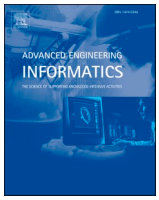
Advanced Engineering Informatics 50 (2021) 101399

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei



Identifying shoplifting behaviors and inferring behavior intention based on human action detection and sequence analysis

Siyeon Kim [a], Sungjoo Hwang [a,*], Seok Hwan Hong [b]

[a] Department of Architectural and Urban Systems Engineering, Ewha Womans University, 52 Ewhayeodae-Gil, Seodaemun-Gu, Seoul 03760, Republic of Korea
[b] Dudaji, Inc., 16, Maeheon-ro, Seocho-gu, Seoul 06771, Republic of Korea

ARTICLE INFO

Keywords:
Abnormal behavior detection
Action sequence analysis
Behavioral intention
Shoplifting
Video surveillance
Public safety

ABSTRACT

Identification of abnormal behaviors affecting public safety (e.g., shoplifting, robbery, and stealing) is essential for preventing human casualties and property damage. Many studies have attempted to automatically identify abnormal behaviors by detecting relevant human actions by developing intelligent video surveillance systems. However, these studies have focused on catching predefined actions associated explicitly with the target abnormal behavior, which can lead to errors in judgment when such actions are undetected or inaccurately detected. To better identify abnormal behaviors, it is essential to understand a series of performed actions to capture behaviors' pre- and post-indications (e.g., repeatably looking around and spotting CCTVs) and infer the intentions underlying such behaviors. Thus, in the present study, we propose a framework to identify abnormal behaviors through deep-learning-based detection of non-semantic-level human action components segmented with a window size of several seconds (e.g., walking, standing, and watching) and performing sequence analyses of the detected action components to infer behavior intentions. Then, we tested the applicability of the framework to the specific scenario of shoplifting, one of the most common crimes. Analysis of actual incident data confirmed that shoplifting intentions could be effectively gauged based on distinct action sequence features, and the intention inference results are continuously updated with the accumulated series of detected actions during the course of the input video stream. The results of this study can help enhance the ability of intelligent surveillance systems by providing a new means for monitoring abnormal behaviors and deeply understanding the underlying intentions.

1. Introduction

In recent years, the identification of abnormal behaviors (e.g., shoplifting, robbery, and stealing) has become essential from the viewpoint of managing public safety in cities and buildings [4,7,49]. These abnormal behaviors are associated with crimes or accident signals that cause human casualties and property damage [5,24]. By detecting such behaviors in advance, the response time to emergencies that affect public safety can be reduced [25,28]. Therefore, abnormal human behavior detection is attracting increasing attention in various domains such as building security, surveillance of shoplifters and terrorists, healthcare, and disaster management [43].

In this regard, closed-circuit televisions (CCTVs) are the most widely used means to monitor abnormal behaviors. However, it is difficult to manually supervise multiple CCTV video screens because of the limitations of human attention and the fatigue that sets in when continuously monitoring abnormal human behaviors that are relatively rare and do not occur frequently [7]. Moreover, CCTVs are mainly used for follow-up after an accident or a crime instead of detecting abnormal behaviors in real-time, which makes it challenging to react in a timely manner to abnormal behaviors [44]. Recently, an intelligent video surveillance system that can automatically detect abnormal behaviors and raise the alarm has been introduced [7,10,16], Nguyen and Meunier [38,41,42]. Advanced video processing methods that employ deep learning and machine learning, statistics, data mining, information theory, or spectral theory have been mainly used to create intelligent video surveillance systems [43,51]. These methods generally focus on the detection of predefined specific classes of human actions (e.g., hitting, picking up a thing) that are explicitly related to abnormal behaviors (e.g., fighting, shoplifting) by labeling existing data and training classification models. However, such actions, which are generally related to crime or safety problems, occur spontaneously, thereby making it difficult to detect

* Corresponding author.
E-mail addresses: kimxy@ewhain.net (S. Kim), hwangsj@ewha.ac.kr (S. Hwang), shhong@dudaji.com (S.H. Hong).

https://doi.org/10.1016/j.aei.2021.101399
Received 21 February 2021; Received in revised form 18 August 2021; Accepted 18 August 2021
Available online 27 September 2021

them accurately. Moreover, because the perpetrators perform these actions as carefully as possible, the difficulty of accurate action detection might be aggravated.

In this regard, the importance of analyzing contextual information related to abnormal behaviors for detecting abnormal actions better and understanding the underlying intentions is growing because an abnormal behavior can be identified by acquiring a deeper understanding of the overall situation rather than detecting it at a specific point in time[42]. People's behaviors are not discrete events but continuous flows that consist of a series of actions. This implies that people tend to perform related series of actions when they engage in a certain behavior. For example, a shoplifter checks to see whether someone is looking at them before stealing. Scoping for people in the vicinity is one of the important pre- and post-indications of shoplifting. Based on this sign, it is possible to approximately identify an abnormal behavior by capturing the pre-and post-indications (e.g., looking around, spotting CCTV) without directly detecting the abnormal action itself (e.g., picking up an object and putting it in a pocket). Thus, the identification of pre- and post-indications facilitates the judgment of suspicious behaviors or scenarios. As another example, falling is an abnormal action, but it can be considered normal behavior when a person stands up and walks right after falling. Therefore, to detect abnormal behaviors, it is essential to understand the whole process of the behavior holistically and capture the pre- and post-indications of abnormal behaviors.

Therefore, in this study, we propose a framework for identifying abnormal behaviors and understanding their underlying intentions, which includes video-based detection of individual actions and analysis of detected action sequences. More specifically, we first detect non-semantic-level actions (i.e., action components which do not have context information regarding specific abnormal behaviors: non-semantic-level action components include walking, running, standing, and watching), which are generally applicable to various abnormal behaviors, by using deep-learning-based action detection algorithms (in particular, three-dimensional convolutional neural network (3D-CNN) algorithm, which is useful for detecting human actions). Then, by analyzing action sequences through the use of combinations of detected actions (e.g., an action sequence such as walking-standing-watching-walking- repeatably watching- …), non-semantic-level actions are transformed into semantic-level information to infer the intention underlying abnormal behaviors. Then, the applicability of the developed framework to a specific abnormal behavior, namely the detection of shoplifters in retail stores, is tested because shoplifting is one of the common and costly crimes [14].

2. Research trends on abnormal behavior identification

In recent days, intelligent video surveillance systems are being developed and improved to identify abnormal behaviors or scenarios. Video surveillance can be used in many scenarios, such as security; crime prevention; traffic control; accident detection; and monitoring of patients, the elderly, and children at home [17,19,22,26,32,34,53,55]. The deployment of a network of cameras across a city, especially in public places, makes it easier to collect large amounts of data on a daily basis. Therefore, a more advanced video surveillance system has been developed to use these data for public safety management by extracting useful information efficiently [17,22,53,55]. The primary purpose of an intelligent surveillance system is to automate the process of recognizing people or objects, tracking them, and alerting the authorities or citizens when an accident occurs. According to Adrian et al. [1], an automated system can reduce the cost of and labor requirement for surveillance, in addition to eliminating human errors in the process of raising the alarm. Intelligent surveillance systems are expected to be used in diverse applications that are closely related to everyday life, especially in scenarios that involving enhancing safety in cities.

Image or video processing using machine learning or deep learning algorithms is a key technology for developing intelligent surveillance systems because this technology can be used to develop the necessary abnormality detection models based on training datasets [39]. Supervised learning is the most widely used method for detecting human actions associated with abnormal behaviors, and this method requires external knowledge and data tagged with one or more labels for classification [2]. Classification is a type of supervised learning that involves automatically making new decisions based on previous decisions [8]. For example, Mehran et al. [37] detected abnormal scenarios in cities, such as people escaping from panic scenarios, protesters clashing, and crowds fighting, by using the social force model. Wang et al. [52] detected abnormal scenarios using high-frequency and spatio-temporal features multiple hidden Markov models, such as a crowd of people scattering in panic and fighting. More recently, Arunnehru et al. [6] presented a 3D-CNN-based action recognition scheme for detecting actions such as walking, running, jumping, and bending by using the KTH and the Weizmann datasets. Sultani et al. [51] and Kim et al. (2020) trained and modeled the video frames of both normal and abnormal data by using a deep learning algorithm with a focus on shoplifting, robbery, and stealing. Mehmood [36] accurately detected falling, loitering, and violence in uncrowded videos by using a two-stream 3D-CNN. Although considerable effort is involved in predefining all scenarios and collecting and training labeled data, the use of labeled data to detect abnormal scenarios helps to achieve the required high-level classification accuracy for predefined behaviors.

