INTERNATIONAL JOURNAL OF SMART SYSTEMS
Volume 1 Issue 1 Year 2023 Pages 40 - 45
e–ISSN XXXX–XXXX
Url: https://ijss.etunas.com/index.php/ijss/index

Anomaly Detection in Surveillance Videos through Object-Oriented Analysis

Auni Khadijah[1]*, Abid Dilnawaz[1]
1 Computer Science, Alyamamah University, Saudi Arabia

*Correspondence to: auni.khadijah@gmail.com

Abstract: Detecting and pinpointing irregularities in surveillance videos has remained a persistent challenge. The current approaches, which are based on patches or trajectories, do not have a semantic understanding of the scenes and may split the targets into fragments. To address this issue, this research proposes a new and efficient algorithm that combines deep object detection and tracking, while fully leveraging spatial and temporal information. A dynamic image is introduced by integrating both appearance and motion information and then fed into an object detection network, which accurately detects and classifies objects, even in crowded and poorly lit scenes. Based on the detected objects, an effective and scale-insensitive feature, named Histogram Variance of Optical Flow Angle (HVOFA), is developed together with motion energy to identify abnormal motion patterns. To further detect missing anomalies and reduce false detections, a post-processing step is carried out with abnormal object tracking. This algorithm outperforms existing methods on established benchmarks.

Keywords: Surveillance videos, Object detection, Tracking, Anomaly detection

Article info: Date Submitted: 30-March-2023 | Date Revised: 30-March-2023 | Date Accepted: 1-April-2023

INTRODUCTION

The detection of unusual activities or incidents in surveillance videos is a crucial and pressing concern for public safety. The advancement of computer vision technologies has led to growing attention towards automated anomaly detection [1],[2],[3]. Nonetheless, detecting anomalies is a formidable task because of its high scene-dependency [4]. Most studies [5],[6] that focus on anomaly detection follow these steps: extracting features from normal training samples during the training phase, fitting a reference model on these features, and then during the testing phase, identifying input data features that do not fit the reference model well as anomalies. Current methods for anomaly detection can be broadly classified as patch-based or trajectory-based [7].

Patch-based approaches circumvent object detection and instead extract features, such as the Histogram of Gradient (HOG) [8] and the Histogram of Optical Flow (HOF) [9], from image patches [10, 11, 12]. Nevertheless, processing all patches using fixed strides is a time-intensive process. Sabokrou et al. [13] suggested modeling discriminative patches surrounding areas of interest. Nonetheless, this method lacks a semantic understanding of the scenes and may split single targets into fragments, which is unsuitable for analyzing behavior. On the other hand, trajectory-based methods, such as those employed by Ma et al. [14] and Piciarelli et al. [15], utilize visual tracking to capture the trajectories of normal objects and develop a model of normal trajectories that can detect distant objects and obtain global motion information. Nevertheless, tracking all targets is a time-consuming process.

The aim of this paper is to introduce a novel algorithm for detecting anomalies in surveillance videos that incorporates deep object detection and tracking methods to fully utilize spatial and temporal information. The recognition of objects in surveillance videos based on appearance is challenging due to its unclear nature. To overcome this, we introduce a dynamic image, which combines the angle and magnitude of optical flow with the input image's intensity, allowing the object detection network to identify the foreground objects. Using object detection not only identifies positions but also target classes, facilitating the discovery of appearance anomalies. Traditional hand-crafted features like HOF are not universal for depth-of-field, so we propose a new scale-insensitive feature called histogram variance of optical flow angle (HVOFA) to detect motion anomalies. Moreover, this work considers location anomalies where people or objects appear in inactive areas. We identify active areas using the training data and determine whether the object is within that area. Since object detection may not identify distant targets, and HVOFA is a local feature, we use object tracking to track the abnormal candidates and extract their full trajectories as global features to filter out false detections, which is more efficient.

