thus adaptive background modelling methods have to be utilized to overcome these shortcomings.

In the case of an adaptive GMM, each background pixel can be modeled by a mixture of Gaussian distributions where different Gaussians can be assumed to represent different colors. The weight coefficients of the mixture can relate to the time over which the colors stay on the scene. The colors which stay longer and appear to be more static have a higher probability to be a background color. An update scheme is applied to allow the model to adopt to illumination changes. Thus the background model can be said to be estimated from a training set. In order to adapt to the changes in the scene this training set is constantly updated by adding new samples and discarding old ones. After adding the new samples to the training set, the density is re-estimated.

The generated output of the GMM based background subtraction yielded a binary image with undesirable noise in the foreground as well as the background. Thus a post processing stage is necessary after the BM needs to be performed in order to eliminate the noise generated. The post processing stage consists of utilizing morphological operations. A square shaped structuring element was thus used to perform a standard opening operation to filter out the noise in the binary image. The opening operation consist of using an erosion operation followed by the dilation operation. This results in a relatively clean binary image with the necessary foreground objects represented by while pixels grouped together while the background is represented using black pixels. After performing the necessary morphological operation, a blob extraction procedure is then initiated. The blob can be defined as a group of pixels which can effectively represent an object present in the scene. Distinct blobs in the scene are then marked using bounding boxes.

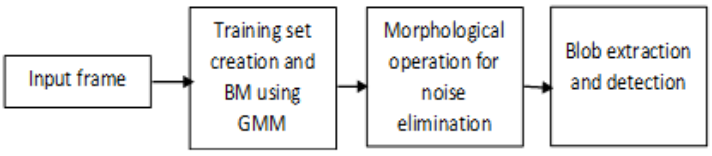


Fig. 2. Object detection

B. Object tracking

Object tracking block deals with the detection of objects in the subsequent frames of a video and to estimate its location throughout the video while differentiating it from the other objects under consideration. In terms of tracking strategy, the classification of object tracking results in two categories, namely, deterministic and

probabilistic tracking. In the case of deterministic tracking, an example is that of the mean shift tracker (MS tracker) due to its simplicity and relative effectiveness. But presence of occlusion or similar objects in the scene degrades the performance of the MS tracker.

The tracking procedure starts as soon as the object enters the frame and is correctly detected by the detection block. The detection block is responsible for providing the correct object mask which helps to separate the object from the background pixels. This mask is supplied to the Kalman filter which works on one of the two models- constant velocity model or the constant acceleration model. The measurement noise and the process noise parameters can also be set during the configuration of the Kalman filter.

The Kalman filter procedure executes two stages, prediction and correction. The prediction step uses the past states, i.e. the location of the object of interest in past frames, to foresee the present state which is the present object location. The correction step or the update step utilizes the current measurement of the objects location to correct the predicted location. A track is assigned to each object to maintain the state of a tracked object. From the binary mask obtained from the detection block, the centroid and bounding boxes are utilized to predict the centroid of each track in the current frame. Assigning object detections to existing tracks is done using a minimization procedure based on a cost function. The cost utilizes the motion model used while configuring the Kalman filter and the observed likelihood function. The detection is assigned to the track with the minimum cost. The cost takes into account the Euclidean distance between the predicted centroid of the track and the centroid of the detection. If the cost is high for assigning the detection to an existing track, a new track is created for the detection. Over each input frame the tracks are thus updated and maintained. If any track has been invisible for too many frames, then that track is deleted and the object is assumed to have exited the scene.

Fig. 3. Object tracking

C. Object Recognition

The object recognition block uses an offline model for classifying the objects based on a database. This process utilizes a pre trained SVM classifier. A database was created from the object categories and images available from the publicly available Caltech database. The training set is constructed from the labelled images present in the database and the SVM classifier is trained based on the HoG features extracted from the images. Different cell sizes can be used for the HoG feature like the 2x2, 4x4 and 8x8 sizes. The 2x2 cell size includes maximum shape information at the cost of increased dimensionality of the HoG feature. The 8x8 has the advantage of decreased dimensionality but minimum shape information. A multiclass classifier using binary SVMs are used for the classification purpose where each image is associated with a definite HoG feature vector. During the training phase the classifier is correctly provided with the training set and labelled input images. This trained classifier is utilized in the test phase where the objects associated with the binary mask during detection phase is provided as the test inputs to the classifier. The recognized category out of the available category is then used to define the object that was detected and tracked.

In order to provide the minimum amount of clear detected image masks and to minimize object recognition time, an optimization algorithm is utilized. Here the object is assumed to be in the 'optimal region' of the frame after 40 frames while using the constant velocity model for tracking purposes. The distinct mask from the tracked object is first segregated and then passed as test input to the object recognition block where its features are extracted and then matched with the trained features. The predicted label is then generated and is then used to identify the object of interest based on the resultant prediction. This method can be practiced irrespective of the nature of the object classification block and the nature of the features involved and also interfering much with the detection and tracking process which takes place independently.

