An Automatic Shoplifting Detection from Surveillance Videos (Student Abstract)

U-Ju Gim,[1] Jae-Jun Lee,[1] Jeong-Hun Kim,[1] Young-Ho Park,[2] Aziz Nasridinov[1,*]
[1]Dept. of Computer Science, Chungbuk National University, Cheongju, 28644, South Korea, +82-43-2613597
{kwj1217, leejj, etyanue, aziz}@chungbuk.ac.kr
[2]Dept. of IT Engineering, Sookmyung Women's University, Seoul, 04310, South Korea
yhpark@sm.ac.kr
*Corresponding Author

Abstract

The use of closed circuit television (CCTV) surveillance devices is increasing every year to prevent abnormal behaviors, including shoplifting. However, damage from shoplifting is also increasing every year. Thus, there is a need for intelligent CCTV surveillance systems that ensure the integrity of shops, despite workforce shortages. In this study, we propose an automatic detection system of shoplifting behaviors from surveillance videos. Instead of extracting features from the whole frame, we use the Region of Interest (ROI) optical-flow fusion network to highlight the necessary features more accurately.

Introduction

In shopping centers, CCTV cameras are mostly utilized to prevent shoplifting related behaviors. However, even with an increased amount of CCTV cameras, an estimated 27 million people annually shoplift. This amounts to at least US $13 billion worth of goods shoplifted each year and more than $100 billion annually (Lasky, Fisher, and Jacques 2017). In other words, CCTV cameras are underutilized because there are not enough human resources to monitor every region of the shopping centers. Thus, there is a need for automatic video surveillance that detects shoplifting related events.

Several methods have been proposed to detect abnormal behavior from surveillance videos. For example, (Jang, Park, and Nasridinov 2019) proposed AVS-NET, which simultaneously performs image processing and object detection to determine abnormal behavior. The authors also proposed a relational module that is used to determine the relationship between the objects in a frame. However, the main problem of this method is that the authors determine abnormal behavior from a single frame when, considering the entire video, that behavior may not be abnormal. (Sultani, Chen, and Shah 2018) and (Zhu and Newsam 2019) propose real-world abnormal behavior detection systems that use deep learning. These methods model normal and abnormal events as instance bags, train these instances using a deep anomaly-ranking model, and predict high anomaly scores for anomalous video segments. Here, abnormal behaviors are detected by extracting features related to the amount of change of actions in the whole frame. While these methods are effective for large-scale explosions and car accidents, they may not be suitable for temporal events such as shoplifting. This is because, unlike other abnormal behavior, shoplifting occurs quickly, meaning that there is not much change in the action of the user.

This study proposes an automatic shoplifting detection method from surveillance videos. Unlike other methods, the proposed method considers the flow of user movements to detect abnormal behaviors. Also, instead of extracting features from the whole frame, we first extract a person object as an ROI using Mask-R-CNN(He et al. 2017) and then determine shoplifting behavior using the amount of change in the ROI person object using optical-flow.

Proposed Method

Figure 1 shows the overall diagram of the proposed method that contains two modules: image segmented annotation and ROI optical-flow fusion network.

Image Segmented Annotation. This module is used to extract three-frame shoplifting and normal behavior from the entire video to define the sequence of an ROI. To do this, we first changed the scale of all surveillance videos to 30x1280x720 and then divided them into a list of frames. We extracted a person object as an ROI using Mask-R-CNN with a pre-trained dataset. If the person object is detected in the video, the frame may have $MultiObject = (Object_1, Object_2, ..., Object_n)$. Frames that contain no person objects are excluded. Here, local coordinates of each $MultiObject$ in the frame are used as coordinate values for the bounding box. With the extracted bounding box, we divide each video into features of the three-frame sequence of movements as optical-flow features.

At this moment, the reason for extracting a person object as an ROI is that we can extract multiple-segmented person objects in one frame. In other words, by extracting a person object, we can avoid checking the features related to the degree of change of actions in the whole frame in determining shoplifting behaviors and instead, focus on detecting shoplifting behavior using the amount of change in the ROI

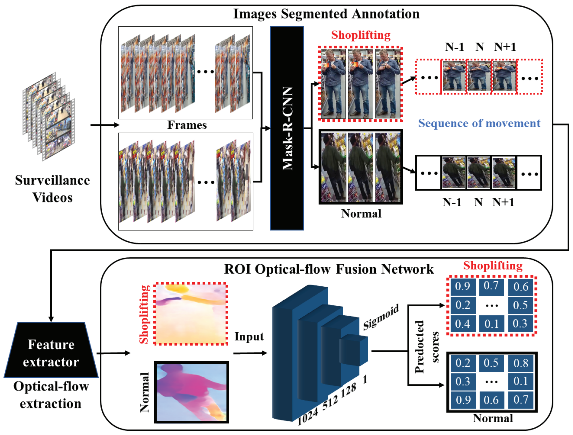

Figure 1: The overall diagram of the proposed method.

