Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Інфокомунікацій
(повна назва)
Кафедра Інфокомунікаційної інженерії імені В.В. Поповського
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

Рівень вищої освіти другий (магістерський)

Аналіз методів виявлення злочинних дій в системах відеоспостереження на основі нейронних мереж
Analysis of methods for detecting criminal actions in video surveillance systems based on neural networks
(тема)

Виконав:
студент 2 курсу, групи АМСЗІімв-20-1
Крістін Шахук
(прізвище, ініціали)

Спеціальність: 125 Кібербезпека
(код і повна назва спеціальності)
Тип програми: освітньо-наукова
(освітньо-професійна або освітньо-наукова)
Освітня програма: Адміністративний менеджмент у сфері захисту інформації
(повна назва освітньої програми)

Керівник: професор кафедри ІКІ ім. В.В. Поповського
Радівілова Т.А.
(посада, прізвище, ініціали)

Допускається до захисту
Зав. кафедри _____ Лемешко О.В.
(підпис) (прізвище, ініціали)

2023 р.

Харківський національний університет радіоелектроніки

Факультет Інфокомунікацій
(повна назва)
Кафедра Інфокомунікаційної інженерії імені В.В. Поповського
(повна назва)
Рівень вищої освіти другий (магістерський)
Спеціальність 125 Кібербезпека
(код і повна назва)
Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)
Освітня програма Адміністративний менеджмент у сфері захисту інформації
(повна назва)

ЗАТВЕРДЖУЮ
Зав. кафедри_____
(підпис)
«____» _____2023 р.

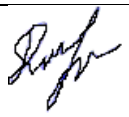
ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студенту Крістін Шахук .
(прізвище, ім'я, по батькові)

1. Тема роботи: Аналіз методів виявлення злочинних дій в системах відеоспостереження на основі нейронних мереж
затверджена наказом по університету від «31» жовтня 2022р. №1434Ст.
2. Термін подання студентом роботи до екзаменаційної комісії 31.01.2023р.
3. Вихідні дані до роботи: нейронні мережі, класифікація, злочинні дії, системи відеоспостереження, відеофрагменти, датасет, аналіз рухів людини
4. Перелік питань, що потрібно опрацювати в роботі:
1) Аналіз підходів виявлення аномальної поведінки людини в реальному часі у відеопослідовності
2) Формування вхідного набору даних
3) Вибір нейронної технології для інтелектуальної системи відеоспостереження
4) Розробка експериментів виявлення злочинних дій інтелектуальними системами відеоспостереження на базі нейронних мереж

5) Перелік графічного матеріалу із зазначенням креслень, плакатів, комп'ютерних ілюстрацій: демонстраційний матеріал у вигляді ppt-презентації.

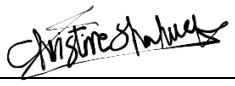
6. Консультанти розділів роботи

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		(підпис)	(дата)
Основна частина	професор Радівілова Тамара Анатоліївна		12.01.2023

Календарний план

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Отримання завдання	05.09.2022	Виконано
2.	Збір матеріалів для дослідження	08.10.2022	Виконано
3.	Розробка 1 розділу	25.10.2022	Виконано
4.	Розробка 2 розділу	12.11.2022	Виконано
5.	Розробка 3 розділу	23.11.2022	Виконано
6.	Розробка 4 розділу	03.12.2022	Виконано
6.	Розробка 5 розділу	14.12.2022	Виконано
7.	Оформлення кваліфікаційної роботи	20.12.2022	Виконано

Дата видачі завдання 1 вересня 2022 року

Студент  Крістін Шахук.

(підпис) (прізвище, ініціали)

Керівник роботи  професор Радівілова Т.А.

(підпис) (посада, прізвище, ініціали)

Робота не містить відомості заборонених до відкритого публікування

Студент  **Крістін Шахук**

Керівник  **Тамара РАДІВІЛОВА**

РЕФЕРАТ

Пояснювальна записка: 68 с., 26 рис., 33 джерела.

ВІДЕОСПОСТЕРЕЖЕННЯ, НЕЙРОННІ МЕРЕЖІ, АНАЛІЗ, ІНТЕЛЕКТУАЛЬНА СИСТЕМА, РОСПІЗНАВАННЯ, ОБРАЗ, ОПТИМІЗАЦІЯ, ЗОБРАЖЕННЯ, ЗЛОЧИННІ ДІЇ, ВІДЕОПОТІК, ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ.

Об’єкт дослідження – процес прийняття рішень в інтелектуальній системі відеоспостереження на базі нейронних мереж.

Мета роботи – дослідити особливості та перспективи використання систем відеоспостереження з використанням нейронних мереж.

Предмет дослідження – моделі, методи, технології розпізнавання даних в інтелектуальній системі відеоспостереження на основі нейронних мереж.

Методи досліджень – методи сегментації зображень, методи навчання нейронних мереж, методи оптимізації.

Кваліфікаційна робота присвячена аналізу підходів виявлення аномальної поведінки людини в реальному часі у відеопослідовності та аналізу сучасних прикладних систем розпізнавання образів.

В роботі досліджується проектування архітектури інтелектуалізованої системи відеоспостереження, нейронні мережі при обробці зображень в інтелектуальних системах спостереження.

ABSTRACT

The report contains: 68 pages, 26 figures, 33 sources.

VIDEO SURVEILLANCE, NEURAL NETWORKS, ANALYSIS, INTELLIGENT SYSTEM, RECOGNITION, IMAGE, OPTIMIZATION, IMAGERY, CRIMINAL ACTIONS, VIDEO FLOW, CONVOLUTIONAL NEURAL NETWORKS.

The object of research is the decision-making process in an intelligent video surveillance system based on neural networks.

The subject of research is models, methods, technologies of data recognition in an intelligent video surveillance system based on neural networks.

The goal of the work is to investigate the features and prospects of using video surveillance systems using neural networks.

Research methods – image segmentation methods, neural network training methods, optimization methods.

The qualification work is devoted to the analysis of approaches to detecting anomalous human behavior in real time in a video sequence and the analysis of modern applied pattern recognition systems.

The work explores the design of the architecture of an intelligent video surveillance system, and neural networks for image processing in intelligent surveillance systems.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS, SYMBOLS, UNITS, AND TERM

3DCNN - 3D Convolutional Neural Networks
AIS - automated identification system
ARS - Automated Response System
B2NM - Back to Normality Moment
BB - bounding box
C3D - technique that can build 3-D graphics out of 2-D images
CCM - Comprehensive Crime Moment
CCTV - Closed circuit television
CES - Crime Evidence Segment
CNN - convolutional neural networks
DL - Deep Learning
FAM - First Appearance Moment
FN - false negatives
FP - false positives
FRCI - Fibrous Refractory Composite Insulation
FV - feature value
FVA - Funding valuation adjustment
FVM - finite volume method
GAN - Generative Adversarial Network
HSV - hue, saturation, value
HVAC - Heating, ventilation and air-conditioning
LBPH - Local Binary Pattern Histogram
LSTM - Long Short-Term Memory
MIS - Manual Identification System
ORL - Olivetti Research Laboratory
PCA - Principal Component Analysis
PCB - printed circuit board
PCB - pre-crime Behavior
PCBS - Pre-Crime Behavior Segment
POD - Police Observation Devices
PODS - Portal Overt Digital Surveillance
RNN - recurrent neural networks

SBS - Suspicious Behavior Segment
SCM - Strict Crime Moment
SSD - single-shot detector
SVM - Support Vector Machine
TN - true negative
TP - true positive
UCF-Crime - large-scale dataset of 128 hours of videos
UFI - Unconstrained Facial Images
VAE - Variation Auto encoder
VFOA - Visual Focus of Attention

INTRODUCTION

Crime detection and their prediction is a fundamental process to reduce criminal activities before they actually happen. Moreover, the detection method is vital since can it potentially can save the victim's life, avoid all-time strain, and harm to the public/private property. In addition, it can be useful in predicting the possible terrorist activities. Crime detection using deep learning models is an attention-grabbing research area. Detecting and reducing the criminal activities is imperative to develop a peaceful society. Video surveillance automates the hazardous situations and enables a law enforcement system to take effective steps towards public safety.

Applications in various areas, including crime prevention, automatic smart visual monitoring and road safety, need for considerable attention to anomaly in event detection in video surveillance.

In recent decades, an enormous number of surveillance cameras are installed in both private and public locations for effective real-time monitoring to prevent malfunctions and protect public safety. Most cameras, however, offer just passive logging services and are not capable of monitoring. The number of these films grows every minute, making it easy for human specialists to comprehend and analyses them. Similarly, monitoring analysts have to wait hours for abnormal occurrences to be captured or seen for immediate reports.

Because there are few anomalous events in the real world, video anomaly detection is studied as a one-class issue, in which the model is trained on typical films and a video is tagged as anomalous when odd patterns appear. All the typical real-world monitoring events cannot be cumulated in one dataset. Different typical actions may thus be distracted from regular training events and may ultimately produce false alarms.

A major purpose of video surveillance is the detection of unusual situations such as traffic accidents, robberies, or illicit activity. Human operators and manual examination are still required by most existing monitoring systems (prone to disturbances and tiredness). As a result, effective computer vision techniques for automatically detecting video anomalies/violence are becoming increasingly relevant. Building algorithms that detect specific anomalous occurrences, such as violence detectors, fight action detectors, and traffic accident detectors, is a tiny

step toward resolving detection of anomalies. In recent years, video action recognition has gotten a lot of attention after achieving very promising results by leveraging CNN's incredible robustness.

In most businesses and sectors, installing CCTVs for ongoing surveillance of people and their interactions is a widespread practice. Every day, every person in a developed country with a population of millions is photographed by a camera. Constant monitoring of these surveillance films by police to determine whether or not the occurrences are suspicious is practically impossible, as it necessitates a workforce and their undivided attention. As a result, we're developing a demand for high-precision automation of this process. It is also vital to show which frame and which parts of it include unexpected activity, as this aids in determining whether the unusual activity is abnormal or suspicious. This will aid concerned authorities in finding underlying cause of anomalies while also saving time as well as effort that would otherwise be spent manually searching the records.

ARS is a real-time monitoring system that recognizes and records evidence of offensive or disruptive behavior in real-time. Using a variety of Deep Learning models, this study seeks to detect and characterize high movement levels in frame. Videos are divided into portions. A detection alert is raised in event of a threat, displaying suspicious behavior at a specific point in time. The movies are divided into classes: threat and safe. Burglary, Abuse, Explosion, Fighting, Shooting, Shoplifting, Arson, Road Traffic Accidents, Robbery, Assault, Stealing and Vandalism are amongst the 12 uncommon actions we recognize. As a result of these irregularities, people would feel safer.

1 DETECT CRIMINAL ACTS IN VIDEO SURVEILLANCE SYSTEM BASED ON NEURAL NETWORKS

Video surveillance systems are commonly used for security and surveillance purposes in various environments, such as public spaces, transportation systems, and private properties. However, manual monitoring of these video feeds can be time-consuming and ineffective, especially in the case of large amounts of footage. Neural networks, a type of machine learning algorithm, can be used to automatically detect criminal acts in video surveillance systems, thus improving the efficiency and accuracy of security and surveillance. The use of neural networks in video surveillance systems can improve the efficiency and accuracy of identifying criminal acts, and can also help to reduce the workload for human operators.

There are several methods that can be used to detect criminal acts in video surveillance systems using neural networks. These methods rely on the ability of neural networks to learn patterns and features from large amounts of data, and can be trained to identify specific objects, actions, and individuals in the video footage.

However, the success of these methods heavily relies on the quality and quantity of data used to train the neural networks, as well as the specific circumstances and environment of the surveillance. Additionally, it is crucial to consider the ethical and legal implications of using these technologies, such as privacy concerns and potential biases. The goal of this analysis is to explore the various methods and techniques used to detect criminal acts in video surveillance systems based on neural networks.

1.1 Detect criminal acts using video surveillance

Reasons why we need to Detect criminal acts using video surveillance

Detecting criminal acts in video surveillance refers to the process of identifying and analyzing unusual or abnormal behavior in video footage captured by surveillance systems. This can include identifying specific actions, such as fighting, stealing, or loitering, as well as identifying specific individuals or vehicles. The goal of detecting criminal acts in video surveillance is to improve security, deter criminal activity, and aid in investigations and apprehensions. This can be achieved through the use of advanced technologies such as neural networks

and machine learning algorithms, which can analyze and process large amounts of video data in real-time, and can be trained to recognize and classify specific objects, activities, and anomalies.

There are several reasons why it is important to detect criminal acts in video surveillance:

– Security and safety: Video surveillance systems are often used to monitor public spaces and private properties to improve security and safety. By detecting criminal acts in video surveillance footage, it is possible to identify and deter criminal activities, such as theft, vandalism, and violence.

– Investigation and prosecution: Video surveillance footage can be used as evidence in criminal investigations and prosecutions. Detecting criminal acts in video surveillance footage can aid in the identification of suspects and the reconstruction of events.

– Resource management: Video surveillance systems generate large amounts of video footage, which can be difficult for human operators to effectively monitor and analyze. Automating the process of detecting criminal acts in video surveillance footage can improve the efficiency and accuracy of identifying criminal activities, and can also help to reduce the workload for human operators.

– Public trust: By detecting criminal acts and increasing security, it can help to build public trust in the surveillance systems, which can be a valuable tool for law enforcement and public safety.

