Improving the Performance of Video Tracking Using SVM

Ms. G.Anusha[#1], Mrs. E. Golden Julie[*2]

[#]Student,[*]Assistant Professor
Department of Computer Science and Engineering
Regional Centre of Anna University
Tirunelveli (T.N) India

***Abstract*— Object tracking is the process of separating the moving object from the video sequences. Tracking is essentially a matching problem in object tracking. In order to avoid this matching problem, object recognition is done on the tracked object. . Kalman filter cannot separate the moving object in multiple object tracking so we use background separation techniques. . Background separation algorithm separate moving object from the background based on white and black pixels .Support Vector Machines classifier is used to recognize the tracked object. SVM classifier are supervised learning that associates with machine learning algorithm that analyse and recognize the data used for classification.**

***Keywords*—Object tracking, Support Vector Machines, Background separation, Gabor filter, Object recognition.**

Introduction

Tracking is closely related to constructing correspondences between frames. Traditional tracking approaches focus on finding low-level correspondences based on image evidence. Online models for low-level correspondences are generally employed to adapt to the changing appearances of the target. However, one notable shortcoming of these online models is that they are constructed and updated based on the previous appearance of the target without much semantic understanding. Image processing is any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image. Image processing usually refers to digital image processing, but optical and analog image processing also are possible. This article is about general techniques that apply to all of them. The acquisition of images or producing the input image in the first place is referred to as imaging. Image processing refers to processing of a 2D picture by a computer. Difficulties in tracking objects can arise due to abrupt object motion, changing appearance patterns of both the object and the scene, non-rigid object structures, object-to-object and object-to-scene occlusions, and camera motion.

Tracking is usually performed in the context of higher-level applications that require the location and/or shape of the object in every frame. There are three in video analysis: Deteection of moving objects of interest tracking of such objest from frame to frame to frmae. Analysis of object tracked to recognize their behavior. The use of object tracking is pertinent int the task of motion-based recognition, automated surveillance,video indexing,human computer interaction etc. First progress presentation on video object tracking with classification and recognition of object. Video object tracking combines two phases of analysis: Recognition and classification of moving objects and tracking of moving objects. Object recognition - in computer vision is the task of finding and identifying objects in an image or video sequence. Humans recognize a multitude of objects in images with little effort, despite the fact that the image of the objects may vary somewhat in different view points, in many different sizes / scale or even when they are translated or rotated. Objects can even be recognized when they are partially obstructed from view. This task is still a challenge for computer vision systems. Object detection in videos involves verifying the presence of an object in image sequences and possibly locating it precisely for recognition. Object tracking is to monitor an object spatial and temporal changes during a video sequence, including its presence, position, size, shape, etc. This is done by solving the temporal correspondence problem, the problem of matching the target region in successive frames of a sequence of images taken at closely-spaced time intervals. These two processes are closely related because tracking usually starts with detecting objects, while detecting an object repeatedly in subsequent image sequence is often necessary to help and verify tracking.

A.Video Tracking

Video tracking is the process of locating a moving object (or multiple objects) over time using a camera. It has a variety of uses, some of which are: human-computer interaction, security and surveillance, video communication and compression, augmented reality, traffic control, medical imaging and video editing. Video tracking can be a time consuming process due to the amount of data that is contained in video. Adding further to the complexity is the possible need to use object recognition techniques for tracking. The objective of video tracking is to associate target objects in consecutive video frames. The association can be especially difficult when the objects are moving fast relative to the frame rate. Another situation that increases the complexity of the problem is when the tracked object changes orientation over time. For these situations video tracking systems usually employ a motion model which describes how the image of the target might change for different possible motions of the object. Examples of simple motion models are: When tracking planar objects, the motion model is a 2D transformation (affine transformation or homography) of an image of the object (e.g. the initial frame).When the target is a rigid 3D object, the motion model defines its aspect depending on its 3D position and orientation. For video compression, key frames are divided into macroblocks. The motion model is a disruption of a key frame, where each macroblock is translated by a motion vector given by the motion parameters. Filtering and data association is mostly a top-down process, which involves incorporating prior information about the scene or object, dealing with object dynamics, and evaluation of different hypotheses. These methods allow the tracking of complex objects along with more complex object interaction like tracking objects moving behind obstructions. Additionally the complexity is increased if the video tracker (also named TV tracker or target tracker) is not mounted on rigid foundation (on-shore) but on a moving ship (off-shore), where typically an inertial measurement system is used to pre-stabilize the video tracker to reduce the required dynamics and bandwidth of the camera system.