A few researchers have used unsupervised and semi-supervised learning methods to identify abnormal human behaviors or scenarios. Unsupervised learning classifies normal and abnormal scenarios by using the statistical properties of unlabeled data and reduces the need for human labor or prior knowledge. However, it involves a considerable number of calculations and requires large computational resources [7,30,33]. Semi-supervised learning, which is a combination of supervised and unsupervised learning, is trained using only normal data or usual observations, and it detects outliers or unusual observations that deviate considerably from the trained data [7,12,18,38]. For example, Nguyen et al. [38] used a deep CNN network to train only normal videos and estimate the abnormality scores for videos of an unknown event. Unsupervised learning and semi-supervised learning are less effort-intensive in terms of data training compared to supervised learning, but these methods cannot be used to obtain detailed information about abnormal data clusters or outliers.

In sum, unsupervised and semi-supervised learning approaches have been effectively applied for identifying abnormal behaviors or scenarios. Despite the additional effort required for data processing and computation, supervised classification provides more detailed information about abnormal scenarios. With advancements in computing capability, recent video surveillance systems aim to detect predefined abnormal actions of interest at specific points in time. In this regard, Ben Mabrouk & Zagrouba [7] reviewed previous studies related to abnormal scenario detection and proposed a framework that can generalize the process used in existing studies. The proposed framework involves defining an action that is explicitly associated with abnormal scenarios and creating a supervised-learning-based classification/recognition model for that action. However, abnormal behaviors or scenarios can be identified not only by detecting a single action at a specific point in time but also by understanding the overall context to infer their intention [3,10]. The rational choice theory (RCT), which adopts a utilitarian belief that humans including criminals are rational in their decision-making on performing behaviors, supports the importance of understanding the overall context of the scenario, especially when abnormal behaviors or scenarios are related to crimes or safety issues. According to the RCT, criminals are ordinary people and act with risks and rewards in mind [9,13]. This theory suggests that there are pre-and post-indications of abnormal behaviors such as crime because criminals need to be aware of the scenario and evaluate their risks, which can provide clues to identify abnormalities. For example, Dabney et al. [14] explained that

shoplifters often exhibit behavioral signs before concealing goods, such as scanning the store and tampering with products. Therefore, to better identify abnormal behaviors from data, it is important to understand the overall context of the performed actions. Moreover, such an understanding is helpful for inferring the intentions underlying such behaviors, especially when spontaneously occurring actions explicitly related to an abnormality are detected improperly.

Recently, Martínez-Mascorro et al. [35] proposed a shoplifting intention detection algorithm based on a 3D-CNN-based deep learning algorithm by training it with video frames related to normal behaviors (i.e., pre-crime behavior segments) and video frames related to shoplifting intention (i.e., suspicious behavior segments in which segments containing direct evidence of shoplifting are removed from the entire abnormal video), and they achieved an accuracy of 75% in suspicious behavior detection. Despite the usefulness of this method in classifying shoplifting intention from normal behaviors, the entire video stream lasting 1–2 min was required as an input in their classification scheme, which limits the inference of intention of behavior early on or in the middle of the input video stream. Moreover, classification based on the frames of an entire video and black box learning algorithms make it difficult to understand the details pertaining to a series of actions associated with the intention of a behavior (e.g., when, how many, and what kind of suspicious actions a person performed in detail). Thus, an understanding of a series of detailed action components performed is required to better capture pre- and post-indications of abnormal behaviors and the underlying intentions during the course of an input video. In this study, therefore, we aim to develop and test a framework for identifying abnormal human behaviors and inferring the intentions underlying such behaviors through deep-learning-based detection of non-semantic-level human action components by using the data of only a few seconds of the unit of analysis (e.g., walking, running, standing, and watching that are the generally applicable characteristics of various behaviors), and consequently, performing the sequence analysis of a series of detected suspicious action components.

3. Proposed framework for identifying abnormal behaviors

A new framework for identifying abnormal behaviors is developed in this study by modifying the general framework to detect abnormal behaviors proposed by Ben Mabrouk & Zagrouba [7], as shown in panels (a)–(d) in Fig. 1. By extending this framework, the framework proposed herein describes a new video surveillance mechanism that consists of two processes, each of which is redefined with action detection for identifying abnormal behaviors (see panels (a)–(d) in Fig. 1) and performing sequence analysis of the detected actions to infer the intentions underlying suspicious behaviors (see panels (e)–(h) in Fig. 1).

First, in the action detection process, non-semantic or low-semantic actions, which are segmented with a window size of several seconds (e.g., walking, running, standing, and picking objects up) and can subsequently constitute semantic action sequences are modeled. Low-level features, such as optical flow, shape, texture, speed, and body movement, are selected for non-semantic action detection (panel (a) in Fig. 1). Each action is defined using suitable low-level features (panel (b) in Fig. 1). Then, an action recognition model is developed using a classification algorithm (e.g., a deep learning algorithm, such as the 3D-CNN) that can distinguish between the various actions defined using the low-level features (panel (c) in Fig. 1). After an essential human action component (e.g., shoplifting, illegal dumping, fighting) is detected, information about the detected action is used to determine the occurrence of abnormal behaviors (panel (d) in Fig. 1). However, the method is occasionally likely to fail to detect such actions that are explicitly related to abnormal behaviors. Moreover, it is difficult to implement any

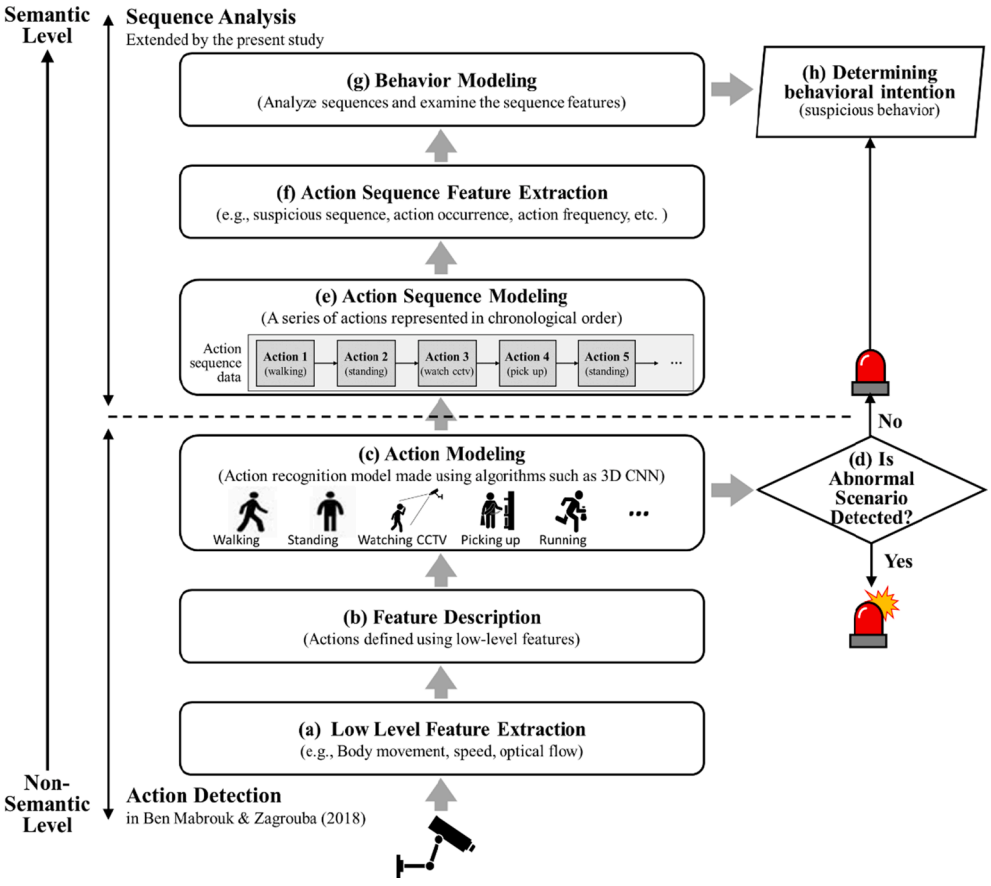


Fig. 1. Framework for identifying abnormal behaviors and inferring the intentions underlying specific behaviors.

preparatory steps for abnormal actions before such actions occur. A means to understand the intentions underlying behaviors is, therefore, necessary to supplement the limitations of the abovementioned abnormal action detection.