– Legal responsibilities: Depending on the location and context, video surveillance systems may have legal responsibilities to detect and report criminal acts, such as in public spaces where the security of the public is in the hands of the authorities.

At the result: detecting criminal acts in video surveillance footage is an important tool for improving security, safety, and aiding in the investigation and prosecution of criminal activities, while also balancing the legal and ethical responsibilities of the surveillance systems.

Methods to detect criminal acts in video surveillance

There are several methods that can be used to detect criminal acts in video surveillance systems using neural networks. Some common approaches include:

– *Object detection:* This method uses convolutional neural networks (CNNs) to detect and classify objects in the video, such as people, cars, and weapons.

This method involves training a neural network to detect specific objects in a video. The network can then be used to identify these objects in real-time surveillance footage.

– *Activity recognition*: This method uses recurrent neural networks (RNNs) to analyze the sequence of events in the video and identify specific actions, such as a person carrying a bag or running away, fighting, stealing, or loitering.

This method involves training a neural network to recognize specific actions or movements. The network can then be used to detect these actions in real-time surveillance footage.

– *Anomaly detection:* This method uses unsupervised learning techniques to identify unusual or abnormal behavior in the video, such as a person walking into a restricted area or a vehicle parked in a suspicious location.

– *Facial recognition:* This method uses deep learning algorithms to identify and match faces in the video to a database of known individuals.

– *Video prediction:* this method uses deep learning algorithms to predict the next frames of the video which can help identify unusual activities.

– *Scene understanding:* This method involves training a neural network to understand the overall context of a scene, such as the presence of a crowd or a building. The network can then be used to detect abnormal or suspicious behavior in real-time surveillance footage.

– *Deep learning-based anomaly detection:* This method uses deep learning algorithms to detect anomalies in surveillance footage, such as a person entering a restricted area or a vehicle driving on the wrong side of the road.

It is important to note that these methods rely heavily on the quality and quantity of the data used to train the neural networks. Additionally, the accuracy of these methods can be affected by factors such as lighting conditions, camera angles, and occlusions.

These methods can be used alone or in combination to create a robust system for detecting criminal acts in video surveillance footage. It is important to note that the accuracy of these methods can be affected by factors such as camera quality, lighting conditions, and the presence of occlusions.

These methods can also be improved by having a large and diverse dataset to train the model, and also having regular check and update the model to adapt to the new situation and criminal acts.

Overall, neural networks can be a powerful tool for detecting criminal acts in video surveillance systems, but it requires a combination of various techniques and appropriate dataset to improve the performance.

The better method to detect criminal acts in video surveillance

It is difficult to say which method is better for detecting criminal acts in video surveillance, as it depends on the specific circumstances and environment of the surveillance. Different methods may be more appropriate for different types of criminal acts and environments.

– Object detection is commonly used to detect and classify objects in video footage, such as people, cars, and weapons, and can be useful for identifying specific individuals or vehicles.

- Activity recognition is useful for analyzing the sequence of events in video footage and identifying specific actions, such as fighting, stealing, or loitering.
- Anomaly detection is useful for identifying unusual or abnormal behavior that deviates from normal patterns in the video, such as a person walking into a restricted area or a vehicle parked in a suspicious location.
- Facial recognition is useful for identifying and matching faces in video footage to a database of known individuals, and can help to identify suspects or missing persons.
- Video prediction can be useful for identifying unusual activities by predicting the next frames of the video.

It is important to evaluate the performance of each method in the specific context of the surveillance and select the method that works best.

Which method accuracy is higher to detect criminal acts in video surveillance

The accuracy of detecting criminal acts in video surveillance can vary depending on the specific circumstances and environment of the surveillance, the quality and quantity of data used to train the model, and the method used.

These two methods have a higher accuracy to detect criminal acts:

– Deep learning-based methods such as object detection, activity recognition, and facial recognition, have been shown to achieve high accuracy in detecting criminal acts in video surveillance footage. These methods can be trained on large datasets and can handle high-dimensional and complex data.

– Anomaly detection methods also can achieve high accuracy in detecting criminal acts in video surveillance footage by identifying unusual or abnormal behavior that deviates from normal patterns.

It's also important to note that, regardless of the method used, the accuracy of detecting criminal acts in video surveillance can be improved by using high-quality data and incorporating domain knowledge about the specific context of the surveillance.

1.2 Anomaly detection systems in the video sequence

Anomaly detection systems in video sequences are used to identify abnormal or suspicious behavior in CCTV footage. These systems use various techniques such as motion detection, background subtraction, and object tracking to analyze video footage and detect abnormal events.

One popular method for anomaly detection in video sequences is to use a background subtraction algorithm, which compares each frame of the video to a pre-defined background model and identifies any pixels that differ significantly from the model. This can be useful for detecting moving objects, such as people or vehicles, that are not part of the normal scene.

Another method is to use a combination of object tracking and behavior analysis, where the system tracks individual objects (e.g. people or vehicles) in the video and analyzes their movements and interactions over time to detect abnormal behavior. This can be useful for detecting suspicious behavior such as loitering or tailgating.

Deep Learning based approaches, like Auto encoders, GANs and One-class SVM are also used to detect abnormal activity in video sequences, these models are trained with normal behavior data and then detect any deviation from normal as abnormal.

Overall, anomaly detection systems in video sequences are useful for improving the security and surveillance of public spaces, and can be integrated

with other security systems such as access control and alarm systems to provide a more comprehensive security solution.

Methods for detecting anomalies in video sequences

There are many methods for detecting anomalies in video sequences that utilize neural networks. Some of the most popular methods include:

– *Recurrent neural networks (RNNs):* RNNs can be used to analyze video sequences and identify patterns or anomalies in the behavior of individuals or groups. The RNNs are trained on normal behavior patterns, and then used to detect any deviation from these patterns.

– *Convolutional neural networks (CNNs):* CNNs can be used to analyze individual video frames and detect any suspicious activities. This can be done by training the CNN on a dataset of normal and abnormal video frames and then using it to classify new frames as normal or abnormal.

– *Auto encoders:* Auto encoders are a type of neural network that can be trained to identify unusual patterns or events in the video by reconstructing the normal patterns and identifying any differences between the reconstruction and the original video.

– *One-class SVM*: One-class Support Vector Machine (SVM) is a machine learning algorithm that can be used to detect anomalies in video sequences. It models the normal data distribution and then detects any new data point that is not similar to the modeled distribution as anomalous.

– *Deep learning-based approaches:* Deep learning-based approaches such as LSTM, CNN-LSTM, and GANs can be used to detect anomalies in video sequences by learning the characteristics of the normal behavior patterns in the videos and then identifying any deviation from these patterns as anomalous.

– *Statistical methods:* These methods use statistical models to identify patterns in the video data and detect any deviations from those patterns.

– *Machine learning-based methods:* These methods use algorithms such as support vector machines, decision trees, and k-nearest neighbors to classify video frames as normal or abnormal.

– *Deep learning-based methods:* These methods use neural networks such as auto encoders, variation auto encoders, and convolutional neural networks to detect anomalies in the video data.

– *Hybrid methods:* These methods combine multiple techniques, such as statistical methods and deep learning, to improve the accuracy of anomaly detection.

These methods have been found to be effective in detecting unusual events and criminal activities in video surveillance systems, but the performance of these methods can be affected by factors such as the quality and resolution of the video, the lighting conditions, and the presence of occlusions and background clutter.

The best method to use to detecting anomalies in video sequences

Anomaly detection methods are very useful for fraud detection or disease detection case studies where the distribution of the target class is highly imbalanced. Anomaly detection algorithms are also to further improve the performance of the model by removing the anomalies from the training sample.

The best method for detecting anomalies in video sequences will depend on the specific circumstances and environment of the surveillance. However, some methods have been shown to have high accuracy and good performance in detecting anomalies in video sequences.

One such method is using deep learning-based approaches, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). These methods can be trained to learn the normal patterns in the video data and identify deviations from those patterns as anomalies. This approach can be very effective in detecting anomalies in video sequences, especially if the data is labeled and the model is fine-tuned accordingly.

1.3 Development of experiments to detect anomalies in the video sequence of surveillance cameras

There are many ways to develop experiments that can be designed to detect anomalies in the video sequence of surveillance cameras using neural networks. Some possible steps include:

– *Unsupervised anomaly detection:* One experiment could involve training an unsupervised neural network, such as an auto encoder, on a dataset of normal video sequences from the surveillance cameras. The network could then be used to identify abnormal behavior in new video sequences by comparing the reconstruction error of the input to a threshold.

– *Semi supervised anomaly detection:* Another experiment could involve training a semi supervised neural network, such as a Generative Adversarial Network (GAN), on a dataset of both normal and abnormal video sequences. The network could then be used to identify abnormal behavior by comparing the likelihood of new video sequences to the likelihood of the training data.

– *Supervised anomaly detection:* A third experiment could involve training a supervised neural network on a dataset of both normal and abnormal video sequences. The network could then be used to classify new video sequences as normal or abnormal.

– *Video prediction*: A fourth experiment could involve training a deep learning model on a dataset of normal video sequences, then using the model to predict the next frame of the video, abnormal activities can be detected by comparing the difference between the prediction and the real frame.

– *Transfer Learning:* Another experiment could involve fine-tuning pre-trained models on a dataset of the surveillance cameras, this could improve the performance of the model on the specific camera's data.

– *Data collection*: Collect a large dataset of surveillance video footage from different cameras and environments. This dataset should include both normal and abnormal (anomalous) behavior for training and testing purposes.

– *Preprocessing:* Preprocess the video data by resizing, cropping, and normalizing the frames. Additionally, annotate the data by labeling the normal and abnormal behavior.

– *Model selection:* Select an appropriate neural network architecture for anomaly detection, such as an auto encoder or a Variation Auto encoder (VAE).

– *Training:* Train the selected model on the preprocessed and annotated dataset.

– *Evaluation:* Evaluate the performance of the trained model by testing it on a separate dataset of surveillance video footage. Measure the model's accuracy in detecting anomalies using metrics such as precision, recall, and F1 score.

– *Fine-tuning*: Fine-tune the model by adjusting the parameters, such as the number of layers and neurons, or changing the architecture.

– *Deployment:* Deploy the final model on the surveillance cameras to detect anomalies in real-time.

It is important to note that for all these experiments, the quality and quantity of the data used to train the neural networks, as well as the specific circumstances and environment of the surveillance cameras, will greatly affect the accuracy of the anomaly detection.

The success of the model heavily relies on the quality and quantity of data used for training and testing. Additionally, it is important to consider the specific context and environment of the surveillance camera and adjust the model accordingly.

1.4 The relationship between cyber security and Detect Criminal Acts in Video Surveillance

The relationship between cyber security and detecting criminal acts in video surveillance is closely related. Video surveillance systems are typically connected to networks and use digital storage for the recorded footage, making them vulnerable to cyber-attacks.

– Cybersecurity threats: Video surveillance systems can be targeted by cyber-attacks, such as malware, ransomware, or unauthorized access attempts. This can compromise the integrity and availability of the recorded footage, rendering the surveillance system ineffective in detecting criminal acts.

– Data protection: Video surveillance footage may contain sensitive personal information, such as facial images, license plate numbers, and audio recordings. To protect this data from unauthorized access, disclosure or alteration, it is essential to implement appropriate cyber security measures, such as encryption, access controls and monitoring.

– Compliance: Video surveillance systems may be subject to various legal and regulatory requirements related to data protection and cyber security. It's important to ensure that the systems are designed and operated in compliance with these requirements in order to avoid legal and reputational risks.

– Network security: Video surveillance systems are often connected to other systems such as access control systems, intercoms, and alarms. It's essential to ensure that these systems are properly secured and that the video surveillance system is not used as a way to gain unauthorized access to other systems.

– Incident response: Video surveillance systems are a key element in incident response and investigations. It's important to have a plan in place to respond to and recover from cyber-attacks, and to ensure that the recorded footage is preserved as evidence.

Overall, cyber security is a critical aspect of detecting criminal acts in video surveillance systems, as it ensures the integrity, availability and confidentiality of the recorded footage, as well as compliance with legal and regulatory requirements.

Police use various methods to detect criminal acts in video surveillance footage, including:

– Manual review: Police officers manually review video footage to identify and analyze suspicious behavior or activities. This method is time-consuming and can be affected by human error or bias.

– Automated analysis: Police can use automated analysis tools, such as video analytics software, to process and analyze video footage. These tools can be used to detect specific behaviors or activities, such as loitering, parking in a restricted area, or crossing a virtual boundary.

– Facial recognition: Police can use facial recognition technology to match faces in video footage to a database of known individuals. This can be used to identify suspects or missing persons.

– Object recognition: Police can use object recognition technology to detect and classify objects in video footage, such as vehicles, weapons, and clothing.

– Video search and retrieval: Police can use video search and retrieval tools to quickly and efficiently search large amounts of video footage to identify specific individuals or vehicles.

– Multi-modal approaches: Police can use multiple modalities such as visual, audio, and metadata information to detect criminal acts.

– Crowd-sourced analysis: Police may use crowd-sourced analysis platforms that allow members of the public to flag suspicious activity or individuals in video footage.

It's important to note that these methods can be used in combination, and the police prefer to use a combination of manual and automated methods to detect criminal acts in video surveillance footage. Additionally, it's essential to consider

the ethical and legal implications of using these technologies, such as privacy concerns and potential biases before deploying any method.

