Anomaly Detection in Topic Based Analysis Of Surveillance Videos

Poster presented by: Rahul Sankhwar

Mentor: Dr. A. Mukherjee

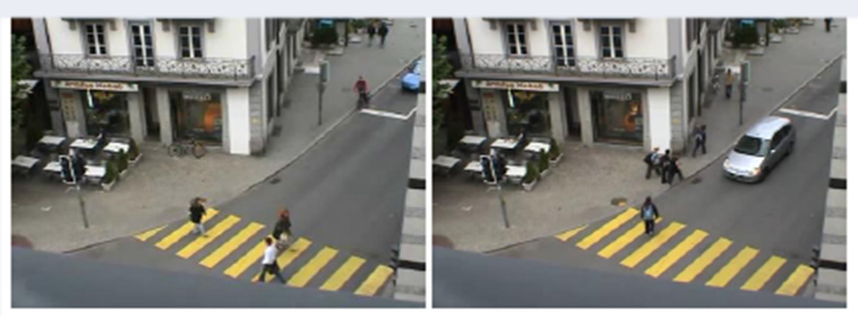
1. INTRODUCTION

In the present era of big data, the presence of large volume of un-annotated data over the web has motivated the research in unsupervised data classification, recognition and segmentation. Owing to enhanced security mechanisms and deployment of surveillance cameras, the need for automated analysis of videos to detect abnormal and anomalous events has recently given rise to an active research in computer vision and machine learning community in this field.

[1] addresses the problem of analysing surveillance videos to identify unusual or anomalous events.

The term **"anomaly"** is defined as the events which are not 'usual' in the video i.e. after modelling the dominant behaviour in the video, the events which are not prevalent are being classified as anomalous.

Source: [4] BTP Report of Deepak et. al



(a) Traffic Junction Dataset [20]. The left image is usual while right is an anomaly - car stops after the stop-line.

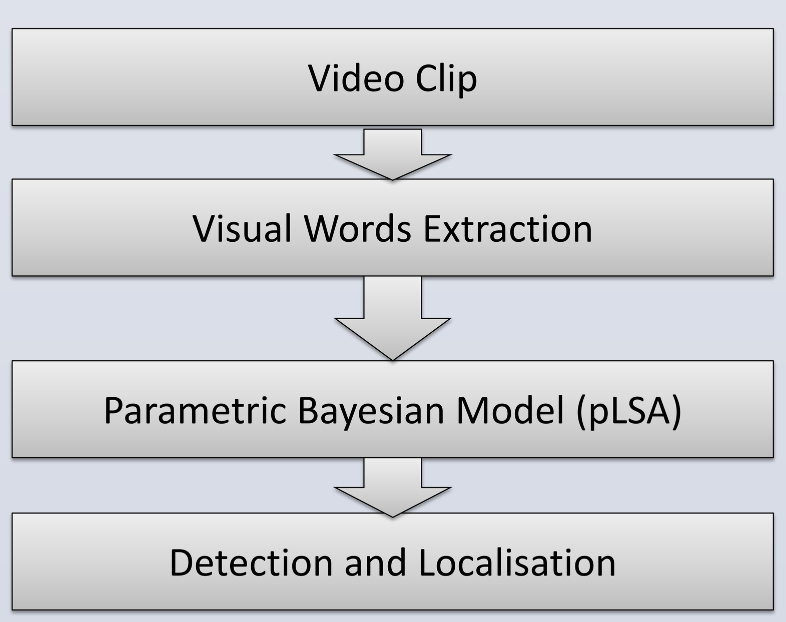
Image Source: [4]

Related work

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 [2].

In [5], author proposed Gaussian Mixture Model for foreground extraction.

2.1 MODELLING



Text Analysis	Video Analysis
Vocabulary of words	Vocabulary of visual words
Text documents	Video clips
Topics	Actions/Events

Visual Word

Location :

- Each frame of dimension m x n is divided into blocks of 20 x 20

HOG - HOF descriptor :

- For each block, a foreground pixel was selected at random and spatio-temporal descriptor was computed around it.
- From the descriptors obtained from the training set, 200,000 descriptors were randomly selected. 20 cluster centres were obtained from these descriptors by k-means clustering.
- Each descriptor was assigned to one of these centres.