To this end, in the next process, action sequence analysis is performed to identify the abnormality of a series of detected actions involving human behaviors or a scenario. As shown in panel (e) of Fig. 1, human actions can be represented in the form of a sequence table in chronological order by combining the non-semantic-level single actions defined in the first process (i.e., action detection process in panels (a)–(d) of Fig. 1). After a behavior sequence is modeled using a series of detected actions, the features of the behavior sequence dataset can be extracted (as in panel (f) of Fig. 1). The features that can be considered for identifying abnormal scenarios include the occurrence of the types of action components included in the sequences, precedence relationships between action components (e.g., looking around and performing a suspicious action, falling and then standing up to walk again, falling and remaining on the ground), and the frequency or repeatability of certain action components (e.g., the number of times an individual looks for CCTVs). Based on these features, we can identify whether a modeled human action sequence (panel (g) in Fig. 1) represents the intention of abnormal behaviors (panel (h) in Fig. 1). Because the important features of action sequences were extracted, the intentions underlying abnormal behaviors can be scored; otherwise, a machine- or deep-learning-based classifier can be applied to determine the intentions underlying normal or abnormal behaviors.

By using the proposed framework, non-semantic-level or lows-semantic-level actions can be combined to obtain semantic information regarding behaviors and the underlying intentions. The key element of the proposed framework is the analysis of action sequences, which can help to not only detect the actions that are directly associated with abnormal behaviors but also understand the pre- and post-indications associated with such behaviors to better infer the intentions underlying these behaviors. The following examples explain the importance of understanding a series of actions in the sequence data. A series of several normal actions, such as walking, changing direction, and running, can be judged as abnormal behavior when combined. The “running” action component is not an abnormal action by itself, but running after looking around can be considered an abnormal action. As such, a lack of understanding of action sequences may lead to an incorrect judgment of scenarios. In addition, checking the repeatability of a particular action can help with inferring abnormal scenarios. “Looking around” once is not necessarily an indicator of an abnormal scenario. However, repeated “looking around” increases the likelihood of suspicious and abnormal behavior. Such behavior, especially in stores, increases the likelihood of shoplifting.

In sum, because various types of actions and various factors are employed to determine abnormality, it is difficult to identify an abnormal scenario by detecting a single action. This difficulty is more pronounced when the detection accuracy of a single target action (e.g., picking up an object) is unsatisfactory, which requires a more nuanced understanding of the pre- and post-indications (e.g., looking around). The framework developed in this study, which aims to identify abnormal behaviors and the underlying intentions based on the sequence analysis of performed actions, is expected to address the problems discussed earlier in the paper (i.e., lack of an understanding of the context information pertaining to abnormal behaviors).

4. Application of developed framework to shoplifting behaviors

4.1. Overview of a case of abnormal behavior identification: Shoplifting

In this study, we demonstrate the applicability of the proposed framework with a focus on shoplifting behaviors, which is one of the common and costly crimes [14]. The retail industry suffers huge losses due to shoplifting. Recently, owing to labor and wage issues, the emergence of unstaffed stores, such as Amazon Go and Bingo Box [29], has become inevitable, and among retail stores, it is expected that unstaffed stores will suffer more from shoplifting than traditional staffed retail stores.

There are several ways to operate an unstaffed store to overcome potential loss due to shoplifting. For example, radio frequency identification (RFID) tags can be attached to products, and multiple sensors and cameras can be installed in stores to track objects and customers [15]. Although the use of RFID tags is convenient from the operational viewpoint, the cost increases because of the costs of RFID tags [40] and the labor required to attach them to objects. Moreover, there is the risk of damage to the tag. In terms of sensor- or camera-based object tracking, the initial cost of installing various sensors and devices is high. In the case of Amazon Go, the operating system uses multiple cameras, microphones, antennas, projectors, and weight sensors [27]. Given that the additional cost of implementing such a system is high, utilization of the existing CCTV infrastructure in retail stores can reduce costs. Thus, it is possible to identify shoplifting behaviors or to check for potential shoplifters without additional investment. The proposed framework in Fig. 1 is applied to identify shoplifting based on actual sequence data. Then, the advantages of using sequence analysis in this identification process are discussed.

4.2. Preliminary action analysis and data collection

Before performing an action sequence analysis of shoplifting, we conducted a preliminary analysis of retail shop customers’ behaviors by observing actual retail store CCTV data and investigating the previous literature on shoplifters’ behaviors[14,46,47]. The CCTV data included both normal customers’ and shoplifters’ behaviors. Then, by arranging the types of actions observed in the CCTV data, we identified 11 representative types of non-semantic-level action components characteristic of shoplifters’ behaviors. As listed in Table 1, these action components include entering the store, walking, scanning the store, spotting CCTVs, picking objects up, putting objects into a pocket, placing objects into a bag, placing objects in a shopping basket, putting objects down, standing, and concealing objects in arms. The action of entering the store is included because the observation of a person starts from the main entrance of the store [14]. The actions of walking and standing are basic components in human action recognition [47]). In addition, it is commonly accepted that the representative behavioral cues of shoplifting intention are scanning the store or watching CCTVs [14,23,46]. Picking objects up; putting them in a pocket, bag, or shopping basket; or putting the object down are essential visual attributes for shoplifting prediction [46]. Notably, altering or removing the packaging of an object can be one of the signs of shoplifting intention, but in our observation, this action did not occur. Therefore, we excluded this action from the list of actions.

Among the selected 11 action components, five actions, namely scanning the store, spotting CCTVs, putting objects into a pocket, putting objects into a bag, and concealing objects in arms, were rarely observed

Table 1
Human actions in a convenience store.

No.	Action	Normal customer	Shoplifter
1	Entering the store	O	O
2	Walking	O	O
3	Scanning the store	–	O
4	Spotting CCTVs	–	O
5	Picking objects up	O	O
6	Putting objects in a pocket	–	O
7	Placing objects in a bag	–	O
8	Placing objects in a shopping basket	O	O
9	Putting objects down	O	O
10	Standing	O	O
11	Concealing objects in arms	–	O

among normal customers. The theoretical investigation also confirmed that the shoplifters exhibited actions that normal customers rarely did [14]. Among the actions of the shoplifters, putting objects into a pocket, placing objects into a bag, and concealing objects in arms are directly associated with stealing, which is critical for the identification of shoplifting. By contrast, scanning the store and spotting CCTVs can be considered pre- and post-indications of stealing. The analysis of action sequences considering these two actions can enhance the performance of abnormal behavior intention inference.

With regard to single-action detection in panels (a)–(d) of Fig. 1, we aimed to develop and test a deep-learning-based action detection model. To this end, we collected more than 1,000 video data representing 11 actions from various sources, as shown in Fig. 2. First, we created data for 11 actions by recruiting 19 subjects. Each subject performed 11 actions in a physically simulated retail store, and their actions were recorded using three CCTVs (total 627 data: 19 subjects performing 11 actions recorded by three cameras). Additionally, all of the subjects performed continuous series of actions as if they were buying goods from the retail store. We divided these videos of continuous actions into smaller units representing individual actions to obtain 200 additional data. Moreover, we collected 200 more single-action data from real shoplifting CCTV videos recorded in a retail store housed in a gas station; these data were provided by the Gas Station Encounters YouTube channel (https://www.youtube.com/c/GasStationEncounters) and the UCF-Crime dataset [51]. The window size of each single-action video was approximately 5 s. In sum, the number of data for each single action was 90–95, and the total number of data for 11 actions exceeded 1000. These data were used for training the action detection model.

