ANOMALY DETECTION IN URBAN AREAS

Abdulrahman Alreshidi
University of Ha'il
Ha'il, Saudi Arabia

Hina Afridi, Wilayat Khan
COMSATS University
Wah Cantt, Pakistan

ABSTRACT

In this paper we propose a novel approach to detect anomalies in urban areas. This is achieved by analyzing the crowd behavior by extracting the local binary patterns (LBP) and Laplacian of Gaussian (LoG) features. We integrate both features together. These features are used to train an MLP neural network during the training stage, and the anomaly is inferred on the test samples. Considering the difficulty of tracking individuals in dense crowds due to multiple occlusions and clutter, in this work we extract LBP and LoG features and consider them as an approximate representation of the anomaly. These features well match the appearance of anomaly and their consistency, and accuracy is higher both in structured and unstructured urban areas compared to other detectors. In the current work, these features are used as input prior to train the neural network. The MLP neural network is subsequently used to highlight these features that can reveal the anomaly. The experimental evaluation is conducted on a set of benchmark video sequences commonly used for anomaly detection.

Keywords; Local binary pattern; Laplacian of Gaussian; Anomaly; Motion

INTRODUCTION

The gatherings of people at different events present challenges of paramount importance to public safety [1][2][3]. These situations are dangerous in presence of anomalies in terms of riots and chaotic acts of crowds in urban areas [4][5]. It is important to detect the occurrence of anomalies early since it can reduce the potential dangerous consequences and can also alert a human operator for monitoring the ongoing situation more effectively. However, the monitoring of video generally requires human operators to watch the screens, which often leads to fatigue, inattention, and failure to identify the occurrence of abnormal events [6][7]. On the other hand, significant challenges arise with the substantial amount of surveillance video data, which are difficult and time-consuming for manual analysis [8][9][10]. Considering these problems, an automated anomaly detection system gains increasing interest from both academia and industry.

Early research related to anomaly detection mainly focuses on specific tasks. For example, Rota et al. [25] used influence of particles to detect different groups, whereas Ullah et al. [11][12] employed a unified model to detect abnormal behavior. However, general video anomaly detection presents difficulty due to the unclear definition of anomaly in practical conditions. One may think that the presence of a vehicle on a pedestrian pathway is normal, but others may consider it an anomaly. In fact, an anomaly is an observation that does not maintain consistency with other observations over time. Considering the inconsistency between normal and abnormal patterns, one can design normal patterns in an unsupervised or semisupervised manner, and the pattern that deviates from the model is considered as anomaly. Significant research has been carried out on feature modeling for normal patterns. In rural areas where classic target tracking can be well achieved, high-level features, such as corner features, can be used for anomaly detection. However, in urban areas where not all targets can be accurately tracked, low-level features including histogram of oriented optical flow and gradients, social force models [3][13], and motion scale-invariant feature transform are robust for extraction, are often used to detect anomalous events in videos.

Despite the capability of these features for other solving other problems, they are easily influenced by the background. Thus, these featuers cannot focus on the object of interest, for example, anomaly. In this work, we present a unified model that integrate widely tested features for different computer vision problems. In the proposed work, we integrate Laplacian of Gaussian and local binary patterns. These features are extracted from individual images. Then we create a feature vector concatenated both the features. The feature vector is fed to MLP neural network that detects anomaly in challenging videos. The rest of the paper is categorized into related work, proposed method, experiments, and conclusion.

RELATED WORK

Some methods define the crowd scene videos as Spatio-temporal Volume (STV), which combines global video dynamics into a three-dimensional feature space. For example, [29][30] introduced a motion labelling method based on the co-occurrence of features. The model is designed as a potential function in the Markov Random Field (MRF) process for anomaly detection. The method in [31][32] introduced STV-based motion patterns in volumetric environment to represent the spatial-temporal statistical characteristics of pedestrians in crowded scenes. The methods [33][34][35][36] developed a STV-based anomaly location detection approach through using localised cuboids in an unsupervised learning model. These methods are valid approaches when considering spatio-temporal features without parameter settings. Some researchers [37][38] consider deep learning based methods for event detection.