The best way to form the input dataset to detect criminal acts in video surveillance depends on the specific circumstances and environment of the surveillance, as well as the method used for detecting criminal acts. However, here are some general guidelines for forming the input dataset in an optimal way:

– Collect a large and diverse dataset: Collect a large and diverse dataset of video footage from surveillance cameras that includes criminal acts. This will ensure that the model is trained on a wide range of data variations and can generalize well to new data.

– Annotate the data: Annotate the video footage with labels indicating the presence or absence of criminal acts, such as fighting, stealing, or loitering. This will enable the model to learn from the data and make predictions about criminal acts.

– Use high-quality data: Use high-quality data, such as footage from cameras with high resolution and good lighting conditions. This will ensure that the model can accurately detect criminal acts in the footage.

– Consider privacy: Consider the privacy of individuals in the footage and remove any sensitive information, such as facial images, license plate numbers, or audio recordings, that is not necessary for the task.

– Take into account the legal and regulatory requirements: Take into account the legal and regulatory requirements related to data protection and privacy and ensure that the input dataset is compliant with them.

Additionally, it's important to note that Forming the input dataset to detect criminal acts in video surveillance is a complex task that depends on the specific circumstances and environment of the surveillance, as well as the method used for detecting criminal acts.

2 CLOSED CIRCUIT TELEVISION

Closed circuit television (CCTV) is a surveillance technology. More specifically, it is “a system in which a number of video cameras are connected in a closed circuit or loop, with the images produced being sent to a central television monitor or recorded. Technological advancements now allow CCTV systems to work on wireless networks, operated remotely, and be watched from several locations. The term closed circuit television was originally used to differentiate between public television broadcasts and private camera-monitor networks. These days CCTV is used as a generic term for a variety of video surveillance technologies including Police Observation Devices (POD) or Portal Overt Digital Surveillance Systems (PODS).

2.1 Closed circuit television Technology

Although some systems are extremely sophisticated, employing bullet-proof casing, color recording, night-vision capability, motion detection, gunshot detection, and advanced zoom and automatic tracking capacities, many existing systems are more rudimentary. More common CCTV installations include a number of cameras connected to either a control room where human operators watch a bank of television screens or an unmonitored data storage system.

Closed Circuit Television Systems (CCTV) are becoming more and more popular and are being deployed in many housing estates, offices, and also in most public spaces. As the number of camera views a single CCTV operator can handle is limited by human factors, such monitoring systems makes for an enormous load for the CCTV operators.

The purpose of the Intelligent Surveillance System is to alert the human operator when: (1) Presence of a dangerous act, after setting zones of interest and danger zones within those zones of interest, the danger is detected when an object trespasses the danger zone, which can reduce the number of accidents in places, (2) An abnormal behavior of a person such as handling some weapons or act of abuse or molestation is detected, which might be a potential threat.

The public (and offenders) can clearly see the surveillance camera and determine the direction in which it is facing. Although most CCTV schemes

employ overt cameras, which are obvious (see Figure 2.1), it is possible to find systems in which cameras are mounted into protective shells or within frosted (polycarbonate) domes. Often termed semi-covert, these camera systems make it more difficult for people under surveillance to determine if they are being watched, as it is usually difficult or impossible to figure out in which direction the camera is facing (see Figure 2.2).

Figure 2.1 – Overt CCTV camera

Some cameras employ dummy lenses to conceal the surveillance target. The advantage of using a one-way transparent casing is that it provides for the possibility of retaining the overt impression of surveillance and hence a deterrent capacity without having to place a camera in every housing or to reveal to the public (and offenders) the exact location under surveillance.

Figure 2.2 – Semi-covert CCTV camera that may have a crime prevention advantage over an overt system because offenders can never be sure in which direction that camera is facing

There is a range of CCTV configurations available. A complete CCTV system (for the purposes of this report) comprises:

– One or more cameras that view a public area

– A mechanism to transmit video images to one or more monitors

– Video monitors to view the scene usually accompanied by recording devices such as a time-lapse video recorder or computer hard drive for digital images

– A viewer or camera operator, such as a police officer or security guard

Variations to this basic configuration include:

– The ability to transmit images across the Internet

– Motion sensors that activate the camera when activity is detected

– Normal or infrared lighting to enhance picture quality at night

– A pan-and-tilt capacity that allows an operator to change the camera's viewing direction, zoom, and focus

More-advanced systems can include limited facial recognition technologies or estimate the location of firearm incidents, though more advanced systems often rely on other technology. For example, a facial recognition program is of limited value unless it is linked to a computer database of suspect photos. Intelligence systems that can detect unusual activity (such as fights in the street) are also under development.

In addition to determining if you want to install a CCTV system (and what type), you should consider how sophisticated you want it to be and if you have the resources to support it.

2.2 How CCTVs Prevent Crimes

The real question: Can CCTVs help you detect crime and potential security breaches before they happen and prevent serious losses?

CCTV's Need Trained Security Guards at the Helm

How can CCTV prevent crime? The answer is through the person observing their output. Though CCTVs operate automatically, they're at their most effective with a person behind the monitor. Indeed, no CCTV can stop a break-in by itself, but with the aid of an alert security guard, they can provide advanced notice that will help stop a crime in progress or before it even begins.

Location Matters for CCTV Success

Though people are an important part of any security system, specific situations and environments matter too. Suppose your business is situated in a high-crime area, a place where people have grown used to seeing cameras peeping out at them. In such places, a CCTV system won't do much to deter break-ins. In a low-crime area, though, it would be an entirely different matter. The mere thought of a lens recording a perpetrator's actions stops many break-ins before they even start. It seeks to change offender perception so the offender believes if he commits a crime, he will be caught.

How CCTVs can Help Reduce Crime Outside Your Property

If CCTV reduces crime (and it does), just how does it manage it? The analysis conducted by Criminology & Public Policy showed that such systems have multiple effects. CCTVs force criminals to think twice before attempting a crime, make security personnel and bystanders more vigilant, and assist police in apprehending criminals.

How can CCTV Security Systems Protect the Business?

Although the number of crimes solved by surveillance continues to rise, there's more to surveillance cameras than simply documenting crimes caught on camera. Part of the reason why security cameras are important goes beyond the technology itself (although well-functioning systems certainly provide vital protection). Simply prepping your property for proper CCTV installation helps reduce crime. Consider that you'll need to remove visual clutter, keep outside areas well lit, and have hedges and trees carefully pruned. Such steps both help CCTVs reduce burglary and facilitate CCTV crime prevention.

Real-life Application of CCTV Security Systems

CCTV security systems do a lot more than merely deter theft. In addition to providing coverage of retail and multi-family residential establishments, they also offer the following benefits:

– Ensuring employee compliance with safety standards

– Deterring the theft of fixed assets such as copper tubing, HVAC units, construction equipment, etc.

– Monitoring animals in zoos, wildlife refuges, national parks, and safaris

– Examining operations to improve productivity

– Bolstering public safety in parks and other public spaces

– Monitoring highways to improve traffic flow and manage accidents

The Impact of Feeling Safer with a CCTV System

Another benefit of CCTV is that it helps people feel safer. For instance, a study by European Journal of Criminology that focused on publicly placed surveillance cameras stated that the public felt more comfortable in the city after dark. More research conducted by the University of Virginia's Curry School of Education and Human Development stated that high-school students felt safer when cameras were trained on parking lots, entrances, and exits. CCTVs generally have a powerful psychological effect.

Advanced Technology: Not All CCTVs Are Equal

The fact that CCTVs come with advanced features also has an impact. In addition to making people feel better, these high-tech options increase their effectiveness. Once a break-in begins, the presence of equipment that detects sounds, arms when alerted to movement, and syncs with mobile devices helps alert property owners virtually instantaneously.

2.3 Benefits about using CCTV

Aid to Police Investigations

Regardless of the potential for a CCTV system to have a role in crime prevention, it can still make a contribution in a detection role. There are numerous examples of CCTV tapes aiding in an offender's conviction. Camera footage can also help identify potential witnesses who might not otherwise come forward to police, help investigators narrow the time window when a crime occurred, establish a sequence of events, capture images of getaway vehicles, and help locate weapons used during the crime. CCTV camera evidence can be compelling, though issues of image quality are a factor if CCTV images are used for identification purposes. If the cameras record an incident, and police respond rapidly and make an arrest within view of the camera (and the offender does not leave the sight of the camera), the recording of the incident can help investigators gain a conviction, usually through a guilty plea. The potential to assist in police investigations may also drive offenders away from committing offenses that take time, as they run a greater risk of capture.

Provision of Medical Assistance

As a community safety feature, CCTV camera operators can contact medical services if they see people in the street suffering from illness or injury as a result of criminal activity (such as robberies and assaults) or non-crime medical emergencies. The ability to summon assistance is a public safety benefit of CCTV. In addition to summoning assistance, when live feeds are being monitored, operators can direct responding personnel to the individual once they are at the scene and can help them avoid any potential dangers. Squires found that police are called about 10 to 20 times for every 700 hours of observation.

Place Management

CCTV can be used for general location management. The cameras can be used to look for lost children, to monitor traffic flow, public meetings, or demonstrations that may require additional police resources, or to determine if alarms have been activated unnecessarily thus removing the need for a police response. Actively monitored CCTV systems also allow for spotting a crime that is occurring before someone calls the emergency, promoting a quicker response. Quicker responses at locations may also reduce the escalation of crime as it happens. Some police commanders claim that assaults on police have reduced because the cameras allow them to determine the appropriate level of response to an incident, either by sending more officer to large fights, or by limiting the number of officers to a minor incident and avoid inflaming the situation.

Information Gathering

Cameras can also be used to gather intelligence and to monitor the behavior of known offenders in public places (such as shoplifters in public retail areas). Camera operators often come to know the faces of local offenders, and the cameras become a way to monitor their movements in a less intrusive manner than deploying plainclothes police officers. For example, officers in one city were able to gather intelligence on the behavior of individuals selling stolen goods. This intelligence was gathered remotely by CCTV cameras and enabled police to interdict in an organized and coordinated manner. Although intelligence gathering is a potential benefit of CCTV, the use of intelligence gathered from CCTV to control public order through surveillance is perceived by some to be a threat to civil liberties.

2.4 Kinds of CCTV Cameras

Overt Systems

Overt camera systems are common. The cameras are in view of the public and are often accompanied by signs indicating that people are now in a CCTV surveillance area. Overt systems have a strong crime prevention rationale but are more vulnerable to tampering and vandalism.

Semi-Covert Systems

These systems are in public view, but the cameras are concealed behind a one-way transparent casing. This approach retains most of the preventative rationale of the overt system, but the cameras have some protection. It also prevents the public from determining who is under surveillance and allows you to conceal the exact number of cameras in a system, as you are not required to install a camera in every casing.

Covert Systems

With these systems, the aim is to hide camera locations. These systems are particularly well suited to crime detection; however, without public signage or a publicity campaign, they have little crime prevention function until word spreads within the offender community. These cameras are fairly immune to tampering.

3 CRIMINAL DETECTION AT EARLY STAGES BY USING 3D CONVOLUTIONAL NEURAL NETWORKS

Crime generates significant losses, both human and economic. Every year, billions of dollars are lost due to attacks, crimes, and scams. Surveillance video camera networks generate vast amounts of data, and the surveillance staff cannot process all the information in real-time. Human sight has critical limitations. Among those limitations, visual focus is one of the most critical when dealing with surveillance. For example, in a surveillance room, a crime can occur in a different screen segment or on a distinct monitor, and the surveillance staff may overlook it.

3.1 Usage of video surveillance

The main idea is to focuses on crimes by analyzing situations that an average person will consider as typical conditions, but may eventually lead to a crime. While other approaches identify the crime itself, we instead model suspicious behavior (the one that may occur before the build-up phase of a crime) by detecting precise segments of a video with a high probability of containing a crime.

Real-time analysis of surveillance cameras has become an exhaustive task due to human limitations. The primary human limitation is the Visual Focus of Attention (VFOA). The human gaze can only concentrate on one specific point at once. Although there are large screens and high-resolution cameras, a person can only pay attention to a small segment of the image at a time. Optical focus is a significant human-related disadvantage in the surveillance context. A crime can occur in a different screen segment or on a different monitor, and the staff may not notice it. Other significant difficulties may be related to attention, boredom, distractions, lack of experience, among others.

Vigilance camera networks generate vast amounts of video screens, and the surveillance staff cannot process all the available information as fast as needed. The more recording devices become available, the more complex the task of monitoring such devices becomes.

Defining what can be considered suspicious behavior is usually tricky, even for psychologists. In this work, the mentioned behavior is related to the commission of a crime, but it does not imply its realization (Figure 1). For this

research, I define suspicious behavior as a series of actions that happen before a crime occurs. In this context, our proposal focuses on crime scenarios, particularly before the build-up phase situations that an average person may consider as typical conditions. The crimes usually take place in supermarkets, malls, retail stores, and other similar businesses. Many of the models for addressing this problem need the suspect to commit a crime to detect it. Examples of such models include face detection of previous offenders and object analysis in fitting rooms.

In this work, we will see an approach to support the monitoring staff to focus on specific areas of screens where crime is more likely to happen. While existing models identify the crime itself, we model suspicious behavior as a way to anticipate a potential crime. In other words, we identify behaviors that usually take place before a crime occurs. Then, the system can label a video: as containing suspicious or normal behavior. By detecting situations in a video that may indicate that suspicious behavior is present, the system indicates that a crime is likely to happen soon.

