International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Object Detection And Tracking Using Instant Online Feature Extraction

Mr. Suraj R.Jaronde
Dept. of Electronics &Telecommunication Engineering
Yashwantrao Chavan College of Engineering, Nagpur, India

Abstract— It is not always possible to provide labeled data for training because it requires substantial human effort, expensive tests, disagreement among experts. Labeling is not possible at instance level.To overcome these problem multiple instance learning (MIL) method is introduced which actively trained the data in online manner and combine with discriminative classifier which separate the object from its background and provide positive and negative bags. The fisher information criteria is used to train dataset in online manner which perfectly describe the label of positive content in positive label bag and negative content in negative label bags. The use of actively trained classifier helps to improve the efficiency of tracking object in motion.
Keywords— Active learning, fisher information, multiple instance learning (MIL)

I. INTRODUCTION

Object detection and tracking is an important challenging task within the area in Computer Vision that try to detect, recognize and track objects over a sequence of images called video. It helps to understand, describe object behavior instead of monitoring computer by human operators. It aims to locating moving objects in a video file or surveillance camera. Object tracking is the process of locating an object or multiple objects using a single camera, multiple cameras or given video file. Invention of high quality of the imaging sensor, quality of the image and resolution of the image are improved, and the exponential computation power is required to be created of new good algorithm and its application using object tracking.
In Object Detection and Tracking we have to detect the target object and track that object in consecutive frames of a video file Object detection and tracking is a one of the challenging task in computer vision. Mainly there are three basic steps in video analysis: Detection of objects of interest from moving objects, Tracking of that interested objects in consecutive frames, and Analysis of object tracks to understand their behavior. Simple object detection compares a static background frame at the pixel level with the current frame of video. The existing method in this domain first tries to detect the interest object in video frames. One of the main difficulties in object tracking among many others is to choose suitable features and models for recognizing and tracking the interested object from a video. Some common choice to choose suitable feature to categories, visual objects are intensity, shape, color and feature points. In this thesis, we studied about multiple instant learning tracking based on the fisher criteria, optical flow tracking based on the intensity and motion. Preliminary results from experiments have shown that the adopted method is able to track targets with translation, rotation, partial occlusion and deformation. The related work about the previous research is explained in section II. Section III proposed architecture of research work. Existing work and conclusion is given in V and VI.

II. RELATED WORK

In [1], Kaihua Zhag, Lei zhang,Qinghua Hu propose an active feature selection approach Motivated by the active learning method that is able to select more informative features than the MIL tracker by using the Fisher information criterion to measure the uncertainty of the classification model.The Fisher information is a way of measuring the amount of information that an observable random variable X carries about an unknown parameter θ upon which the probability of X depends. Formally, it is the variance of the score, or the expected value of the observed information. In Bayesian statistics, the asymptotic distribution of the posterior mode depends on the Fisher information and not on the prior (according to the Bernstein–von Mises theorem, which was anticipated by Laplace for exponential families.[2] The role of the Fisher information in the asymptotic theory of maximum-likelihood estimation was emphasized by the statistician R. A. Fisher (following some initial results by F. Y. Edgeworth. The Fisher information is also used in the calculation of the Jeffreys prior, which is used in Bayesian statistics.The Fisher-information matrix is used to calculate the covariance matrices associated with maximum-likelihood estimates. It can

also be used in the formulation of test statistics, such as the Wald test.Statistical systems of a scientific nature (physical, biological, etc.) whose likelihood functions obey shift invariance have been shown to obey maximum Fisher information. The level of the maximum depends upon the nature of the system constraints