PROPOSED METHOD

The Laplacian is a measure of the second spatial derivative of an image. The Laplacian of an image shows regions of rapid intensity change and is therefore often used for feature extraction. The Laplacian is generally used in connection with a Gaussian smoothing filter to smooth the image and to reduce its sensitivity to noise. The Laplacian of the Gaussian (LoG) is a very popular and robust technique for extracting features from the images. Given an input image f(x,y), this image is processed by a Gaussian kernel as formulated in Eq. (1).

$$G(x,y,t)=\frac{1}{2\pi t}e^{-\frac{x^2+y^2}{2t}} \qquad (1)$$

The result of LOG is obtained according to the formulation provided in Eq. (2).

$$L(x,y,t)=G(x,y,t)*f(x,y) \qquad (2)$$

The LOG sometime results in strong positive responses for dark areas of a specific radius and strong negative responses for bright areas of similar size. A significant challenge when applying LOG at a single scale is that the response is strongly dependent on the relationship between the size of the area structures in the image domain and the size of the Gaussian kernel used for pre-smoothing. In order to automatically capture anomalous areas of different sizes in the image domain, a multi-scale approach is therefore necessary.

A simple technique to obtain a multi-scale area detector with automatic scale selection is to use the scale-normalized Laplacian extractor and to detect scale-space maxima/minima, that are pixel points in accordance with both space and scale. Considering the given input image, we find scale-space volume. We nominate a point as a bright if the value at this point is greater than the value in all its neighbors. This method provides a concise operational definition, which directly leads to an efficient and robust method for feature extraction. The important characteristics from scale-space maxima of the normalized Laplacian technique are the covariant responses with translations, rotations and rescaling in the image under observation. It is worth noting that if a scale-space maximum is considered at a point under a rescaling of the image by a scale factor, there will be a scale-space maximum at in the rescaled image. This feature implies that besides the specific topic of Laplacian point detection, local maxima/minima of the scale-normalized Laplacian are also used for scale selection in other contexts, such as in corner detection, scale-adaptive feature tracking, scale-invariant feature transform as well as other image descriptors for image matching and object recognition.

To further enhance the robustness of the extracted feature, we concatenate local binary patterns (LBP) with LOG. The LBP is a type of visual descriptor used for classification in computer vision. It is a powerful feature for texture classification; we have further determined that when LBP is combined with the LOG, it improves the anomaly detection performance considerably. We model the LBP feature vector in a few simple steps. We divide the examined window of an image into cells and for each pixel in a cell, we compare the pixel to each of its 8 neighbors. We note 0 where the center pixel's value is greater than the neighbor's value, otherwise, we note 1. This provides an 8-digit binary number. We then compute the histogram, over the cell, of the frequency of each number occurring (i.e., each combination of which pixels are smaller and which are greater than the center). Thus, this histogram can be seen as a 256-dimensional feature vector. We normalize the histogram and concatenate histograms of all cells. This provides a feature vector for the entire window.

We reduce the length of the feature vector and code a simple rotation invariant descriptor since some binary patterns occur more commonly in anomaly images than others. A local binary pattern is called uniform if the binary pattern contains at most two 0-1 or 1-0 transitions. For example, 00010000 (2 transitions) is a uniform pattern, and 01010100(6 transitions) is not a uniform pattern. In the calculation of the histogram, the histogram has a separate bin for every uniform pattern, and we assign all non-uniform patterns to a single bin. Using uniform patterns, the length of the feature vector for a single cell reduces from 256 to 59. The 58 uniform binary patterns correspond to the integers 0, 1, 2, 3, 4, 6, 7, 8, 12, 14, 15, 16, 24, 28, 30, 31, 32, 48, 56, 60, 62, 63, 64, 96, 112, 120, 124, 126, 127, 128, 129, 131, 135, 143, 159, 191, 192, 193, 195, 199, 207, 223, 224, 225, 227, 231, 239, 240, 241, 243, 247, 248, 249, 251, 252, 253, 254 and 255.

