

Anomalous Events Detection in Frequent Sequence Video via Object Segmentation Using Motion Pixel Compensation and Gradient

S. Tamilselvan¹, B. Monika²

¹Assistant Professor, Department of CSE, Arulmigu Meenakshi Amman College of Engineering, Vadamavandal

²ME (CSE) – Final Year, Arulmigu Meenakshi Amman College of Engineering, Vadamavandal

Abstract: *The surveillance takes major part in the present world. The collection of data through surveillance is not the thing required by the people but to use it effectively for decision support. Manual processing of huge amount of surveillance data is not feasible; a support to analyse and process all incoming data is needed by the operator of the surveillance system. In this proposal, an approach to intelligent video surveillance is introduced with significance on finding behavioural anomalies. The proposed technique of robust and efficient anomaly detection is capable of dealing with crowded scenes. Initially, the foreground segmentation of input frames is done to analyse the foreground objects and to effectively ignore irrelevant background dynamics. The input frames are split into non overlapping cells, and then the features are extracted based on motion, size and texture from each cell. Each and every feature type is independently analysed to detect the presence of an anomaly. An effective estimate of object motion is obtained by computing the optical flow of only the foreground pixels. To process even large training datasets, the motion and size feature are modelled by an approximated version of kernel density estimation.*

Keywords: Gradient image, Motion estimation, Object detection, Segmentation, Tracking

1. Introduction

The anomaly detection techniques for Automated Teller Machine (ATM) surveillance are considered in this paper. ATM transactions are quick and convenient, but the machines and the surrounding areas can be susceptible to criminal activity if not properly protected. By using the algorithm of anomaly detection in a surveillance video we can depict the criminal activity at the earliest and provide early actions to the areas of view. Fig. 1 shows the working steps of this algorithm. The criminal activity can be intimated to the areas of view by an alert.

ATM surveillance videos capture the behavioural activities of the objects accessing the ATM system. Some of the behaviours are frequent sequence of events and some deviate from the known frequent sequences of events. These events are termed as anomalies and may be susceptible to criminal activities. In the past, work was based on discovering the known abnormal events [1]. Here, the unknown abnormal activities are to be detected and alerted such that early actions can be taken.

One of the challenging tasks in video processing is the process of detecting and tracking moving objects in a video scene. This requires the detection of all objects present in a video scene so that the moving objects can be isolated from them. Object detection can be done by clustering similar pixels based on their intensity gradient. In a given video, moving object detection tends to identify different moving object regions with respect to the background. Especially, moving object detection in a video frame is the process of detecting those objects whose movements will create an immediate and dynamic variation in the scene. This can be achieved by the following two different ways: 1) motion

detection and 2) motion estimation.

Moving object detection or motion detection is the process of identifying changing and unchanging regions from the image frames extracted from the given video. This will be done effectively with the help of thresholding [2]. Here the moving object detection is done when the camera is fixed and the objects are in motion. For motion estimation, motion vectors of the object regions are computed to estimate the positions of the moving objects from one frame to another frame. This can be done with the help of block matching algorithm [3], [4]. In a video scene, both the objects and the camera may move. After identifying the moving objects from the extracted image frames, it is needed to track them. By tracking the moving object in a video sequence, it is easy to find the position of object at different time intervals [5].

Moving object detection is restricted by the use of a reference frame where the object is not present. This can be done based on moving object detection algorithm by using the intensity difference. In the absence of reference frame, if there is a large amount of movement of an object exists from frame to frame, then it can be tracked completely by generating a reference frame [6]. However, when the reference frame is not available and the object do not have large amount of movement from one frame to the subsequent frame, or the object move and stop for some amount of time and move again, the identification of moving object becomes difficult. This difficulty can be overcome by segmenting the regions based on spatial and temporal aspects [7].

A robust video image segmentation algorithm is essential in order to effectively identify moving objects. Coarse-fine segmentation with respect to spatial and temporal aspects [8] provides an efficient way to segment the images. By using

the segmented images background dynamics can be eliminated and the moving objects in the foreground can be identified. The normal events that will occur on an ATM

have trained and several features of the moving object will be extracted to compare with the trained activities [9], [10].

