Object Recognition using Lucas-Kanade Technique and Support Vector Machine Based Classification in Video Surveillance Systems

Raviprakash M L, C S Pillai, Ananda babu J

Abstract— Object recognition in video surveillance systems is the primary and most significant challenge task in the field of image processing. Video Surveillance systems provides us continuous monitoring of the objects for the enhancement of security and control. This paper presents novel approach recognizing the objects using Shi-Tomasi approach for detecting the corners of the object and then applies the Lucas-Kanade techniques to extract the features of the objects. The main objective of this paper is providing precise recognition of objects and estimation of their location from an unknown scene. Whenever the object is recognized from extracted frames of the input video the background subtraction will be applied. Then the classification of the objects into their respective categories can be achieved using support vector machine classifier by supervised learning. In case of multiple objects of different classes in a single frame, a vector containing the classes of all the detected in that frame is produced as output. The results of this work are drawn in the MATLAB tool by considering the input video dataset taken from various sources and extracting the frames from the input video for the detection then the efficiency of the proposed techniques will be measured.

Keyword: Object recognition , video surveillance systems, Shi-Tomasi approach, Lucas-Kanade techniques, MATLAB

I. INTRODUCTION

Object recognition is one the critical challenge that majority of the computer vision research is focused on. Mainly, object detection and tracking is very popular with growing demands in surveillance, defense front, traffic control and medical imaging. It is also a result of the powerful algorithms in sub-fields of machine learning, computer vision as well as due to innovations in hardware that support highly data intensive computations in magnitude of couple of minutes. The ultimate purpose underlying vision based object detection is to comprehend the kind of objects in the image, the characteristics of those, their location in space and additionally involve motion or trajectory of the object while moving. Of the many objects commonly considered for detection, human as an object has got a lot of attention and produced many applications.

The more general application of multiple object detection that may include targets like chair, car, and tree i.e. non-living objects in addition to humans or animals is more complex one.

This task has the challenges of accurately detecting the object categories as a result of inherent difficulties of multiple objects, varying size in the scene, overlapping of objects and deformities in the target object that vary across images.

To overcome this problem when attempting a large detection mission of objects spanning a number of classes, the cue that helps is contextual recognition.

Public benchmark datasets for object detection and classification available Online, involve millions of images with objects belonging to thousands of different categories. Some of those datasets are ImageNet database, Microsoft COCO database and Pascal VOC database. The datasets have annotations made already to each object present. The object classification datasets may have one or more of the three different kinds of images- iconic target dataset, which contain target object visual occupying the central part without other background (context); non-iconic dataset, which contain object instances along with contextual elements and humans; last, iconic scene dataset that involve various objects present in the frame either belonging to the target class or not but absence of humans. Though iconic image types seem to be easier to detect and also track in case the video contains only iconic images of one specific target, it is the non-iconic image data are good candidates for generalization containing contextual information. Having context data in images makes the object recognition algorithm to learn well and output good classification rate when it is tested against a new image dataset containing many generic naturally captured clicks containing various objects.

The problem of detection and tracking of objects from video data has been implemented using many approaches. The main challenges involved are accurate detection of object in occlusions, ambiguous background object movement and real time tracking without comprising quality of output produced. There is much scope in detection of generic objects from videos. In order to detect an object accurately, the object features must be extracted and matched (feature extraction) or the object structure must be learnt automatically through training (machine learning).

Many works have exploited different feature set to achieve good detection performance. Few of the researches involve tracking of objects by means of detection, wherein the mobile object is continuously detected in frames and labeled. One such algorithm makes use of color histogram probability and centroid features for object detection [1].

Object Recognition using Lucas-Kanade Technique and Support Vector Machine Based Classification in Video Surveillance Systems

Optical flow motion vectors [2] of moving objects like vehicle, pedestrian have been chosen to detect objects, both living and non-living while in movement. This technique works based on motion vector comparison between two successive frames corresponding to the same object to find the object movement assuming intensity of object remains fixed.