We process the feature vector using neural network to detect anomalies. We integrate both features, and train an MLP neural network. The motivation for exploiting MLP is in its substantial ability, through backpropagation, to resist to noise, and the dexterity to generalize. The extracted features are fed as an input to the MLP. The output is obtained by propagating the features as an input vector through the hidden layers. In MLP networks, there are L + 1 layers of neurons, and L layers of weights. During the training stage, the weights W and biases b are updated so that the actual output becomes closer to the desired output. For this purpose, a cost function is defined as in Eq. (3).

$$E(W,b)=\frac{1}{2}\sum_{i=1}^{n_l}(d_i-y_i^L)^2 \qquad (3)$$

This function measures the squared error between the desired and actual output vectors and the backpropagation

is gradient descent on the cost function in Eq. (3). Therefore, during the training stage, weights and biases updates are equivalent to Eq. (4) and Eq. (5), respectively.

$$\Delta W_{ij}^{l}=-\eta\frac{\partial E}{\partial W_{ij}^{l}} \qquad (4)$$

$$\Delta b_{ij}^{l}=-\eta\frac{\partial E}{\partial b_{ij}^{l}} \qquad (5)$$

The backpropagation algorithm begins with the forward pass where the input vector is converted into output. The difference between the desired output and the actual output is computed to estimate the error. During the backward pass, the estimated error at the output units is propagated backwards through the entire network. The weights and biases are updated, for l = 1 to L using the results of the forward and backward passes. The learned weights and biases from the training stage are used to detect anomaly.

EXPERIMENTS

We evaluate the performance of our proposed method for anomalies detection on UMN benchmark dataset available publicly. The UMN dataset consists of normal and abnormal crowd videos from the University of Minnesota. It consists of three different indoor and outdoor representing 11 different scenarios of escape events. There are total 7739 frames of 320x240 pixels. Each video begins with the normal behaviors of people walking and standing. There are also students moving across two buildings lasting for 12 and 5 minutes, respectively. Each sequence is segmented into two different subsequences with people mainly moving in a horizontal direction in the scene. This dataset defines anomaly as the deviations from what has been observed beforehand. This anomaly represents any pedestrian moving in the opposite direction of the general flow of the pedestrians. The detection of anomalies in terms of panic situation is shown for the four sample videos in the Figure.

Figure 1. UMN Dataset: Four sample frames from four videos are shown in the first row. The second row shows the results.

We also presented quantitative results in term of F-scores for the four sequences of UMN dataset in Table 1 below. We achieved very good performance.

TABLE 1. F-scores for the UMN dataset are presented.

(IJRST) 2019, Vol. No. 9, Issue No. I, Jan-Mar e-ISSN: 2249-0604, p-ISSN: 2454-180X

Dataset	Sequences	F-scores
UMN	***S1***	***0.56***
	S2	0.78
	S3	0.66
	S4	0.71
Average		0.67

REFERENCES

[1] Ullah, H., Altamimi, A. B., Uzair, M., & Ullah, M. (2018). Anomalous entities detection and localization in pedestrian flows. Neurocomputing, 290, 74-86.

[2] Khan, W., Ullah, H., Ahmad, A., Sultan, K., Alzahrani, A. J., Khan, S. D., ... & Abdulaziz, S. (2018). CrashSafe: a formal model for proving crash-safety of Android applications. Human-centric Computing and Information Sciences, 8(1), 21.

[3] Ullah, H., Ullah, M., & Uzair, M. (2018). A hybrid social influence model for pedestrian motion segmentation. Neural Computing and Applications, 1-17.