The former gives the surveillance staff more opportunities to act, prevent, or even respond to such a crime. In the end, it is the security personnel who will decide how to proceed in each situation.

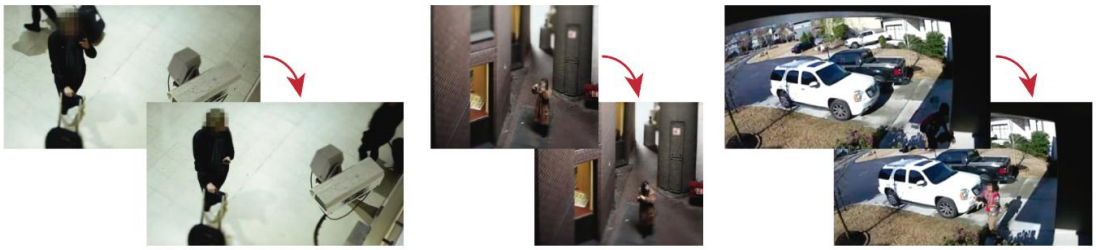


Figure 3.1 – Different situations may be recorded by surveillance cameras. Suspicious behavior is not the crime itself. However, particular situations will make us distrust a person if we consider their behavior to be “suspicious”

Overall, there's method to extract segments from videos that feed a model based on a 3D Convolutional Neural Network (3DCNN) for classifying behavior (as normal or suspicious). Once we train the model with such segments, it accurately classifies the behavior on a video dataset composed of daily action samples and shoplifting samples. The results suggest that the proposed approach has applications in crime prevention in places cases.

As a summary, this work contributes to the literature mainly in three aspects.

– It describes a methodology, the PCB method, to unify the processing and division of criminal video samples into useful segments that can later be used for feeding a Deep Learning (DL) model.

– It represents the first implementation of a 3DCNN architecture to detect criminal intentions before an offender shows suspicious behavior.

– It provides a set of experiments to validate the results, confirming that the proposed approach is suitable for such a challenging task: to detect criminal intention even before the suspect begins to behave suspiciously.

3.2 Background and Related Work

A surveillance environment must satisfy a particular set of requirements. Those requirements have promoted the creation of specialized tools, both on equipment and software, to support the surveillance task. The most common approaches include motion detection, face recognition, tracking, loitering detection, abandoned luggage detection, crowd behavior, and abnormal behavior. Prevention and reaction are two primary aims in the surveillance context.

Prevention requires forestalling and deterring crime execution. The monitoring staff must remain alert, watch as much as possible, and alert the ground personnel. Reaction, on the other hand, involves protocols and measures to respond to a specific event. The security teams take action only after the crime or event has taken place.

Most security support approaches focus on crime occurrence. Tsushita and Zin presented a snatching-detection algorithm, which performs background subtraction and pedestrian tracking to make a decision. Their approach divides the frame into eight areas and searches for a speed shift in one tracked person. Unfortunately, Tsushita and Zin’s algorithm can only alert when a person has already lost their belongings. proposed a violence detection framework combining a trained MobileNet-SSD model for person detection and a C3D model. Besides, they optimize the trained model with the OPENVINO toolkit. They test their model with three different violence datasets: violent crowd, violence in movies, and hockey fight.

Presented a real-world anomaly detection approach, training 13 anomalies, such as burglary, fighting, shooting, and vandalism. They use a 3DCNN for feature

extraction and label the samples into two categories: normal and anomalous. Their model includes a ranking loss function and trains a fully connected neural network for decision-making. In a similar context.

Presented a deep anomaly detection approach. They implemented a bilateral background subtraction, use the pertained C3D model for feature extraction, and attached a fully connected network to perform regression. Using the UCF-Crime dataset, they trained their model on 11 complete classes and tested their results on “robbery”, “fighting”, and “road accidents”. Ishikawa and Zin proposed a system to detect loitering people.

Their system combines grid-based analysis, direction-based analysis, distance-based analysis, acceleration based analysis, and a decision-fusion stage of the people shown in the video to make a decision.

Convolutional Neural Networks (CNN) have shown a remarkable performance in computer vision and other different areas in the last recent years. Particularly, 3DCNNs an extension of CNN focus on extracting spatial and temporal features from videos. Some interesting applications that have been implemented using 3DCNN include object recognition, human action recognition, gesture recognition, and —particularly related to this work— behavior analysis from customers in the banking sector. Although all the works mentioned before involve using a 3DCNN, each one has a particular architecture and corresponding set of parameters to adjust. For example, concerning the number of layers, many approaches rely on simple structures that consist of two or three layers, while others require several layers for exhaustive learning.

Concerning a crime in places, the current literature is somewhat limited. Surveillance material is, in most cases, a company’s private property. The latter restricts the amount of data available for training and testing new surveillance models. For this reason, several approaches focus on training to detect normal behavior. Anything that lies outside the cluster is considered abnormal. In general, surveillance videos contain only a small fraction of crime occurrences. Then, most of the videos in the data are likely to contain normal behavior. Many approaches have experienced problems regarding the limited availability of samples and their unbalanced category distribution. For this reason, some works have focused on developing models that learn with a minimal amount of data.

The work aims at developing a support approach for crime prevention. Our model detects a person that, according to their behavior, is likely to commit a

crime in places. We achieve the latter by analyzing the people's comportment in the videos before the crime occurs.

3.3 Methodology

As part of this work, I propose a methodology to extract segments from videos where people exhibit behaviors relevant to crime in the places. The methodology considers both normal and suspicious behaviors, being the task to classify them accordingly. The following lines describe the dataset used and how split it for experimental purposes, the pre-crime Behavior (PCB) method, and the 3DCNN architecture used for feature extraction and classification.

Description of The Dataset

Among the many works related to surveillance security, the analysis of non-verbal behavior is one of the less researched areas. This generates a lack of enhancement of security protocols and available information. Many works build their datasets using actors. However, they cannot catch the essential behavioral cues that an offender may show in a stressful situation. Some types of crimes have been more explored, such as crowd behavior, vandalism, fights, or assaults. For non-violent crimes, such as shoplifting, pickpocketing, or theft, it is harder to detect the crime in public places and get access to the videos.

The dataset includes scenarios from several people and locations, which are grouped into 13 classes such as "abuse", "burglary", and "explosion", among others. We extracted the samples used in this investigation from the "crime" and "normal" classes from the UCF-Crime dataset.

To feed this model, they require videos that show one or more people whose activities are visible before the crime is committed. Due to these restrictions, not all the videos in the dataset are useful. Suspicious behavior samples were extracted only from videos that exhibit a crime, but to be used by this system, such samples must not contain the crime itself. Conversely, normal behavior samples were extracted from the "normal" class. Thus, it is important to stress that the model we propose is a behavior classifier (normal or suspicious) and not a crime classifier.

Data augmentation techniques aim to increase the number of useful examples in the training dataset, producing variations of the original images that the model is likely to see. Examples of these techniques include flipping, rotation, zoom, and brightness. It is relevant to mention that many of these techniques are

not useful in our work. For example, vertical flipping an image makes no sense in our system as the videos will never be watched upside down. Rotation turns the image clockwise an arbitrary number of degrees, but it may drop pixels out of the image and produce areas with no pixels, which have to be filled in somehow.

Zoom augmentation either adds new pixels around the image (zoom out) or leaves out part of the original image (zoom in), leading to losing or altering the scene's information. The situations derived from using such data augmentation techniques could potentially do more harm than good and, for that reason, were not considered for this work.

The Pre-Crime Behavior Method

Video sample segmentation does not follow a specific methodology in criminal intentions and suspicious behavior analysis. This makes it unreliable for creating a benchmark and testing a model across different video sets. For example, in some investigations, the segmentation is left to the experts' judgement. In others, the researchers select the frame before the criminal act. In some particular cases, there is no segmentation at all.

The Pre-Crime Behavior (PCB) method arises as a new proposal to unify moments, such as the build-up phase and the crime itself, and provide a new segment to the analysis, the suspect's behavior before any aggression attempt. It is composed of four steps that allow the identification of four specific moments in the video sample. The PCB method is described as follows.

1- Identify the instant where the offender appears for the first time in the video. We refer to this moment as the First Appearance Moment (FAM). The analysis of suspicious behavior starts from this moment.

2- Detect the moment when the offender undoubtedly commits a crime. This moment is referred to as the Strict Crime Moment (SCM). This moment contains the necessary evidence to argue the crime commission.

3- Between the FAM and the SCM, find the moment where the offender starts acting suspiciously. The Comprehensive Crime Moment (CCM) starts as soon as we detect that the offender acts suspiciously in the video.

4- After the SCM, locate the moment where the crime ends (when everything seems to be ordinary again). If the video sample started from this instant, we would have no evidence of any crime committed in the past. This moment is known as the Back to Normality Moment (B2NM).

Please note that, as a sample video from the UCF-Crime dataset may contain more than one crime, the PCB method is applied once for each crime occurrence. Then, sometimes we can extract various suspicious behavior samples from the same video in the UCF-Crime dataset.

The output of the PCB method comprises four moments per crime in the input video. These four moments divide each sample into three relevant segments, as described below.

Pre-Crime Behavior Segment (PCBS). The PCBS is the video segment between the FAM and the CCM. This segment has the information needed to study how people behave before committing a crime, even acting suspiciously. Most human observers will fail to predict that a crime is about to occur by only watching the PCBS.

Suspicious Behavior Segment (SBS). The SBS is the video segment contained between the CCM and the SCM. The SBS provides specific information about an offender's behavior before committing a crime.

Crime Evidence Segment (CES). The CES represents the video segment included between the SCM and the B2NM. This segment contains the evidence to accuse a person of committing a crime.

For the sake of clarity, we present the four moments and the three segments derived from the PCB method graphically, as depicted in Figure 3.2.

Figure 3.2 – Video segmentation by using the moments obtained from the Pre-Crime Behavior Segment (PCB) method

To extract the samples from the videos, we follow the process depicted in Figure 3.3. Given a video that contains one or more places crimes, we identify the precise moment when the offense is committed. After that, we label the different suspicious moments, moments where a human observer doubts what a person in the video is doing. Finally, we select the segment before the suspect is preparing to

commit the crime. These segments become the training samples for the Deep Learning (DL) model.

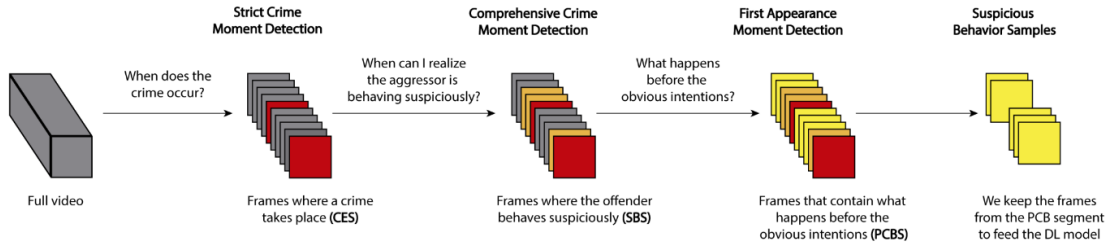


Figure 3.3 – Graphical representation of the process for suspicious behavior sample extraction

In a video sample, each segment has particular importance regarding the information it contains (see Figure 3.2). The PCB segment has less information about the crime itself, but it allows us to analyze the suspect's normal-acting behavior when they appear for the first time, even far from a potential crime. The SBS allows us to have a more precise idea about who may commit the crime, but it is not conclusive. Finally, the CES contains the doubtless evidence about a person committing a crime. If we remove both the SBS and the CES from the video, the result will be a video containing only people shopping, and there will be no suspicion or evidence that someone commits a crime. That is the importance of the accurate segmentation of the video. From the end of a CES until the next SBS, there is new evidence about how a person behaves before attempting a crime.

For experimental purposes, we only use the frames from the PCBS in this work. As these segments lack specific criminal behavior, they have no information about any transgression. The PCB segments are ideal for feeding our 3DCNN model, aiming to characterize the people's behavior. The objective of the model is to identify when such behavior is suspicious, which may indicate that a crime is about to be committed.

3D Convolutional Neural Networks

For this work, we use a 3DCNN for feature extraction and classification. 3DCNN is a recent approach for spatio-temporal analysis that has shown remarkable performance in processing videos in different areas, such as moving objects action recognition, gesture recognition, and action recognition. We decided to implement a 3DCNN in a more challenging context, such as searching for patterns in video samples, which lack suspicious and illegal visual behavior.

We should employ a basic structure to explore the performance of the 3DCNN for behavior classification. The model comprises four Conv3D layers (two pairs of consecutive convolutional layers for capturing long dependencies), two max-pooling layers, and two fully connected layers. As a default configuration, in the first pair of Conv3D layers, we apply 32 filters, and for the second pair, 64 filters. All kernels have a size of 3 × 3 × 3, and the model uses an Adam optimizer and cross-entropy for loss calculation. The graphical representation of this model is shown in Figure 3.4.

The last part of the model contains two dense layers with 512 and two neurons, respectively. This architecture was selected because it has been used for similar applications, and it seems suitable as a first approach for behavior detection in surveillance videos.