In [2], Li Sun, Guizhong Liu they provides visual object tracking which provide combination of local scale-invariant feature transform (SIFT) description and global incremental principal component analysis (PCA) representation in bad conditions. Our algorithm for local descriptors (termed PCA-SIFT) accepts the same input as the standard SIFT descriptor: the sub-pixel location, scale, and dominant orientations of the key point. We extract a 41×41 patch at the given scale, centered over the key point, and rotated to align its dominant orientation to a canonical direction.1 PCA-SIFT can be summarized in the following steps: (1) pre-compute an eigen space to express the gradient images of local patches; (2) given a patch, compute its local image gradient; (3) project the gradient image vector using the eigen space to derive a compact feature vector. This feature vector is significantly smaller than the standard SIFT feature vector, and can be used with the same matching algorithms. The Euclidean distance between two feature vectors is used to determine whether the two vectors correspond to the same key point in different images. Principal Component Analysis (PCA) [7] is a standard technique for dimensionality reduction and has been applied to a broad class of computer vision problems, including feature selection (e.g., [5]), object recognition (e.g., [15]) and face recognition (e.g., [17]). While PCA suffers from a number of shortcomings [8,10], such as its implicit assumption of Gaussian distributions and its restriction to orthogonal linear combinations, it remains popular due to its simplicity. The idea of applying PCA to image patches is not novel (e.g., [3]). Our contribution lies in rigorously demonstrating that PCA is well-suited to representing key point patches (once they have been transformed into a canonical scale, position and orientation), and that this representation significantly improves SIFT's matching performance. PCA-SIFT is detailed in the following subsections. This paper introduced an alternate representation for local image descriptors for the SIFT algorithm. Compared to the standard representation, PCA-SIFT is both more distinctive and more compact leading to significant improvements in matching accuracy (and speed) for both controlled and real-world conditions. We believe that, although PCA is ill- suited for representing the general class of image patches, it is very well-suited for capturing the variation in the gradient image of a key point that has been localized in scale, space and orientation. We are currently extending our representation to color images, and exploring ways to apply the ideas behind PCA-SIFT to other key point algorithms.

In [3], Wei Zhong, Huchuan Lu proposed collaborative appearance model and develop a sparse discriminative classifier (SDC) and sparse generative model (SGM) for object tracking. They develop a simple yet robust model that makes use of the generative model to account for appearance change and the discriminative classifier to effectively separate the foreground target from the background their approach If the input feature vector to the classifier is a real vector $\vec{x}$, then the output score is

$$y = f(\vec{w} \cdot \vec{x}) = f\left(\sum_j w_j x_j\right),$$

where $\vec{w}$ is a real vector of weights and *f* is a function that converts the dot product of the two vectors into the desired output. (In other words, $\vec{w}$ is a one-form or linear function mapping $\vec{x}$ onto **R**.) The weight vector $\vec{w}$ is learned from a set of labeled training samples. Often *f* is a simple function that maps all values above a certain threshold to the first class and all other values to the second class. A more complex *f* might give the probability that an item belongs to a certain class.For a two-class classification problem, one can visualize the operation of a linear classifier as splitting a high-dimensional input space with a hyperplane: all points on one side of the hyperplane are classified as "yes", while the others are classified as "no".

$$\min_{\mathbf{s}} \left\| \mathbf{A}^{\cdot}\, \mathbf{s} - \mathbf{p} \right\|_2^2 + \lambda \left\| \mathbf{s} \right\|_1$$

A linear classifier is often used in situations where the speed of classification is an issue, since it is often the fastest classifier, especially when $\vec{x}$ is sparse. Also, linear classifiers often work very well when the number of dimensions in $\vec{x}$ is large, as in documentation classification, where each element in $\vec{x}$ is typically the number of occurrences of a word in a document . In such cases, the classifier should be well regularised.

$$\rho = \left[\beta_1^{\cdot}, \beta_2^{\cdot}, \cdots, \beta_M^{\cdot} \right]^{\cdot}$$

It also help in occlusion condition as follows:

For handeling occlusions, we constructed histogram and modify it in order to exclude the occluded patches when deal with the target object. If patch is largely reconstructed then error is regarded as occlusion and the corresponding sparse coefficient vector is set to be zero

$$\varphi = \rho \square \ \mathbf{o}$$

$$o_i = \begin{cases} 1 & \varepsilon_i < \varepsilon_0 \\ 0 & \text{otherwise} \end{cases}$$

Finally with the help of SDC and SGM they develop collaborative method within the particle filter framework , and the tracking result is the candidate with the highest probability.The generative model is effective to account for appearance change. The discriminative classifier is effective to separate the foreground target from the background. Our method exploits the collaborative strength of both schemes using Equation

$$\begin{aligned} p_c &= H_c L_c \\ &= \exp\left(-\left(\varepsilon_f - \varepsilon_b\right)/\sigma\right)\left(\sum_{j=1}^{J\times M} \min\left(\varphi_c^j, \psi^j\right)\right) \end{aligned}$$

this paper conclude that an effective and robust tracking method based on the collaboration of generative and discriminative model. The SDC module can effectively deal with continuous changing background. The SGM module is capable of handling heavy occlusion.