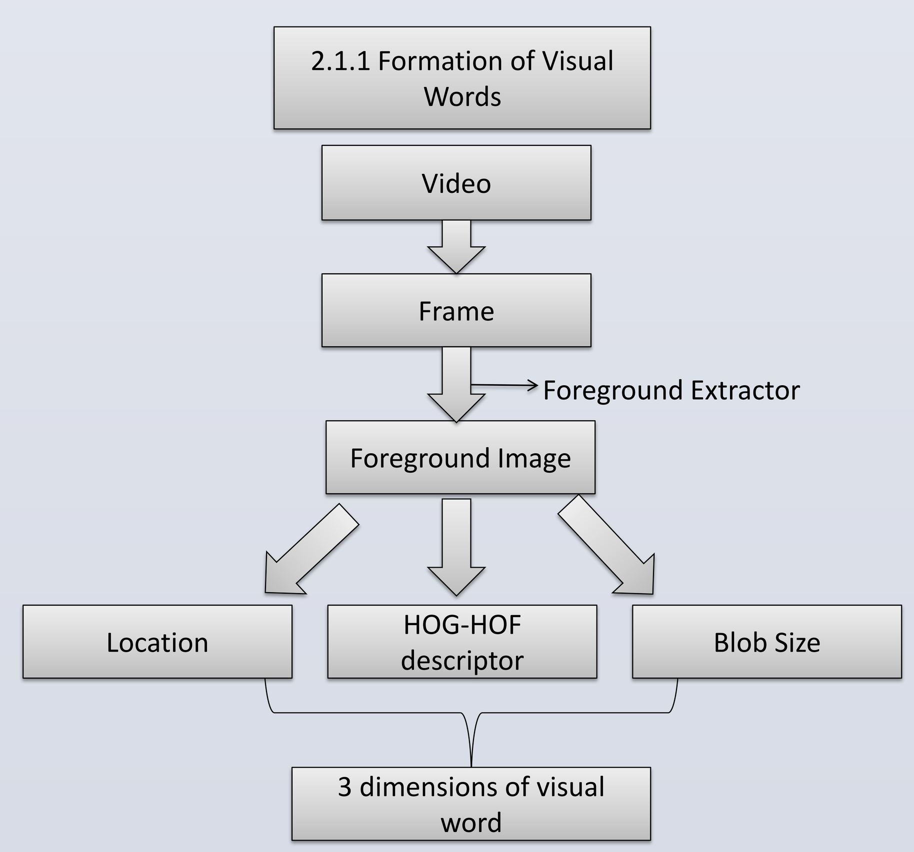
Size :

- In each block , we compute the connected components of the foreground pixels
- The size of the connected components is quantised to two values: large and small

Source: [4] BTP Report of Deepak et. al

(a) Sample Frame (b) Foreground Marked (c) HOG-HOF Marked

Image Source: [4]



Parametric Bayesian Model (pLSA)

Fixed number of topics : $\{z_1, z_2 \ \dots z_k\}$. Each word in the vocabulary is attached with a single topic.