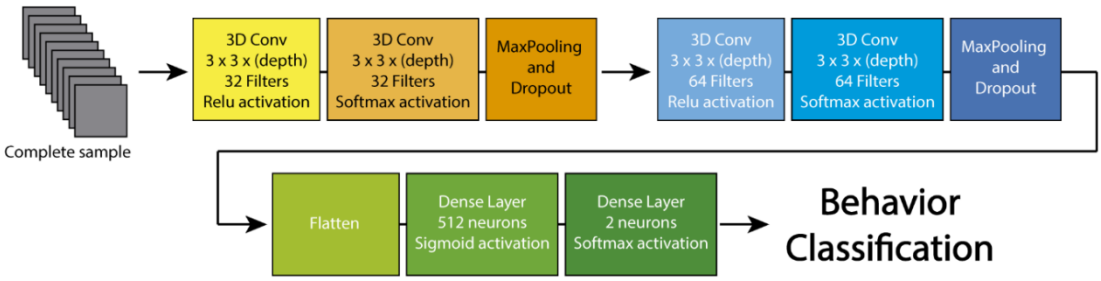


Figure 3.4 – Architecture of the DL Model used for this investigation. The depth of the kernel for the 3D convolution is adjusted to 10, 30, or 90 frames, according to each particular experiment

Metrics

As the decisive metric to analyze the results, we considered the accuracy (Equation (1)). It considers the correct hits, true positive (TP) plus true negative (TN), over the total number of samples evaluated (FP and FN represent false positives and false negatives, respectively).

$$\text{accuracy} = \frac{TP+TN}{TP+TM+FP+FN}. \qquad (3.1)$$

As accuracy shows the general performance of the model, we complement its information by presenting the confusion matrices of the best runs. These matrices allow checking, in detail, the model capability to classify suspicious and

normal behavior. We used two additional metrics for adequately analyzing the results from the confusion matrices: precision (Equation (2)) and recall (Equation (3)). Precision indicates the proportion of samples classified as suspicious that are, in fact, suspicious, a model with a precision of 1.0 produces no FP. Recall indicates the proportion of actual suspicious samples that were correctly classified by the system, a model with a recall of 1.0 produces no FN.

$$\text{precision} = \frac{TP}{TP+FP}, \tag{3.2}$$

$$\text{recall} = \frac{TP}{TP+FN}. \tag{3.3}$$

3.4 Experiments and Results

We conducted a total of six experiments in this work. These experiments (six experiments) are divided into two categories: preliminary and confirmatory. The first four experiments are preliminary as they focus on exploring the effect of different configurations under different scenarios, aiming to find some suitable configurations that may lead to better model performance. I refer to the last two experiments as confirmatory as we tested the system on more challenging configurations derived from the preliminary experiments, to validate the approach. Among all these experiments, a total of 708 models were generated and tested. Although the specific details of each experiment are detailed in its corresponding description, for the ease of the reader, have provided an overview of our experimental setup in Figure 3.5.

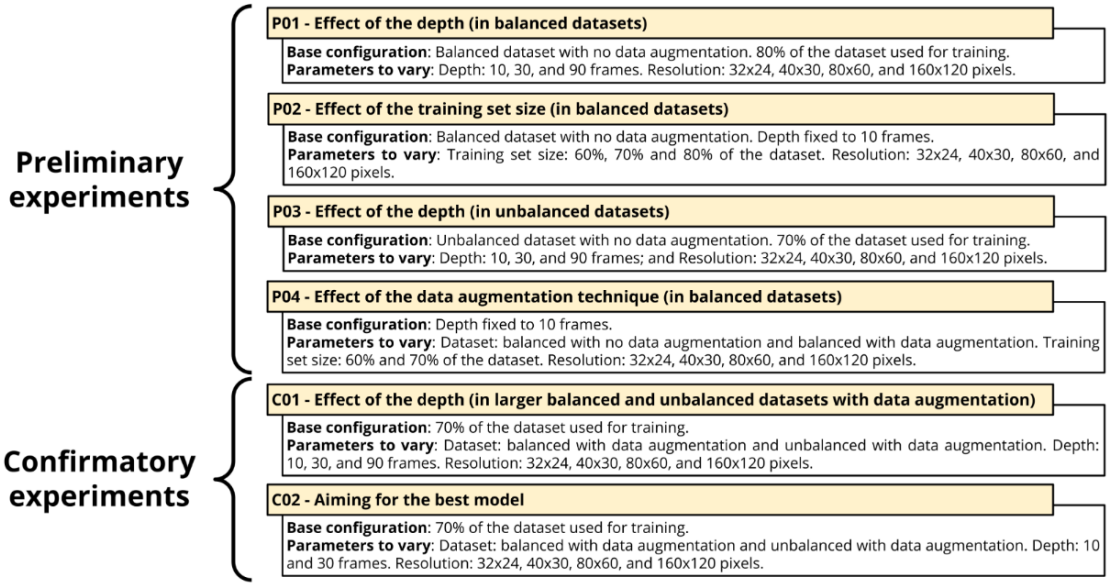


Figure 3.5 – Overview of the experimental setup followed in this work. For a detailed description of the parameters and the relation of the samples considered for each experiment

3.5 Summary section 3

For this section, I have focused on the behavior performed by a person during the build-up phase of a crime. The neural network model identifies the previous conduct, looking for suspicious behavior, and not recognizing the crime itself. This behavior analysis is the principal reason why we remove the committed crime segment from the video samples, to allow the artificial model to focus on decisive conduct and not in the offense. We implement a 3D Convolutional Neural Network due to its capability to obtain abstract features from signals and images, based on previous action recognition and movement detection approaches.

The final intention of this chapter is to develop a tool capable of supporting the surveillance staff, presenting visual behavioral cues, and this work is a first step to achieve the mentioned goal.

In these experiments, used a selected number of videos from the UCF-Crimes dataset. As future work, and aiming at testing this model in a more realistic simulation, we should increase the number of samples, preferably the normal behavior ones, to create a bigger sample imbalance between classes. Another exciting aspect of the development of this project is expanding our behavior detection model to other contexts. It exists many situations where we can find suspicious behavior, such as stealing, arson intents, and burglary. will gather

videos of different contexts to strengthen the capability to detect suspicious behavior. Finally, the automation of the PCB method for video segmentation stands out as an interesting point to explore. This will reduce the preprocessing time, which would allow analyzing a larger amount of data. For this reason, we consider this an important path for future work derived from this investigation.

4 DETECTION OF ABNORMAL BEHAVIOR BY A SURVEILLANCE CAMERA IMAGE

At present, an enormous amount of accidents and terrorisms has been occurred all over the world. Due to the spread of security cameras, the number of occurrences of theft and robbery incidents has been decreasing more and more. Nonetheless, the arrest rate has not improved so much and improvement and rising of the arrest rate are required.

4.1 Abnormal Behavior

The objective of this section is detection of snatching that involves an event between two persons, and we made an effort to detect snatching in various kinds of situations by using some video scenarios. This video scenario includes the scene of snatching with a bicycle and the scene of non-snatching with normal pedestrian passing. the proposed methods consist of several steps: background subtraction, pedestrian tracking, feature extraction, and snatch theft detection.

We should have focused on the feature extraction process in details and used weighted decision fusion system based on these parameter, area feature, motion feature, and appearance feature in the paper. in this work attempt to detect the snatching event from diverse features.

Snatching is a criminal act that occurs frequency comparing with other serious criminals, and it is characterized by that socially vulnerable people such as the elderly and women are easily targeted.

In this section, we aimed to develop a surveillance camera system that automatically detects the occurrence of snatches on the street. Surveillance cameras have been wide spreading, and active detection of incidents is also required for cost reduction of security problems.

Objectives

In this section, made the system of detecting the snatch theft. The main purpose is to detect the theft such as snatching in various situations. Moreover, we would like to detect the events of snatching with bicycle and the events of snatching in various environments in real time from the occurrence situations of snatching recently.

I believe that development of this research is necessary because it leads to cost reduction of security problems and prevention of accident incidents.

4.2 Proposed Method

In this section, I proposed an effective method for detecting snatch thieves. The overall system flowchart is shown in Fig.4.1. This proposed system consists five steps to detect snatching event. Moreover, introduced the point system to detect snatching event efficiently by using three feature parameters, area, motion, and appearance. We need to find out its characteristics, identify it as abnormal behavior and recognize it. These features are used in this study to detect whether snatching event has occurred and to distinguish between normal pedestrian and suspicious person.

Fig.4.1 – The overall flowchart to detect snatch theft

Background Subtraction

Foreground image is made by difference between background and input image. It is not accurately extracted if we calculate the amount of white pixel without removing of the shadow of person silhouette. That's why we cut the silhouette about 10% from the bottom of bounding box (BB). The process result is shown in Fig. 4.2

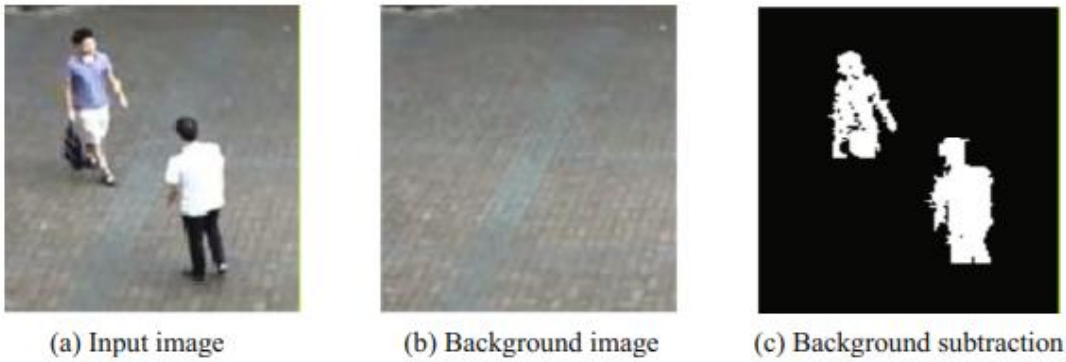


Figure.4.2 – The process of background subtraction and cutting the shadow

Feature Extraction

This step is main process in the proposed system to detect the snatching. As mentioned in some of the previous chapters, it is possible to judge abnormal behavior by extracting a person from the input image and considering the characteristic of the person.

We compared with the intersection events before and after, and the change amount compared with before and after, is used as feature value (FV). We describe these features in details from following.

Spatial Feature

This feature value is most weighted in the point system. This feature is used to gain the information in which direction the two persons in the video frame are moving. Also focused only on frames after intersection has occurred in this area feature. we divided the area where a person moves in eight areas (Fig.4.3). Then if the moving directions of the two persons after the intersection are the same, it is identified that there is a high possibility that snatching has occurred in the spatial feature.

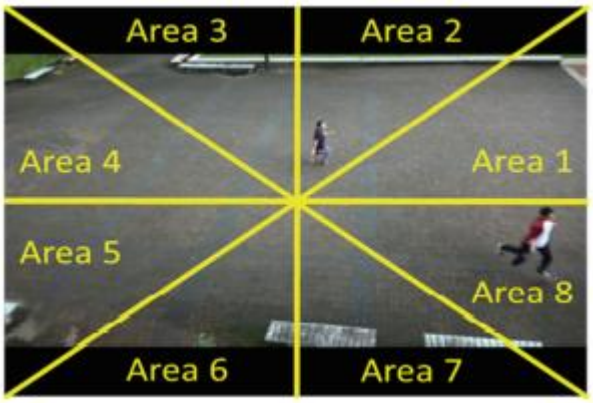


Figure.4.3 – The frame is divided into eight areas

Motion Feature

The feature of the movement used this time is not just the moving velocity but the moving acceleration of a person. In the case of a person who is just walking in the frame, no significant change in the moving acceleration can be seen because of walking at a constant speed in the frame. On the other hand, after a suspect rubs a bag, its feature value of acceleration is estimated to take high value because a suspect runs away. The difference of acceleration between walking and running

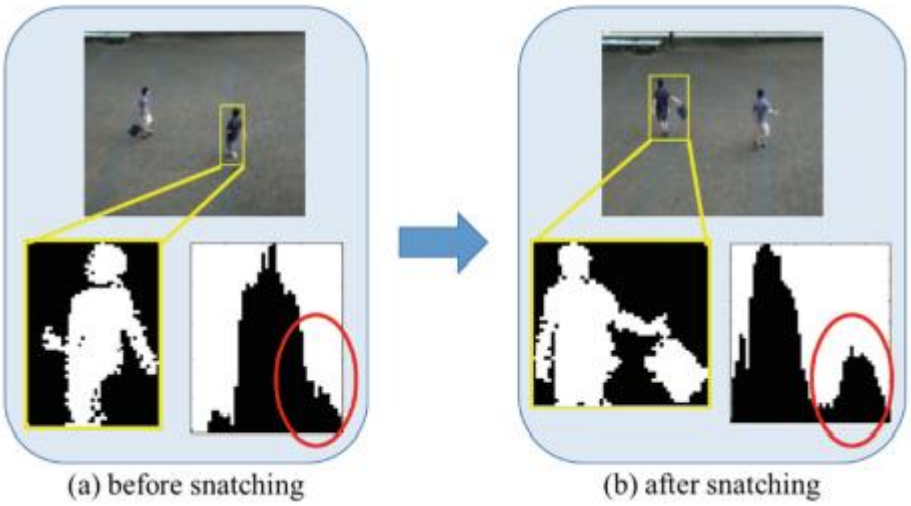
person is shown in Fig.4.4. Moreover, we focused on the frames before and after intersection in this feature. The average values are taken using the feature values every five frames before and after intersection. The equation for calculating the feature value, FVM of the moving acceleration for every five frames is shown in below.