[4] Ahmad, F., Khan, A., Islam, I. U., Uzair, M., & Ullah, H. (2017). Illumination normalization using independent component analysis and filtering. The Imaging Science Journal, 65(5), 308-313.

[5] Ullah, H., Uzair, M., Ullah, M., Khan, A., Ahmad, A., & Khan, W. (2017). Density independent hydrodynamics model for crowd coherency detection. Neurocomputing, 242, 28-39.

[6] Khan, S. D., Tayyab, M., Amin, M. K., Nour, A., Basalamah, A., Basalamah, S., & Khan, S. A. (2017). Towards a Crowd Analytic Framework For Crowd Management in Majid-al-Haram. arXiv preprint arXiv:1709.05952.

[7] Saqib, M., Khan, S. D., Sharma, N., & Blumenstein, M. (2017, December). Extracting descriptive motion information from crowd scenes. In 2017 International Conference on Image and Vision Computing New Zealand (IVCNZ) (pp. 1-6). IEEE.

[8] Ullah, M., Ullah, H., Conci, N., & De Natale, F. G. (2016, September). Crowd behavior identification. In Image Processing (ICIP), 2016 IEEE International Conference on(pp. 1195-1199). IEEE.

[9] Khan, S. (2016). Automatic Detection and Computer Vision Analysis of Flow Dynamics and Social Groups in Pedestrian Crowds.

[10] Arif, M., Daud, S., & Basalamah, S. (2013). Counting of people in the extremely dense crowd using genetic algorithm and blobs counting. IAES International Journal of Artificial Intelligence, 2(2), 51.

[11] Ullah, H., Ullah, M., Afridi, H., Conci, N., & De Natale, F. G. (2015, September). Traffic accident detection through a hydrodynamic lens. In Image Processing (ICIP), 2015 IEEE International Conference on (pp. 2470-2474). IEEE.

[12] Ullah, H. (2015). Crowd Motion Analysis: Segmentation, Anomaly Detection, and Behavior Classification (Doctoral dissertation, University of Trento).

[13] Khan, S. D., Bandini, S., Basalamah, S., & Vizzari, G. (2016). Analyzing crowd behavior in naturalistic conditions: Identifying sources and sinks and characterizing main flows. Neurocomputing, 177, 543-563.

[14] Shimura, K., Khan, S. D., Bandini, S., & Nishinari, K. (2016). Simulation and Evaluation of Spiral Movement of Pedestrians: Towards the Tawaf Simulator. Journal of Cellular Automata, 11(4).

[15] Khan, S. D., Vizzari, G., & Bandini, S. (2016). A Computer Vision Tool Set for Innovative Elder Pedestrians Aware Crowd Management Support Systems. In AI* AAL@ AI* IA (pp. 75-91).

[16] Saqib, M., Khan, S. D., Sharma, N., & Blumenstein, M. (2017, August). A study on detecting drones using deep convolutional neural networks. In 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)(pp. 1-5). IEEE.

[17] Khan, S. D., Vizzari, G., Bandini, S., & Basalamah, S. (2015, October). Detection of social groups in pedestrian crowds using computer vision. In International Conference on Advanced Concepts for Intelligent Vision Systems (pp. 249-260). Springer, Cham.

[18] Khan, S. D., Porta, F., Vizzari, G., & Bandini, S. (2014, September). Estimating Speeds of Pedestrians in Real-World Using Computer

Vision. In International Conference on Cellular Automata (pp. 526-535). Springer, Cham.

[19] Khan, S. D., Crociani, L., & Vizzari, G. (2013). Integrated Analysis and Synthesis of Pedestrian Dynamics: First Results in a Real World Case Study. From Objects to Agents.

[20] Khan, S. D., Crociani, L., & Vizzari, G. PEDESTRIAN AND CROWD STUDIES: TOWARDS THE INTEGRATION OF AUTOMATED ANALYSIS AND SYNTHESIS.