4.3. Two-stream 3D-CNN-based single-action component detection

Fig. 3 presents an overview of the deep-learning-based action component detection process. In the data preprocessing step, we used the You Only Look Once (YOLO) algorithm v3 [45] for real-time detection of people in video streams. To improve the object detection accuracy when multiple people are present in the video, we combined YOLO-v3 with a Siamese algorithm (i.e., a neural network that contains two identical sub-networks useful for analyzing the similarity of inputs by using their feature vectors: [11] for accurate object detection and person re-identification (panel (a) in Fig. 3). In this manner, we can perform real-time multiple human detection given a video scene.

Then, we used the two-stream 3D-CNN algorithm to detect the actions performed by the detected person. A two-stream convolutional network, which combines image recognition by using spatial stream convolutional networks and temporal stream convolutional networks by using multi-frame dense optical flow, exhibited outstanding performance in video action recognition [6,21,36,48,54]. The performance of the developed model was tested using the UCF101 human action class datasets, which are widely used to test action recognition[50]. We achieved an action recognition accuracy of 96% on the UCF101 datasets, which is similar to or higher than the accuracy levels achieved in previous studies [21,54] (panel (b) in Fig. 3).

Although many deep-learning-based action recognition models have been developed and are suitable for detecting large movements (e.g., UCF101 datasets including archery, biking, standing, etc.), but they are inaccurate when applied to detect fine movements, such as picking objects up, spotting CCTVs, or putting objects in a pocket, which are typical shoplifting behaviors. Among them, the two-stream 3D-CNN performed the best when detecting 11 actions related to shoplifting, as presented in Table 1.

We confirmed that the average recognition rate for these 11 actions was 85%. Table 2 presents detailed test results including precision (i.e., the ratio of correctly predicted positive observations to the total predicted positive observations), recall (i.e., the ratio of correctly predicted positive observations to all observations in actual class), and F1-score (i.e., the weighted average of Precision and Recall) values for each single actions.

As shown in Fig. 4, the developed model was tested using the sample shoplifting videos extracted from the UCF-Crime dataset [51], which reaffirmed its object detection and action recognition accuracy. This accuracy can be improved by using additional training data and more refined classifiers. However, in this study, we hypothesize that abnormal shoplifting behaviors can be better identified through further action sequence analysis by acquiring a better understanding of contextual information (i.e., pre- and post-indications). Moreover, abnormal shoplifting behavior intentions can be inferred through sequence

Fig. 2. Overview of collected action video data from the simulated and real retail stores.

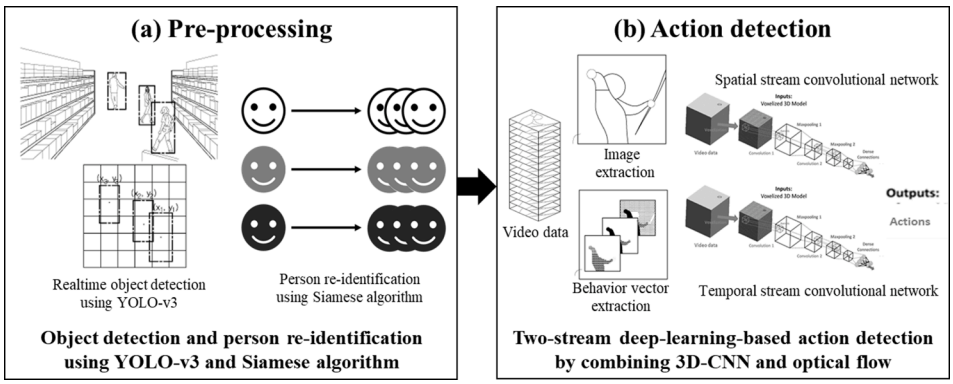


Fig. 3. Overview of action detection process using 3D-CNN algorithm.

Table 2
Classification results of 11 non-sementic-level action components using 3D-CNN-based action detection.

Model	Precision	Recall	F1-score
Entering the store	0.98	0.73	0.83
Walking	0.95	0.99	0.97
Scanning the store	0.85	0.77	0.81
Spotting CCTVs	0.83	0.93	0.88
Picking objects up	0.84	0.79	0.82
Putting objects in a pocket	0.72	0.70	0.71
Placing objects in a bag	0.75	0.73	0.74
Placing objects in a shopping basket	0.96	1.00	0.98
Putting objects down	0.86	0.79	0.83
Standing	0.93	0.94	0.94
Concealing objects in arms	0.76	0.91	0.83

analysis, even when the direct detection of abnormal actions (e.g., putting objects in a pocket, placing objects in a bag, and concealing objects in arms) fails.

4.4. Sequence data collection and feature extraction

For sequence analysis, we collected 173 normal customer videos and 89 shoplifting videos from the CCTV data of retail stores (262 videos in total). The dataset comprising the 173 normal customer videos was collated from the real CCTV data of a retail store. In the dataset comprising the 89 shoplifting videos, 43 real shoplifting CCTV videos were obtained from the Gas Station Encounters YouTube channel, and 46 videos were extracted from the UCF-Crime dataset [51]. The sequence data were reconstructed by accumulating the detected actions in the sequence table in chronological order while omitting the sex or age of the subjects for their privacy. When creating sequence datasets, the window size of walking and standing actions was five seconds, and the window size of the other actions was one second. For example, when the walking action lasted 10 s, it was recorded as two walking actions. Examples of these sequence datasets are shown in Fig. 5.

The recorded action datasets can be divided into two adjacent units to analyze the relationships between actions. From the analysis results, pre-and post-indications can be identified, and significant sequence features can be inferred. The following parallel graphs show which actions people engaged in following each action (Fig. 6). In the graphs, the thick and thin lines represent the action sequences that occur more frequently and less frequently, respectively.

A comparison of the normal customers' and shoplifters' graphs indicated that the sequence patterns of the normal customers were less complex because the normal customers' performed fewer actions than the shoplifters. By contrast, the shoplifters' sequence patterns were complex because they performed more actions, resulting in more diverse combinations of actions. Moreover, there was no dominant sequence in the shoplifters' actions, thus indicating that all of their actions were distributed across various sequences. Therefore, action sequences that characterize shoplifters can be identified by excluding the normal customers' sequences from the shoplifters' sequences. These sequence features can be used to distinguish shoplifting scenarios from normal scenarios. Table 3 lists the features used in this study and describes what each feature indicates.

(a) Examples of shoplifter sequence datasets

Seq. 14	1	2	10	2	10	5	5	10	10	2	3	6	10	2	2
Seq. 19	1	2	2	10	5	3	2	2	10	10	5	6	2	2	
Seq. 40	1	3	2	2	10	5	5	2	11	11	2	2			

(b) Examples of normal customer sequence datasets

Seq. 70	1	2	2	10	10	5	10	2	10	2	10	10	5	2	2
Seq. 175	1	2	2	10	10	10	10	10	5	10	2	2			
Seq. 208	1	2	2	2	10	10	10	10	10	10	10	5	2	2	

[Note] #1 – Entering the store, #2 – Walking, #3 – Scanning the store, #4 – Spotting CCTVs, #5 – Picking objects up, #6 – Putting objects in a pocket, #7 – Placing objects in a bag, #8 – Putting objects in a shopping basket, #9 – Putting objects down, #10 – Standing, #11 – Concealing objects in arms

Fig. 5. Examples of sequence datasets.

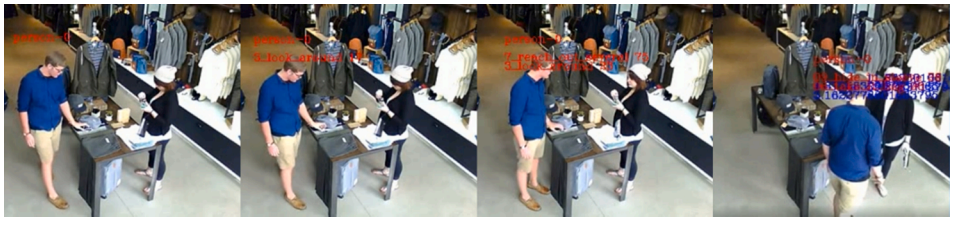


Fig. 4. Test of action detection model by using a shoplifting video from the UCF-Crime dataset.