(a) walking pattern (b) running pattern

Figure 4.4 – The difference in acceleration between walking and running person

Appearance Feature

The reason for adopting the feature of appearance originally is that the silhouette of the bag increases and decreases in that case the snatching event happens. The histogram of before and after snatching event is shown in Fig.4.5. Even with appearance feature, we focused on the frames before and after intersection as well as the feature of motion.



(a) before snatching (b) after snatching

Figure 4.5 – Comparing with the histogram in a victim case

Also we take the average value of each before and after intersection by using the feature value every five frames. The equation for calculating the feature value, FVA of the appearance for every five frames is shown in below.

Weighted Point in the System

As a method of determining the specific gravity, it was decided considering the occurrence situation and various environments snatching event happens in diverse scenes. This proposed system has 10 points in total, and we classified into three outputs “Snatching”, “Potential Snatching”, “Non Snatching”. The flowchart of classification is shown in Fig.4.6.

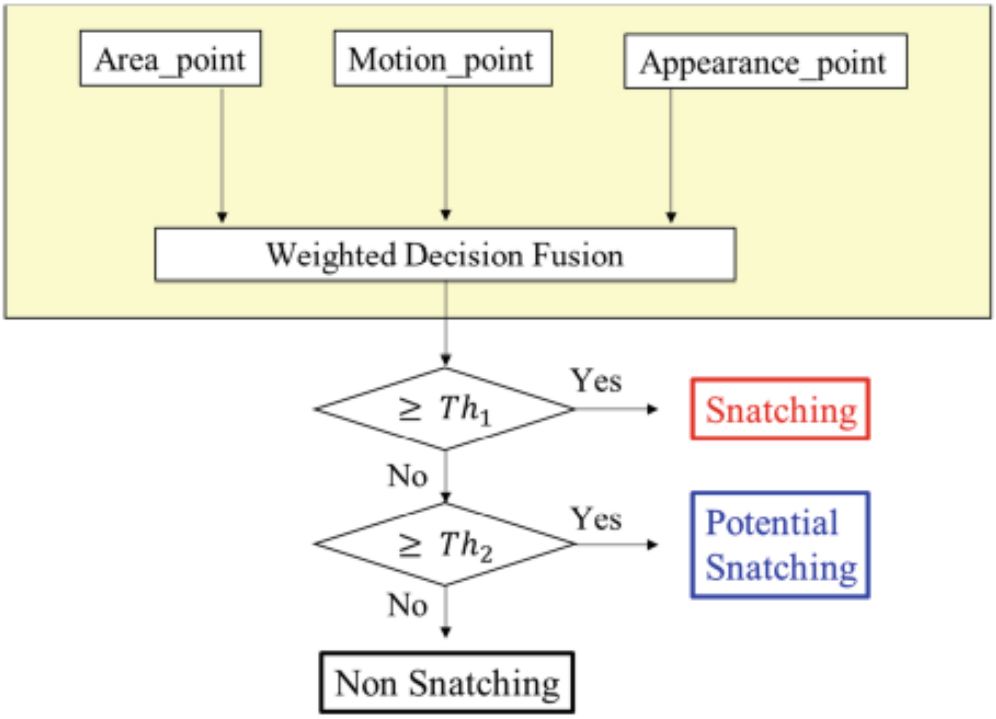


Figure 4.6 – The flowchart of classification into each output

Setting the Threshold

I used 4 training videos to set the threshold in this work. The contents of the events that occur within the training videos are as follows: snatching with bicycle, snatching without bicycle, non-snatching (crossing each other), and non-snatching (passing the bag). The state of the events that occur in the training videos is shown in Fig.4.7. By including all these different kinds of scenes in the training videos and setting threshold values, it was made possible to correspond to various kinds of test video. Each feature value in each of training videos is shown in Table 4.1.

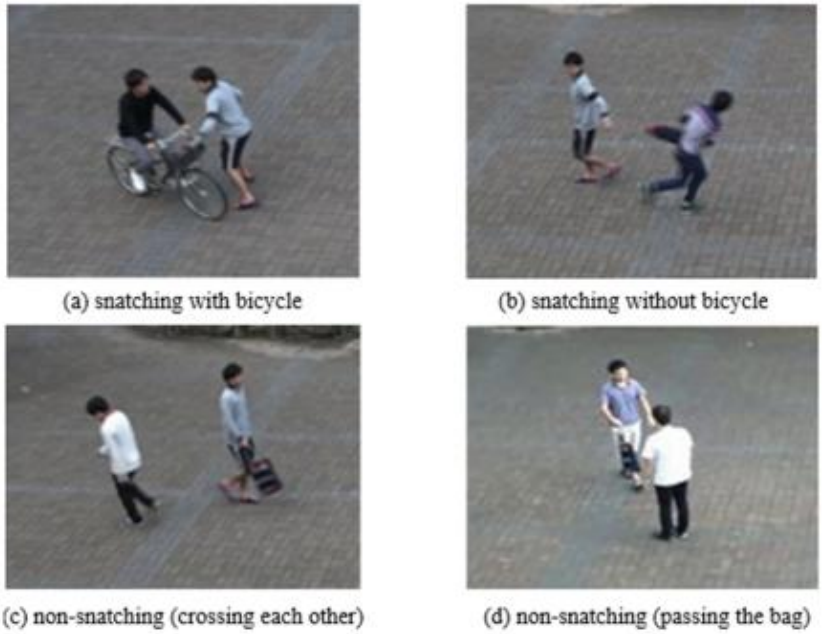


Figure 4.7 – The state of the events in 4 training videos

Table 4.1 – Each feature value in 4 training videos

Data	**Person**	**Area**	**Motion**	**Appearance**
Training 1	Suspect	4	0.051	16.9
Training 1	Victim (with bicycle)	4	0.111	-17.9
Training 2	Victim	1	0.034	-12.1
Training 2	Suspect	1	0.164	26.8
Training 3	Pedestrian A	1	0.034	12.2
Training 3	Pedestrian B	5	0.054	-14.5
Training 4	Pedestrian A	5	0.064	-0.4
Training 4	Pedestrian B	1	-0.020	-1.1

Training 1 and 2: Data with snatching

Training 3 and 4: Data without snatching

Experimental Results

In this experiment, the total number of videos used is 19 videos. They consist 4 videos for training to determine the thresholds and 15 videos for test whether snatching is detected correctly or not. The test data contained 9 video data with snatching events, and 6 video data without snatching events. The video data with snatching includes scenes using bicycles and scenes where snatching was occurred between humans. On the other hand, the video data without snatching includes scenes where two persons pass each other and scenes of handover of bags. Thus we examined the snatching detection in various situations, and gained correct classification result. The result is shown in Table 4.2

Table 4.2 – The classification result in total test video data

	Classification Result			Total
	Snatching	Potential Snatching	Non Snatching	
Video with snatching	8	1	0	9
Video without snatching	0	0	6	6

4.3 Summary Section 4

In this section proposed and tried the methods to effectively detect snatching event in the modern life where the surveillance cameras are spreading more and more. As the method of introducing the point system, conducted research to help solve the snatching incidents that occur in various environments and situations. Detection of snatching was attempted by using mainly three kinds of feature parameters. Also taking into consideration the features before and after intersection of two persons, also focused on the amount of change in each feature parameters.

The video data without snatching includes scenes where two persons pass each other and scenes of handover of bags. Thus we examined the snatching detection in various situations, and gained correct classification result.

In recent snatching cases, there are many different types of snatching cases such as using cars and motorcycles. It is also predicted that snatching event will occur not only during the day time but also in the evening and night time. Thus solving the theft case is a demand in the future, it becomes a challenging research area.

5 CRIMINAL IDENTIFICATION SYSTEM USING FACIAL RECOGNITION

We all know that our Face is a unique and crucial part of the human body structure that identifies a person. Therefore, we can use it to trace the identity of a criminal person. With the advancement in technology, we are placed CCTV at many public places to capture the criminal's crime. Using the previously captured faces and criminal's images that are available in the police station, the criminal face recognition system of can be implemented.

5.1 Facial Recognition

In this paper, we propose an automatic criminal identification system for Police Department to enhance and upgrade the criminal distinguishing into a more effective and efficient approach. Using technology, this idea will add plus point in the current system while bringing criminals spotting to a whole new level by automating tasks. Technology working behind it will be face recognition, from the footage captured by the CCTV cameras; the system will detect the face and recognize the criminal who is coming to that public place. The captured images of the person coming to that public place get compared with the criminal data we have in our database. If any person's face from public place matches, the system will display their image on the system screen and will give the message with their name that the criminal is found and present in this public place. This system matching more than 80% of the captured images with database images.

Criminal identification is the most important task for the Police who are finding the criminals, but it is the difficult and most time-consuming task as they have to find it everywhere. It will be more difficult in cities or public places with high people density. In some cases, manual type of identification gives chance for getting more information related to criminals. Hence this section proposes an automatic criminal identification system by detecting the face of criminals. This will help Police to identify and catch the criminals in public places.

Criminal identification can be done in two ways, which is shown in figure.5.1. In Manual Identification System (MIS), identification is done by the Police officers searching them at public places. It takes a lot of time to give the proper attention and it also has the chances of skipping criminals as they will be

alerted by seeing cops easily gets escape from there. Since the MIS is in the process of taking more time and we will not properly focus on everyone. But when it comes to an automated identification system (AIS) there is no need for observation going in a public place. Here all the process involved in this system is automated.

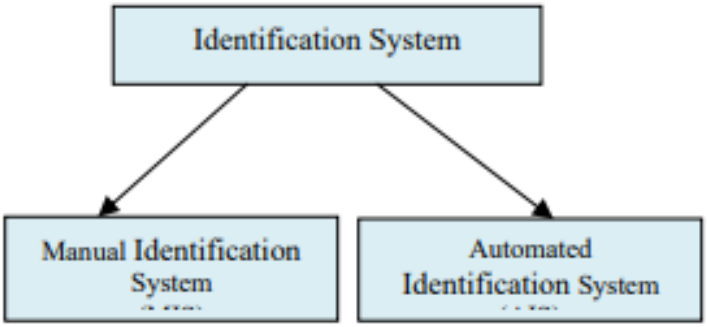


Figure 5.1 – Types of Identification system

Automated Criminal identification monitoring system's some important things shown below:

Criminal Enrolment: criminal images with their name to photos are added to the criminal database so that we can compare the captured images with database one.

*CCTV Connectivity:*CCTV Cameras should be connected to the system on which we are having a criminal database and program where we are running.

Criminal Confirmation: If a person is found from a public place by using this system, then check who was the criminal using a special folder available on the desktop.

5.2 Ease of Use

Face Detection

The primary function of this step is to capture the faces of the people who are available in front of the camera. The outputs from this step are patches that contain each face in the input image. To design a perfect and preferable face recognition system. Face alignment is performed to rationalize the scales and orientation of these patches. Further Next step after the face detection step is human face patches are extracted.

Face Recognition

Face recognition is a method of identifying or verifying the identity of an individual using their face. The step after the representation of faces is to identify them. In this comparison of the detected face image with the images, we have in our database based on face encodings. A facial recognition system maps facial expressions from an image or video using biometrics. To find known faces match from the database, it compares the details to a database. Facial recognition may aid in the identification of personal identity, but it also introduces privacy concerns. Commercial applications use facial recognition as well as it is used for a variety of purposes ranging from security to promotions.

Figure 5.2 – Facial mapping on a captured image

5.3 Literature Survey

In this section, I have collected some surveys from some of the Literature who have proposed ways to help identify crime by using facial recognition:

In this article [1], they are taking help of the CCTV footage and comparing the images from the footage with criminal database if they didn’t find any fingerprint from the crime scene. This system consists of five stages where the first stage is planning in which the why and how the system is made are discussed. The second stage of Requirement analysis discussed the requirement to design the system. Design, the third stage where they defined system design and its workflow. The fourth ultimate important stage is Implementation and testing, system is implemented using Principal Component Analysis (PCA) Technique and tested. The last stage is maintenance; this phase hadn’t undertaken due to this system was developed in a controlled environment. For criminal identification, authors had used PCA Technique for finding similar features of images available in the

database with captured images of footage. The machine will use a database that contains the person's personal information so that if FRCI identifies a face, it can display the person's information.

This article [2] consists of four steps, the first one is real-time image training and the second one is Harr-classifier using for face detection. The third step is the comparison of Surveillance camera captured images with real-time images and last, is the result part based on the comparison. The authors are using the Haar-classifier on Open-CV for face detection; Haar-cascading is one of the algorithms for face detection. On the open-CV platform, face tracking is taken with help of Harr-like classifiers. More than one person is identified in this system and it can be used to find the suspects whom we are finding. The accuracy of the proposed system is very high as compared to the previous model.

In this article [4] Using the Passport database, they are identifying whether the traveler is an authorized passport holder or not. In this, they are using image processing techniques as well as LBPH (Local Binary Pattern Histogram) mathematical model. This method consists of six steps for airport security purpose that are: a) Capture image using webcam b) Captured image is sent to the Django server c) Using LBPH feature set is taken from image d) Image is compared with database image by applying classifier e) If matching is done user details are fetched from database f) The predicted details of the user are sent to the admin via mail. They are using webcam images for LBPH processing and then applying the classifiers comparing them with database images. This will also help to catch the criminals who travel from one country to another and also detect if the traveler having a loan from the bank then traveler's detailed information will be sent for verification to the police station.

