A Survey on Video Anomaly Detection

Monika Singh
ECE Dept.
AIACT&R, Delhi, India

Abstract— **Anomaly detection has become an important issue that has been researched in the vision based intelligence surveillance application domain and research areas. Anomaly detection is the process whereby a normal behavior is established in the context of computer vision. It is important to detect abnormal behavior patterns and recognize the normal events. So, to achieve this, Anomaly detection techniques have been developed and structured for such application domain. I conduct a survey to provide an overview of the research on anomaly detection techniques based on image processing. I have assembled the existing techniques based on the sparse representations for single object trajectories and joint sparsity model for multiple objects for anomaly detection and also provide the applications, advantages and disadvantages of the techniques in this domain. I hope, this survey will provide a better understanding of the different researches in this field.**

Keywords—Anomaly Detection; Sparsity model; Sparse Representation; Joint Sparsity Model

I. INTRODUCTION

There is an increasing desire and need in video surveillance applications for a proposed solution to be able to analyse human behaviors and identify subjects for standoff threat analysis and determination. The main purpose of this survey is to look at current developments and capabilities of visual surveillance systems and assess the feasibility and challenges of using a visual surveillance system to automatically detect abnormal behavior, detect hostile intent, and identify human subject.

Visual surveillance technologies, CCD cameras, thermal cameras and night vision devices, are the three most widely used devices in the visual surveillance market. Visual surveillance in dynamic scenes, especially for humans, is currently one of the most active research topics in computer vision and artificial intelligence. It has a wide spectrum of promising public safety and security applications, including access control, crowd flux statistics and congestion analysis, human behavior detection and analysis, etc. Visual surveillance in dynamic scene with multiple cameras, attempts to detect, recognize and track certain objects from image sequences, and more importantly to understand and describe object behaviors. Most visual surveillance systems start with motion detection. Motion detection methods attempt to locate connected regions of pixels that represent the moving objects within the scene; different approaches include frame-to-frame difference, background subtraction and motion analysis using optical flow techniques. Motion detection aims at segmenting regions corresponding to moving objects from the rest of an image. The motion and object detection process usually involves environment (background) modelling and motion segmentation. Subsequent processes such as object classification, tracking, and behavior recognition are greatly dependent on it. Most of segmentation methods use either temporal or spatial information in the image sequence.

Abnormality is the sense of something deviating from the normal or differing from the typical is a subjectively defined behavioral characteristic, assigned to those with rare or dysfunctional conditions. Behavior is considered abnormal when it is atypical, out of the ordinary, causes some kind of impairment, or consists of undesirable behavior.

Generally, patterns of behaviors in a scene can be constructed by supervised or unsupervised learning. Based on learned patterns of behaviors, we can detect anomalies and predict object behaviors. When a detected behavior does not match the learned patterns, it is classed as an anomaly.

Many techniques have been developed for the anomaly detection but it is impossible to gather the large number of the training samples for anomalous events, so, the method of sparsity based anomaly detection has been developed. But this method gives the anomaly detection for the single object then the new approach of joint sparsity model has been developed by Xuan Mo for multiple object trajectories.

In this paper the anomaly detection has discussed. Section II describes the related work in this field. Section III gives a brief description sparsity based anomaly detection. Section IV details the joist sparsity model. Section V gives the applications and future work. The last section concludes the paper.

II. RELATED WORK

Bayesian networks are useful for giving the information in vision domain. They have been used in so many applications of audio-visual speaker detection and image and video indexing. bayesian networks are also used for automatic traffic analysis. In [1] Bayesian networks used for classification of targets in videos taken from fixed cameras. Shape parameters of an object depend upon position of object in the image.

Bayesian network classifier for inferring target class. Bayesian network used for target classification. Each node is a variable and target class is the root node. Here, they use the seven measurement nodes x_i, y_i (x, y coordinates of the target in image space), V_x and V_y (x, y coordinates of the targets motion in image space taken from tracking), a and b are the major and minor axis of the ellipse modelling the target and a/b is the aspect ratio of the ellipse.

The velocities V_x and V_y are dependent upon both the target class and the image position of the target, (x_i, y_i). Similarly, the size of the target represented by a and b is made dependent upon the position of the target and its type. The aspect ratio, a/b, measured for a target is dependent on its position and target type. In future development of our work we will consider network structure learning algorithms for better classification of the targets.

ratio, measured for a target is dependent on its position and target type. In future development of our work we will consider network structure learning algorithms for better classification of the targets.

Pankaj Kumar [2] computes local motions as low level features and compute the intensity difference between two successive frames, if the pixel is above threshold then it is a moving pixel and each pixel has two features i.e. position and direction of motion. Then the video is divided into short clips of each 10 seconds in length. These clips are treated as documents and moving pixels are treated as words and the feed these documents to the hierarchical Bayesian models for activity analysis. By Bayesian model, we can detect the abnormal video clips and localise the abnormalities.

