Human Anomalous Activity Detection: Shape and Motion Approach in Crowded Scenes

B.Prabha*, N.R.Shanker†, M.Priya‡ and E.Ganesh§

*Department of Computer Science and Engineering, Koneru Lakshmaiah Educational Foundation, Vaddeswaram, India.
†Department of Computer Science and Engineering, Aalim Muhammed Salegh College of Engineering Chennai, India.
‡Department of Physics, Saveetha Engineering College, Chennai, India.
§Department of Electronics and Communication Engineering, Saveetha Engineering College, Chennai, India.

E-mail: jemi.prabha@gmail.com, nr_phd@yahoo.co.in, priyam7373@gmail.com, eganesh90@gmail.com

Abstract: Detecting anomalous activities in crowded scenes is a very challenging task in computer vision. An enhanced video anomaly detection framework is proposed for frame-wise anomalous activity detection in crowded scenes that is based on both shape and motion based features. The Histogram of Oriented Gradients (HOG) is used to represent the shape based features of the video frames and for representing the motion, Histogram of Oriented Optical Flow (HOOF) is used. These features are modeled using two-class Support Vector Machines (SVM) to detect abnormal events in every frame. The proposed method is modeled with both normal and abnormal behaviors which are learnt from the training data and it is capable of detecting abnormal activities in a live surveillance video. To evaluate the performance of the proposed work, experiments are conducted on the standard benchmark UCSD data set and the results are compared with the HOOF feature bin values.

Keywords: Anomalous activity detection, crowded scenes, video surveillance, Histogram of Oriented Gradients, Optical Flow, computer vision

1. INTRODUCTION

Intelligent video systems (IVS) and video analytics (IVA) are challenging and complex areas that have received a growing demand in recent years. The aim of developing IVS is to detect, track, classify and recognize the events and then the system will generate alarm when any abnormal activity happens. Computer vision based intelligent system concentrates more on human behavior analysis such as activity recognition, anomalous pattern discovery, motion tracking and behavior understanding. Applications are not only limited for video surveillance [1, 3], but also includes intelligent vehicles [3], human-computer interfaces [3], fraud detection [3, 4, 5], and multimedia semantic annotation and indexing [8].

Anomalous activity detection in video has become a significant research aspect in the video surveillance due to the rapid increase in security needs. Some widely used targets include individual, crowd and traffic monitoring. In an intelligent surveillance system, the human activities are monitored and the system will generate alert when any anomalous activity happens. The system is modeled to learn patterns of normal/abnormal activities using methods of human activity

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Published under licence by IOP Publishing Ltd
recognition. An activity that differs from normal patterns is identified as anomaly. Oluwatoyin et al. [1], Honghai et al. [3], Manish et al. [4] and Chandola et al. [5] provide an extensive overview of anomaly detection techniques in various applications. Some of the challenges involved in identifying an abnormal activity in the crowded scenes scenario are listed below:

1. Detection of anomalous activities in the dense crowded scenes is very difficult due to the speed variations, deviations in direction and occlusion.
2. Defining the boundaries of normal/abnormal activity in the crowded scenes cannot be precise.
3. An event may be considered as normal in one perspective while abnormal in another. Therefore model should incorporate context sensitive information.
4. Insufficient samples of abnormal events will affect the robustness of the anomaly detection system.

As observed, anomalous activity detection in crowded scenes can be really challenging. This has motivated us to develop an anomaly detection system to detect the abnormal behaviors in the crowded scenes. Most of the existing methods in the literature on anomaly detection have concentrated on either motion based or shape based approach. In [9,10] motion based features are represented using the trajectory information of moving objects such as speed, direction etc. These features are more suitable in case of less cluttered scenes with only a few objects, but in crowded environments it is very difficult to achieve better tracking due to the presence of occlusions. Whereas the shape based features [7, 11, 20] rely mainly on local features such as color, size and texture of the objects in the frame. These features can obtain better results where the scene is comprised of few objects and less motion, but in dense crowd the motion information are very much essential for accurate anomalous activity detection. To address these limitations, the proposed method presents an enhanced anomalous activity detection system which combines both motion and shape based features. The proposed method which employs robustness in abnormal activity identification on crowded pedestrian scenes where traditional shape and tracking approaches tend to fail.

In order to capture the shape based abnormal activity in the surveillance video, HOG [12,15] and SFTA [13] are used. HOG is a feature vector which captures local of a pedestrian/human; it can be represented by distribution of edge or gradient directions. SFTA extracts the texture of objects in the frame to describe segmented texture patterns. To estimate the motion: Horn and Schunck (HS) [14] method is used to compute optical flow and to calculate histogram of optical flow, the HOOF [16] function is used. Finally, these feature representations are combined together to provide a powerful framework to model normal and abnormal activities in the surveillance video.