The authors of this article [5] are presenting an automatic face recognition system for attendance monitoring. They are capturing faces by using a camera and the captured image is compared with images that are already present in their database. They are using machine learning technology with an SVM (Support Vector Machines) classifier for name detection and gradient-oriented Histogram for face detection. They are using open-CV for image detection & recognition, Tkinter for GUI application creation, and Numpy to work with arrays as those are libraries of python. To develop and test the application using the Xampp server, as it is a free open source server. There proposed model has achieved an accuracy of 99.38%. Using Cloud feasibility of the system can be increased.

In this research [6], the authors had discussed that an attendance monitoring system is very important in the teaching and learning process. The student who is entering in classroom his/her image is captured. Preprocessing and Face region extraction take place using that captured image for further process. They are using a face recognition algorithm for marking present if the student came to school or absent if the student is not coming to school. They are capturing the student's image using a camera and after preprocessing comparing with their student database and marking attendance.

In this paper [7], the authors have presented a face identification system that uses the fast algorithm. This model uses two datasets: 1) Olivetti Research Laboratory (ORL), 2) Unconstrained Facial Images (UFI). Captured image converted to HSV system (see fig.5.3) and after that force field features is extracted from that image. Classification is done by using three distance methods that are: Manhattan, Euclidean, and Cosine. By comparing these methods, they got the best resolution and achieved an accuracy of 99.9% for the datasets ORL and UFI.

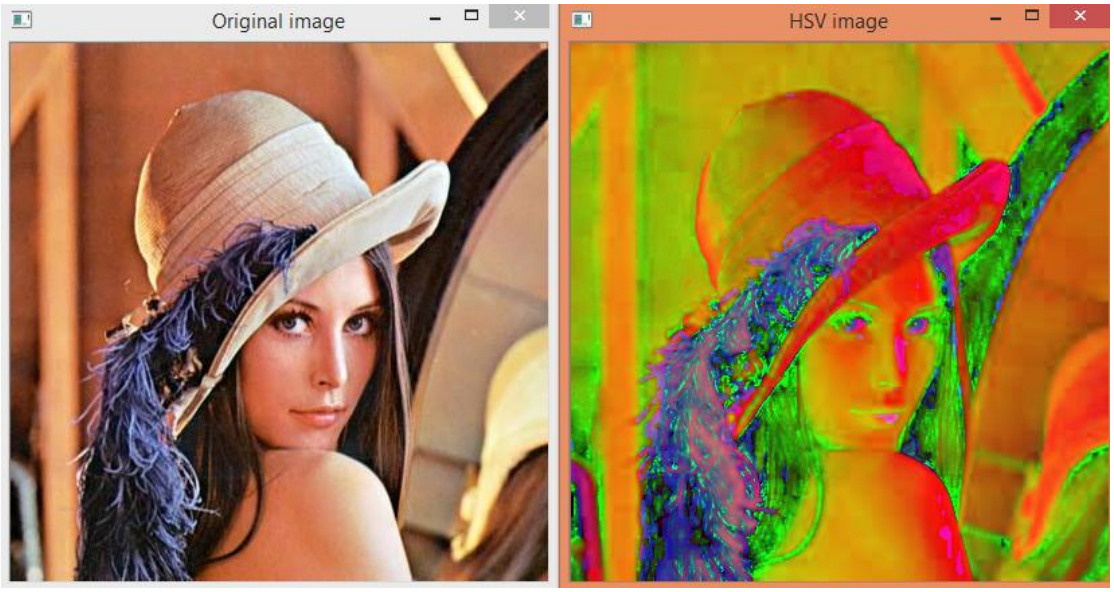


Figure 5.3 – Output of the system, showing the original image the and image in the HSV color space

5.4 Proposed Work

In this section we are Using CCTV cameras which are continuously working in a public place. In the Implementation of the system, we already saved criminal's

images data with their names on photographs in the database. We are processing those images and extracting features from them and in feature extraction; we are taking the face encodings of the present images and saving them into one file. Using open-CV while capturing the footage in CCTV and captured images face encodings taken placed and comparing with our saved face encodings of the criminal database if any match is found then automatically on screen it will display an image of that criminal whose face matches and display the message with his name that criminal found, and his captured image will be saved into special folder police will go and catch him from that public place even if he once captured in the CCTV footage. As we saved the person's image in the special folder from there police will come to know whose image was matched with captured once.

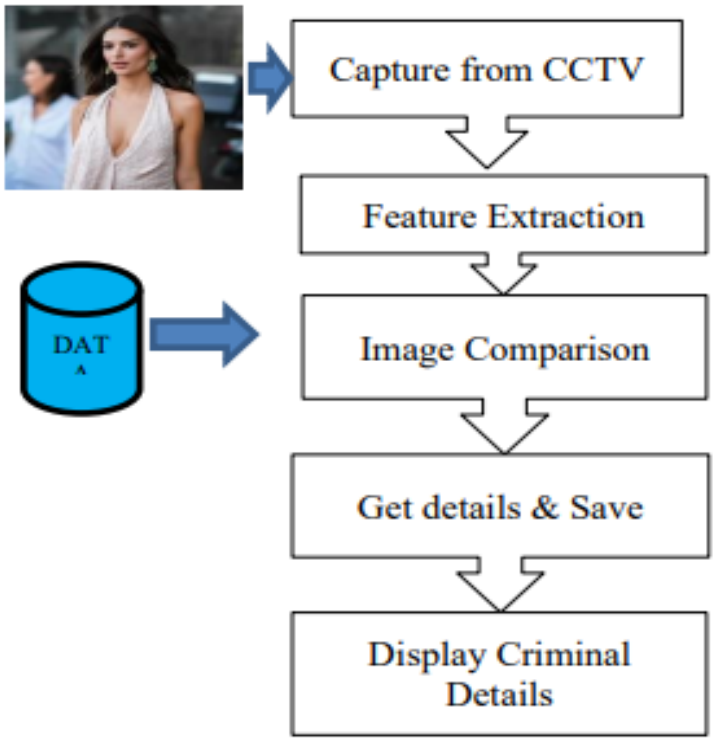


Figure 5.4 – Execution Flow of System

Criminal enrolment is the initial step in which we are storing the criminal information using the criminal database. Here criminal names with their image will get stored in the database. By using this image and information, face recognition and identification will take place. After the face detection process i.e., using face encodings of images will take place. Criminal Enrolling is the database

process, but the main process starts from face detection. Face detection is done considering 68 landmarks available on the face as shown in figure 5.5.

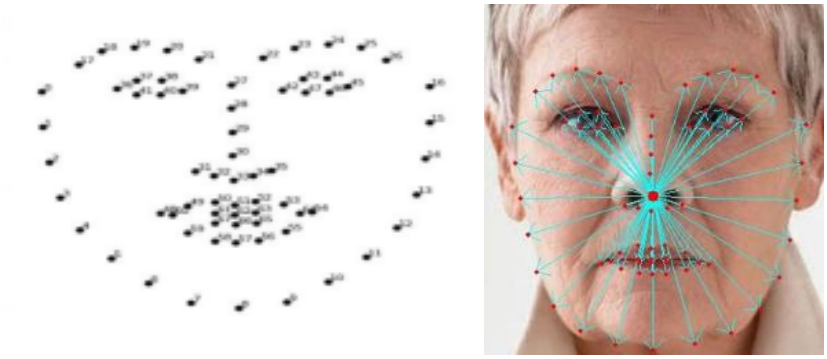


Figure 5.5 – 68 Landmarks on Every Face

CCTV camera captured footage or image is stored and its features i.e., encodings are extracted after that it will be compared with the image encodings which are available in the database. The process of matching the face will take place in the database; the name and the criminal found message will be displayed with the criminal image on the screen present in CCTV Room.

How the system Working: we should be using CCTV Camera to capture the images of the Public so that we can identify the proper person who is having a criminal record in the database to catch.

– First, we will find out the face encodings of the criminal database images and save those into one list, and splitting the name which is saved with criminal image saving into another list.

– Now we are using CCTV to capture Public images so that we can identify the criminal who is present in a public place and easily can catch

– Extracting the features from captured images i.e., taking the face-encodings of the captured images.

– Comparing Captured image encoding values with our database image encoding values.

– If the encoding values matching with captured image encoding values, then criminal image, name, and criminal found message will be displayed on the screen.

– Image of that person will be saved into a special folder on the desktop, so that police can easily identify the criminal who differs from other peoples present in a public place.

Applications

We can use this system Anywhere for Security Purpose for ex. in the Jewelers shop, Banks, Hotels, Airports, Restricted areas, Religious Places and Famous places where maximum crowd use to gather.

5.5 Experimental Results

We have proposed a promising Criminal Detection system for Face Videos. CCTV Cameras are used for continuous capturing of the video and images; we will get the information on our main screen that which image from the database is matching. When the database image matches with CCTV captured image then on the main screen the name of the criminal with the criminal found message will be displayed as shown below in figure 5.6.

Figure 5.6 – Camera images with name and Criminal found Message

When a criminal is identified using a CCTV image, then which criminal found from the database that criminal image will be saved in Special folder on desktop as shown in figure 5.7. Police will come to know that who is found in a public place because he will not be there in front of a CCTV camera for a long time. So this approach more helpful when criminals come for few minutes in front of a CCTV camera that one will be identified.

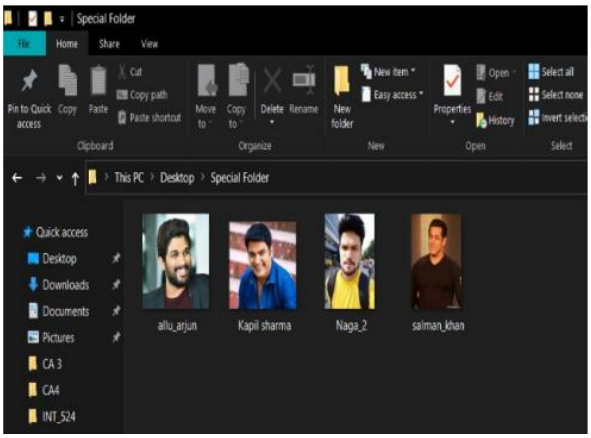


Figure 5.7 – Database images which are saved in a special folder when criminal found

5.6 Summary section 5

This upgraded version of the criminal detecting system not only provides a huge convenience to the Police in the identification of criminals but also saves time for them as processes are automated in the system. The novelty of this Research Paper is face detection done by using Face Encodings.

also maybe in future work, can add the Alarms to the criminal detection system. It will range only when matches are found so that if anyone is not there to keep watch in the CCTV room, they will come to know that someone is found from the database in that public place. This paper presents a surveillance system that will give us alerts when any controversy, fight, or intruder is detected.

6 VIDEO SURVEILLANCE FOR HUMAN MOTION DETECTION

Video surveillance for human motion detection is a field of research that involves using cameras and computer algorithms to detect and track the movement of people in a specific area. This technology has a wide range of applications, including security, crowd management, and traffic analysis.

The use of video surveillance systems for human motion detection has been growing rapidly in recent years due to the increasing availability of high-resolution cameras and the advancement of computer vision algorithms. These systems can use a variety of techniques, such as object detection, motion analysis, deep learning, and multi-modal approaches, to detect and classify human motion in video footage.

One of the main advantages of using video surveillance for human motion detection is its ability to automate the monitoring of large areas and detect suspicious or abnormal behavior in real-time. This can aid in the prevention of crime and the quick response to emergencies.

6.1 System to detect human motion

Detecting human motion accurately in a video is one of the important topic of research due to its application in security purposes. It is quite challenging to process the image obtained from a surveillance video due to its low Resolution and night vision. There is a system allows user to intrigue the mobile platform from a remote location. The detection process in the system shown on fig.6.1, occurs in four steps: Image Acquisition Image Segmentation, Image Subtraction, Recognition Engine and Action. The system provides feedback to the end user via a mail on the registered email address which contains the visuals of the captured human motion thus can be useful for spying purposes.

The proposed system has four main components:

- Image Acquisition
- Image Segmentation and Image Subtraction
- Recognition Engine
- Action

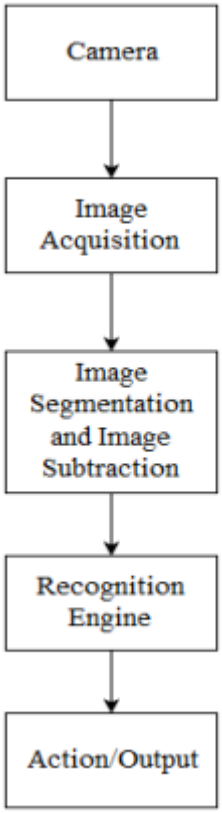


Figure 6.1 – System Overview

Image Acquisition: This is the most initial step for the motion detection algorithm. This block accepts the image from the camera or the recorded sequence.

Image Segmentation and Image Subtraction shown on fig.6.2:

Image Subtraction: In this part, the two subsequent images are compared and processed using absolute arithmetic subtraction. This block it separates the image into three planes. It then performs the arithmetic operations on each plane. The results are then combined back to form the color image. The result is in inverse color format because the subtraction on pixel values is performed.

Image Processing: In this step, the image processing operations are performed on the result of the previous step. There are two outputs from this block. The first is the result obtained from the threshold function. The result of this is further used for the recognition purposes since it filters the human body shape better than the other output. The other output from this block is the eroded and dilated image. This function removes the small pieces of noise that may be present because of the camera signal noise or small pixel changes.