Texture features like Local Binary Pattern (LBP) [6] of the image have been considered for feature extraction for detecting objects.

LBP features are computed by means of pixel neighborhood operations. A popular feature descriptor for object detection involves Histogram of Oriented Gradients. HOG features are shape descriptors that represent an object in terms of intensity gradients in specific directions. In [7], the authors have exploited HOG features citing their properties of invariance with respect to transformations like rotation, deformities as well as lighting conditions. In another work [8], Gabor wavelet has been utilized. It produces magnitude responses of high frequency parts of a binary image. To detect multiple objects of same class but varying in color, Gabor wavelet is estimated on binarized images.

In this work, we have used the corner features that are most suitable to be detected in videos using Shi-Tomasi detector and Lucas-Kanade algorithm. For object class recognition, multi-class Support Vector Machine is utilized.

II. RELATED WORK

This section briefly describes the existing systems for object detection and tracking in videos and the techniques used by their authors for the same.

Jansari et.al [1] proposed a system for detection based mobile object tracking using color probability features. The methodology consists of two different detection techniques namely background subtraction and optical flow based detection executed against video inputs containing multiple mobile objects with occlusions taken into account. Tracking based on centroid and average RGB pixel measure of each detected object is used. Comparison of the features in consecutive frames to compute the differences is employed. Thresholding is used to predict whether the track belongs to a new object or current.

Evian et.al [2] in their paper use automatic K-means clustering technique to cluster the feature points of moving objects in order to detect and track them. The work includes two feature extraction methods- first Shi-Tomasi method detects the feature points of objects whose movement can easily be tracked; second the Lucas-Kanade method extracting the same features but in the successive frame exploiting its presumptions of minimal temporal motion, brightness and spatial similarity. Along with these object features, the direction and rate of movement are also included. The elimination of feature points related to the scene is done to avoid tracking undesirable objects. The tracking follows consecutively detecting the above features with automatic k-means clustering.

Fu et.al [3] presented an object recognition system for video data based on a combined SVM and binary tree classifier. The object detection is achieved using Gaussian background modeling and frame differencing for the video frames. The authors have exploited shape features which involve a ratio of object's upper and lower half's duty cycle, aspect ratio, eccentricity and compactness measures of the mobile object for recognition. The recognition is implemented by a binary decision tree, in which SVM classifier is used at each parent node. The system is tested against objects of two categories namely humans and vehicles in motion.

Lychkov et.al [4] proposed a novel object tracking system that addresses the issue of losing track of object's feature points. The authors use the concept of regeneration in living beings to detection of objects in video. The sparse optical flow features of the object in motion are extracted in order to detect the movement. The tracking is based on matching object location at present instant with the trajectory being generated as per observed relative displacement. The epimorphosis based regeneration technique of object's feature points is exploited with elimination of feature points less than user defined threshold and new feature point replacement. The work includes the affine transform to adapt the tracking process with scaling as well as rotation.

Swalaganata et.al [5] has used a hybrid method for tracking of moving objects from video inputs, which is combination of Camshift algorithm and Kalman filter. The tracking is based on detection per frame of the video. The bounded box is used to mark the target object to be tracked as part of detection process. The first tracking technique is Camshift based on color histogram probability and centroid establishment with Kalman filter as an enhancement to it. Since the occlusions in the frames are not handled by Camshift method, the authors thus rely on improving tracking by prediction using Kalman filter.

Runze et.al [6] constructs a new framework for tracking of mobile objects called Tracking Learning Detection framework. Tracking module is used for frame-wise tracking of object movement using optical flow cues and the detection module works by using ensemble and nearest neighbour classifier for recognition based on texture features. The cross correlation concept in the detection module for matching the object model with target in the video is also employed. Track updating with consecutive detections of the object in consecutive frames is done by the learning module.