I. Related Work

A. *Object Tracking : A Survey*

Object tracking is an important task within the field of computer vision. There are three keys steps in video analysis: detection of interesting moving objects, tracking of such objects from frame to frame and analysis of objects tracks to recognize their behaviour.

Object representation:

Points. The object is represented by a point, that is, the centroid or by a set of points. In general, the point representation is suitable for tracking objects that occupy small regions in an image.

Primitive geometric shapes. Object shape is represented by a rectangle, ellipse etc. Object motion for such representations is usually modeled by translation, affine, or projective. This representing is suitable for tracking non rigid object.

Object contour. Contour representation defines the boundary of an object.

Articulated shape models. Articulated objects are composed of body parts that are held together with joints. For example, the human body is an articulated object with legs, hands, head, and feet connected by joints.

Skeletal models. Object skeleton can be extracted by applying medial axis transform to the object contour. This model is commonly used as a shape representation for recognizing objects.

Color. The apparent color of an object is influenced primarily by two physical factors,1) the spectral power distribution of the illuminant and 2) the surface reflectance properties of the object. In image processing, the RGB (red, green, blue) color space is usually used to represent color.

Edges. Object boundaries usually generate strong changes in image intensities. Edge detection is used to identify these changes.

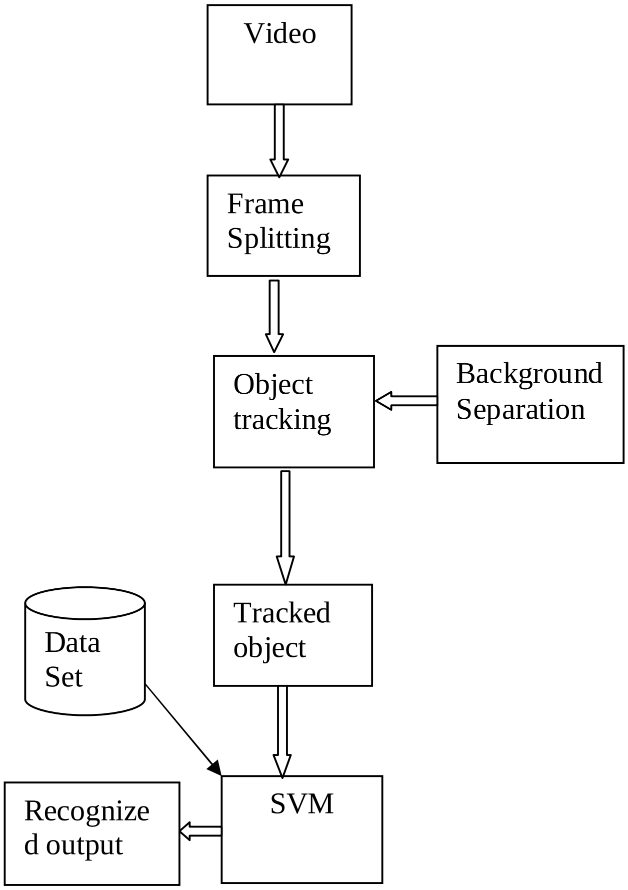
Optical Flow. Optical flow is a dense field of displacement vectors which defines the translation of each pixel in a region.

Texture. Texture is a measure of the intensity variation of a surface which quantifies properties such as smoothness and regularity.

Video tracking and recognition uses the Appearance-Adaptive Models. It incorporates appearance-based models in a particle filter to realize robust visual tracking and recognition algorithms. In conventional tracking algorithms, the appearance model is either fixed or rapidly changing.The motion model is simply a random walk with fixed noise variance. More complex models, namely adaptive appearance model, adaptive-velocity transition model, and intra- and extra-personal space models, are introduced to handle appearance changes between frames and between frames and gallery images. Many algorithms and technologies have been developed to automatically monitor pedestrians, traffic or other moving objects. One main difficulty in object tracking, among many others, is to choose suitable features and models for recognizing and tracking the target. One of the suitable feature is SIFT. SIFT is an approach for detecting and extracting local feature descriptors that are reasonably invariant to changes in illumination, scaling, rotation, image noise and small changes in viewpoint.