RELATED WORK

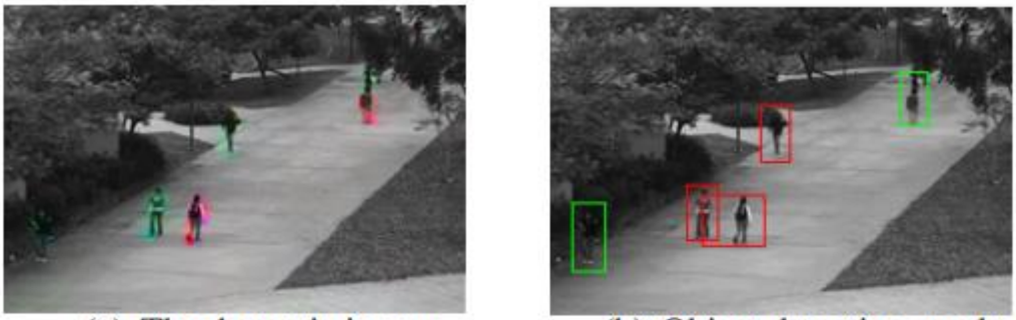
There have been several related works in the field of anomaly detection in surveillance videos. Some of these works use traditional methods such as background subtraction, optical flow, and trajectory analysis. However, these methods have limitations in handling complex scenes with multiple objects, varying lighting conditions, and occlusions. To address these challenges, recent works have focused on using deep learning methods for anomaly detection in surveillance videos. These methods use deep convolutional neural networks (CNNs) to learn high-level features from the video frames and identify anomalies based on these features. Another approach is to use object-oriented analysis, which focuses on detecting anomalies based on the behavior of objects in the scene. This approach involves identifying and tracking objects in the scene and analyzing their trajectories and interactions to detect anomalies. Overall, the related works in this field aim to improve the accuracy and efficiency of anomaly detection in surveillance videos by using advanced techniques such as deep learning and object-oriented analysis.

METHODS

In this section, we provide a detailed explanation of our proposed algorithm, which comprises three modules, as depicted in Fig. 1. Firstly, we perform object detection based on a dynamic image to extract objects. Then, we extract object category, HVOFA, and motion energy to identify appearance and motion anomalies, and detect location anomalies using the background model. Lastly, we employ tracking on the identified abnormal candidates to locate missing targets that cannot be detected by object detection. Additionally, we apply a post-processing step to eliminate false positive anomalies based on the extracted trajectories. To obtain the final score maps, we assign abnormal scores to the pixels belonging to the abnormal candidates.

Dynamic Image Object Detection

Most existing algorithms for extracting foreground patches or clusters rely solely on optical flow, which is time-consuming and lacks semantic understanding of the scene. However, with the rapid development of deep learning technologies in computer vision, we propose using a deep object detection network to extract objects more effectively. However, existing object detection algorithms are limited by poor lighting conditions and visual quality of surveillance videos, making it difficult to recognize objects with similar appearances to the background. To overcome this challenge, we introduce a new dynamic image that fuses appearance and motion information together. The proposed dynamic image can be directly fed into existing object detection algorithms with minimal changes on network structures. Specifically, we calculate the optical flow of the input image and assign the angle and magnitude of the optical flow as the first and second channels of the dynamic image. The intensity of the original image is the third channel.



(a) The dynamic image (b) Object detection results

Figure 1: The image transformation technique for object detection is illustrated in Fig. 1. The dynamic image is created by combining optical flow information with the intensity of the original image. As shown in Fig. 1(a), the resulting dynamic image highlights the moving objects, making them distinguishable from the background. The object detection results from the dynamic image are shown in Fig. 1(b), where objects that are difficult to detect in the original RGB image are successfully detected and highlighted with green bounding boxes.

The structure of the dynamic image is like an Hue-Saturation-Intensity (HSI) image:

$$\begin{cases} H = \text{ Angle of the optical flow,} \\ S = \text{ Magnitude of the optical flow,} \\ I = \text{ Image intensity} \end{cases} \tag{1}$$

To display the dynamic image, it is converted to an RGB image by treating it as an HSI image. The resulting image, shown in Figure 1(a), has colors that correspond to object motion. For object detection, we use the region-based fully convolutional network (RFCN) with ResNet-101 pre-trained on ImageNet, which is accurate and efficient. We fine-tune the model on labeled traffic surveillance videos. Figure 1(b) illustrates an example of object detection using our proposed method. The green bounding boxes show objects that cannot be detected from the original RGB image, but can be well detected using the dynamic image, indicating that our approach achieves better detection results in low-light conditions.

Appearance and Motion Anomaly

Based on the detected objects, we extract the features and judge whether their appearance or movement is abnormal.

Appearance Anomalies. As a result of object detection, we can derive the object category and the corresponding confidence score. We check whether each detected object belongs to the normal class. If not and the confidence score exceeds 0.9, it will be considered an anomaly. The confidence score will be considered as an anomaly score.

Movement Anomalies. In addition to appearance, objects can also become abnormal due to inappropriate movements. For example, running is not permitted in some scenarios such as a hospital. To detect such motion anomalies, we define the so-called motion energy to reflect the speed of the object's movement:

$$E_m = \sum_{i=1}^{n} \frac{v_i^2}{N} \quad (2)$$

where N is the number of pixels of an object, v_i is the magnitude of optical flow at each pixel.