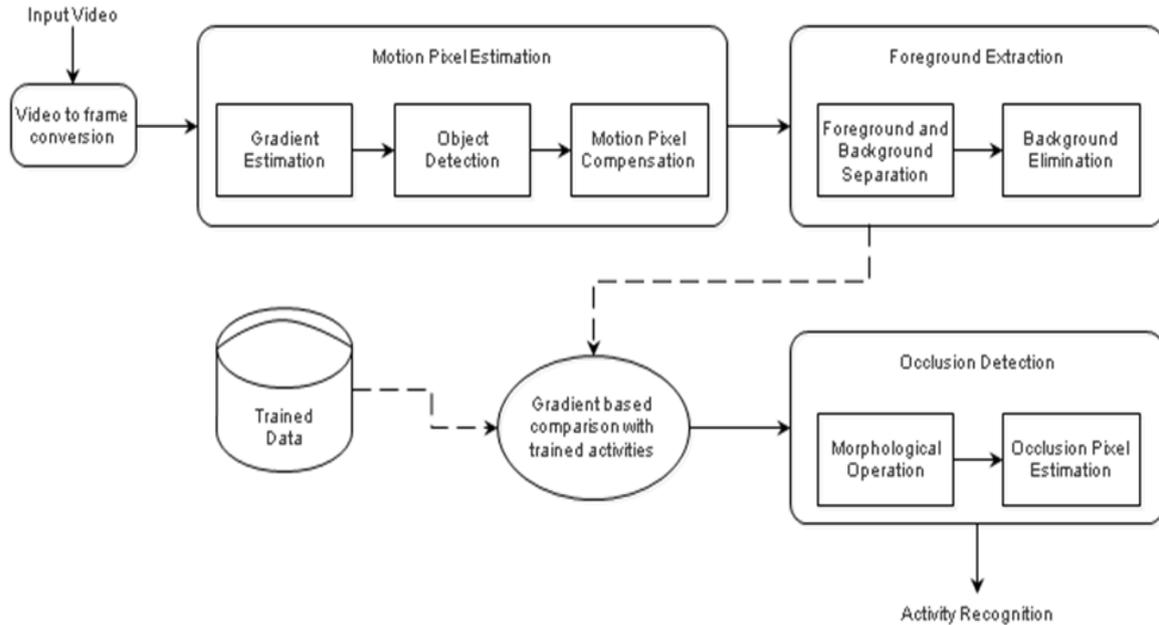


Figure 1: Architecture Diagram

Comparison of the features of detected moving objects with the trained features discovers the presence of occlusion in the normal sequence of events and intimates it with an alert.

2. Object Detection

In this module, the given video will be converted into frames. In each frame, the pixels are clustered using k-means clustering algorithm. Then the images are transformed to frequency domain using DTCWT, from which intensity gradient image can be obtained using Gaussian gradient estimation. Based on the gradient, all object regions present in the area of view can be detected. Using motion compensation, only a set of frames which are not similar to one another are extracted.

A. K-means Clustering Algorithm

The K-means clustering algorithm partitions an image frame into K clusters. It is an iterative technique. The algorithm works through the following steps:

1. Choose K cluster centres randomly.
2. Compute similarity between the cluster centre and the neighbouring pixel.
3. Add the pixels which has no or minimum difference with the cluster centre into the cluster.
4. After merging a pixel into the cluster, recompute the cluster centre by taking the average of all the pixels in the cluster.
5. Iterate from step 2 to 4 until there is no overlap between the clusters.

Here the difference between the cluster centre and the neighbouring pixels is computed based on intensity and colour. The quality of this clustering algorithm is highly depends on the K value and the set of clusters partitioned initially.

B. Dual-Tree Complex Wavelet Transform (DTCWT)

The sparse representation of the image is provided by the DTCWT. It helps to focus quickly on object regions where the objects may found. Thus the additional features can be computed using this Complex Wavelet Transform for those object regions only. The detection of smaller objects can be accurately done through these additional features. The complex transform of a signal can be calculated by DTCWT. This is achieved by two Discrete Wavelet Transform decompositions. The real and imaginary coefficients are obtained from the Discrete Wavelet Transform.