II. Proposed System

The object of interest is initialized by a user-specified bounding box, but its category is not provided. This target may or may not have a semantic meaning. Therefore, in the first few frames when the tracker does not know the target category, tracking the target only relies on the online target model, which is the same as traditional tracking. Meanwhile, video-based object recognition is applied on the tracked objects. When the target is recognized properly, the offline target model will be automatically incorporated to provide more information about the target. At time t, we denote the target state by $x_{t,}$ and the target category by c_t. Denote the image sequence by I_t=$\varphi\{I_{1,......,}I_t\}$, where I_t is the input image at time t. So the target measurement at time t is Z_t=$I_t(x_t)$.



System Architecture of Video Tracking

A.Frame Splitting

Frame splitting is the process of splitting the given video into number of frames. In one second 24 frames are generated. Frames are stored in a separate file. Alternate Frame Rendering (AFR) is a technique of graphics rendering in personal computers which combines the work output of two or more graphics processing units (GPU) for a single monitor, in order to improve image quality, or to accelerate the rendering performance. The technique is that one graphics processing unit computes all the odd video frames, the other renders the even frames. This technique is useful for generating 3D video sequences in real time, improving or filtering textured polygons and performing other computationally intensive tasks, typically associated with computer gaming, CAD and 3D modeling.

B. Object Tracking

Extracting the background image from sequences of frames is a very important task in order to help tracker detect motion.This task is repeated from time to time in order to incorporate any changes in the illumination of the tracking scene. There are several methods used to extract the background image from a sequence of frames but three are the most popular. These are based on statistical characteristics on the pixels of the frames: mean, median and highest appearance frequency methods. In all methods, every pixel of the background image is separately calculated using the mean or the median or the highest appearance frequency value from the series of frames.

Background Separation

Background subtraction, also known as Foreground Detection, is a technique in the fields of image processing and computer vision wherein an image's foreground is extracted for further processing (object recognition etc.). Generally an image's regions of interest are objects (humans, cars, text etc.) in its foreground. After the stage of image preprocessing (which may include image denoising etc.) object localisation is required subtraction is mostly done if the image in question is a part of a video stream. which may make use of this technique. Background subtraction is a widely used approach for detecting moving objects in videos from static cameras. The rationale in the approach is that of detecting the moving objects from the difference between the current frame and a reference frame, often called “background image”, or “background model”.

Using Frame Differencing

Frame difference (absolute) at time $t+1$ is

$$D(t+1)=|V(x,y,t+1)-V(x,y,t)|$$

The background is assumed to be the frame at time *t*. This difference image would only show some intensity for the pixel locations which have changed in the two frames. Though we have seemingly removed the background, this approach will only work for cases where all foreground pixels are moving and all background pixels are static. A threshold "Th" is put on this difference image to improve the subtraction.

$$|V(x,y,t+1)-V(x,y,t)|>Th$$

The accuracy of this approach is dependent on speed of movement in the scene. Faster movements may require higher thresholds.

Mean Filter

For calculating the image containing only the background, a series of preceding images are averaged. For calculating the background image at the instant *t*,

$$B(x, y) = 1/N \sum_{i=1}^{N} V(x, y, t-1)$$

where *N* is the number of preceding images taken for averaging. This averaging refers to averaging corresponding pixels in the given images. *N* would depend on the video speed (number of images per second in the video) and the amount of movement in the video. After calculating the background *B(x,y)* we can then subtract it from the image *V(x,y,t)* at time *t*=t and threshold it. Thus the foreground is

$$|V(x, y, t) - B(x, y) > Th$$

where *Th* is threshold. Similarly we can also use median instead of mean in the above calculation of *B*(*x*,*y*).

When the video is given as input to the system, it is first splitted into frames. The frames are stored in a folder for further usage. These frames are then taken consecutively and the frames are subtracted. i.e., frame 1 is subtracted from frame 2 and frame 2 is subtracted from frame 3 and so on. After subtraction, the difference is then compared with the threshold. If the difference is greater than the threshold, then the moving object is present. Else, there is no moving object. The moving object is represented in white, whereas the others in black.