In addition, there may be cases where objects appear to be normal based on their appearance and motion speed, but in reality, they are deceptive. For example, a person riding a skateboard may be classified as a normal pedestrian. To handle such cases, some hand-crafted features such as HOF have been proposed. However, HOF is sensitive to depth-of-field and requires additional processing as it is calculated using weights based on magnitudes [17]. Therefore, we propose a new feature called Histogram Variance of Optical Flow Angle (HVOFA), which is effective and scale-insensitive. HVOFA counts the frequency of different directions based solely on the angles of optical flow. The HOFA feature is defined as follows:

$$HOFA = [f_1, f_2, \ldots, f_B], \quad \sum_{i=1}^{B} f_1 = N \quad (3)$$

In the above equation, B represents the number of directions in HOFA, fi represents the number of pixels in a direction, and N represents the total number of pixels in an object. Therefore, HVOFA can be defined as the histogram variance of the optical flow angles.

Location Anomaly

To address location anomalies such as walking on the grass, a background model is created using principal component analysis that considers the background and foreground as a low-rank matrix and a sparse error matrix, respectively. Foregrounds are extracted from successive frames and combined to obtain the active region. Instead of considering objects as anomalies if most of their pixels are not in the active region, we label objects with the lower part of their body outside of the active region as anomalies, with a location anomaly score (sl) set to 1. This method is effective in avoiding false positive detections, such as objects standing next to lawns.


(a) Ped1


(b) Ped2

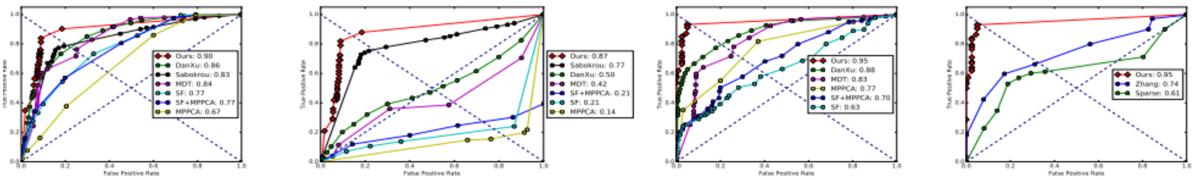
Figure2: Anomaly detected

RESULT AND DISCUSSION

Several experiments were conducted to evaluate the effectiveness of each component of our algorithm. The results indicate that each part of the proposed algorithm has improved the performance in various scenes, particularly the D-RFCN with dynamic image and tracking. These findings also demonstrate the versatility of the proposed algorithm and the efficacy of the developed features.

Notably, the Ped2 dataset lacks location anomalies, and thus the background model has no impact on its performance.

The performance of each component of the proposed algorithm was evaluated through several experiments. As indicated in Table 1, different parts of the algorithm improved the performance in different situations, particularly the proposed Dynamic Region-based Fully Convolutional Network (D-RFCN) and tracking. The effectiveness of the algorithm and the developed features were validated through this. The background model did not make any difference in Ped2 since there were no location anomalies. Comparing the proposed method with other state-of-the-art methods, as presented in Fig. 3 and Table 1, it showed better results in both scenes. Furthermore, the algorithm's ability to detect and localize anomalies accurately was demonstrated by the fact that the performances on pixel level and frame level were almost the same. The algorithm was also tested on Scene02 of ShanghaiTech Campus dataset, and the AUC/EER for frame level was 0.85/0.19, which is better than the 0.71/0.33 obtained by Conv-AE [23].



(a) Frame Level on Ped1 (b) Pixel Level on Ped1 (c) Frame Level on Ped2 (d) Pixel Level on Ped2

Figure 3: ROC curves and AUC (As far as we know, only [20, 21] have provided pixel-level ROC curves for Ped2)

Our algorithm has three advantages that contribute to its good performance. Firstly, object detection using deep neural networks can learn the semantic information of videos, which helps to avoid the splitting of moving targets and provides precise localization. Moreover, the proposed dynamic image incorporates both motion and appearance information, which enhances the accuracy of detection even further. Secondly, the HVOFA feature can distinguish between abnormal and normal patterns using only the angle of optical flow, making it insensitive to scale. Thirdly, tracking makes use of temporal information, which reduces the number of false positives and missed targets. Combining object detection with tracking improves performance since the multi-frame detection mechanism can handle tracking drift.

CONCLUSION

This paper presents an algorithm for detecting and localizing anomalies in surveillance videos that focuses on objects. The algorithm proposes a new dynamic image for object detection that improves detection accuracy in real surveillance videos. The HVOFA feature is introduced as an effective and scale-insensitive feature for detecting motion anomalies. A background model is also proposed to detect location anomalies. Tracking is utilized to reduce false positives and pick up missing anomalies. The paper provides extensive experiments to demonstrate the effectiveness of each part of the algorithm and its superior overall performance.