The overview of this paper is as follows: Section 2 reviews the related work on anomalous activity detection in crowded scenes. In Section 3 the proposed system is described in detail. Section 4 evaluates the performance and comparison of this system and Section 5 concludes the paper with important observations of anomaly detection in crowded scenes.

2. RELATED WORK

This section review related works on abnormal event detection in the crowded scenes which are reported in [11, 15, 16, 17, 18, 20, 21]. Many of the existing work focused either on motion or shape based features. For example, in [17], a hierarchical clustering method is used to group salient points based on a set of features which include position and optical flow direction. For each cluster, the main type of motion pattern is represented by a dominant force. Abnormal crowd events are detected when the crowd direction is not consistent with the predicted one. A limitation of this method is that only direction is taken into account, such that the changes in crowd speed cannot be detected. Mehran et al. [18] proposed the social force model to detect and localize crowd behavior. This model computes optical flow of interactions of individuals.
Kim et al. [21] introduced a space-time Markov random field (MRF) model to detect abnormal activities in a local and global context. In this approach, optical flow were extracted at each frame and then used a mixture of probabilistic principal component analyzers (MPPCA) to identify the typical patterns, and construct a space-time MRF to enable inference at each local site. Weixin et al. [20] proposed an anomaly detector that accounts for both scene shape and dynamics, spatial and temporal context, and multiple spatial scales using a set of mixture of dynamic texture models [19] which detects the anomalous behaviors (vehicles or group of people moving in different directions) in crowded pedestrian walkways. Tian et al. [16] introduced a histogram of oriented optical flow as a descriptor, to detect abnormal events in video by using SVM classifier. In [15] Yingying et al. presented a framework to jointly model related activities with both motion and context information for activity recognition and anomaly detection. This work used the Spatio-temporal interest point (STIP) detector developed to generate concatenated HOG and histogram of optical flow (HOF) features for motion regions surrounding the bounding boxes of the moving objects.

Vikas et al. [11] presented a framework to detecting anomalies (bikers, skateboarders etc) based on size, texture and motion. This approach detects speed and out-of-place objects, but it does not detect direction based anomalies. Depending on either shape or motion alone for anomalous activity detection is not sufficient, because some anomalies may exhibit as normal in one perspective but abnormal in another. Furthermore, motion may affect from the aperture problem. In order to increase the sensitivity of anomaly detection, combine the shape features with motion features will enhance the performance of the system.

3. OVERVIEW OF THE PROPOSED WORK

The main contribution of this work is to show, how the combination of both shape and motion based feature representations can be exploited for anomalous activity detection in video surveillance. The discriminative features that captures the shape based and the motion based abnormal activities are also focused in this work.

![Fig.1. Framework of proposed anomaly detection system](image-url)
The proposed work has the following components: Feature extraction, Modeling events and classifying the frames as normal or abnormal.

3.1. Feature Extraction
The proposed anomaly detection framework is shown in Figure 1. In this, the input video is split into frames. For each frame, extract shape based features such as HOG and SFTA and motion based feature HOOF. The details of these features are given below.

3.1.1. Histogram of Oriented Gradients
HOG is used to detect the local object shape within a frame that can be expressed by the distribution of edge direction and intensity gradients. These detectors can be achieved by partitioning the image into 16x16 blocks and each block consists of 2x2 cells with size of 8x8. For each cell, collect a histogram of edge orientations/gradient direction. The Gaussian window size sigma value is set to 8. The following processes are involved in the HOG feature vector computation.

1. Gradient computation with centered derivative horizontal and vertical directions.
2. Detection of each window size is 64x128.
3. Quantize the gradient orientation using 9 bins.

Combinations of these values represent the feature descriptor. Finally the histogram values are normalized to achieve the better pedestrian detection and invariance to changes in illumination or shadowing.

3.1.2. Segmentation-based Fractal Texture Analysis
The HOG feature for anomaly detection might be insufficient as certain anomalies may be considered as normal due to slow motion of vehicles etc. To address this issue, the HOG along with the texture feature is extracted from the frame. This in turn increases the accuracy of the anomaly detection in the crowded scene. This texture decomposes the frames into various frames with thresholds, using several sets of lower and upper threshold values. The threshold segmentation which was applied using the following representation

\[ I_b(x, y) = \begin{cases} 
1: & t_l < I(x, y) < t_u \\
0: & \text{otherwise}
\end{cases} \]

where \( I_b(x, y) \) is the \( x, y \) co-ordinates of an image, \( t_l \) and \( t_u \)are lower and upper threshold values. By using this threshold, the feature vector which is corresponds to the texture information is extracted from the frames in order to achieve the robust anomaly detection.

