International Journal of Engineering Research in Electronic and Communication Engineering (IJERECE) Vol 3, Issue 5, May 2016

Improvement in object detection using Super Pixels

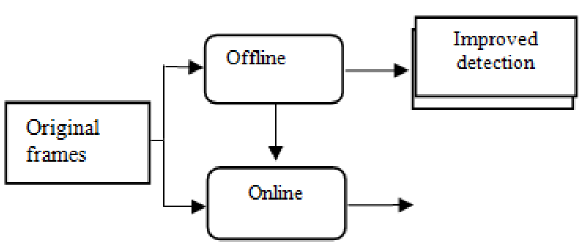
[1] Shruti D Kadam [2] H.Mallika
Dept. of Electronics and communication
M. S. Ramaiah Institute of Technology, Bangalore, Karnataka
[1] chachchi92@gmail.com [2] mallika.h@msrit.edu

***Abstract*— Most of the object detectors performance is degraded due to change in illumination, variant background and camera position. A method to enhance the detection performance of an offline generic detector is proposed in the paper. In this approach, all the detections are represented in Bag of Word fashion considering super pixels as its feature of classification, combining super pixels extraction and bag of word improves the object detection of a generic offline detector, object shape extraction from its background is segmented using graph cut algorithm. In standard, proposed approach takes the detection bounding box generated by a generic detector as input and improves the detection with better average precision. Bounding boxes are reduced with the objects shapes giving better performance using graph cut algorithm.**

***Keywords*—BoW, superpixels, segmentation**

I. INTRODUCTION

Video surveillance plays very important role in public security. Computer aide surveillance can be automatic and non- stop for the surveillance. Human detection is one of the computer aid surveillance which is a challenging task due to their variable appearances and wide range of poses. There are two major object detectors one is state-of-art object detectors [1] and another is object detection using Histograms of Gradients (HOG) or Haar-like features[2]. When applied to a particular video there performance degrades due to lack of pre trained examples in all unconstrained video environment and fails to fully extract all the information from the video.



1. This figure shows our online approach to enhance the object detection of an offline model. In an improved detection improved object detection frames will be available.

In present paper we aim to improve such problems for the detection. Online learned model[3, 4]used to refine output of a detectors. Proposed model works as online model.In a surveillance video each individual changes their poses and move from one location to other but their color feature remains same.Makinguse of this consistent color feature improves the detection of a generic detector.Superpixel [5] information is extracted rather than taking a pixel to make this task faster and easy for segmentation, Superpixel helps in extracting information as a sub regions some times its brittle to handle all the pixels individually, it reduces the complexity of image processing tasks and miss-alignment of features for different poses are reduced.Superpixel is faster than the pixels. Reduction of memory for storage becomes lesser.

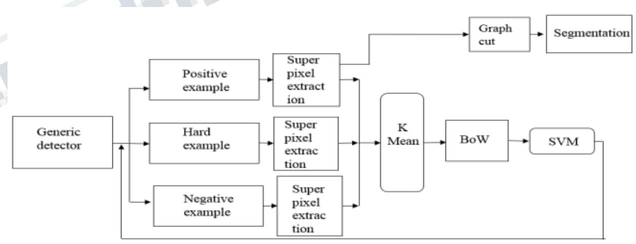


Figure 2. Demonstration of our project

Online proposed approach is illustrated in figure 2.First step is applying a simple generic detector on every frame of a video and obtain a detection examples.Those examples are classified initially according to its detection score as positive and hard examples. Negative examples are manually collected from the background. Next step is extracting the superpixel features from all the examples and then represent these in a bag-of-word fashion later applying SVM classifier to refine the classification of positive hard and negative examples according to its score. Iteratively refining these classes and get a proper and fixed set of examples stored at each classes. Graph cut [6]algorithm is used for the segmentation only on positive examples to extract object shape.

In the proposed methodpedestrian detection is considered for the implementation. In section 2, related work is briefly introduced. In section 3 approached method is briefly described. Section 4, draws the experimental results and in section 5 paper is concluded.