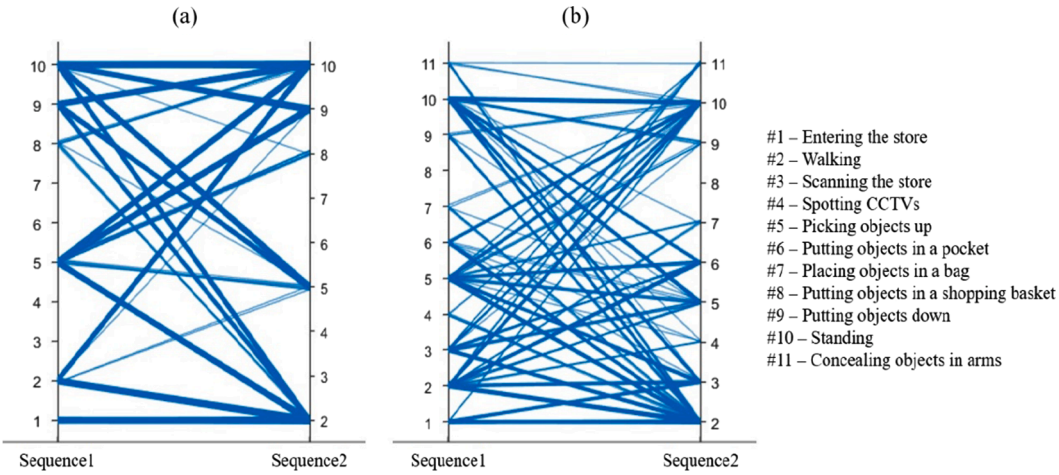


Fig. 6. Pattern of sequences: (a) Normal customers and (b) shoplifters.

Table 3
Features used for classification.

	Feature Type	Features
A	Repetitions of each action in the action sequence dataset (Action Occurrence)	Repetitions of "walking" action Repetitions of "scanning the store" action Repetitions of "spotting CCTVs" action Repetitions of "picking objects up" action Repetitions of "putting objects in a pocket" action Repetitions of "placing objects in a bag" action Repetitions of "placing objects in a shopping basket" action Repetitions of "putting objects down" action Repetitions of "standing" action Repetitions of "concealing objects in arms" action
B	The ratio of each action to the total number of actions (Action Frequency)	The ratio of "walking" action The ratio of "scanning the store" action The ratio of "spotting CCTVs" action The ratio of "picking objects up" action The ratio of "putting objects in a pocket" action The ratio of "placing objects in a bag" action The ratio of "placing objects in a shopping basket" action The ratio of "putting objects down" action The ratio of "standing" action The ratio of "concealing objects in arms" action
C	Suspicious Action Sequences	Entering the store → Scanning the store Walking → Scanning the store/spotting CCTVs Scanning the store/spotting CCTVs → Picking objects up Picking objects up → Scanning the store/spotting CCTVs Scanning the store/spotting CCTVs → Putting objects in a pocket Putting objects in a pocket → Scanning the store/spotting CCTVs

4.5. Inferring shoplifting intention

To infer shoplifting intention, the intention can be scored by extracting important features from action sequences or machine- or deep-learning based classifier can be applied. From a preliminary analysis, we discovered the existence of strong explanatory features in action sequence data that can be used to distinguish between normal and abnormal behaviors. In addition, the action sequence data and features were well-structured by the developed framework. Therefore, we hypothesize that simple machine learning classifiers (e.g., decision tree and logistic regression) would be adequate to identify abnormalities accurately. The collected 262 sequence data were labeled "normal (173 datasets)" and "shoplifting (89 datasets)," and the features listed in Table 3 were extracted. Various classification models ranging from simple ones (e.g., decision tree and logistic regression) to more sophisticated ones (e.g., support vector machine (SVM), subspace discriminant, and neural network) were applied. Then, to evaluate the classification models, we divided the collected data into a training set and a test set and performed 10-fold cross-validation.

Based on the classification of normal and shoplifting scenarios using various features and an evaluation of the classification results, we found that the type-B features in Table 3 (i.e., action frequency) were the most explanatory with regards to shoplifting classification. This explanatory power was ascribed to the fact that the type-B features explained the proportions of actions directly associated with shoplifting (e.g., putting objects into a pocket, placing objects in a bag, and concealing objects in arms) and actions representing behavioral cues of shoplifting intention (e.g., scanning the store or spotting CCTVs). Table 4 lists the accuracy, precision, recall, and F1-score of the classification models. Notably, the single-action detection accuracy obtained in Table 2 (i.e., 3D-CNN-based detection results of non-semantic-level action components) was applied to create action sequence data in this analysis. The classification accuracies of the models using the five algorithms were 90.1%–98.9%.

The actions that are explicitly associated with shoplifting (i.e., putting things in a pocket, putting things in a bag, and concealing things in arms) usually occur rapidly, thus making it difficult to detect them

Table 4
Classification results of normal and shoplifting action sequence data obtained using the ratios of all actions as features.

Model	Accuracy	Precision	Recall	F1-score
Decision Tree	97.3%	0.988	0.933	0.960
Logistic Regression	98.9%	1.000	0.966	0.983
SVM	96.9%	0.918	1.000	0.957
Subspace Discriminant	90.1%	1.000	0.708	0.829
Neural Network	98.5%	1.000	0.955	0.977

well compared to the other actions. Therefore, we attempted to confirm whether normal and shoplifting intention scenarios can be classified using different features, such as indirectly related suspicious actions (e.g., frequency of "scanning the store" and "spotting CCTVs"). Table 5 lists the accuracy, precision, recall, and F1-score of the classification models by excluding the actions directly associated with stealing. The classification accuracies of the resulting models were slightly lower than those of the classification models that employed all action features (Table 3); nevertheless, at 88.5%–93.1%, the accuracy levels were rather high. This result indicates the strong possibility of identifying a shoplifting intention, even when the actions directly associated with stealing cannot be detected.

Moreover, we classify normal and abnormal scenarios based on the number of actions performed by a person since they enter the store. While classification based on the entire video frame makes it difficult to infer shoplifting intentions in the early stages, analysis based on individual action components is expected to solve this challenge. This analysis aimed to determine whether shoplifting intention can be inferred in the initial phases of a series of actions. In this analysis, as shown in Fig. 7, the ratios of each action in the sequence were calculated over the initial 5, 10, 15, and 20 actions, separately, based on the fact that the average number of customers' action components in the analyzed sequences is approximately 20.

Fig. 7(a) summarizes the classification accuracy results obtained using the five classification models as a function of the number of actions performed after entering the store. By using the initial five actions to classify normal and abnormal scenarios, all five models achieved accuracies higher than 75.2%. Similarly, by using the first 10 actions to classify the normal and abnormal scenarios, all models achieved accuracies higher than 83.2%, which represents a significant improvement compared to the results obtained using the initial five actions. When the classification was performed using the initial 15 and 20 actions, the accuracies improved to 86.6%–98.1%. In sum, the results indicate that the greater the number of actions performed after entering the store, the higher is the model accuracy. Thus, it is possible to infer the abnormality of human behaviors to some extent merely by checking the initial actions in an individual's action sequence, and the identification accuracy increasingly improves as more actions are considered.

Fig. 7(b) lists the accuracy results of the five classification models as a function of the number of actions performed after entering the store when excluding actions directly associated with stealing (i.e., putting objects into a pocket, placing objects in a bag, and concealing objects in arms). This analysis aimed to confirm the possibility of shoplifting intention through an investigation of the initial sequence in the case when the instantaneous actions directly associated with stealing were not properly detected. When performing classification using the initial five actions of an individual after they enter the store, the classification accuracies were the same as those when the actions directly associated with stealing were included. When using the initial 10, 15, and 20 actions, the classification accuracies increased to more than 85.9%–91.6%, even though they were somewhat lower than the accuracies achieved when the actions directly associated with stealing were included. According to the results, the greater the number of actions performed, the higher were the classification accuracies. In addition, based on the results, it was possible to infer shoplifting to some degree without considering the number of all actions in the sequence and without proper detection of the actions directly associated with stealing.

Table 5
Classification results obtained by excluding the ratios of direct stealing actions.