3.1.3. Histogram of Oriented Optical Flow
SFTA and HOG features relies on shape information, but anomalies occurring due to motion may not be detected this will reduce the accuracy of modeling. To address this limitation, combine motion with HOG and SFTA features. In [14] proposed an algorithm computing the optical flow by introducing a global constraint of smoothness. The Horn-Schunck optical flow method is applied to acquire lowlevel features at each pixel of training frames. Adopt this optical method combining a data term with a spatial term. The data term assumes constancy of the same image property, and the expected flow variation is modeled by the spatial term. The optical flow is formulated to minimize the following global energy functional

\[ E = \int \left[ \left( L_{xx} + L_{yy} + L_t^2 \right) + \alpha^2 \left( \nabla^2 \mathbf{v} \right) \right] \, dx\,dy \]

Where \( L_x , L_y \) and \( L_t \) are the derivatives of the image intensity along the x, y and time t dimension, \( u \) and \( v \) are the horizontal and vertical directions of the optical flow, \( \alpha \) is the parameter representing the weight of the regularization term. The original OF is not useful when the number of pixels in an object is changes overtime. In addition, of computations must have noisy background data, changes in
scale and direction. To avoid such issues, adopt HOOF of the training frames. There is no prior information is required to extract this feature. Initially, OF is calculated at each frame in the video. Every HOOF flow vector is binned with angle and magnitude.

\[-\frac{\pi}{2} + \pi \left( \frac{b-1}{B} \right) \leq \theta \leq -\frac{\pi}{2} + \pi \left( \frac{b}{B} \right)\]

Where \( b \) is a sum in bin, \( b \) lies between \( 1 \) to \( B \) \((0 \leq b \leq B)\), \( B \) is a out of a total of \( B \) bins and \( \theta = \tan^{-1}\left(\frac{y}{x}\right) \).

Finally, the histogram values are sum of 1. The advantage of HOOF is a scale invariant and also direction is independent.

3.2. Modeling events using SVM classifier

Features extracted from the training frames alone are considered for the modeling process. As described in the section 3.1 there are three features types: HOG, SFTA and HOOF. Each frame is sequentially checked to detect, whether the frame is anomalous or not. For modeling, the shape and motion based features are extracted and the frames that can be classified into normal or abnormal by using SVM classifier. SVM is a supervised learning method, which learns the training examples of both normal and abnormal classes. Then the labels will be assigned to all the training frames. Given a set of testing data, model will classify the frame as normal or abnormal.

(a) Cyclist riding Slowly (Original Frame)

(b) Detected anomaly is marked in the red box.
4. EXPERIMENTAL STUDIES

In order to assess the performance of the proposed approach, experiments are conducted on crowd anomaly detection dataset UCSD. The dataset were split into two subsets which contains different video scenes (Ped1 and Ped2), both with anomalies including vehicles, skateboarders, bikers and also people walking on the lawn. The Ped1 has 34 training and 36 test image sequences. To achieve the better results, experiments are performed only for frame level anomaly detection, with the samples of (normal = 150 training, 50 test images and for abnormal = 112 training, 38 test images) Ped1 training and test image sequences. Examples of sampled Ped1 test sequences are shown in Figure 2.

| Bin Size | Ped1 (Frame-level and EER) |
|----------|-----------------------------|
| Bin = 4, α = 1, Iteration = 3 | 32.5%  |
| Bin = 8, α = 1, Iteration = 3 | 34.5%  |
| Bin = 50, α = 1, Iteration = 3 | 28%  |
| Bin = 70, α = 1, Iteration = 3 | 48.3%  |
| Bin = 30, α = 1, Iteration = 3 | 16%  |

From Table 1. It can be observed that the proposed approach outperforms when the size of the bin value is changed in the HOOF feature. The proposed approach has the ability to detect anomalies (like cyclist) present in the frame and it is also absolutely detects the missing anomaly which is reported in [9], the cyclist was riding slowly by matching his speed with the neighbouring pedestrian. The selection of bin size for HOOF is very crucial. Since the proposed approach is based on the accurate shape (HOG and SFTA) and motion (HOOF) detection features, the rate of detecting anomalies happens in a faster manner.
5. CONCLUSION

In proposed method that combines both shape and motion based approach to enhance the anomalous activity detection in crowded scenes. These features are modelled by using the two-class SVM, which is computationally efficient even on large datasets. Experiments on publicly available dataset UCSD shows that the proposed approach attained better results with low complexity, when varying the bin size in HOOF feature (EER=16.2 %).