Topics are hidden variables. Used for modelling the probability distribution

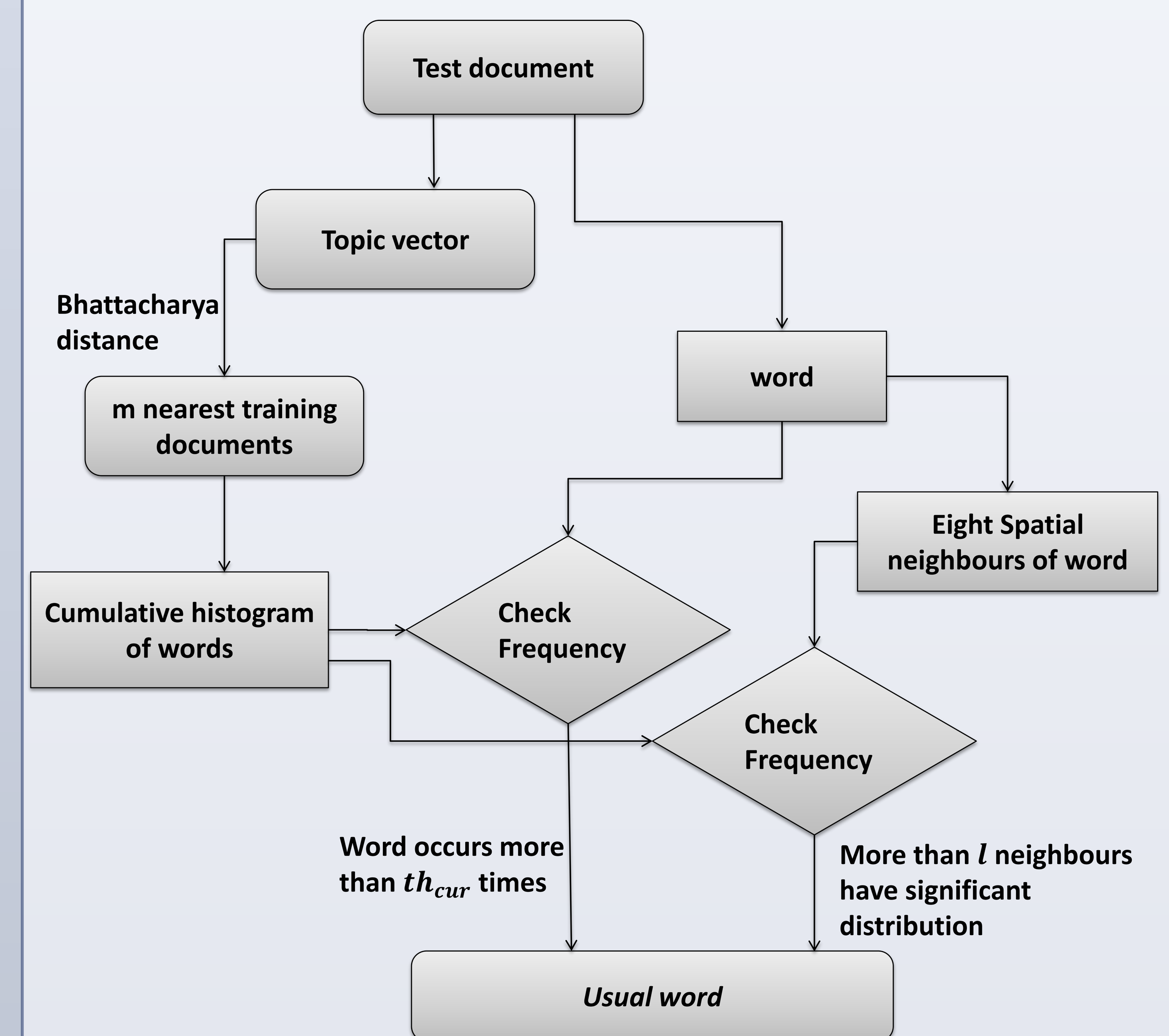
Computation

- Marginalise over hidden variables
- Conditional independence assumption: p(w|z) and p(d|z) are independent of each other

$$P(d,w) = P(d)P(w|d) = P(d)\sum_{z\in Z} P(w,z|d)$$
$$= P(d)\sum_{z\in Z} P(w|z,d)P(z|d)$$
$$= P(d)\sum_{z\in Z} P(w|z)P(z|d)$$

2.2 DETECTION

Projection Model Algorithm



2.3 Localization

Spatial Localization :

Every word has location information in it. Therefore we can directly localize the anomalous words in test document to their spatial locality.

Temporal Localization :

If we maintain a list of frame numbers corresponding to document-word pair, we can tag the frames with anomalous words.

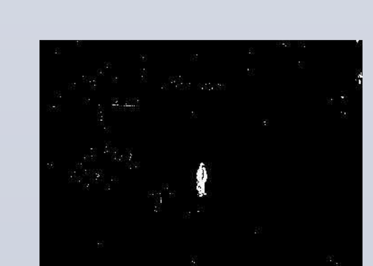
Source: [4] BTP Report of Deepak et. al

Results

Comparison between ViBe foreground extraction and Gaussian Mixture Model foreground Extraction in traffic video dataset.