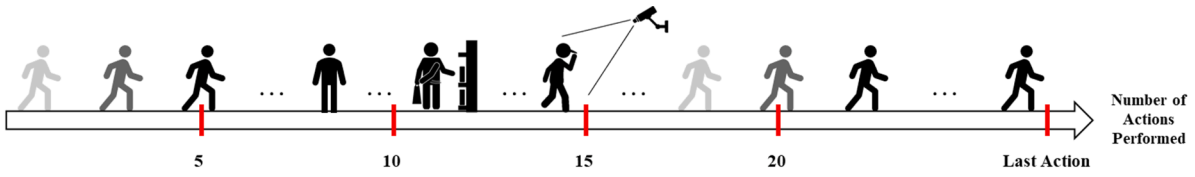
Model	Accuracy	Precision	Recall	F1-score
Decision Tree	88.5%	0.847	0.809	0.828
Logistic Regression	93.1%	0.961	0.831	0.892
SVM	88.9%	0.857	0.809	0.832
Subspace Discriminant	86.3%	1.000	0.596	0.746
Neural Network	90.1%	0.854	0.854	0.854

5. Discussions

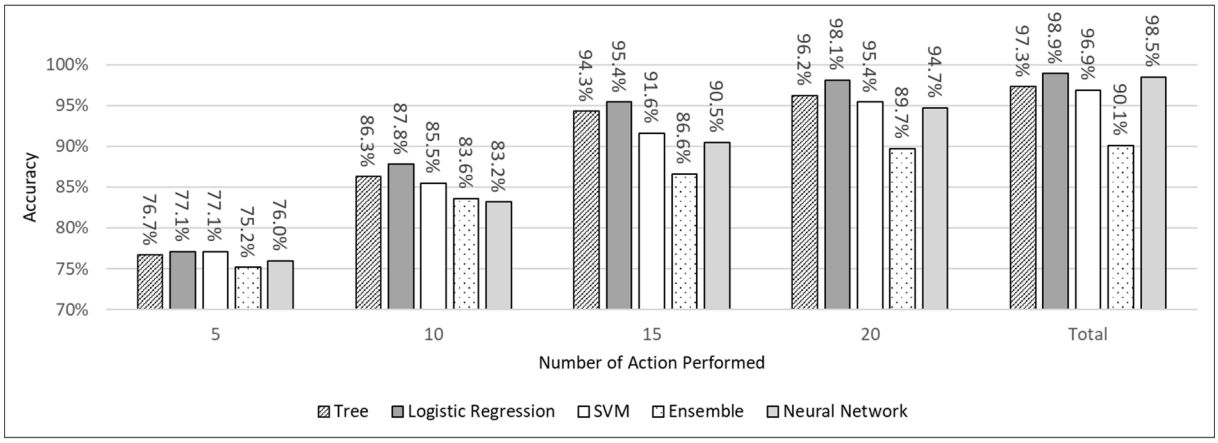
To apply the proposed framework to identifying shoplifting behaviors, we first investigated the series of performed actions of normal customers and shoplifters. Normal customers seldom performed actions that were unrelated to their shopping purpose. They mostly performed actions related to shopping, such as entering the store, looking at goods, picking up objects, and putting them down. By contrast, shoplifters performed additional actions unrelated to shopping, such as scanning the store and spotting the CCTVs, as well as stealing objects. Shoplifters' action sequences were more complex and diverse, which made them useful for extracting features that can be used to distinguish sequence patterns between normal customers and shoplifters, thereby better understanding their intention. The actions involved in the shoplifters' abnormal sequences were divided into two types: actions explicitly associated with shoplifting (e.g., putting objects into a pocket, placing objects in a bag, and concealing objects in arms) and pre-and post-indications of shoplifting (e.g., scanning the store and spotting CCTVs). Abnormal behaviors and intention can be identified more efficiently by considering the actions explicitly related to abnormal behaviors and the action sequences including pre-and post-indications of abnormal behaviors.

In shoplifting, we used various useful features of the action sequence data (e.g., action occurrence and frequency). Despite the 85% of accuracy of the 3D-CNN-based single-action component detection scheme and the use of simple classification algorithms with action sequence data, the classification accuracies for suspicious behavior intentions were 90.1%–98.9% or higher because of a deeper understanding of the details of action sequences. In addition, abnormal shoplifting behavior intentions were inferred in this study even without information about the actions directly associated with stealing. After excluding the actions directly related to stealing, the classification accuracies of the models were still 88.5%–93.1%. The result indicated that the occurrence frequency of suspicious actions (e.g., scanning the store and spotting CCTVs) is an important feature for inferring shoplifters' behavior intention. Although it is difficult to definitively detect shoplifting based on this inference result alone, it can be used as a useful hint by CCTV supervisors to keep monitoring such suspicious customers across numerous CCTV monitors.

Moreover, we identified that abnormal behaviors could be inferred to some degree in the initial phases of a customer's action sequences after entering the store. The classification accuracies of the models using the initial five actions from a sequence dataset were around 75%–77%. The classification accuracies of the models considering 10, 15, and 20 actions indicated that the accuracy increased as the number of actions performed increased to over 86.3% when the actions directly related to stealing (e.g., putting objects into a pocket, placing objects in a bag, and concealing objects in arms) were excluded. Previously, Martínez-Mascorro et al. [35] proposed a shoplifting intention detection scheme by classifying video streams with a window size of several minutes for normal behaviors and shoplifting intention by using deep learning, and they achieved a classification accuracy of 75%. To the best of our knowledge, this previous work is the only one pertaining to the inference of shoplifting intention, and it shares a similar research motivation as that of our study. Compared to this previous study, we achieved higher accuracy (i.e., more than 90%) in inferring shoplifting intention. In addition, compared to the classification based on several minutes of the entire video scene as a unit of analysis, action sequence analysis based on every action component segmented over several seconds as the unit of analysis enabled us to infer intention at an early stage before the entire video frame could be input. Furthermore, action-component-based sequence analysis provides a deeper understanding of a video scene by detailing the series of actions associated with the intentions underlying shoplifting behaviors. These results indicated that by using



(a) Classification results as a function of the number of actions performed.



(b) Classification results as a function of the number of actions performed when excluding the actions directly associated with stealing.

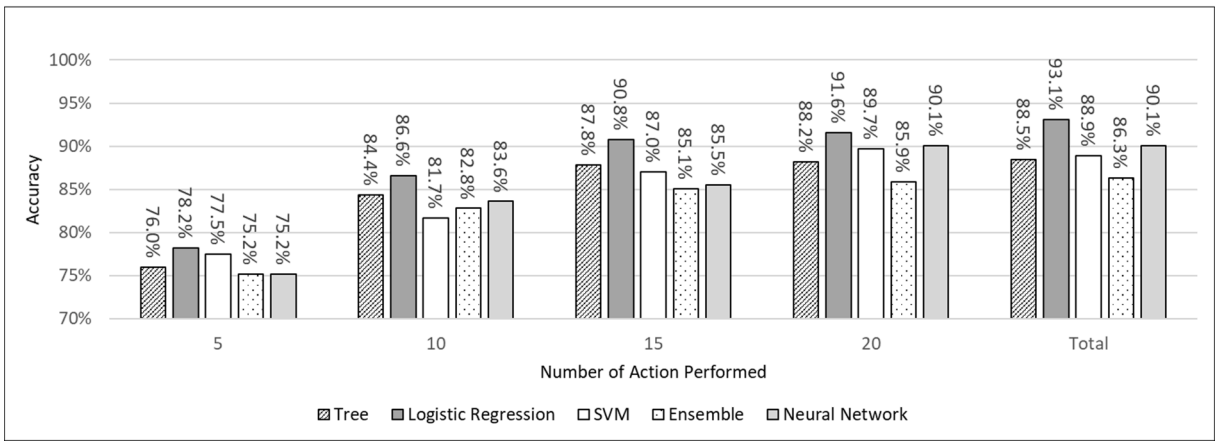


Fig. 7. Timeline of shoplifting intention inference: (a) classification results as a function of the number of actions performed, and (b) classification results as a function of the number of actions performed when excluding the actions explicitly related to shoplifting.

the framework proposed herein, it was possible to infer shoplifting behaviors and the underlying intentions without considering the entire sequence of actions. When the detection of irrefutable evidence of shoplifting (e.g., putting objects into a pocket, placing objects in a bag, and concealing objects in arms) fails, action-based intention inference can be used to inform CCTV supervisors about potentially suspicious customers early on so that the supervisors can carefully monitor such customers.