Jaganatthan et.al [7] proposed a framework for real-world 3D objects detection based on Histogram of Oriented Gradients (HoG) features and subsequently mapping them into respective object classes (classification) for use in energy efficient embedded devices. The dataset included three object categories of 'pedestrian', 'vehicle' and 'traffic signal'. HoG features relating to each object are extracted from feature planes of the image constructed. AdaBoost cascaded tree classifier along with meanshift algorithm is used for classification. In addition, second classifier i.e. 7 layer CNN is also implemented. The classification accuracy is tested and compared in both classifiers used. The CNN classifier outperformed the AdaBoost classifier.

III. METHODOLOGY

The objective of our work is to implement an object detection system which is well trained to detect multiple generic objects in video frames and can generalize the detection in presence of occlusions in real time based on good feature descriptors and classifier.

System Architecture

The proposed technique adopts the feature extraction technique used in [2], which consists of two methods namely Shi-Tomasi algorithm and Lucas-Kanade method. The classification is achieved using multi-class Support Vector Machine (SVM). Fig. 1 describes the system architecture of system architecture of our generic object detection and classification system.

The Shi-Tomasi algorithm detects objects based on extracting corner features [9]. Lucas-Kanade method [10] is used as it is based on object detection using optical flow features, suitable for detecting static objects as well as tracking moving objects in videos.

Following is description of the proposed methodology:

3.1.1 Object Detection and Classification System

Step 1. Read the input video containing multiple objects including moving and static for detection.

Step 2. Pre-processing: Pre-process the input video by performing the two steps below:

a. Extract and store individually every frame of the video from the beginning.

b. Apply morphological filtering operation to remove noise and refine the frame.

Step 3. Feature Extraction: Extraction of distinct object features to compute the feature vector for detection in successive frames of the video.

a. Extract the object corner features using Shi-Tomasi corner detection technique in frame i at time instant (t-1).

b. Extract the corresponding corner features again using Lucas-Kanade algorithm in next frame i+1 at time instant t, that is based on optical flow vectors.

Step 4. Classification: To classify the detected objects into their respective categories with the help of a classifier. Here, multi-class SVM classifier is used to output object class by supervised learning. This involves two steps-

a. Training the SVM with feature vectors and target classes of objects detected from training video dataset.

b. Testing the SVM classification performance over videos from testing video dataset.

The output of the classification is the class that a detected object belongs to. In case of multiple objects of different classes in a single frame, a vector containing the classes of all the detected in that frame is produced as output.
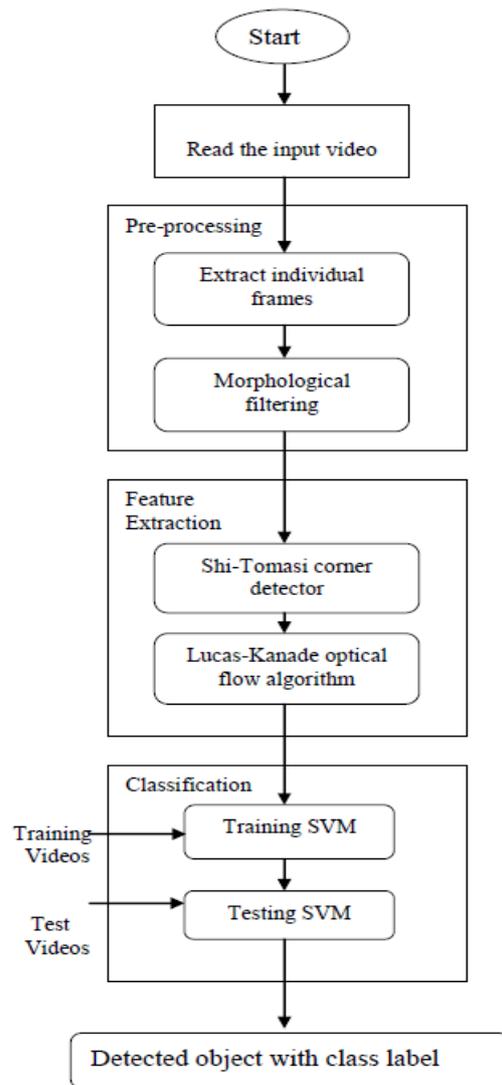

Fig. 1: System Architecture of Object Detection and Classification

3.1 Feature Extraction for Recognition of objects

This section describes in detail the theoretical as well as mathematical background related to the techniques used for extracting the features and object recognition based on those features in our system.