REFERENCES

[1] Vijay Mahadevan, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos, “Anomaly detection in crowded scenes,” in CVPR. IEEE, 2010, pp. 1975–1981.

[2] [2] Yachuang Feng, Yuan Yuan, and Xiaoqiang Lu, “Learn ing deep event models for crowd anomaly detection,” Neurocomputing, vol. 219, pp. 548–556, 2017.

[3] [3] Asimenia Dimokranitou, Adversarial Autoencoders for Anomalous Event Detection in Images, Ph.D. thesis, Purdue University, 2017.

[4] [4] Borislav Antic and Bj ´ orn Ommer, "Video parsing for ¨ abnormality detection," in ICCV. IEEE, 2011, pp. 2415–2422.

[5] [5] Jefferson Ryan Medel and Andreas Savakis, "Anomaly detection in video using predictive convolutional long short-term memory networks," arXiv preprint arXiv:1612.00390, 2016.

[6] [6] Roberto Leyva, Victor Sanchez, and Chang-Tsun Li, "Video anomaly detection with compact feature sets for online performance," TIP, 2017.

[7] [7] Serhan Cos¸ar, Giuseppe Donatiello, Vania Bogorny, Carolina Garate, Luis Otavio Alvares, and Franc¸ois Bremond, "Toward abnormal trajectory and event detection in video surveillance," TCSVT, vol. 27, no. 3,pp. 683–695, 2017.

[8] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in CVPR. IEEE, 2005, vol. 1, pp. 886–893.

[9] Navneet Dalal, Bill Triggs, and Cordelia Schmid, "Hu□ man detection using oriented histograms of flow and ap□ pearance," in ECCV. Springer, 2006, pp. 428–441.

[10] Tan Xiao, Chao Zhang, and Hongbin Zha, "Learning to detect anomalies in surveillance video," IEEE Signal Processing Letters, vol. 22, no. 9, pp. 1477–1481, 2015.

[11] Mohammad Sabokrou, Mahmood Fathy, Mojtaba Ho□ seini, and Reinhard Klette, "Real-time anomaly de□ tection and localization in crowded scenes," in CVPR,2015, pp. 56–62.

[12] Dan Xu, Yan Yan, Elisa Ricci, and Nicu Sebe, "Detect□ ing anomalous events in videos by learning deep repre□ sentations of appearance and motion," CVIU, vol. 156, pp. 117–127, 2017.

[13] Mohammad Sabokrou, Mohsen Fayyaz, Mahmood Fathy, and Reinhard Klette, "Deep-cascade: Cascad ing 3d deep neural networks for fast anomaly detection and localization in crowded scenes," TIP, vol. 26, no. 4, pp. 1992–2004, 2017.

[14] Ke Ma, Michael Doescher, and Christopher Bodden,"Anomaly detection in crowded scenes using dense tra□ jectories," .

[15] Claudio Piciarelli and Gian Luca Foresti, "On-line tra jectory clustering for anomalous events detection," Pat□ tern Recognition Letters, vol. 27, no. 15, pp. 1835–1842, 2006.

[16] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun, "R-fcn: Object detection via region-based fully convolutional networks," in NIPS, 2016, pp. 379–387.

[17] Ziping Zhu, Jingjing Wang, and Nenghai Yu, "Anomaly detection via 3d-hof and fast double sparse representa□ tion," in ICIP. IEEE, 2016, pp. 286–290.

[18] John Wright, Arvind Ganesh, Shankar Rao, Yigang Peng, and Yi Ma, "Robust principal component anal ysis: Exact recovery of corrupted low-rank matrices via convex optimization," in NIPS, 2009, pp. 2080–2088.

[19] Qiankun Liu, Bin Liu, and Nenghai Yu, "Tccf: Tracking based on convolutional neural network and correlation filters," in ICIG. Springer, 2017.

[20] Ying Zhang, Huchuan Lu, Lihe Zhang, and Xiang Ruan, "Combining motion and appearance cues for anomaly detection," Pattern Recognition, vol. 51, pp. 443–452, 2016.

[21] Yang Cong, Junsong Yuan, and Ji Liu, "Abnormal event detection in crowded scenes using sparse representa□ tion," Pattern Recognition, vol. 46, no. 7, pp. 1851–1864, 2013.

[22] Weixin Luo, Wen Liu, and Shenghua Gao, "A revisit of sparse coding based anomaly detection in stacked rnn framework," ICCV, Oct, 2017.