C. Gaussian Gradient Estimation

The gradient images are obtained from the given video frame using Gaussian gradient estimation. The gradient image is created by measuring the change in intensity in the original video frame. Each pixel in the gradient image computes the intensity change of the same point in the given video frame in a particular direction.

The gradient of image is obtained by the formula,

$$\Delta f = \frac{\partial f}{\partial x} \hat{x} + \frac{\partial f}{\partial y} \hat{y} \quad (1)$$

where,

$\frac{\partial f}{\partial x}$ is the gradient in x- direction

$\frac{\partial f}{\partial y}$ is the gradient in y- direction

The gradient direction is obtained by the formula,

$$\theta = \text{atan2} \left(\frac{\partial f}{\partial y}, \frac{\partial f}{\partial x} \right) \quad (2)$$

3. Moving Object Detection

Each extracted frames are correlated based on spatial and temporal aspects. Those correlated frames can be segmented based on coarse-fine segmentation which follows region

growing algorithm where similar pixels are identified and grouped as a cluster which may segment the textures and moving objects separately.

Using canny edge detection algorithm, the edges of the object in each frame can be identified. By the segmented and edge detected frames, the moving object can be identified.

a) Coarse-Fine Segmentation

The Coarse-Fine segmentation follows region-growing algorithm. Here the spatial and temporal aspects are considered during segmentation. This segmentation starts with a single region R_1 and iterate through the following steps:

1. Choose an unallocated neighboring pixel.
2. Measure of intensity is done through the difference between the intensity value of the pixel and the mean δ of the region.
3. If $\min(\delta)$ is less than a predefined threshold T , then the pixel is added to the corresponding region R_i .
4. If the above condition fails, the pixel is said to be different from all the current regions. Thus a new region R_{n+1} is created with this pixel.

The above steps iterate as long as all the pixels are assigned to a region.

b) Canny Edge Detection Algorithm

The Canny edge detection algorithm is a multi stage algorithm which is used to detect wide range of edges in a given image.

Canny edge detection algorithm consists of the following steps

1. This step convolves the raw image with a Gaussian filter which results in the blurred version of the original image that is free of any noisy pixel.
2. The gradients' intensity and direction is obtained by applying the gradient operator.
3. The vertical, horizontal and diagonal edges in the blurred version of image can be detected by the use of filters for each direction.
4. The value for first derivative in both horizontal (G_x) and vertical (G_y) direction is returned by the edge detection operator.
5. From the above returned value, the gradient and the direction of edge can be determined.

$$G = \sqrt{G_x^2 + G_y^2} \quad (3)$$

$$\theta = \text{atan2}(G_y, G_x) \quad (4)$$

where G is the edge gradient and θ is the edge direction angle.

6. For each pixel, if its magnitude is less than the two neighbours in the gradient direction, then the non-maximum suppression suppresses the pixels' edge strength by setting its value to zero. This step determines a better candidate pixel for the edge. The beginning and ending of edges are determined through hysteresis thresholding. It starts by

applying high threshold which helps to mark out the edges which are perfect. Finally, a binary image is obtained where each pixel is marked either as an edge pixel or a non-edge pixel.

4. Motion Estimation

Kalman filter and BMA are used to estimate the motion of object. Using kalman filter, the object tracking is performed by predicting the object's position from the previous information and verifying the existence of the object at the predicted position.

The BMA uses the motion vector of the macro block to its immediate left to predict its own motion vector. Based on the estimated motion which occurs between the current frame and the reference frame, motion vector for each block will be calculated. From the calculated motion vector, the motion pixels are identified and the unwanted background dynamics are eliminated.

a) Kalman Filter

The kalman filter is used to estimate the position of object using the past and present information about the object position.

The tracking of moving object is done by kalman filter. Initially, the frame number from which the object tracking needs to be started should be provided. On the selected frame, the position of mask has to be set on the object to be tracked and it can be tracked on subsequent frames.