[21] Ullah, H., Ullah, M., & Conci, N. (2014, September). Dominant motion analysis in regular and irregular crowd scenes. In International Workshop on Human Behavior Understanding(pp. 62-72). Springer, Cham.

[22] Saqib, M., Khan, S. D., & Blumenstein, M. (2017, February). Detecting dominant motion patterns in crowds of pedestrians. In Eighth International Conference on Graphic and Image Processing (ICGIP 2016) (Vol. 10225, p. 102251L). International Society for Optics and Photonics.

[23] Ullah, H., Ullah, M., & Conci, N. (2014, March). Real-time anomaly detection in dense crowded scenes. In Video Surveillance and Transportation Imaging Applications 2014(Vol. 9026, p. 902608). International Society for Optics and Photonics.

[24] Ullah, H., Tenuti, L., & Conci, N. (2013, March). Gaussian mixtures for anomaly detection in crowded scenes. In Video Surveillance and Transportation Imaging Applications (Vol. 8663, p. 866303). International Society for Optics and Photonics.

[25] Rota, P., Ullah, H., Conci, N., Sebe, N., & De Natale, F. G. (2013, September). Particles cross-influence for entity grouping. In Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European (pp. 1-5). IEEE.

[26] Ullah, H., & Conci, N. (2013, September). Structured learning for crowd motion segmentation. In Image Processing (ICIP), 2013 20th IEEE International Conference on (pp. 824-828). IEEE.

[27] Ullah, H., & Conci, N. (2012, November). Crowd motion segmentation and anomaly detection via multi-label optimization. In ICPR workshop on Pattern Recognition and Crowd Analysis.

[28] Khan, W., & Ullah, H. (2010). Authentication and Secure Communication in GSM, GPRS, and UMTS Using Asymmetric Cryptography. International Journal of Computer Science Issues (IJCSI), 7(3), 10.

[29] Ullah, H., Ullah, M., Uzair, M., & Rehman, F. (2010). Comparative study: The evaluation of shadow detection methods. International Journal Of Video & Image Processing And Network Security (IJVIPNS), 10(2), 1-7.

[30] Khan, W., & Ullah, H. (2010). Scientific Reasoning: A Solution to the Problem of Induction. International Journal of Basic & Applied Sciences, 10(3), 58-62.

[31] Uzair, M., Khan, W., Ullah, H., & Rehman, F. U. (2009, December). Background modeling using corner features: An effective approach. In Multitopic Conference, 2009. INMIC 2009. IEEE 13th International (pp. 1-5). IEEE.

[32] Ullah, M., Ullah, H., & Alseadoon, I. M. HUMAN ACTION RECOGNITION IN VIDEOS USING STABLE FEATURES.

[33] Khan, W., Ullah, H., & Hussain, R. (2013). Energy-Efficient Mutual Authentication Protocol for Handhled Devices Based on Public Key Cryptography. International Journal of Computer Theory and Engineering, 5(5), 754.

[34] Saqib, M., Khan, S. D., & Blumenstein, M. (2016, November). Texture-based feature mining for crowd density estimation: A study. In Image and Vision Computing New Zealand (IVCNZ), 2016 International Conference on (pp. 1-6). IEEE.

[35] Khan, A. (2015). Coordinated Unmanned Aerial Vehicles for Surveillance of Targets (Doctoral dissertation, Queen Mary University of London).

[36] Khan, A., Rinner, B., & Cavallaro, A. (2015, September). Multiscale observation of multiple moving targets using micro aerial vehicles. In Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on (pp. 4642-4649). IEEE.

[37] Ahmad, K., Mekhalfi, M. L., Conci, N., Melgani, F., & Natale, F. D. (2018). Ensemble of Deep Models for Event Recognition. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 14(2), 51.

[38] Ahmad, K., Conci, N., & De Natale, F. G. (2018). A saliency-based approach to event recognition. Signal Processing: Image Communication, 60, 42-51.