Trial and Error Method

Trial and error is a fundamental method of solving problems. It is characterised by repeated, varied attempts which are continued until success, or until the agent stops trying. It is an unsystematic method which does not employ insight, theory or organised methodology. Trial and error is also a heuristic method of problem solving, repair, tuning, or obtaining knowledge. In the field of computer science, the method is called generate and test. In elementary algebra, when solving equations, it is "guess and check".

Gabor Filter

In image processing, a Gabor filter, named after Dennis Gabor, is a linear filter used for edge detection. Frequency and orientation representations of Gabor filters are similar to those of the human visual system, and they have been found to be particularly appropriate for texture representation and discrimination. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave.

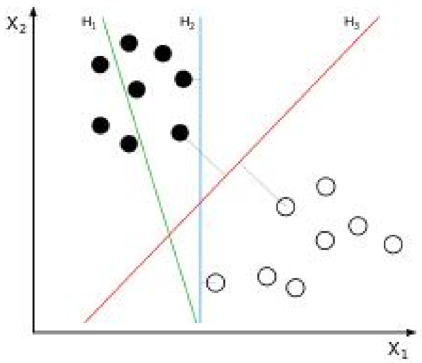
A set of Gabor filters with different frequencies and orientations may be helpful for extracting useful features from an image. Gabor filters have been widely used in pattern analysis applications. Gabor filters are directly related to Gabor wavelets, since they can be designed for a number of dilations and rotations. However, in general, expansion is not applied for Gabor wavelets, since this requires computation of bi-orthogonal wavelets, which may be very time-consuming. Therefore, usually, a filter bank consisting of Gabor filters with various scales and rotations is created. The filters are convolved with the signal, resulting in a so-called Gabor space. This process is closely related to processes in the primary visual cortex. Jones and Palmer showed that the real part of the complex Gabor function is a good fit to the receptive field weight functions found in simple cells in a cat's striate cortex.

C. *Object Recognition*

Object Recognition is the task of finding a given object in an image or video sequence. SVM classifier is used to recognize the tracked object. SVM is a method for the classification of both linear and nonlinear data. Training and testing are the two phases in object recognition. In training phase, object are already trained and saved in a database. Tracked object is compare with training object, if the object is present then it display object is recognized otherwise object cannot be recognized**.**

Support Vector Machines Classifier

SVM is a method for the classification of both linear and nonlinear data. It uses a nonlinear mapping to transform the original training data into a higher dimension. Within this new dimension, it searches for the linear optimal separating hyperplane that is a “decision boundary” separating the tuples of one class from another. With an appropriate nonlinear mapping to a sufficiently high dimension, data from two classes can always be separated by a hyperplane. The SVM finds this hyperplane using support vectors (“essential” training tuples) and margins (defined by the support vectors). SVMs can be used for prediction as well as classification. They have been applied to a number of areas, including handwritten digit recognition, object recognition, and speaker identification, as well as benchmark time-series prediction tests.



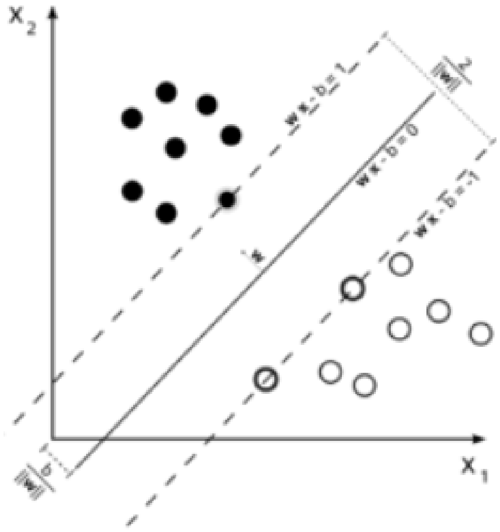
SVM Separation

H_1 does not separate the classes. H_2 does, but only with a small margin. H_3 separates them with the maximum margin.