In [4],This paper gives some theoretical principles for online learning of target model and provide adaptive tracking algorithm which is able to deal with drastic variations in target appearance then there occur some problem in tracking. Once target is extracted in each frame then the frame sample taken from target are first classified into foreground and background using an effective classifier. This paper propose robust, adaptive appearance model for motion-based tracking of difficult background changing object. Adaptive learning that is implemented in the classroom environment using information technology is often referred to as an Intelligent Tutoring System or an Adaptive Learning System. Intelligent Tutoring Systems operate on three basic principles Systems need to be able to dynamically adapt to the skills and abilities of a student.Environments utilize cognitive modeling to provide feedback to the student while assessing student abilities and adapting the curriculum based upon past student performance. Inductive logic programming (ILP) is a way to bring together inductive learning and logic programming to an Adaptive Learning System. Systems using ILP are able to create hypothesis from examples demonstrated to it by the programmer or educator and then use those experiences to develop new knowledge to guide the student down paths to correct answers. Systems must have the ability to be flexible and allow for easy addition of new content. Cost of developing new Adaptive Learning Systems is often prohibitive to educational institutions so re-usability is essential. School districts have specific curriculum that the system needs to utilize to be effective for the district. Algorithms and cognitive models should be broad enough to teach mathematics, science, and language. Systems need to also adapt to the skill level of the educators. Many educators and domain experts are not skilled in programming or simply do not have enough time to demonstrate complex examples to the system so it should adapt to the abilities of educators.

In [5], Robert T. Collins provide online feature selection mechanism for evaluating multiple features while tracking and adjusting the set of features used to improve tracking performance. we generally conclude that the features that can be easily identified can be easily tracked. There are three conventional approaches to moving object detection: temporal differencing [1]; background subtraction [13, 29]; and optical flow (see [3] for an excellent discussion). Temporal differencing is very adaptive to dynamic environments, but generally does a poor job of extracting all relevant feature pixels. Background subtraction provides the most complete feature data, but is extremely sensitive to dynamic scene changes due to lighting and extraneous events. Optical flow can be used to detect independently moving objects in the presence of camera motion; however, most optical flow computation methods are computationally complex, and cannot be applied to full-frame video streams in real-time without specialized hardware. Under the VSAM program, CMU has developed and implemented three methods for moving object detection on the VSAM test bed. The first is a combination of adaptive background subtraction and three-frame differencing (Section 3.1.1). This hybrid algorithm is very fast, and surprisingly effective – indeed, it is the primary algorithm used by the majority of the SPUs in the VSAM system. In addition, two new prototype algorithms have been developed to address shortcomings of this standard approach. First, a mechanism for maintaining temporal object layers is developed to allow greater disambiguation of moving objects that stop for a while, are occluded by other objects, and that then resume motion (Section 3.1.2). One limitation that affects both this method and the standard algorithm is that they only work for static cameras, or in a ”stepand-stare” mode for pan-tilt cameras. To overcome this limitation, a second extension has been developed to allow

background subtraction from a continuously panning and tilting camera (Section 3.1.3). Through clever accumulation of image evidence, this algorithm can be implemented in real-time on a conventional PC platform.region centered at 0. The probability of the feature(color) of the target was modeled by the its histogram with kernel

$$\hat{q}_u = C\sum_{i=1}^{n} k(\| x_i^* \|^2)\delta[b(x_i^*) - u], \quad u = 1,...m \text{ bins}$$