REFERENCES

[1] P.Oluwatoyin, Popoola, Kejun Wan, “Video-Based Abnormal Human Behavior Recognition-A Review”, IEEE Transactions on Systems, Man and cybernetics, Vol.42, No.6, pp.865, 2012.
[2] Paulo Vinicius Koerich Borges, Nicola Conci and Andrea Cavallaro,” Video-Based Human Behavior Understanding : A Survey”, IEEE Transactions on Circuits and Systems for Video Technology, Vol.23, No.1, pp.199,2013.
[3] Honghai Liu, Shengyong Chenand Naoyuki Kubota, “Intelligent Video Systems and Analytics: A Survey”, IEEE Transactions on Industrial Informatics, Vol.9, No.3, pp.1222,2013.
[4] Manish Gupta, Jing Gao, C.Charu, Aggarwal and Jiawei Han, “Outlier Detection for Temporal Data: A Survey”, IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No.1, pp.1, 2014.
[5] V.Chandola, A.Banerjee and V.Kumar, “Anomaly detection: A survey”, ACM Computing Surveys, Vol.41, No.3, pp.15, 2009.
[6] Weiming Hu, Tieniu Tan, Liang Wang, and Steve Maybank, “A survey on visual surveillance of object motion and behaviors”, IEEE Transactions on Systems, Man and Cybernetics, Vol. 34, No.3, pp. 334, 2004
[7] Weixin Li, Vijay Mahadevan and Nuno Vasconcelos, “Anomaly Detection and Localization in Crowded Scenes”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.36, No.1, pp. 18, 2014.
[8] Weiming Hu, Nianhua Xie, Li Li, Xianglin Zeng, and Stephen Maybank, “A Survey on Visual Content-Based Video Indexing and Retrieval”, IEEE Transactions on System, Man and Cybernetics, Vol.41, No.6, pp. 797, 2011.
[9] F. Bashir, A. Khokhar, and D. Schonfeld, “Automatic object trajectory-based motion recognition using Gaussian mixture models,” in Proc. IEEE Conf. Multimedia and Expo, pp.1532, 2005.
[10] Faisal I. Bashir, Ashfaq A. Khokhar and Dan Schonfeld, “Object Trajectory-Based Activity Classification and Recognition using Hidden Markov Models”, IEEE Transactions on Image Processing, Vol.16, No.7, pp.1912, 2007.
[11] Vikas Reddy, Conrad Sanderson and Brian C.Lovell ,” Improved Anomaly detection in crowded scenes via cell-based Analysis of Foreground Speed, Size and Texture”, in Proc. IEEE Conf on Comput Vis and Pattern Recognition, pp.1446, 2009.
[12] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in Proc. IEEE Conf. Comput Vis. Pattern Recognition, pp. 886–893, 2005.
[13] Vivek Nair and Nikhil Parthasarathy “Supervised Learning Methods for Vision Based Road Detection” CS229 Final Project in standford University, pp.1-5, 2013.
[14] B. K. Horn and B. G. Schunck, “Determining optical flow,” Artif. Intell. vol. 17, no. 1, pp. 185–203, 1981.
[15] Yingying Zhu, Nandita M. Nayak, and Amit K. Roy-Chowdhury, “Context-Aware Activity Recognition and Anomaly Detection in Video”, IEEE Transactions on Signal Processing, Vol.7, No.1, pp.91, 2013.
[16] Rizwan Chaudhry, Avinash Ravichandran, Gregory Hager and Rene Vidal, “Histograms of Oriented Optical Flow and Binet-Cauchy Kernels on Nonlinear Dynamical Systems for the Recognition of Human Actions” in Proc. IEEE Conf. Comput. Vis. Pattern Recognition, pp. 1932-1939, 2005.
[17] Duan-Yu Chen, Po-Chung Huang, “Motion-based unusual event detection in human crowds,”
J. Vision Communication and Image Representation, Vol. 22, no. 2, pp. 178-186, 2011.

[18] Ramin Mehran, Alexis Oyama, Mubarak Shah, “Abnormal Crowd Behavior Detection using Social Force Model”, Proceedings in Conference on Computer Vision and Pattern Recognition, pp. 935, 2009.

[19] Antoni B. Chan and Nuno Vasconcelos, “Modeling, Clustering, and Segmenting Video with Mixtures of Dynamic Textures”, IEEE Transactions on Pattern Analysis and Machine Learning, Vol. 30, No. 5, pp. 909, 2008.

[20] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, “Anomaly Detection in Crowded Scenes”, Proceedings in Conference on Computer Vision and Pattern Recognition, pp. 1975, 2010.

[21] J. Kim and K. Grauman. “Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates”. In Computer Vision and Pattern Recognition (CVPR), pages 2921–2928, 2009.