Bin Zhao, Li Fei-Fei and Eric P.Xing [3] using sparse for unusual event detection which is based on the spatio-temporal cuboids, to detect salient points within the video and describe each detected interest point with histogram of gradient (HoG) and histogram of optical flow (HoF). This approach employs a sliding window to define an event. A video is represented as a set of cuboids residing in a sliding window define an event. As the sliding window scans along the spatial and temporal axes, the video is broken into a set of events, each represented by a group of spatio-temporal. Specifically, the video is represented as $X = \{X_1, \ldots X_m\}$, with each event X_i composed of a group of cuboids, i.e., $X_i = \{X_1 i, \ldots, X_{ni}\}$, where ni is the total number of cuboids within the sliding window.

Detecting unusual patterns in video define as a sparse coding. Idea of this approach is to represent the unusual events in the dictionary D, whose columns are bases for reconstructing signals. the input signal in unusual event detection is an event, composed of a group of cuboids $X_i = \{X_1 i, \ldots, X_{ni}\}$. Therefore, the basic unit of input signal is no longer a vector, but instead a group of vectors, with both spatial and temporal location information. In sparsity of the reconstruction weight vectors, need to consider the relationships between these weight vectors imposed by the neighbourhood structure of cuboids that define the event.

III. SPARSITY BASED ANOMALY DETECTION

Sparsity is a powerful prior in this model for multiple reasons:

- When a new collection of trajectories manifests, it is expected to invoke only a few columns of the training dictionary that combine to create it.

- Object-wise correspondence is important in the linear combination for this model to physically meaningful leading to a block-diagonal sparse structure on coefficients.
- The sparse structure conveys information about normal/anomalous event classes in the absence of training data for anomalous events develop and measures on the sparse coefficient matrix that can help with object anomaly detection in unsupervised settings.

Sparse approximation is a sparse vector that solves approximately a system of equations and this is used for finding sparse approximations which widely used in many applications as image processing, audio processing, document analysis etc.

Sparse representation usually offers better performance with its capacity for efficient signal modelling. Research has focused on three aspects of the sparse representation: pursuit methods for solving the optimization problem.

If we have a very large matrix with few non-zero values, it doesn't store the zero values. It can make operations very expensive when the dimensions are large. We can take advantage of having less information by representing your matrix in a way that only uses the non-zero entries. The matrix is converted into a table of index-value pairs. The index (m,n) is two integers representing the row and column, and the value is the entry for that element. We can see how this will save memory, at the very least, if our matrix is huge but only has a few (non-zero) entries. The sparse matrix would be small in memory, while the actual matrix would be huge for storing all the zeros. Sparse matrix multiplication and other operations can also be sped up with this representation.

The concept of sparsity is useful in application areas such as network theory, which have a low density of significant data or connections.

Abnormal behavior detection via sparse reconstruction [4] analysis of trajectory is a recent idea in the field of video anomaly detection. The fundamental underlying assumption is that any new trajectory can approximately be modeled as a (sparse) linear combination of training trajectories.
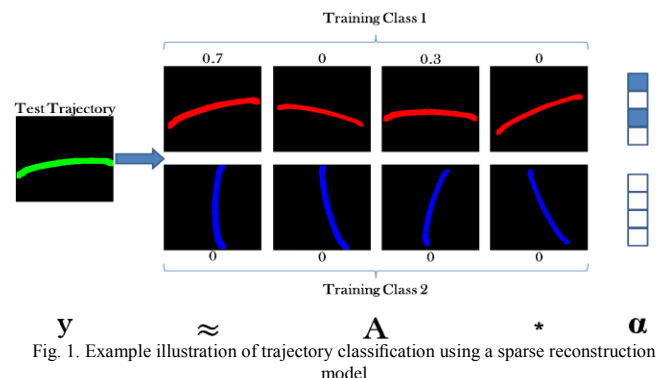

$$\mathbf{y} \approx \mathbf{A} * \boldsymbol{\alpha}$$

Fig. 1. Example illustration of trajectory classification using a sparse reconstruction model

Let each trajectory representation [4] lie in R^n, and T denote the number of training samples (i.e., example trajectory representations) from each of K different classes, that is,

behavior patterns in a video that may be normal or anomalous. The T training samples (trajectory representations) from the ith class are arranged as the columns of a matrix $A_i \in R^{n \times T}$. The dictionary A$\in R^{n \times KT}$ of training samples from all classes is then formed as $= [A_1 A_2 \ldots A_K]$. Given a sufficient number of training samples from the mth trajectory class, a test image y $\in R^n$ from the same class is conjectured to approximately lie in the linear span of those training samples. Any trajectory feature vector is synthesized by a linear combination of the set of all training trajectory samples as

$$y \approx A\alpha = [A_1 A_2 \ldots A_K]\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_K \end{bmatrix}$$

where each $\alpha_i \in R^T$. Typically, for an example trajectory y, only one of the α_i's will be active (corresponding to the class/event from which y is generated).

