Detecting abandoned object in surveillance videos.

Project Report

Rahul Sankhwar(11569)
Supervisor: Prof. A. Mukerjee

April 17, 2015

Abstract:

The motive of this project was to detect and localize anomalies in surveillance videos. The main idea of this work is based on [1] (Anomaly Localization in Topic-based Analysis of Surveillance Videos). The objective was completed by extending [1], by considering a different foreground extraction mechanism such that it also includes static objects as anomalies.

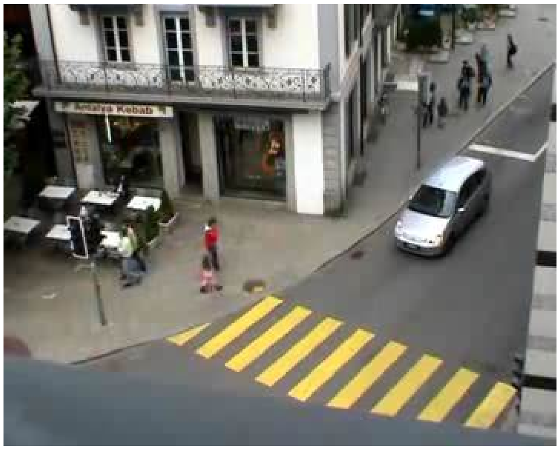


*Figure 1:*Car stopping after parking spot should treated as anomaly.

1. Motivation:

Due to increase in terrorism, terrorists are targeting crowded public places such as traffic junction, bus station etc. Therefore, need for surveillance system is necessary to ensure safety of people and it would also not require much manual assistance. So effective and efficient detection and localization of abandoned objects is very important to prevent attacks.

2. Related work:

As stated in [1], author had used topic based anomaly detection in surveillance videos, by using object-based models, for foreground modeling and low-level feature description. In [1], Pathak et al. used, foreground extraction method, ViBe proposed in [3].

In [2], Stauffer et al. developed a foreground extraction using Gaussian Mixture Model.

In [4], Authors proposed block-based gaussian mixture modeling of the background and used three cascade classifiers for foreground extraction.

3. Approach:

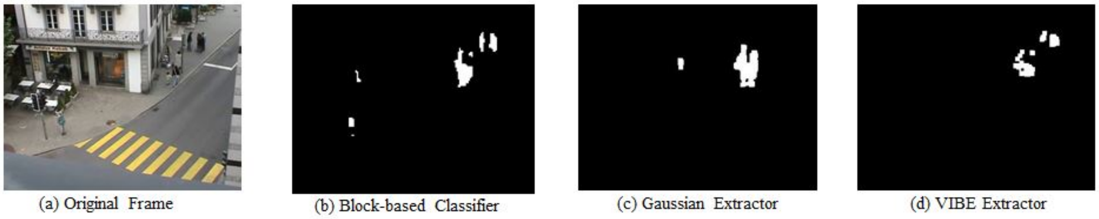
Foreground extraction mechanism proposed in [3], Vibe, is based on motion cues and models abandoned objects/vehicles as foreground for few frames but then this information dies out. Therefore, in [1], problem with abandoned objects is that they loose the foreground characteristic after sometime. Therefore, to give some improvement, thinking along this direction, in order to detect abandoned anomaly, grids over abandoned objects should also give some foreground info.
In order to do so, I implemented foreground extraction mechanism proposed in [2] and [3], which are based on background subtraction.

4. Methodology:

In this section a brief step by step pipeline of this project would be discussed.

4.1 Modeling:

The frames of the video are extracted and then the foreground extraction mechanism in [2], [3],[4] is used.



(a) Original Frame

(b) Block-based Classifier

(c) Gaussian Extractor

(d) VIBE Extractor

Figure 2 (a) is the original frame. (b) shows foreground extracted by [4] (c) shows foreground extracted by [2] (c) shows foreground extracted by [3]

On these frames we then find out context-based three dimensional visual word(figure 3). These three dimensions are:

1. *Location*: Each frame is divided into 20*20 grid. The centre of the grid is the location parameter.

2. *Hog-Hof descriptor*: The gradient information and optical flow is captured using hog hof descriptor.

3. *Blob size*: Each foreground blob is quantized into small and large.