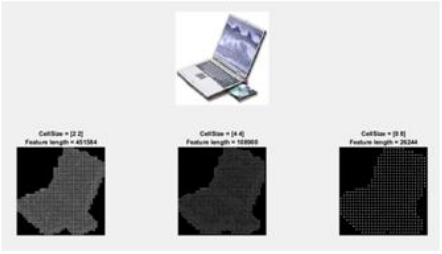


Fig. 4. Input object and HoG feature visualized for 2x2, 4,4, 8x8 cell sizes

II. Results and discussions

A. Object detection results

Different video sequences were used to analyse the performance of the detection block. The result of the GMM based background subtraction is shown below.

Fig. 5. Noisy binary mask result after GMM based background subtraction

In order to eliminate the noise morphological operation was performed on the noisy frames. The resulting frames was devoid of most of the present noise pixels in the background and foreground.

Fig. 6. Binary mask after morphological operation

After obtaining the filtered binary mask blob extraction and detection was performed. Blobs with area less than 400 pixels were rejected to avoid false positives.

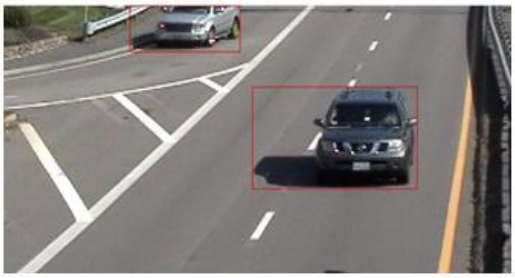


Fig. 7. Object detection output after blob extraction

B. Object tracking results

The binary mask from the detection block is used as input to the tracking block. Kalman filter was used to assign and update tracks. A unique tracking ID was supplied to each object which appears in the scene. The tracks which were invisible for a long duration were deleted since the object was assumed to have exited.

Fig. 8. Object tracking and track assignment on binary mask input

Each bounding box was updated with the unique ID of the track and this can be utilized to get an approximate count of the vehicles also. Multiple objects were tracked simultaneously without much difficulty using this method also.

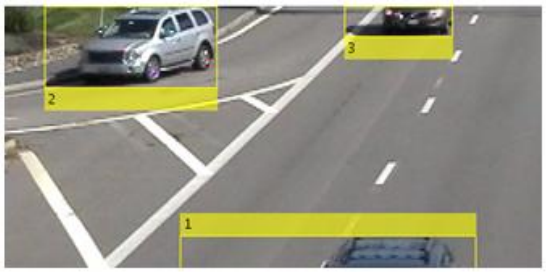


Fig. 9. Object tracking output for traffic monitoring

C. Object recognition results

The binary mask based input is again utilized to obtain the classification results and thus categorize the objects detected. The mask is used to determine the objects to be recognized under minimal occlusion by using the optimal area with chances of minimum occlusion. This greatly reduces error due to false predictions during entry, exit and background occlusions.

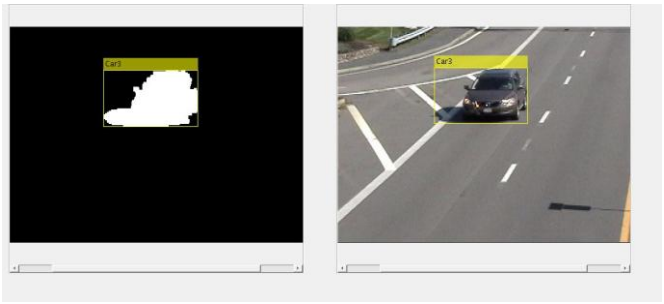
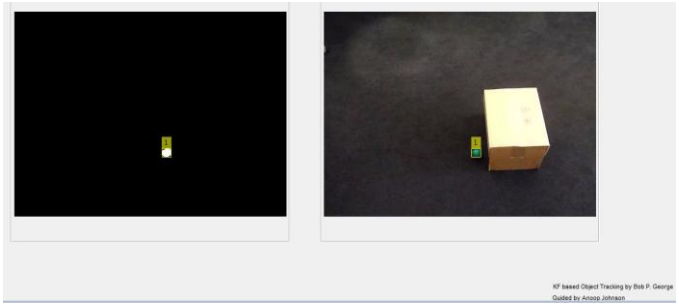