A support vector machine constructs a hyperplane or set of hyperplanes in a high or infinite dimensional space, which can be used for classification, regression, or other tasks. A good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. The vectors defining the hyperplanes can be chosen to be linear combinations with parameters α_i of images of feature vectors that occur in the data base. With this choice of a hyperplane, the points x in the feature space that are mapped into the hyperplane are defined by the equation as:

$$\sum_i \propto_i K(x_{i,},x) = Constant$$

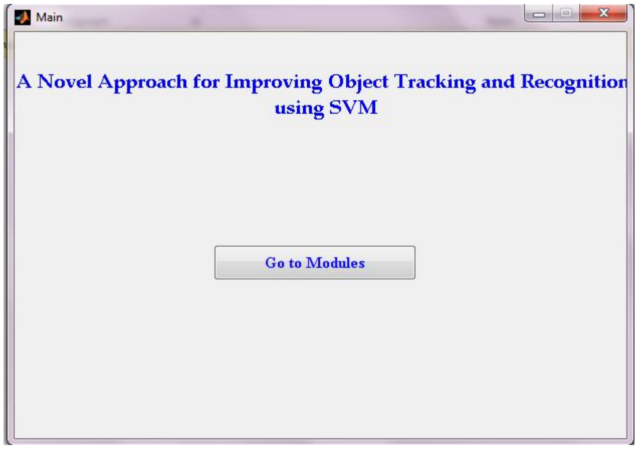
Note that if $K(x,y)$ becomes small as y grows further away from x, each element in the sum measures the degree of closeness of the test point x to the corresponding data base point x_i.



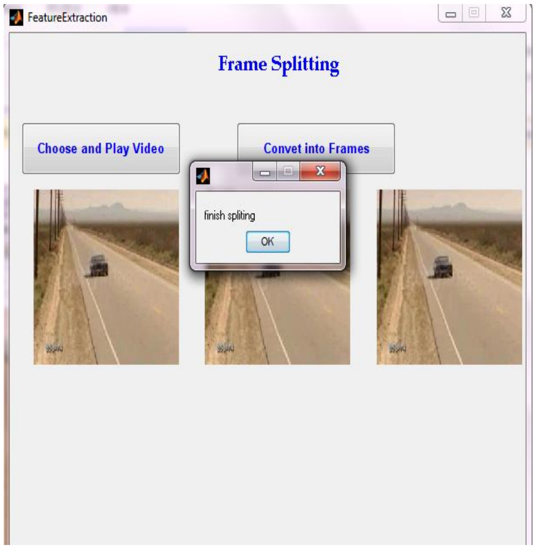
SVM Classification

III. Simulation and Results

Frame splitting is the process of splitting video into numbers of frames. In frame splitting process choose an video from the given dataset and convert into frames. Input video should be in avi(audio video interleave) format. Length of the video is based on the time taken to run on video. Size of the frame is based on the number of frames generated. In one second 24 frames are generated. Frames are stored in a separate folder for furthur usage.