3.2.1 Shi-Tomasi Corner Detector

It is an improvised version of Harris corner detector, which extracts corner features from images that are the regions of high texture variations with small changes in position. A brief description of Harris corner detection technique follows-

The detection is done by scanning the whole image through equal sized small windows. If I(x,y) is the intensity of a pixel located at position (x,y) in an image window and when the window is moved slightly by a displacement (u,v) then, the intensity change is denoted as I(x+u, y+v). Since the aim of this corner detector is to find out regions or windows in the image whose intensity changes greatly with small displacements, it is mathematically expressed as-

$$E(u,v) = \sum_{x,y} w(x,y)[I(x+u, y+v) - I_{x,y}^2 \dots\dots (1)$$

Where 'w' is a weight function. The result E(u,v) produces the high intensity variation windows in the image.

After using Taylor's series and simplification of Equation 1, the following is the result

$$M = \sum w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \dots (2)$$

A score based on Eigen values of matrix M is used to find out suitable corners.

$$S = |M| - k(trace(M))^2 \dots (3)$$

If λ_1, λ_2 **are Eigen values of M**, then $|M| = \lambda_1 \lambda_2$ and **trace (M)** $= \lambda_1 + \lambda_2$

A high score S represents one of the corners of the object in the image.

A small change in score calculation is accommodated in Shi-Tomasi Corner Detector, which leads to finding out only the most suitable 'n' corners of the object rather than extracting every corner feature. Shi-Tomasi score for corner detection follows-

$$S = min(\lambda_1, \lambda_2) \dots (4)$$

If the score, S exceeds a threshold it is considered as a corner [9].

3.2.2 Lucas-Kanade Optical Flow

The estimation of optical flow in an image is carried out to track the motion of objects in an image when either the objects themselves are motion or if the camera is in motion. The optical flow is apparent movement of any entity from one frame to another in a video. Lucas-Kanade algorithm[10] works on the basis of assumptions of optical flow theory- Throughout the video frames, the intensity of any pixel belonging to same object remains unchanged, the pixels in a neighborhood have similar motion and the displacements of pixels between successive frames is very small.

A pixel with intensity I(x,y,t) in a frame at time t after moving with a small displacement (d_x, d_y) in consecutive frame with a difference of time d_t is expressed as follows-

$$I(x,y,t) = I(x + d_x, y + d_y, t + d_t) ..(5)$$

The optical flow equation for movement of objects in an image is given as-

$$\frac{\partial_f}{\partial_x}\frac{d_x}{d_t} + \frac{\partial_f}{\partial_y}\frac{d_y}{d_t} + f_t = 0 \dots (6)$$

Lucas-Kanade method outputs optical flow motion vectors by solving Equation (6):

$$\begin{bmatrix} u \\ v \end{bmatrix} = \Sigma_i \begin{bmatrix} f_{x_i}^2 & f_{x_i} f_{y_i} \\ f_{x_i} & f_{y_i}^2 \end{bmatrix}^{-1} \Sigma_i \begin{bmatrix} -f_{x_i} f_{t_i} \\ -f_{y_i} f_{t_i} \end{bmatrix} \dots (7)$$

Where (u, v) represent the displacement of the object between consecutive frames.

3.2.3 Classification of the Recognized Objects

The last step in our system is the classification of detected objects into respective categories or classes. Multi-class Support Vector Machine classifier is used to perform this classification.

Support Vector Machine is a machine learning, supervised learning algorithm. It relies on training itself with a number of sample inputs in the form of a n vector *(x1, x2, x3, x4,......xn)* along with the target class vector of same size *(c1, c2, c3, c4,....cn)* specified in order to learn the pattern involved in the data. In testing phase, the classifier attempts to generate the target class of the unobserved test input. The SVM constructs a hyper-plane with all inputs that are basically the feature vectors against their classes specified so that the best fit including all the input-output points is produced. The output is generated by locating the position of test feature vector onto the hyper-plane giving the target class prediction.