$\hat{y}_0$ $\hat{y}_1$

With mean shift method, the kernel is recursively moved from the current location to the new location with

$$\hat{y}_1 = \frac{\sum_{i=1}^{n_h} x_i w_i g(\left\| \frac{\hat{y}_0 - x_i}{h} \right\|^2)}{\sum_{i=1}^{n_h} w_i g(\left\| \frac{\hat{y}_0 - x_i}{h} \right\|^2)}$$

For a kernel with a convex and monotonic decreasing kernel profile, it is guaranteed to converge (to local maxima)

In [6], Zulfiqar Hasan Khan, Irene Yu-HuaGu A novel visual object tracking scheme is proposed by using joint point feature correspondences and object appearance similarity. For any object in an image, interesting points on the object can be extracted to provide a "feature description" of the object. This description, extracted from a training image, can then be used to identify the object when attempting to locate the object in a test image containing many other objects. To perform reliable recognition, it is important that the features extracted from the training image be detectable even under changes in image scale, noise and illumination. Such points usually lie on high-contrast regions of the image, such as object edges.Another important characteristic of these features is that the relative positions between them in the original scene shouldn't change from one image to another. For example, if only the four corners of a door were used as features, they would work regardless of the door's position; but if points in the frame were also used, the recognition would fail if the door is opened or closed. Similarly, features located in articulated or flexible objects would typically not work if any change in their internal geometry happens between two images in the set being processed. However, in practice SIFT detects and uses a much larger number of features from the images, which reduces the contribution of the errors caused by these local variations in the average error of all feature matching errors.SIFT can robustly identify objects even among clutter and under partial occlusion, because the SIFT feature descriptor is invariant to uniform scaling orientation and partially invariant to affine distortion and illumination changes.This section summarizes Lowe's object recognition method and mentions a few competing techniques available for object recognition under clutter and partial occlusion.

The joint tracker per-forms better as compared to anisotropic mean shift trackingin and SIFT tracking in followed by the RANSAC.

III. PROPOSED ARCHITECTURE

Fig 1. Show the system architecture Strore the video with image file name storing the video sequences initiate the x and y co-ordinate for the bounding rectangle initialize the width.

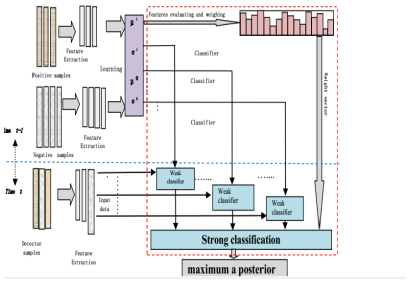
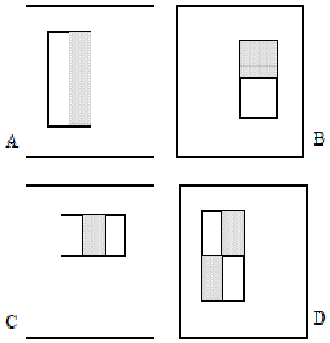


Fig. 1. Example of a figure caption.

IV. PROPOSED MODEL

The process of Active Learning consists of two main stages, an initialization stage and a sample query with retraining stage. The initialization stage is the same process as creating a passive learning function. The Query and retraining stage consists of initially obtaining new data by running the classifier on an unlabeled, independent dataset and a ground truth mechanism (a human) is assigned to label the newly obtained data. This data is then used to retrain the classifier to generate a new learning function (Query by Misclassification). A broad schematic of the Active Learning process is provided in The passively trained classifier obtained initially was evaluated on an independent dataset by using the test rig on a busy highway during a low lighting period. This run produced very high classification accuracy but missed some true positives and produced false positives. These instances were queried and labeled by a human oracle and included for retraining. Thus, the retraining process consisted of some positive samples which included initial training samples along with missed ”true” detection instances from the independent dataset. The main idea of working with features is that it is much faster than a pixel based classification system which is integral to the idea of rapid detection in real time. The weak classifiers (explained later in detail) works with values of very simple features. These features are derivatives of Haar basis functions used by Papageorgiou et. al in his trainable object detection framework. The three kinds of features used in this study are: Two Rectangle Feature : As shown in Figure 3-1, the value of a two rectangle feature is the difference between the sum of pixel values within two rectangular regions in a Region of Interest (ROI). The Region should have the same size and should be horizontally or vertically adjacent. Three Rectangle Feature : Similarly, a three rectangle feature is the sum of the pixels of the two outside rectangles subtracted from the sum of pixels of the center triangle.Four Rectangle Feature : A four rectangle feature is the difference in the sum of pixels of two pairs of diagonally opposite rectangles. The minimum size of the detection window was chose to be 20x20 based on trial runs and given this information, the set of rectangular features is much higher than the number of pixels in the window . Thus this representation of features is over complete and a suitable feature selection procedure has to be integrated into the algorithm to speed up the classification process.