We expected that the framework proposed in this study could be extended to understand the intention of other abnormal behaviors. Detected non-semantic-level actions, such as walking, running, standing, and watching, are also included in other abnormal behaviors (e.g., illegal dumping, robbery, and stealing). For example, actions included in the sequence of shoplifting (e.g., walking, scanning, spotting CCTVs, and standing) can be included in the sequence of illegal dumping, where only the action of throwing objects needs to be additionally trained and modeled. Then, a method for action sequence analysis can be used to infer the intention of illegal dumping.

6. Conclusions

In this study, we proposed an abnormal behavior identification and behavior intention inference framework by combining the deep-learning-based detection of non-semantic-level human action components (e.g., walking, running, standing, and watching) with the sequence analysis of the detected actions to infer the intentions underlying abnormal behaviors by better understanding the contextual information of abnormal behaviors. Although most of the previous studies have focused on increasing the recognition rate and accuracy when detecting certain types of abnormal actions, capturing the actions associated with abnormal behaviors that occur spontaneously has been difficult. Moreover, the mere detection of a particular type of action at a specific point in time does not provide an understanding of the pre- and post-indications of abnormal behaviors. To overcome these challenges, by using the framework proposed in this study, we aimed to analyze entire sequences of actions performed by investigating not only the actions that are directly associated with abnormal behaviors (e.g., picking

up an object and putting it in a pocket or a bag) but also those that are indirectly related to abnormal behaviors (e.g., looking around, spotting CCTVs). With a focus on shoplifting behavior, we demonstrated that the proposed framework performs well. The detection accuracy for the actions directly associated with abnormal shoplifting behaviors was approximately 85% based on the two-stream 3D-CNN. The addition of action sequence analysis to the existing framework enabled us to infer shoplifting intentions with an accuracy of more than 90%.

This study contributes to the detection of abnormal behavior and to improving the inference of behavioral intentions through action sequence analysis beyond mere action recognition, even when action detection is not performed properly or an error occurs. By integrating video-based action detection technologies with theories of human behavior and by using single-action components as the unit of analysis rather than entire video scenes, the proposed framework can provide a more comprehensive and deeper understanding of the contextual information associated with abnormal scenarios in a video scene. In addition, the framework facilitates inference of behavioral intention in the initial stages of a CCTV video stream, and the intention inference results can be continuously updated based on the accumulated series of detected actions during the course of the input video stream. As a practical tool, we expect that the proposed framework can improve public safety in urban areas.

Despite the contributions of this study, further development is necessary to increase the applicability of the given results. In addition to the 11 single-action components in this study, we found additional actions such as altering or removing the packaging of objects, talking with others, or sitting. Training for and detecting such actions can improve the classification accuracy of shoplifting intentions. Because we study aimed to highlight the methodological advancement of the developed framework, we applied simple classification algorithms to the action sequence data. Therefore, the use of more advanced classification algorithms will improve the accuracy of abnormality identification. As described, future studies should be conducted with the above limitations in mind. Ultimately, it is necessary to detect and understand various abnormal scenarios more efficiently and accurately to improve public safety in urban areas.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [Sungjoo Hwang reports administrative support, equipment, drugs, or supplies, and statistical analysis were provided by National Research Foundation of Korea (NRF) and the Ministry of SMEs and Startups (MSS, Korea)].

Acknowledgment

This research was supported by the National Research Foundation of Korea (NRF) grant (2020R1F1A1073178) funded by the Korean government, and the Technology Development Program (S2658843) funded by the Ministry of SMEs and Startups (MSS, Korea). The authors also would like to acknowledge anonymous participants who participated in the survey.

References

[1] A.I. Adrian, P. Ismet, P. Petru, An overview of intelligent surveillance systems development, in: 2018 13th International Symposium on Electronics and Telecommunications, ISETC 2018 - Conference Proceedings, 2018, pp. 1–6, https://doi.org/10.1109/ISETC.2018.8584003.

[2] M. Ahmed, A. Naser Mahmood, J. Hu, A survey of network anomaly detection techniques, J. Network Comput. Appl. 60 (2016) 19–31, https://doi.org/10.1016/j.jnca.2015.11.016.

[3] F. Anwar, I. Petrounias, T. Morris, V. Kodogiannis, Mining anomalous events against frequent sequences in surveillance videos from commercial environments, Expert Syst. Appl. 39 (4) (2012) 4511–4531, https://doi.org/10.1016/j.eswa.2011.09.134.

[4] I. Arora, M. Gangadharappa, A survey of motion detection in image sequences, in: Proceedings of the 2019 6th International Conference on Computing for Sustainable Global Development, INDIACom, 2019, pp. 215–223.

[5] R. Arroyo, J.J. Yebes, L.M. Bergasa, I.G. Daza, J. Almazán, Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls, Expert Syst. Appl. 42 (21) (2015) 7991–8005, https://doi.org/10.1016/j.eswa.2015.06.016.

[6] J. Arunnehru, G. Chamundeeswari, S.P. Bharathi, Human action recognition using 3D convolutional neural networks with 3D motion cuboids in surveillance videos, Procedia Comput. Sci. 133 (2018) 471–477, https://doi.org/10.1016/j.procs.2018.07.059.

[7] A. Ben Mabrouk, E. Zagrouba, Abnormal behavior recognition for intelligent video surveillance systems: A review, Expert Syst. Appl. 91 (2018) 480–491, https://doi.org/10.1016/j.eswa.2017.09.029.

[8] M. Bilal, L.O. Oyedele, J. Qadir, K. Munir, S.O. Ajayi, O.O. Akinade, H.A. Owolabi, H.A. Alaka, M. Pasha, Big Data in the construction industry: A review of present status, opportunities, and future trends, Adv. Eng. Inf. 30 (3) (2016) 500–521, https://doi.org/10.1016/j.aei.2016.07.001.

[9] C. Cardone, R. Hayes, Shoplifter Perceptions of Store Environments: An Analysis of how Physical Cues in the Retail Interior Shape Shoplifter Behavior, J. Appl. Sec. Res. 7 (1) (2012) 22–58, https://doi.org/10.1080/19361610.2012.631178.

[10] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection for discrete sequences: A survey, IEEE Trans. Knowl. Data Eng. 24 (5) (2012) 823–839, https://doi.org/10.1109/TKDE.2010.235.

[11] D. Chicco, Siamese neural networks: An overview, Artificial Neural Networks (2021) 73–94.

[12] J.K. Chow, Z. Su, J. Wu, P.S. Tan, X. Mao, Y.H. Wang, Anomaly detection of defects on concrete structures with the convolutional autoencoder, Adv. Eng. Inf. 45 (2020) 101105, https://doi.org/10.1016/j.aei.2020.101105.

[13] D.B. Cornish, R.V. Clarke, Understanding crime displacement: An application of rational choice theory, Criminology 25 (4) (1987) 933–948, https://doi.org/10.1111/j.1745-9125.1987.tb00826.x.

[14] D.A. Dabney, R.C. Hollinger, L. Dugan, Who actually steals? A study of covertly observed shoplifters, Justice Q. 21 (4) (2004) 693–728, https://doi.org/10.1080/07418820400095961.

[15] K. Domdouzis, B. Kumar, C. Anumba, Radio-Frequency Identification (RFID) applications: A brief introduction, Adv. Eng. Inf. 21 (4) (2007) 350–355, https://doi.org/10.1016/j.aei.2006.09.001.

[16] D. Duque, H. Santos, P. Cortez, Prediction of abnormal behaviors for intelligent video surveillance systems, in: Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2007, Cidm, 2007, pp. 362–367, https://doi.org/10.1109/CIDM.2007.368897.

[17] M. Elbouz, Fuzzy logic and optical correlation-based face recognition method for patient monitoring application in home video surveillance, Opt. Eng. 50 (6) (2011) 067003, https://doi.org/10.1117/1.3582861.

[18] S. Fitriani, S. Mandala, M.A. Murti, Review of semi-supervised method for Intrusion Detection System, Proceedings - APMediaCast 2016 (2017) 36–41, https://doi.org/10.1109/APMediaCast.2016.7878168.