II. RELATED WORK

Section reviews briefly on related works of online learning model, such approaches can be classified as one with semi-supervised learning system [7], second approach is automatically detecting the objects[8], third is detection-by-tracking approach. In first method co-training based approach is used to label the data but requiring a huge amount of labeled examples manually. In automatic labeling of data this approach make use of background subtraction algorithm which fails to work for a moving camera and a complex Camera[7,9] is a non-parametric transfer learning approach these are based on objects with similar appearances fails for the hard examples with large variations.

Third category is detection-by-tracking approaches this category improves the detection by tracking[10, 11, 12]its efficiency is based on trackers. If particular tracker is not suitable for the scenario it lowers its performance and tracker produces lot of noise.

Proposed approachtriessolving the problems in semi-supervised model and detection-by-tracking. Instead of considering a generic detector our approach considers a video specific detector obtaining constant color patterns in a frame. Trackers are not used in this paper so the problems introduced by trackers will not be considered. Giving better performance compared to both.

III. PROPOSED WORK

3.1. Implementing generic detector

We employ Histogram of Oriented Gradients for Human detection [13]as initial detector for our approach, we can make use of any generic detectors like Deformable part based model (DPM)[1]detector which shown excellent performance in static region but HOG based detector is easily available and is used for the experiment. According to the score of the detectors each detection is classified as positive and hard example, positive refers to the examples having true positive detection. Hard examples are those with occluded human detection. Negative examples are randomly collected from the background manually which do not collapse with any positive and hard examples.

3.2. Extraction of super pixels

Initial detector uses HOG or Haar like features for detection. Histogram of Gradient (HOG) fails to leverage the information for different poses of a single object.To overcome such issues we make use of superpixel instead of HOG.This makes the generic detector as video specific detector.

Figure 3. Extraction of super pixel using SLIC algorithm

We use simple linear iterative clustering (SLIC)[12]algorithm to extract the superpixels because it is significantly more efficient compared to competing methods. We cluster each detection as N number of desired super pixel regions' has to be selected carefully so that the color proximity in the sub regions remains same within the boundaries figure 3 shows example of a superpixel extraction.Each superpixel has 5 dimensions $f = (L, a, b, x, y)$in which(L, a, b) are the values of the CIELAB color space and $(x, y)$is the average pixel location. All the Superpixels are clustered using k-means algorithm.These vocabulary is transferred to bag of feature which returns a bag ofvisual features that uses Superpixels to learn its visual vocabulary.

3.3. Classification

We train a support vector machine to classify the hard examples. Initial training examples have limitations because of which obtaining a good decision on classifying hard, positive and negative examples are difficult. Iterative way is used to gradually improve the training sequence so that the output gets refined.We move the examples with high score to positive and low score to negative. After this labeling these positive examples are used for segmentation.

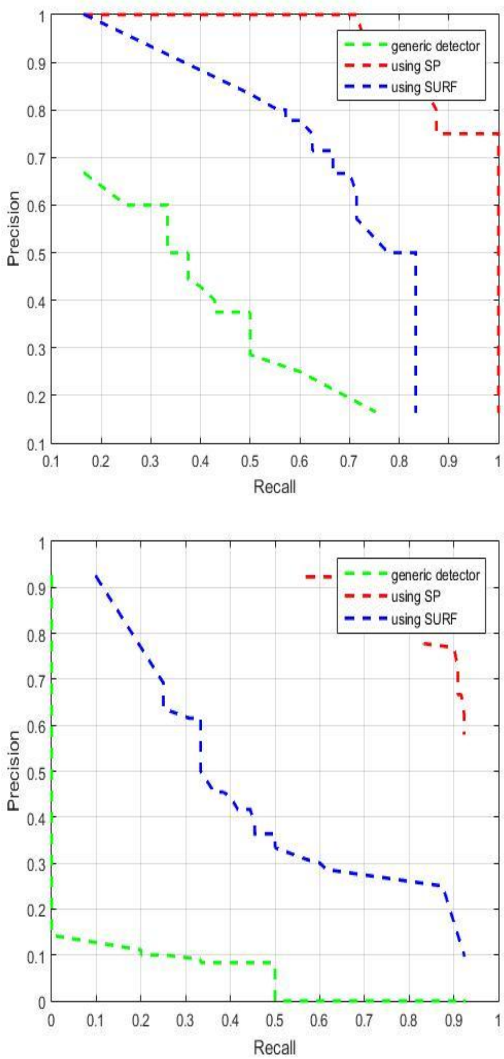


Figure 4.Performance comparison of the generic detector, surf and Superpixel with PETS 2009 database and Oxford town center database.