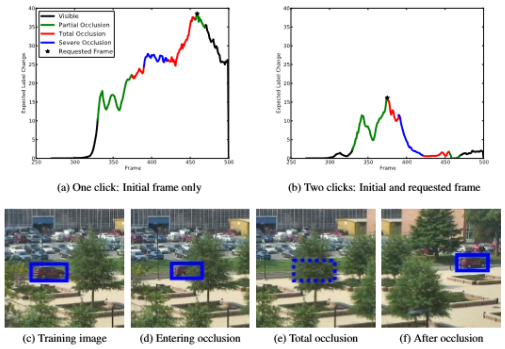


V. IMPACT OF PROPOSED SYSTEM

This system provided the proof of concept implementaion of the Actively trained classifier for the purpose of real time vehicle detection. The classifier was evaluated on publicly available static image datasets as well as on real time video stream collected using the test rig. The classifier was then compared to the Passively trained classifier based on some specific evaluation criteria and results were presented. The full implementation of the multi-object detection and tracking system complete with a feature tracker and distance estimation on a real time scenario is also presented.

VI. EXPECTED OUTCOME

The idea of this project was to present a vision based detection and tracking system which can be implemented in real time systems such as in intelligent vehicles and autonomous cars. The main challenge in addressing this issue was to create a robust, reliable system which was simple to implement. A machine learning based approach was devised to solve this problem where a cascade of classifiers were trained based on Adaboost (working on Rectangular Haar-like features in an image) to rapidly detect Regions of Interest (ROIs) corresponding to cars in a video frame. This classifier was then retrained using Query by Misclassification to produce an Active classifier which was much more sensitive to noise.



(a) One click: Initial frame only (b) Two clicks: Initial and requested frame

(c) Training image (d) Entering occlusion (e) Total occlusion (f) After occlusion

VII. CONCLUSION

Thus we proposed a robust tracker based on online discriminative appearance model.We develop an online active feature selection approach via minimizing Fisher information criterion and show that This method could also be used to selectively sample the independent dataset used for Active Learning and query it for retraining. Integration of lane detection, trajectory learning and pedestrian detection are some other key.

REFERENCES

[1] KaihuaZhag, Lei zhang,QinghuaHu"Robust object tracking via active feature selection", IEEE Ttransaction On Circuit And System For Video Technology, Vol.23, No.11, Nove.2013.

[2] Li Sun, GuizhongLiu."Visual object tracking based on combination of local description and global representation", IEEE Transactions On Circuits And Systems For Video Technology, Vol. 21, No. 4, April 2011.

[3] Wei Zhong, HuchuanLu"Robust object tracking via sparse collaborative appearance model", IEEE Transactions On Image Processing, Vol. 23, No. 5, May 2014.

[4] Peng Wang, Hong Qiao"Online appearance model learning and generation for adaptive visual tracking", IEEE Transactions On Circuits And Systems For Video Technology, Vol. 21, No. 2, February 2011.

[5] Robert T. Collins, YanxiLiu"Online selection of discriminative tracking features", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 27, No. 10, October 2005.

[6] ZulfiqarHasan Khan, Irene Yu-HuaGu"Joint features correspondence and appearance similarity for robust visual object tracking", IEEE Transactions on Information Forensics and Security, Vol. 5, No. 3, September 2010.