IV. RESULTS AND DISCUSSION

The system is implemented in Matlab Tool, which facilitates application of complex mathematical computations through easy built-in functions for performing object detection and classification using its Computer Vision and Machine Learning toolboxes.

The video processing is smoothly done by operating on individual frames after their extraction. The Image Processing toolbox containing a lot of filtering and refining functions needed for pre-processing of an image/frame. A total of 10 videos are collected containing generic objects of different classes. The dataset is divided into two categories-training dataset of 10 videos and testing dataset of 8 videos required for support vector machine classification.

The below figure 2 describes histogram of the frames extracted from the video 1 where different objects are recognized after the extraction of features, in the same manner different objects detected from video 2 and video 3 are shown in figure 3 and figure 4 respectively.
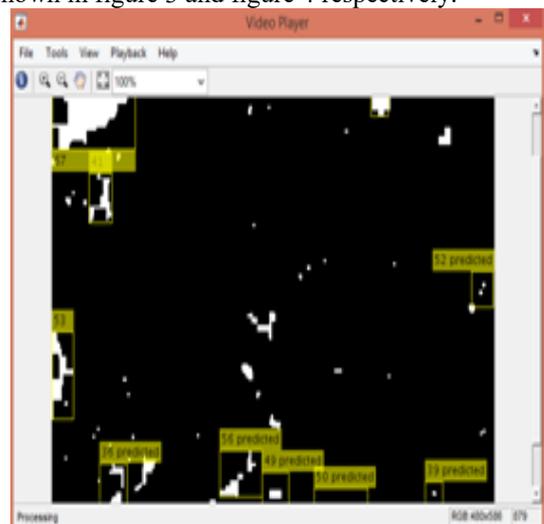

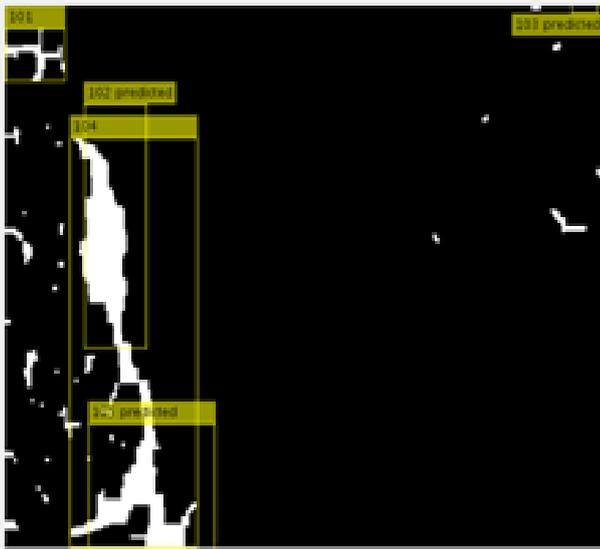**Figure 2: Extracting the frames of the detected objects in the video 1**

Figure 3: Extracting the frames of the detected objects in the video 2

Figure 4: Extracting the frames of the detected objects in the video 3

The dataset contains videos where the objects to be detected can make up background or foreground of the scene like vehicles. Occlusions are also present in the input videos, involving both the scene-object and inter-object occlusions. The objects are detected based on easily tractable corner features, basically high intensity regions readily detected with small movement in the scene. The movement can be due to either mobile object or moving camera. The multi-class support vector machine is given as input the feature vectors of all objects with the respective classes in training. The classifier output obtained when the feature vector of an object in test video is given is the target class that is predicted by support vector machine.

The accuracy of the proposed technique is analyzed with total number of percentage of frames detected with object vs. the total number of frames. Here graph is generated by taking four input video from various sources with different sizes and different background, for each video we could able to get an accuracy of 89%. Where frames for each video vary from 100 to 1000 and accuracy for each video varies from 80% to 90% in recognizing the objects.