Obtaining a better precision-recall compared to any other method.

3.4. Segmentation

Segmentation is done using Graph cut method that minimizes energy function

$$E(l) = \sum_p D(p, l_p) + \sum_{\{p,q\}} V_{pq}(l_p, l_q) \qquad (1)$$

p,q indicates the single pixel, l_p and $l_q$are the labels assigned for the p and q pixels respectively.
First term indicates sum over all pixels p of the penalty (or cost),$D(p, l_p)$ is cost of assigning label $l_p$to pixel p with the observed data, second term is sum over all

Figure 4. Output of the generic detector with thebounding boxes (Left) output after using proposed approach is in red color having object shape.

Figure 5. Output after segmentation using graph cut method pairs of neighboring pixels with distinct penalty$V_{pq}(l_p, l_q)$. Unlike Conditional random field which depends on its initial detector for energy minimization function GC algorithm is independent of initial detector. Graph cut algorithm helps to segment the object boundaries. Table 1.calculation of Average Precision (AP) function is used which is one of the feature detector and used for object recognition and output is driven. We set initial detection threshold to classify the examples as t= 0.1, number of super pixels as N=100 and k=200 as the number of clusters. Proposed approach has achieved better result in both the datasets as shown in figure 4. Average precision (AP) is calculated by averaging all the observed calculations and tabulated in table1.

Database / Types	**Database1 (PETS 2009)**	**Database2 (Town Centre)**
HOG	42.046	8.36
Using SP	81.788	76.04
Using SURF	61.96	39.48

V. CONCLUSION

Proposed approach is an effective model to improve generic detector detection using super pixel and Bag-of-Word hence it is highly distinctive and robust against appearance changes. We extract superpixel from the detection and apply bag-of-Feature we improve the performance as shown in the table and we iteratively improve the process. We use Graph Cut method which helps in segmenting exact object shape instead of bounding boxes.

REFERENCES

[1] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. "*Object detection withdiscriminatively trained part based models*", In PAMI, 2010.

[2] Navneet Dalal and Bill Triggs "*Histograms of Oriented Gradients for Human Detection*", in CVPR 2005.

[3] V. Nair and J. Clark. "*An unsupervised, online learning framework for moving object detection*", In CVPR, 2004

[4] A. Levin, P. Viola, and Y. Freund. "*Unsupervised improvement of visual detectors using co-training*", In CVPR, 2003.

[5] R. Achanta, A. Shaji, P. F. Kevin Smith, Aurelien Lucchi, and S. Susstrunk. "Slic superpixels", In EPFL Technical Report, 2010.

[6]Y. Boykov and Jolly. "*Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images*". In ICCV, 2001.

[7] P.ViolaandM.Jones. "*Rapid object detection using a boosted cascade of simple features*", In CVPR, 2001.

[8]V. Nair and J. Clark. An unsupervised, online learning framework for moving object detection. In CVPR, 2004

[9] A. R. Zamir, A. Dehghan, and M. Shah. Gmcp-tracker "*Global multi-object tracking using generalized minimum clique graphs*", In ECCV, 2012.

[10] G.Shu, A. Dehghan, O.Oreifej, E.Hand, and M.Shah. "*Part based multiple-person tracking with partial occlusion handling*", In CVPR, 2012.

[11] B. Benfold and I. Reid. "*Stable multi-target tracking in real time surveillance video*", In CVPR, 2011.

[12] H. Celik, A. Hanjalic, and E. A. Hendriks. "*Unsupervised and simultaneous training of multiple object detectors from unlabeled surveillance video*", In Computer Vision and Image Understanding, 2009.

[13] Guang Shu, Afshin Dehghan, Mubarak Shah "*Improving an Object Detector and Extracting Regions using Superpixels*",In CVPR,2013