Fig. 10. Object recognition output under minimal occluded conditions

Fig. 11. Object recognition output under occluded conditions

D. Performance under occluded conditions



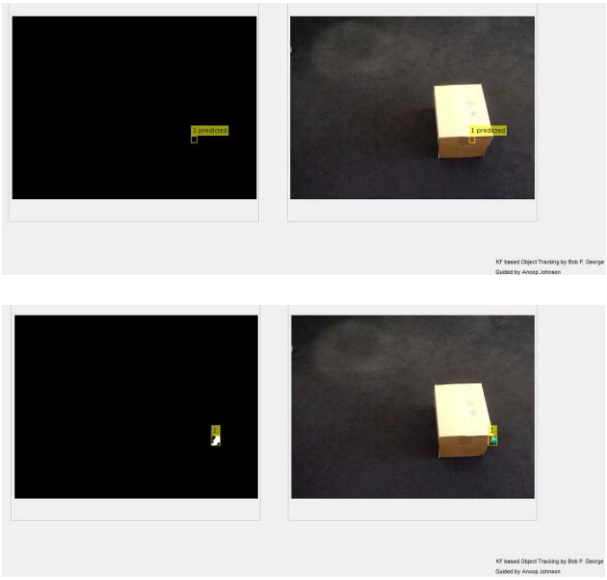


Fig. 11. Object prediction output under fully occluded conditions

The above result is obtained for a video containing a moving object which was occluded for a maximum of 7 frames. The prediction is found to hold true in such a case where the object is fully occluded by a background object during 7 frames. The object upon re-entry was again tracked as before.

TableI. Performance Under Minimal Occlusion

Video input	**Video 1 (traffic movement)**
No. of training frames	**50**
Min. blob area	**400**
Total no. of objects in video	**7**
Total tracks created	**7**
Track normalization frames	**40**

TableII. Performance under occlusion

Video input	**Video 2 (human movement)**
No. of training frames	**50**
Min. blob area	**400**
Total no. of objects in video	**12 (moving)**
Total tracks created	**16**
Track normalization frames	**40**
Occlusion conditions	**Entry, exit, background, foreground and self occlusions**
Object overlap cases	**3**

The above tables illustrates the various observation made for different video inputs with varying occlusion conditions present.

III. Conclusion

In this paper, a system for object detection, tracking and recognition from a single camera video scene was presented. Background modelling using GMM was utilized for background subtraction operation, Kalman filter for tracking purposes and HoG feature for recognition. The advantage of this method with respect to the existing methods is the automatic and adaptive object detection process which can withstand variations in illumination and small changes in background. Also Kalman filter based tracking makes this system perform much better in the presence of occlusions. The HoG based recognition facilitate the use of more categories by updating the classifier with the relevant training set. The recognition rate is improved using only the minimally occluded object masks for recognition at select frames determined by the optimal area of minimum occlusion occurrence.

References

[1] Shuai Zhang, Chong Wang, Shing-Chow Chan, Xiguang Wei, and Check-Hei Ho, "New Object Detection, Tracking, and Recognition Approaches for Video Surveillance Over Camera Network Sensors Journal, IEEE, Volume: 15, Issue: 5, May 2015.

[2] Akhil A. Vijay, Anoop K. Johnson, "An Integrated System for Tracking and Recognition using Kalman Filter", IEEE International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), July 2014

[3] Oytun Akman, A. Aydin Alatan and Tolga C iloglu. "Multi-Camera Visual Surveillance for Motion Detection, Occlusion Handling, Tracking and Event Recognition", Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications - M2SFA2 2008, Oct 2008, Marseille, France

[4] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," ACM Comput. Surv. vol. 38, no. 4, pp. pp. 1–13, 2006.

[5] Chun Yuan, Wei Xu, "Multi-Object Events Recognition from Video Sequences using Extended Finite State Machine", 2011 4th International Congress on Image and Signal Processing.

[6] Charles Bibby, Ian Reid, "Real-time Tracking of Multiple Occluding Objects using Level Sets", 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[7] Vera Kettnaker and Ramin Zabih, "Bayesian Multi-camera Surveillance", 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[8] Xue Mei and Haibin Ling, "Robust Visual Tracking using l1 Minimization", 2009 IEEE 12th International Conference on Computer Vision (ICCV).

[9] Shai Avidan, "Ensemble Tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, No. 2, February 2007

[10] Zoran Zivkovic and Ferdinand van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction", Elsevier, Pattern Recognition Letters 27 (2006) 773–780.

[11] Boris Babenko, Ming-Hsuan Yang and Serge Belongie, "Robust Object Tracking with Online Multiple Instance Learning", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 8, August 2011.

[12] Ilan Shimshoni, Amit Adam and Ehud Rivlin, "Robust Fragments-based Tracking using the Integral Histogram", 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)

[13] N. Dalal and B. Triggs, "Robust Fragments-based Tracking using the Integral Histogram", 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, pp. 886 - 893 vol. 1

[14] Shunli Zhang, Xin Yu, Yao Sui, Sicong Zhao, and Li Zhang, "Object Tracking With Multi-View Support Vector Machines", IEEE Transactions on Multimedia, Vol. 17, No. 3, March 2015