[19] H. Foroughi, B.S. Aski, H. Pourreza, Intelligent video surveillance for monitoring fall detection of elderly in home environments, in: Proceedings of 11th International Conference on Computer and Information Technology, ICCIT 2008, Iccit, 2008, pp. 219–224, https://doi.org/10.1109/ICCITECHN.2008.4803020.

[21] Y. Han, P. Zhang, T. Zhuo, W. Huang, Y. Zhang, Going deeper with two-stream ConvNets for action recognition in video surveillance, Pattern Recogn. Lett. 107 (2018) 83–90.

[22] M. Haque, M. Murshed, Panic-driven event detection from surveillance video stream without track and motion features, in: 2010 IEEE International Conference on Multimedia and Expo, ICME 2010, 2010, pp. 173–178, https://doi.org/10.1109/ICME.2010.5583057.

[23] R. Hayes, Shoplifting Control, Prevention Press Orlando, FL, 1993.

[24] P. Intani, T. Orachon, Crime Warning System using Image and Sound Processing, in: 2013 13th International Conference on Control, Automation and Systems (ICCAS 2013), 2013, pp. 1751–1753.

[25] S.-R. Ke, H. Thuc, Y.-J. Lee, J.-N. Hwang, J.-H. Yoo, K.-H. Choi, A review on video-based human activity recognition, Computers 2 (2) (2013) 88–131, https://doi.org/10.3390/computers2020088.

[26] N. Kiryati, T.R. Raviv, Y. Ivanchenko, S. Rochel, Real-time abnormal motion detection in surveillance video, in: Proceedings - International Conference on Pattern Recognition, 2008, https://doi.org/10.1109/icpr.2008.4761138.

[27] D. Kumar, E.M. Kornfield, A.C. Prater, S. Boyapati, X. Ren, C. Yuan, Detecting Item Interaction and Movement (U.S. Patent Application No. 13/928,345.), 2015.

[28] J. Laufs, H. Borrion, B. Bradford, Security and the smart city: A systematic review, Sustainable Cities Soc. 55 (2020) 102023, https://doi.org/10.1016/j.scs.2020.102023.

[29] S.-H. Lee, An Analysis on the Changes of Logistics Industry using Internet of Things, J. Adv. Informat. Technol. Convergence 9 (1) (2019) 57–66, https://doi.org/10.14801/jaitc.2019.9.1.57.

[30] K. Leung, C. Leckie, Unsupervised anomaly detection in network intrusion detection using clusters, Conf. Res. Practice Informat. Technol. Series 38 (January) (2005) 333–342.

[32] C.W. Lin, Z.H. Ling, Y.C. Chang, C.J. Kuo, Compressed-domain fall incident detection for intelligent home surveillance, in: Proceedings - IEEE International Symposium on Circuits and Systems, 2005, pp. 3781–3784, https://doi.org/10.1109/ISCAS.2005.1465453.

[33] C. Lu, Z. Wang, B.o. Zhou, Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification, Adv. Eng. Inf. 32 (2017) 139–151, https://doi.org/10.1016/j.aei.2017.02.005.
[34] H. Luo, J. Liu, W. Fang, P.E.D. Love, Q. Yu, Z. Lu, Real-time smart video surveillance to manage safety: A case study of a transport mega-project, Adv. Eng. Informat. 45 (October 2019) (2020) 101100, https://doi.org/10.1016/j.aei.2020.101100.
[35] G.A. Martínez-Mascorro, J.R. Abreu-Pederzini, J.C. Ortiz-Bayliss, A. Garcia-Collantes, H. Terashima-Marín, Criminal Intention Detection at Early Stages of Shoplifting Cases by Using 3D Convolutional Neural Networks, Computation 9 (2) (2021) 24, https://doi.org/10.3390/computation9020024.
[36] A. Mehmood, Abnormal behavior detection in uncrowded videos with two-stream 3d convolutional neural networks, Appl. Sci. (Switzerland) 11 (8) (2021) 3523, https://doi.org/10.3390/app11083523.
[37] R. Mehran, A. Oyama, M. Shah, Abnormal crowd behavior detection using social force model, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2010, pp. 935–942. https://doi.org/10.1109/cvpr.2009.5206641.
[38] T.N. Nguyen, J. Meunier, Anomaly detection in video sequence with appearance-motion correspondence, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1273–1283.
[39] S. Omar, A. Ngadi, H. Jebu, Machine Learning Techniques for Anomaly Detection: An Overview, Int. J. Comput. Appl. 79 (2) (2013) 33–41, https://doi.org/10.5120/13715-1478.
[40] D. Parkash, T. Kundu, P. Kaur, The RFID technology and its applications: a review, Int. J. Electron. Commun. Instrument. Eng. Res. Develop. (IJECIERD) 2 (3) (2012) 109–120.
[41] Q.C. Pham, A. Lapeyronnie, C. Baudry, L. Lucat, P. Sayd, S. Ambellouis, D. Sodoyer, A. Flancquart, A.C. Barcelo, F. Heer, F. Ganansia, V. Delcourt, Audio-video surveillance system for public transportation, in: 2010 2nd International Conference on Image Processing Theory, Tools and Applications, IPTA 2010, 2010, pp. 47–53, https://doi.org/10.1109/IPTA.2010.5586783.
[42] O.P. Popoola, K. Wang, Video-based abnormal human behavior recognitiona review, *I*EEE Trans. Syst. Man Cybern. Part C Appl. Rev. 42 (6) (2012) 865–878, https://doi.org/10.1109/TSMCC.2011.2178594.
[43] N.R. Prasad, S. Almanza-Garcia, T.T. Lu, Anomaly detection, Comput. Mater. Continua 14 (1) (2009) 1–22, https://doi.org/10.1145/1541880.1541882.
[44] M.M. Rathore, A. Paul, W.H. Hong, H.C. Seo, I. Awan, S. Saeed, Exploiting IoT and big data analytics: Defining Smart Digital City using real-time urban data, Sustainable Cities Soc. 40 (December 2017) (2018) 600–610, https://doi.org/10.1016/j.scs.2017.12.022.
[45] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779–788.
[46] S. Reid, P. Vance, S. Coleman, D. Kerr, S. O'Neill, Visual Social Signals for Shoplifting Prediction, in: NexComm 2021 Congress, 2021, pp. 37–42.
[47] Y. Sheikh, M. Sheikh, M. Shah, Exploring the space of a human action, in: Tenth IEEE International Conference on Computer Vision (ICCV'05)1, vol. 1, no. 1, 2005, pp. 144–149.
[48] K. Simonyan, A. Zisserman, Two-stream convolutional networks for action recognition in videos, 2014, arXiv preprint arXiv:1406.2199.
[49] K. Singh, S. Rajora, D.K. Vishwakarma, G. Tripathi, S. Kumar, G.S. Walia, Crowd anomaly detection using Aggregation of Ensembles of fine-tuned ConvNets, Neurocomputing 371 (2020) 188–198, https://doi.org/10.1016/j.neucom.2019.08.059.
[50] K. Soomro, A.R. Zamir, M. Shah, UCF101: A dataset of 101 human actions classes from videos in the wild, 2012, arXiv preprint arXiv:1212.0402.
[51] W. Sultani, C. Chen, M. Shah, Real-world anomaly detection in surveillance videos, in: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2018, pp. 6479–6488.
[52] B. Wang, M. Ye, X. Li, F. Zhao, J. Ding, Abnormal crowd behavior detection using high-frequency and spatio-temporal features, Mach. Vis. Appl. 23 (3) (2012) 501–511, https://doi.org/10.1007/s00138-011-0341-0.
[53] X. Wang, Intelligent multi-camera video surveillance: A review, Pattern Recogn. Lett. 34 (1) (2013) 3–19, https://doi.org/10.1016/j.patrec.2012.07.005.
[54] G. Yao, T. Lei, J. Zhong, A review of convolutional-neural-network-based action recognition, Pattern Recognit. Lett. 118 (2019) 14–22, https://doi.org/10.1016/j.patrec.2018.05.018.
[55] K. Yun, H. Jeong, K.M. Yi, S.W. Kim, J.Y. Choi, Motion interaction field for accident detection in traffic surveillance video, in: Proceedings - International Conference on Pattern Recognition, 2014, pp. 3062–3067, https://doi.org/10.1109/ICPR.2014.528.