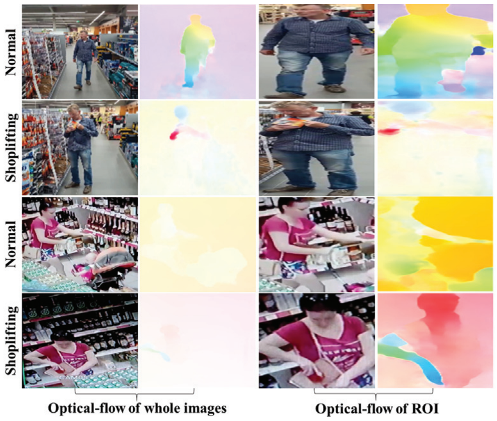

Figure 2: The Optical-flow of whole images extracted and optical-flow of ROI.

person object. To extract the behavior change amount of the ROI for N^{th} frame, we determine the average value of ROI using the $N-1^{th}$, N^{th}, and $N+1^{th}$ ROI of a person feature. The reason why we obtain the average value of ROI is that, if the ROI of a person objects changed even slightly, the feature value of the optical-flow would also change. Thus, we used the average value for the ROI. At this point, our features are divided into a three-frame sequence of movements that are related to our predefined behavior. In other words, we can use a three-frame sequence of ROI features as the input for the ROI optical-flow fusion network.

ROI Optical-Flow Fusion Network. This module uses a three-frame ROI sequence as input for the optical flow extraction and then applies these features to predict shoplifting or normal behavior. Here, we defined shoplifting behaviors as putting things in a bag, arms, and pocket, while normal behaviors were defined as those which are not associated with shoplifting. To detect shoplifting behaviors, we first use the three-frame ROI features as input for the feature extractor and then we extract the optical-flow of the ROI using the three-frame sequence of ROI features. For the sake of comparison, Figure 2 shows the behavior of a person in both an ROI optical-flow and whole image optical-flow. The optical-flows corresponding to normal and shoplifting events are shown in two sets of events. It can be seen that unnecessary features in the whole image can be excluded when using ROI optical-flow. We use the optical-flow of ROI features as input to predict two classes(i.e., normal and shoplifting) from neural networks.

Second, we designed a three-layer fully-connected neural network with the optical-flow of ROI as input. At this time, we normalized the feature values of the optical-flow of ROI between 0 and 1. the shoplifting feature should be close to 1 and the normal feature should be close to 0. The extracted optical-flow of ROI is normalized to 1024x1 and is used for the input of the neural network. The input of the first fully connected layer is 1024 and the three-layer then consists of 512-units, 128-units, and 1-unit. We also avoid overfitting problems by adding a dropout between the 512-units, 128-units, and 1-unit layers. In this case, ReLU activation and Sigmoid activation were used for the 512-units and 1-unit of 3-layer, respectively. Score predictions of normal and shoplifting features are values between 0 and 1. We can use the scores output from the fully connected layer to detect normal and shoplifting behavior.

Conclusions

In this study, we have proposed a model to detect shoplifting behavior as much time and workforce are wasted in monitoring abnormal behavior, such as shoplifting. The optical-flow of ROI can be used to extract the behavior features of shoplifting and normal events that occur in a short time without using features of the entire video frames. In the future, we are planning to demonstrate the efficiency of the proposed method via extensive experiments. We will also demonstrate the scalability of the proposed method when applying it to other fields to detect abnormal behavior.

Acknowledgments

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No.2016-0-00406, SIAT CCTV Cloud Platform).

References

He, K.; Gkioxari, G.; Dollár, P.; and Girshick, R. 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 2961–2969.

Jang, S.; Park, Y.-H.; and Nasridinov, A. 2019. Avs-net: Automatic visual sur-veillance using relation network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 9947–9948.

Lasky, N. V.; Fisher, B. S.; and Jacques, S. 2017. 'thinking thief' in the crime prevention arms race: Lessons learned from shoplifters. *Security Journal* 30(3):772–792.

Sultani, W.; Chen, C.; and Shah, M. 2018. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE Conference on Computer Vision and Pat-tern Recognition*, 6479–6488.

Zhu, Y., and Newsam, S. 2019. Motion-aware feature for improved video anomaly detection. *arXiv preprint arXiv:1907.10211*.