Advantages of sparse representations

When building a representation of a sentence, or an image, it is important to have a form of distributed representation. On the one hand, there are so many different combinations of scenes, or sentences that we simply cannot use cluster based representations. On the other hand, a very dense distributed representation can be difficult to learn. Our representation must mimic the topology of the underlying manifold, and the denser our representation, the less degrees of freedom we have when building our map, and the more non-linear the relationship. A sparse representation provides an intermediate form between a pure cluster based or one-hot representation, and between a purely distributed representation.

Disadvantage

Sparsity based anomaly detection is applicable for single object tracking approach.

Then, the joint sparsity model for video anomaly detection has been developed for multiple object tracjectories.

IV. JOINT SPARSITY MODEL

The sparsity-based approach is useful, but does not find interactions the two or more object anomalies. A new joint sparsity model Xuan Mo [4] for video anomaly detection that gives multiple object trajectories.

The detection of anomalies involving P ≥ 1 objects. Their corresponding P trajectories can be represented as a matrix:

$$Y = [y_1 \, y_2 \ldots y_P] \in R^{n \times P}$$

, where yi correspond to the ith trajectory. The training dictionary can be defined as $A = [A_1 A_2 \ldots A_P] \in R^{n \times PKT}$ where each dictionary $A_i = [A_{i,1} A_{i,2} \ldots A_{i,K}] \in R^{n \times KT}$, $i = 1,2,\ldots P$

, is formed by the concatenation of the sub-dictionaries from all classes belonging to the ith trajectory. The crucial aspect of this formulation is that the training trajectories for any class j, that is, $A_{i,j}$, i=1 ,2,..P, are observed jointly from example videos. The test P trajectories can now be represented as a linear combination of training samples as

$$Y \approx AS$$
$$= [A_{i,1} A_{i,2} \ldots A_{i,K} \ldots \ldots A_{P,1} A_{P,2} \ldots A_{P,K}][\alpha_1 \ldots \alpha_P]$$

where, the coefficient vectors α_i lie in R^{PKT} and S = $[\alpha_1 \ldots \alpha_1 \ldots \alpha_P]$.

V. APPLICATIONS AND FUTURE WORK

Vast amounts of video footage are collected and analyzed for traffic violations, accidents, crime, terrorism, vandalism, and other suspicious activities. Since manual analysis of such large volumes of data is prohibitively costly, there is a desire to develop effective algorithms that can aid in the automatic or semiautomatic interpretation and analysis of video data for surveillance and law enforcement. The problem of finding patterns in data that do not conform to expected behavior and may warrant special attention or action. Two precursors to anomaly detection are an effective encoding of events, and a systematic means of modelling normal events and normal event classes.

VI. CONCLUSION

The study is on the existing methods for anomaly detection. This survey has discussed the different techniques in which the anomaly detection problem has been formulated in related work. The discussion is done on the sparse representation for anomaly detection in image processing but sparsity model cannot be sufficiently optimized for the multiple object anomaly detection then the revised joint sparsity model has been developed to overcome this demerit. I hope to provide the better samples of various techniques in the field of video anomaly detection.

REFERENCES

[1] Unsupervised Activity Perception in Crowded and Complicated Scenes Using Hierarchical Bayesian Models, IEEE transactions on pattern analysis and machine intelligence, vol. 31, no. 3, march 2009

[2] Framework for Real-Time Behavior Interpretation From Traffic Video Pankaj Kumar*, Member, IEEE*, Surendra Ranganath, Huang Weimin, and Kuntal Sengupta IEEE transactions on intelligent transportation systems, vol. 6, no. 1, march 2005

[3] Online Detection of Unusual Events in Videos via Dynamic Sparse Coding Bin Zhao School of Computer Science Carnegie Mellon University, Li Fei-Fei Computer Science Department Stanford University, Eric P. Xing School of Computer Science Carnegie Mellon University

[4] Adaptive Sparse Representations for Video Anomaly Detection, Xuan Mo, *Student Member, IEEE,* Vishal Monga, *Senior Member, IEEE,*, Raja Bala, *Member, IEEE,* and Zhigang Fan, *Senior Member, IEEE,* IEEE transactions on circuits and systems for video technology, vol. 24, no. 4, april 2014