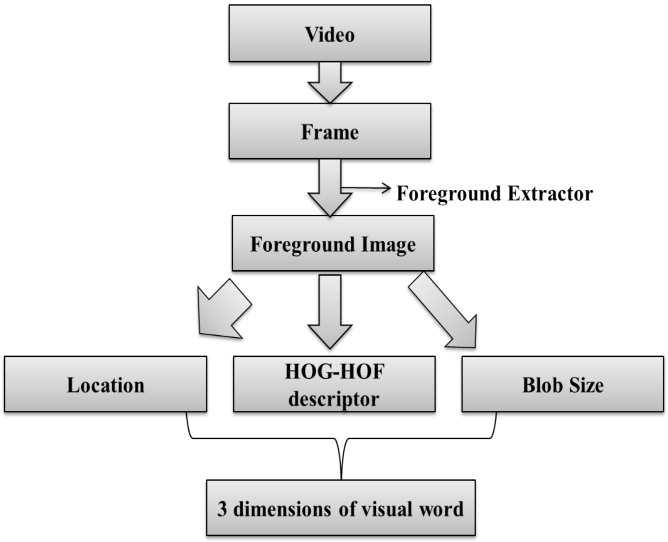


Figure 3 :Formation of visual word

Analogous to NLP we would do topic modeling on these visual words. Video is divided into clips which is treated as documents and then these are represented as histograms over the visual words. Now, parametric Bayesian topic model, pLSA(probabilistic latent semantic analysis) is used to model the term frequency matrix of these document clips and vocabulary. Therefore we get representation of each document as probability distribution over latent topic space.

4.2 Detection:

In[1], the authors proposed an efficient Projection model algorithm for detection of anomaly. As stated in [1], “For a new video document, we investigate the `usualness' of each visual word by comparing with the projected word histograms of nearest train documents in topic space.” Nearest train documents are analogous to video clips that are similar to the test clip. If the test event has occurred in them then they are usual otherwise anomalous.(See figure 4).

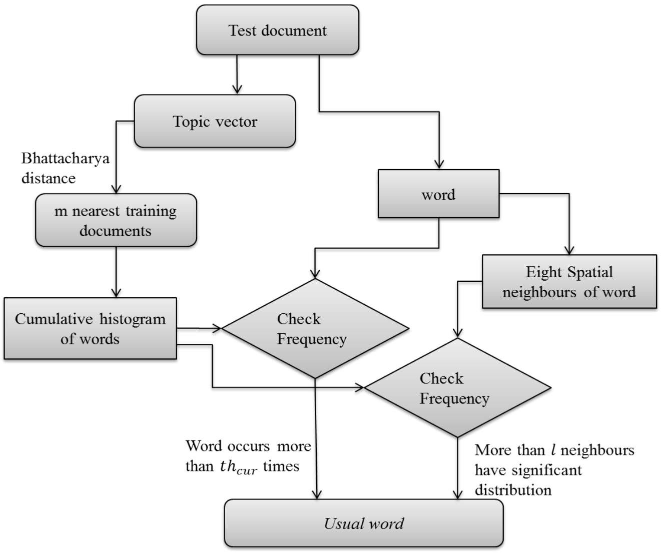

Test document
Topic vector
Bhattacharya distance
m nearest training documents
word
Eight Spatial neighbours of word
Cumulative histogram of words
Check Frequency
Check Frequency
Word occurs more than th_{cur} times
More than l neighbours have significant distribution
Usual word

Figure 4 : Flowchart of projection model algorithm to detect anomalies, proposed in [1]

4.3 Localization:

The spatial location of the anomaly is in the location parameter of the visual word while for getting temporal location we have tagged the frame numbers with each word.(see figure 6)

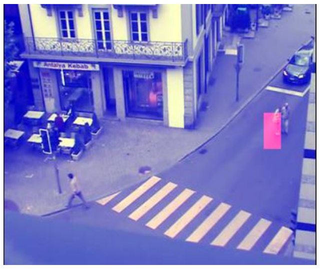
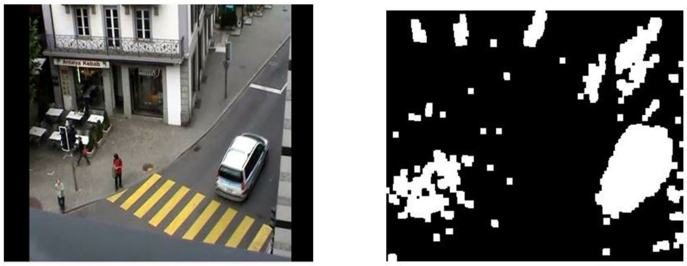


Figure 6 : People jaywalking, localized as anomaly (implemented using block-based extractor [4])



(a) Original frame

(b) Foreground using GMM

Figure 5 : A large noise is observed when GMM foreground extractor[2]

5. Results:

Traffic Junction Dataset, [5] is used in the experimentation. It is a single video of 45 minutes duration of a traffic junction, shot from a static camera from top of a building. Normal events include people crossing the road using zebra crossing, cars stopping before the stop line etc.

5.1 Foreground extraction using Gaussian Mixture Modeling[2]:

We modeled the background using three Gaussian mixtures and set the number of training frames to be 1500. Figure 1(b) shows an example of extracted foreground. As shown in Figure as compared to ViBe,[3], abandoned objects were also detected in it. But this method has a lots of noise in it as seen in figure 5.

5.2 Foreground extraction using block-based method,[4]

In this method block size was kept 8*8 pixels with advancement of 2 pixels. The number of training frames used were 1500. Figure 1(b) shows an example of extracted foreground. In ViBe,[3], abandoned objects were also detected in it. It also has good contour and different objects are not merged in the foreground.