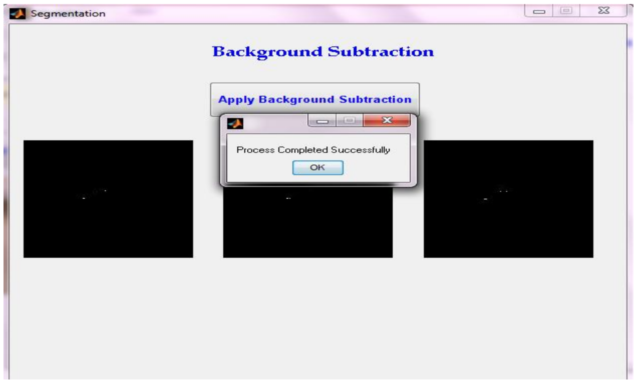


A Novel Approach for Improving Object Tracking and Recognition using SVM



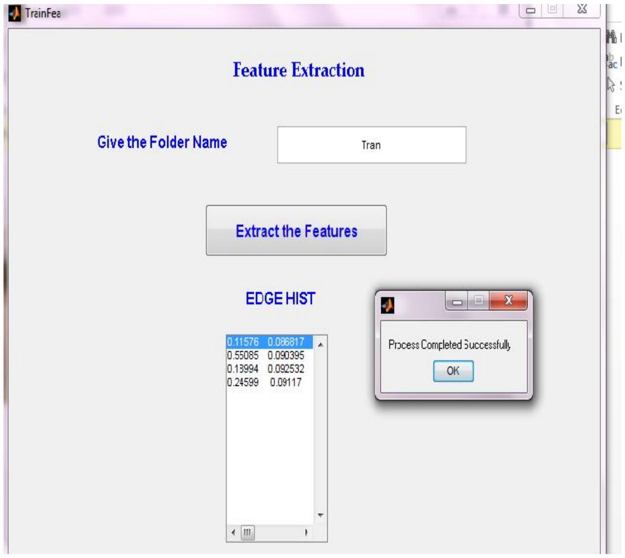
finish spliting

Frame Splitting



Process Completed Successfully

Object Tracking

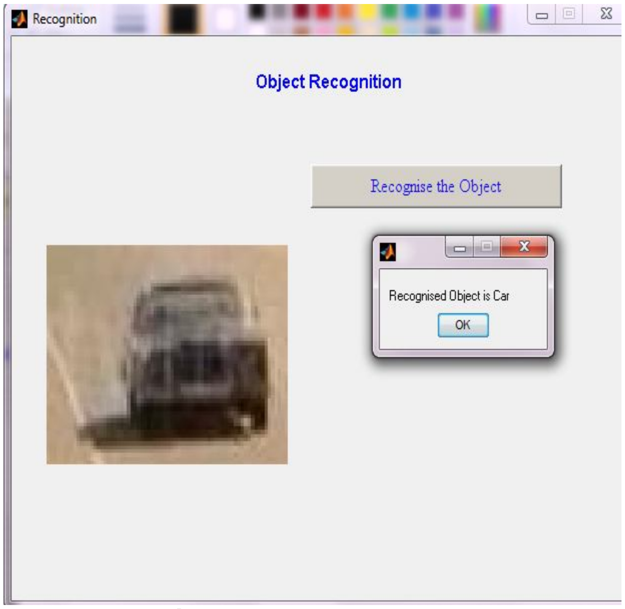


Tran

0.11576	0.086817
0.55085	0.090395
0.13994	0.092532
0.24599	0.09117

Process Completed Successfully

Training Object

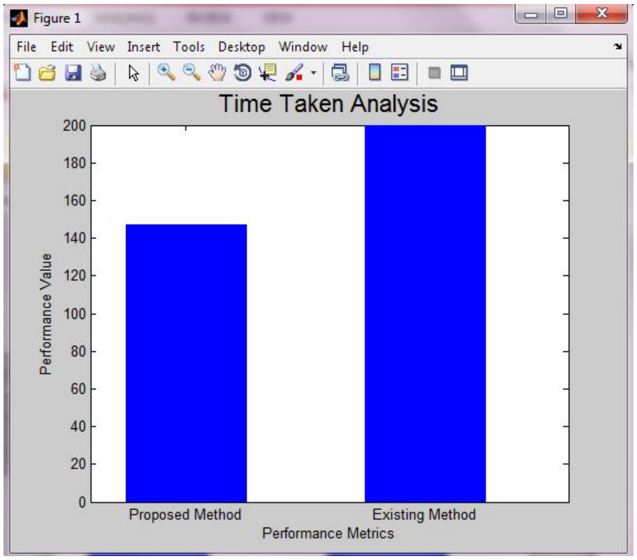


Recognised Object is Car

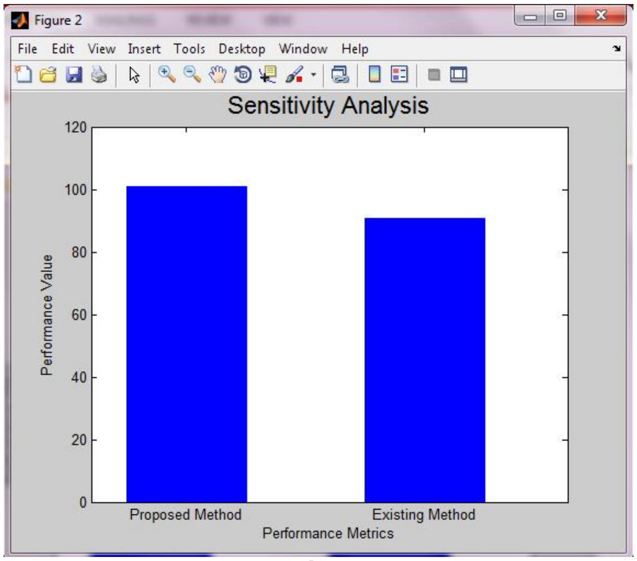
Object Recognition

The performance of the prediction was evaluated in terms of sensitivity, specificity, accuracy. Accuracy measures the quality of the binary classification (two-class). It takes into account true and false positives and negatives. Accuracy is generally regarded with balanced measure whereas sensitivity deals with only positive cases and specificity deals with only negative cases. TP is number of true positives, FP is number of false positives, TN is number of true negatives and FN is number of false negatives A confusion matrix provides information about actual and predicted cases produced by classification system .The performance of the system is examined by demonstrating correct and incorrect patterns.