IV. CONCLUSION AND FUTURE SCOPE

This paper presents a novel approach for Object Recognition using Lucas-Kanade technique and Support Vector Machine based Classification in Video Surveillance Systems. This paper is providing precise recognition of objects and estimation of their location from an unknown place. First the recognition of object is archived using Shi-Tomasi and Lucas-Kanade techniques, whenever the object is recognized from extracted frames of the input video the background subtraction will be applied. Then the classification of the objects into their respective categories can be achieved using support vector machine classifier by supervised learning. Finally the accuracy of detected frames of the object for the input video is measured in this work we achieved an accuracy of 89% by considering different input videos. In the future the algorithm can be applied over tracking of the objects on live surveillance systems using the proposed approach and research work will be extended to develop improve the accuracy of the proposed work, so that we can be able develop intelligent surveillance systems.

REFERENCES

1. D. Jansari, S. Parmar and G. Saha, "Real-time object tracking using colour-based probability matching," 2013 IEEE International Conference on Signal Processing, Computing and Control (ISPCC), Solan, 2013, pp. 1-6.
2. A. Keivani, J. Tapamo and F. Ghayoor, "Motion-based moving object detection and tracking using automatic K-means," 2017 IEEE AFRICON, Cape Town, 2017, pp. 32-37.
3. Y. Fu, D. Ma, H. Zhang and L. Zheng, "Moving object recognition based on SVM and binary decision tree," 2017 6th Data Driven Control and Learning Systems (DDCLS), Chongqing, 2017, pp. 495-500.
4. I. I. Lychkov, A. N. Alfimtsev and S. A. Sakulin, "Tracking of Moving Objects With Regeneration of Object Feature Points," 2018 Global Smart Industry Conference (GloSIC), Chelyabinsk, 2018, pp. 1-6.
5. G. Swalaganata, Muniri and Y. Affriyenni, "Moving object tracking using hybrid method," 2018 International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, 2018, pp. 607-611.
6. W. Runze, W. Yuxing and Z. Jianlin, "Improved object tracking algorithm based on tracking-leaning-detection framework," 2017 7th IEEE International Conference on Electronics Information and Emergency Communication (ICEIEC), Macau, 2017, pp. 74-77.
7. S. Jagannathan et al., "Efficient object detection and classification on low power embedded systems," 2017 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, 2017, pp. 233-234.
8. S. Susan and S. Chandna, "Object Recognition from Color Images by Fuzzy Classification of Gabor Wavelet Features," 2013 5th International Conference and Computational Intelligence and Communication Networks, Mathura, 2013, pp. 301-305.
9. Jianbo Shi and Tomasi, "Good features to track," 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 1994, pp. 593-600.
10. B. D. Lucas, T. Kanade, "An iterative image registration technique with an application to stereo vision," in IJCAI, 1981.
11. C. Tomasi and T. Kanade, "Detection and tracking of point features," Technical Report CMU-CS-91132, Carnegie-Mellon University, 1991.

AUTHORS PROFILE

Mr. Raviprakash M L: is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Computer Science and Engineering, Kalpataru Institute of Technology, Tiptur. His main research work focuses on Image Processing. He has 8 years of teaching experience and 6 years of Research Experience and he is a member of MISTE and CSI

Dr. C S pillai B.E,MTech,,P.hD: is currently working as Associate Professor in Department of Computer Science and Engineering, ACS College of Engineering, Banglore, India. His area of interest includes Image Processing and Query processing. He has 14 years of teaching experience and 10 years of Research Experience and he is a member of MISTE.

Dr. Ananda babu J B.E,MTech,,P.hD is currently working as Associate Professor in Department of Computer Science and Engineering, Malnad College of Engineering, Hassan, India. His area of interest includes Wireless sensor networks, Image Processing. He has 12 years of teaching experience and 08 years of Research Experience. and he is a member of MISTE and CSI