The following steps describes the tracking of single object

- 1) Finding all moving regions in the video frame through background subtraction.
- 2) Once the region is found, the region which has lowest distance from the region selected in previous video frame is selected.
- 3) The time and measurement update equations have been calculated using the centroid and other parameters of the selected region.
- 4) The object position values X obtained for every frame is stored in an array.
- 5) By drawing a line which joins each stored point in every frame, the trajectory of the selected object has been found out.
- 6) Kalman filter falls into two groups such as time update equation and measurement update equation. Here the time update equation can be said as predictor equation which projects the current position estimate ahead in time whereas the measurement update equation can be said as corrector equation which adjusts the estimated position by an actual measurement at that time.

b) Block Matching Algorithm (BMA)

The BMA is used to identify the matching block in a sequence of video frames for the estimation of motion. From the motion vector of a macro block which is located at the left of the desired block, its own motion vector can be predicted. Here the rood pattern search is used which directly starts search from the region that has high probability of finding a good matching block.

The BMA projects motion pixel as the output through the estimated motion vector. Projected motion pixels will be further processed in order to separate the foreground and background and to effectively compare with the training data.

5. Occlusion Detection

The frequent action images are trained and stored as a trained data set. The features extracted from the motion pixel estimated frames are compared against the trained dataset. Here the comparison is done based on the gradient. When there is a mismatch occurs, the frame undergoes morphological operation to eliminate noise and to increase the pixel quality. Thus the presence of occlusion is recognized by those pixels and an alert is made to intimate the anomalous event which occurs in the area of view.

6. Conclusion and Future Work

Abnormal events in video surveillance scenarios can be concluded as something that seldom occurs, or something that is previously unseen. There are many different types of events that can be described as abnormal, but the common description is that these events deviate from most of the action that occur in a scene. To be able to model abnormal events, the work had start off by performing a visual subjective analysis on a video training set. The work had to be able to get the understanding of what was actually happening in the video sequence before, during and after the event of object dropping. By examining several videos with the same type of event, it was possible to define a set of criteria's for the event. The set of criteria's was set to act as threshold values saying that an object dropping had occurred if all these predefined values were fulfilled. Several features from the objects in their entire presence in the scene had been extracted. A combination of these low-level features did contain sufficient information to make a description on what characterizes the abnormal event of object dropping. Those features have been compared with the trained data and the existence of abnormal event has been encountered. The future work is to establish an operational computer vision system for online detection and tracking of moving objects to effectively encounter anomalous events.

References

- [1] W. Lin, MT. Sun, R. Poovandran, and Z. Zhang, "Human Activity Recognition for Video Surveillance", in IEEE International Symposium on Circuits and Systems, 2008, ISCAS 2008
- [2] L. Senni, M. Ricci, A. Palazzi, P. Burrascano, P. Pennisi, and F. Ghirelli, "On-line automatic detection of foreign bodies in biscuits by infrared thermography and image processing", Journal of Food Engineering, 2014 – Elsevier
- [3] CH. Yeh, LW. Kang, YW. Chiou, CW. Lin, and SJF. Jiang, "Self-learning-based post-processing for image/video deblocking", Journal of Visual Communication and Image Representation, 2014 - Elsevier
- [4] LM. Wang, Y. Qiao, and X. Tang, "Motionlets: Mid-Level 3D Parts for Human Motion Recognition", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2013
- [5] S. Huang, and J. Hong, "Moving Object Tracking System Based On Camshift And Kalman Filter", in IEEE International Conference on Consumer Electronics, Communications and Networks (CECNet), April 2011
- [6] L. Wang, L. Cheng, and L. Wang, "Elastic Sequence correlation for Human Action Analysis", IEEE Transactions on Image Processing, Vol. 20, No. 6, June 2011
- [7] D. Metaxas, and S. Zhang, "A review of motion analysis methods for human Nonverbal Communication Computing", Image and Vision Computing, 2013 - Elsevier
- [8] BN. Subudhi, PK. Nanda, and A. Ghosh, "A Change Information Based Fast Algorithm for Video Object Detection and Tracking", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 21, No. 7, July 2011
- [9] N. Bhuptani, and B. Talati, "Variations in Shape Context Descriptor: A survey", in International Journal of Computer Applications (0975 – 8887) Volume 90 – No 12, March 2014
- [10] H. Jiang, MS. Drew, and ZN. Li, "Action Detection in Cluttered Video with Successive Convex Matching", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 20, No. 1, January 2010