Sensitivity TP/(TP+FN)
Specificity TN/(TN+FP)
Time Taken (TP+TN)/(TP+FP+TN+FN)



Time Taken Analysis



Sensitivity Analysis

IV. CONCLUSIONS

In this paper, the mid-level task, visual tracking plays an important role for high-level semantic understanding or video analysis. Meanwhile the high-level understanding (e.g., object recognition) should feed back some guidance for low-level tracking. Motivated by this, we propose a unified approach to object tracking and recognition. In our framework, once the objects are discovered and tracked, the tracking result is fed forward to the object recognition module. The recognition result is fed back to activate the off-line model to and help improve tracking. Extensive experiments demonstrate the efficiency of the proposed method.

References

[1] “What Are We Tracking: A Unified Approach of Tracking and Recognition” Jialue Fan, Xiaohui Shen, Student Member, IEEE, and Ying Wu, Senior Member, IEEE

[2] B. Babenko, M.-H. Yang, and S. Belongie, “Visual tracking with online multiple instance learning,” in Proc. Conf. Comput. Vis. Pattern Recognit., 2009, pp. 983–990.

[3] H. Grabner, M. Grabner, and H. Bischof, “Real-time tracking via on-line boosting,” in Proc. British Mach. Vis. Conf., 2006, pp. 1–10.

[4] Y. Li, H. Ai, T. Yamashita, S. Lao, and M. Kawade, “Tracking in low frame rate video: A cascade particle filter with discriminative observers of different life spans,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 30,no. 10, pp. 1728–1740, Oct. 2008.

[5] B. Wu and R. Nevatia, “Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors,” Int. J. Comput. Vis., vol. 75, no. 2, pp. 247–266, Nov. 2007.

[6] C. Huang, B. Wu, and R. Nevatia, “Robust object tracking by hierarchical association of detection responses,” in Proc. Eur. Conf. Comput. Vis., 2008, pp. 788–801.

[7] B. Leibe, K. Schindler, N. Cornelis, and L. V. Gool, “Coupled object detection and tracking from static cameras and moving vehicles,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 10, pp. 1683–1698, Oct. 2008.

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

[9] M. Isard and A. Blake, “Contour tracking by stochastic propagation of conditional density,” in Proc. Eur. Conf. Comput. Vis., 1996, pp. 343–356.

[10] Z. Kalal, J. Matas, and K. Mikolajczyk, “P-N learning: Bootstrapping binary classifiers by structural constraints,” in Proc. Conf. Comput. Vis. Pattern Recognit., 2010, pp. 49–56.

[11] A. Srivastava and E. Klassen, “Bayesian and geometric subspace tracking,” Adv. Appl. Probab., vol. 36, pp. 43–56, Dec. 2004.

[12] A. D. Jepson, D. Fleet, and T. El-Maraghi, “Learning Adaptive Metric for Robust visual tracking,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 10, pp. 1296–1311, Oct. 2003.

[13] N. Alt, S. Hinterstoisser, and N. Navab, “Rapid selection of reliable templates for visual tracking,” in Proc. Conf. Comput. Vis. Pattern Recognit., 2010, pp. 1355–1362.

[14] X. Mei, H. Ling, and Y. Wu,“Minimum error bounded efficient l1 tracker with occlusion detection,” in Proc. Conf. Comput. Vis. Pattern Recognit., 2011, pp. 1257–1264.

[15] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, “Visual tracking and recognition using probabilistic appearance manifolds,” Comput. Vis. Image Understand., vol. 99, no. 3, pp. 303–331, 2005.

[16] M. Andriluka, S. Roth, and B. Schiele, “People-tracking-by-detection and people-detection-by-tracking,” in Proc. Conf. Comput. Vis. Pattern Recognit., 2008, pp. 1–8.

[17] S. Zhou, R. Chellappa, and B. Moghaddam, “Visual tracking and recognition using appearance-adaptive models in particle filters,” IEEE Trans. Image Process., vol. 13, no. 11, pp. 1434–1456, Nov. 2004.