Abandoned car in the corner

GMM Method: Car detected in the corner.



ViBe Method: Car not detected in the corner.

We performed experimentation on the Traffic junction dataset (i-Lids dataset: http://www.eecs.qmul.ac.uk/~andrea/avss2007_d.html.)

We kept the number of actions in the video to be 20, which served as the number of topics in the document. The document length was l = 4 to l = 10 seconds. Anomalous video clips were separated from the rest of the video clips for testing. From the remaining set, 75% of the clips were used for training and the remaining 25% of the clips were included in the test data along with the anomalous ones.

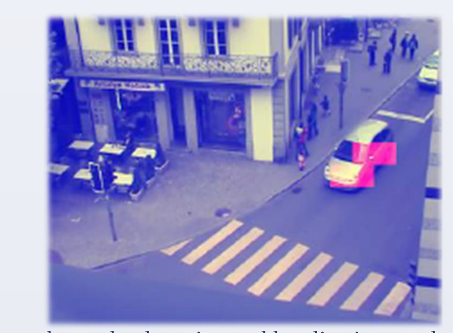
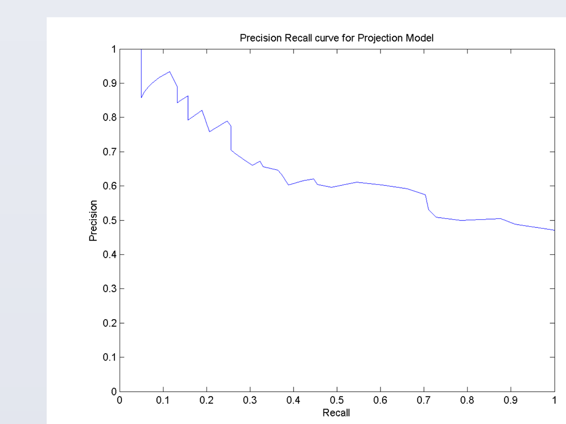
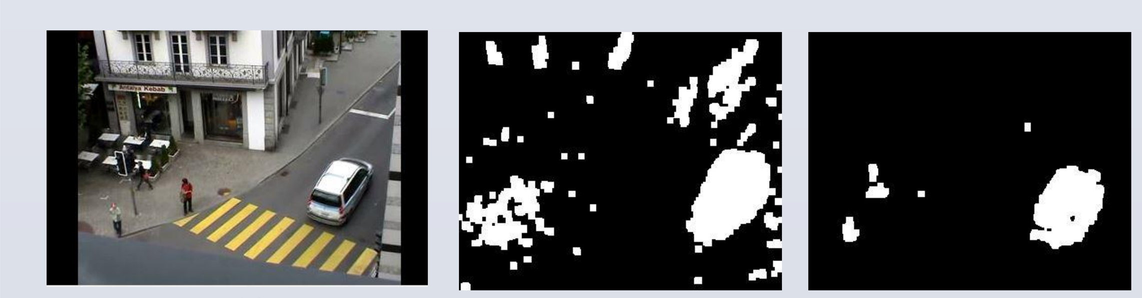


Figure shows the detection and localization results of proposed projection model algorithm for traffic junction dataset.



Precision Recall curve for ViBe Method

When Vibe was used as foreground extraction it localizes 55% of abnormal events in Traffic Junction Dataset, while the GMM foreground extraction localizes only 21% of the anomalies. Also in GMM case there were many false positive case. This arises because GMM foreground extraction contained lots of noise.



Original frame GMM method Vibe method

Conclusion

We can get better result if we improve the GMM foreground extraction method(by removing noise). The noise is resulting into lots of visual words which are not in the video, thus resulting in degraded result.

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).
2. 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.
3. Hofmann, Thomas. "Probabilistic latent semantic indexing." *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1999
4. BTP Report and presentation on, "Unsupervised Modeling, Detection and Localization of Anomalies in Surveillance Videos" D Pathak, A Sharang, A Mukerjee.
5. Stauffer, C. and Grimson, W.E.L,*Adaptive Background Mixture Models for Real-Time Tracking*, Computer Vision and Pattern Recognition,