Out of the three foreground extractors, the block-based extraction seems most relevant to our case, since, it is detecting abandoned objects and foreground extracted is more precise.

5.3 Anomaly detection using ViBe[3] and block-based extractor[4]

As stated in [1], "There are four kinds of anomalous actions in the video, namely: jay walking , car stopping after the stop line on the road, people crossing the road away from the zebra crossing(see Figure 6) and car entering the pedestrian area".
The number of actions in the video were kept to be 20, which are analogous the number of topics in the document. The video clips of 4s duration each were divided, where each clip is analogous to as document in NLP case. Anomalous video clips were removed for testing. From the remaining set of the rest of the video clips, for training 3/4 of the clips were used. For the test data the remaining quarter of the clips were included in along with the anomalous clips. The anomalous clips served as positive examples and the non-anomalous clips were served as negative cases.

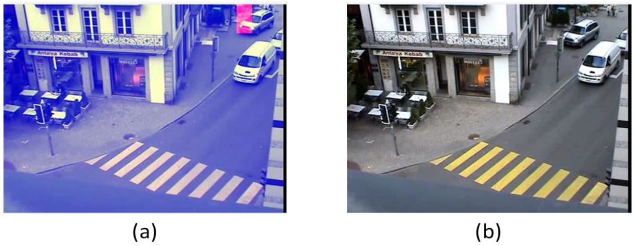
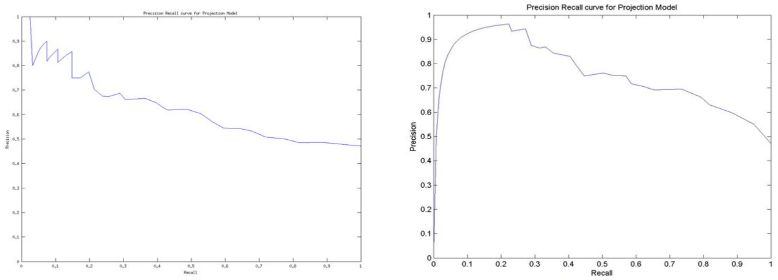


Figure 7: (a) shows car on sidewalk as anomaly in block based method while when ViBe was used it is not the case as seen in (b)

Figure 7 shows car parked in pedestrian area detected as anomaly in when block-based extractor was used but not detected as anomaly when ViBe was used. Precision and recall curve for the two cases is shown in figure 7. Table 1 shows optimum precision, recall and area under precision and recall curve for the two cases. For demo clips visit: http://home.iitk.ac.in/~rahulsan/cs365/project/demo.zip

Statistics	VIBE	Block-based Classifier
Optimal Recall	0.36364	0.35537
Optimal Precision	0.66667	0.84314
Area Under Precision Recall Curve	0.61689	0.75674

*Table 1:*Comparison when ViBe and Block-based extractor was used



(a) VIBE Used (b) Block-based extractor used

Figure 8: Comparision of precision recall curve for the two cases

Conclusion:

1. We changed the foreground extraction technique deployed, and used two other techniques for it. One is Gaussian mixture model foreground extractor and other is Block based classifier foreground extractor.
2. The Gaussian mixture model foreground extractor produces very noisy images and does not capture discrete objects properly with too large contour size.
3. The performance with the Block based classifier foreground extractor is better than the one with VIBE foreground extractor. The precision-recall curve obtained is indicative of this fact.

Future work:

Implementation of hierarchical LDA instead of pLSA.

Acknowledgment:

I would like to thank Prof. Amitabha Mukherjee and Deepak Pathak for their humble guidance.

References:

1. D Pathak, A Sharang, A Mukerjee, “*Anomaly Localization in Topic-based Analysis of Surveillance Videos*” IEEE Winter Conference on Applications of Computer Vision (WACV 2015).
   http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7045912
   Source code: http://www.cse.iitk.ac.in/users/abhisg/btp/abhisg.zip

2. Stauffer, C. and Grimson, W.E.L,*Adaptive Background Mixture Models for Real-Time Tracking*, Computer Vision and Pattern Recognition
   http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5571035&tag=1
   Source code: Available in vision toolbox of Matlab.

3. O. Barnich and M. Van Droogenbroeck. “*Vibe: A universal background subtraction algorithm for video sequences.*” Image Processing, IEEE Transactions on, 20(6):1709–1724, 2011.
   http://dl.acm.org/citation.cfm?id=2333806

Source code: http://www.motiondetection.org/

4. Vikas Reddy, Conrad Sanderson, Brian C. Lovell.*"Improved Foreground Detection via Block-based Classifier Cascade with Probabilistic Decision Integration."*
   University of Queensland, School of ITEE, QLD 4072, Australia
   Source code: http://arma.sourceforge.net/foreground/

5. Traffic Surveillance dataset: http://www.eecs.qmul.ac.uk/~andrea/avss2007_d.html