Contour Finding: This operation is performed on the eroded and dilated image obtained from the above step. The contours are displayed with different

colors for the contours which are different from each other. The contours having same color are considered to be connected.

Bounding Rectangle Calculations: In this step the operations are performed to remove the overlaying boxes or rectangles when drawn to the source image. The rectangles and boxes which are near and almost crosses one-another edges are joined together to form the large rectangle.

Binary Image Processing: In this step, the processing is performed on the threshold image. The image is further enhanced to fill the empty spaces inside the binary region to detect the object in motion. An algorithm it scans through the vertical line and filling up each vertical line's first and last pixels'. Also on the horizontal lines the same task is performed. After we obtain the resultant image from these algorithm, we use the AND operator to combine both the result to form a better representation of moving object shape.

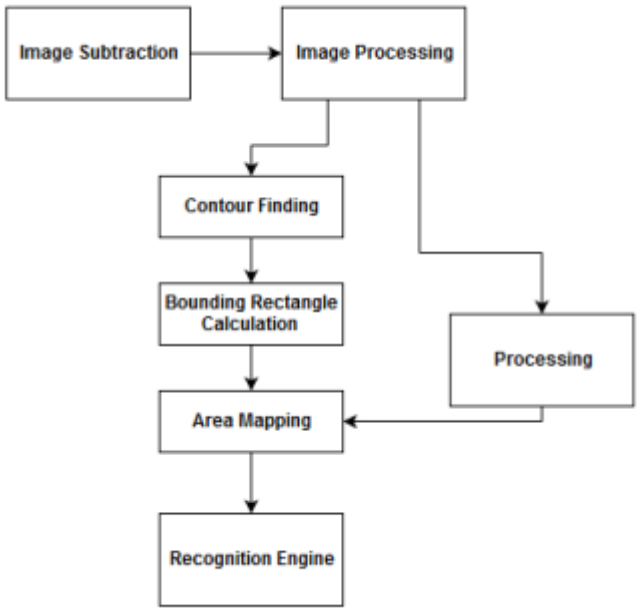


Figure 6.2 – Image Segmentation and Image Subtraction

Area Mapping: After the bounding rectangles are identified, the position is then mapped to the source image and the rectangle being drawn there. Mapping is done for the area of the bounding boxes drawn in the source and also the corresponding area in the binary processed image. The area from this processed binary images are the ones to be used for the recognition engines. This project does not cover the recognition engine and the output component. The recognition engine is implemented from the work of the partner in this project.

Recognition Engine: In this block, .NET framework, which contains different classes for motion detection is used to extract the object under motion clearly. Once the object is extracted it is classified whether it is an object or human.

Action: In this block, if the motion is done by the human then the alarm is triggered and the mail is send on the registered email address which contains the image of the area where the motion is detected.

6.2 System Result

If in front of the camera the human is present or a clip is given that contains human, then the system gives as output Object is human (fig.6.3, 6.4).

Figure 6.3 – Input 1

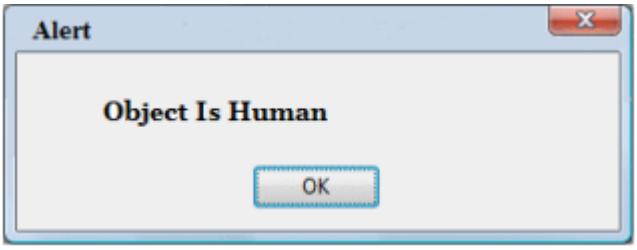


Figure 6.4 – Output 1 (Human/Object)

If in front of the camera the object is present or a clip is given that contains non-human, then the system gives as output Object is non-human (fig.6.5, 6.6).

Figure 6.5 – Input 2

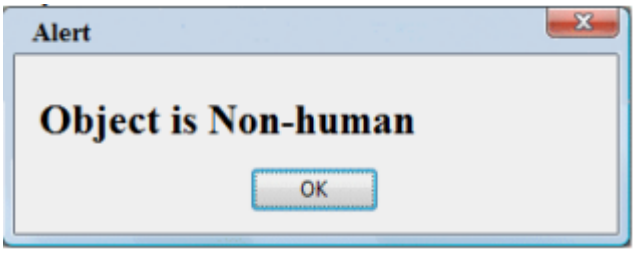


Figure 6.6 – Output 2 (Human/Object)

6.3 System Statement

Traditional video surveillance system needs huge amount of storage space. All the recorded videos were saved which requires excessive storage and thus it limits effectiveness of the system. In order to solve the problem, only the videos which contains necessary information i.e. the videos that contains motions are stored and all rest are ignored.

Several problems exist in human motion analysis and this includes: Unable to provide an optimal method of dimensionality reduction to achieve higher recognition rate.10 Human motion tracking cannot be done (from the training set) by combining the Eigen faces alone. The weighting factors need to be more adaptive to achieve better results. Less scalability exists for detecting human motion. Image block matching, gradient constraints, phase conservation or energy models are bottlenecks.

Once the motion within the video is detected, the object is further classified whether it human or an object. As soon as the human motion is detected, the alarm will be triggered and the captured video will be send on registered email address.

CONCLUSION

In conclusion, neural networks can be effectively used to detect criminal acts in video surveillance systems using various methods such as object detection, activity recognition, anomaly detection, facial recognition, and video prediction. These methods have shown promising results in accurately detecting and classifying criminal activities in real-world scenarios. video surveillance systems can play an important role in detecting criminal acts by providing valuable visual information. However, the sheer volume of video footage generated by these systems can make it difficult for human operators to effectively monitor and identify criminal activities.

However, it is important to note that the accuracy of these methods can be affected by factors such as the quality and quantity of training data, as well as environmental conditions. Furthermore, it is important to consider ethical and privacy issues when implementing these systems.

Video surveillance can play an important role in detecting criminal acts by providing valuable visual information. Video surveillance systems can be used to monitor public spaces and private properties, and can help to identify and deter criminal activities. The use of advanced technologies such as neural networks can improve the efficiency and accuracy of identifying criminal acts in video surveillance footage.

It's important to evaluate the performance of the system, fine-tune it accordingly and deploy it in a responsible manner. Overall, video surveillance can be an effective tool for detecting criminal acts, but it must be used with caution and consideration for the rights and privacy of individuals.

REFERENCE

1. Nurul Azma Abdullah, Md. Jamri Saidi, Nurul Hidayah Ab Rahman, Chuah Chai Wen, and Isredza Rahmi A. Hamid's "FACE RECOGNITION FOR CRIMINAL IDENTIFICATION: AN IMPLEMENTATION OF PRINCIPAL COMPONENT ANALYSIS FOR FACE RECOGNITION", AIP Conference Proceedings (2017); https://doi.org/10.1063/1.5005335

2. Apoorva. P, Impana. H.C, Siri. S.L, Varshitha. M.R and Prof. Ramesh. B's "AUTOMATED CRIMINAL IDENTIFICATION BY FACE RECOGNITION USING OPEN COMPUTER VISION CLASSIFIERS", Third International Conference on Computing Methodologies and Communication (ICCMC 2019), DOI:10.1109/ICCMC.2019.8819850

3. Prof. Rupali T. Umbare, Ms. Janhavi S. Takalgavankar, Ms. Harshada S. Yadav, and Ms.Pruthvi A. Tilekar's "AIRPORT SECURITY USING FACE-RECOGNITION", International Journal of Future Generation Communication and Networking, vol. 13, No. 3 (2020).

4. Vikram Mohanty, David Thames, Sneha Mehta, and Kurt Luther, "Photo Sleuth: Combining Human Expertise and Face Recognition to Identify Historical Portraits", Conference: the 24th International Conference, March 2019, https://doi.org/10.1145/3301275.3302301

5. Dr. Asif Ali, Radhika Mandhanya, Shraddha Birla, Ujjwal Mandloi and Vipul Jain's "Automatic Face Recognition Attendance System using Python and OpenCV", GRD Journals Global Research and Development Journal for Engineering | Volume 6 | Issue 4 | March 2021

6. P. Kowsalya, J. Pavithra, G. Sowmiya and C.K. Shankar's "ATTENDANCE MONITORING SYSTEM USING FACE DETECTION & FACE RECOGNITION", International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue: 03 | Mar 2019.

7. Kian Raheem Qasim and Sara Salman Qasim's, "Force Field Feature Extraction Using Fast Algorithm for Face Recognition Performance", Iraqi Academics Syndicate International Conference for Pure and Applied Sciences, https://doi.org/10.1088/1742-6596/1818/1/012195.

8. Ratcliffe, J. (2006). Video surveillance of public places. Washington, DC: US Department of Justice, Office of Community Oriented Policing Services

9. Anderson, J., and A. McAtamney (2011). Considering Local Context When Evaluating a Close Circuit Television System in Public Spaces. Trends & Issues in Crime and Criminal Justice, 430. Canberra: Australian Government Australian Institute of Criminology.

10. Bowers, K. J., and S. D. Johnson (2003). "Measuring the Geographical Displacement and Diffusion of Benefit Effects of Crime Prevention Activity." Journal of Quantitative Criminology 19(3):275–301.

11. Brown, B. (1995). CCTV in Town Centres: Three Case Studies. Crime Detection and Prevention Series, Paper 68. London: Home Office

12. Clancey, G. (2009). Consideration for Establishing a Public Space CCTV Network. Research in Practice Resource Manual, No. 8. Canberra: Australian Institute for Criminology.

13. Harris, C., P. Jones, D. Hillier, and D. Turner (1998). "CCTV Surveillance Systems in Town and City Centre Management." Property Management 16(3):160–165.

14. Martínez-Mascorro, G. A., Abreu-Pederzini, J. R., Ortiz-Bayliss, J. C., Garcia-Collantes, A., & Terashima-Marín, H. (2021). Criminal intention detection at early stages of shoplifting cases by using 3D convolutional neural networks. *Computation*, *9*(2), 24.

15. Ling, T.S.; Meng, L.K.; Kuan, L.M.; Kadim, Z.; Baha'a Al-Deen, A.A. Colour-based Object Tracking in Surveillance Application. In Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 (IMECS 2009), Hong Kong, China, 18–20 March 2009; Volume 1

16. Geng, X.; Li, G.; Ye, Y.; Tu, Y.; Dai, H. Abnormal Behavior Detection for Early Warning of Terrorist Attack. In *AI 2006: Advances in Artificial Intelligence*; Sattar, A., Kang, B.H., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1002–1009.

17. Devroye, L., & Wise, G. L. (1980). Detection of abnormal behavior via nonparametric estimation of the support. *SIAM Journal on Applied Mathematics*, *38*(3), 480-488.

18. Sjarif, N. N. A., Shamsuddin, S. M., & Hashim, S. Z. (2012). Detection of abnormal behaviors in crowd scene: a review. *Int. J. Advance. Soft Comput. Appl*, *4*(1), 1-33.

19. Penmetsa, S., Minhuj, F., Singh, A.: Autonomous UAV for suspicious action detection using pictorial human pose estimation and classification. Electron. Lett. Comput. Vis. Image Anal. 13(1), 18–32 (2014)

20. Development of a High-Performance Next Generation Intellectual Security Camera System for Automatic Detection of Crime Situation. https://kaken.nii.ac.jp/ja/grant/KAKENHIPROJECT-26350454/. Accessed 25 Dec 2017

21. Ibrahim, N., Mokri, S.S., Siong, L.Y., Mustafa, M.M., Hussain, A.: Snatch theft detection using low level features. In: Proceedings of the World Congress on Engineering, London, UK, vol. 2, pp. 862–866, July 2010

22. Tsushita, H., Zin, T.T.: An effective method for detecting snatch thieves in video surveillance. In: International Conference on Artificial Life and Robotics (ICAROB 2017), Miyazaki, Japan, pp. 303–306, January 2017

23. Chang, Y.-H., Lin, P.-C., Jeng, L.-D.: Automatic motion trajectory analysis for dual human interaction using video sequences. Int. J. Comput. Electr. Autom. Control Inf. Eng. 9, 1294– 1301 (2015). World Academy of Science, Engineering and Technology

24. van Huis, J.R., Bouma, H., Baan, J., Burghouts, G.J.: Track-based event recognition in a realistic crowded environment. In: Proceedings of SPIE, vol. 9253, pp. 92530E-2–92530E-7. Copyright Society of Photo-Optical Instrumentation Engineers (SPIE), The Netherlands, September 2014

25. Yang, H., Du, Q., Ma, B.: Weighted decision fusion for supervised and unsupervised hyperspectral image classification. In: Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Honolulu, USA, pp. 875–879, December 2010

26. Aherwadi, N. B., Chokshi, D., Pande, D., & Khamparia, A. (2021, July). Criminal Identification System using Facial Recognition. In Proceedings of the International Conference on Innovative Computing & Communication (ICICC).

27. Federation, N.R. 2020 National Retail Security Survey; National Retail Federation: Washington, DC, USA, 2020

28. Ba, S.O.; Odobez, J. Recognizing Visual Focus of Attention From Head Pose in Natural Meetings. IEEE Trans. Syst. Man, Cybern. Part B (Cybernetics) 2009, 39, 16–33

29. Nayak, N.M.; Sethi, R.J.; Song, B.; Roy-Chowdhury, A.K. Modeling and Recognition of Complex Human Activities. In Visual Analysis of Humans: Looking at People; Springer: London, UK, 2011; pp. 289–309

30. Nurhopipah, A.; Harjoko, A. Motion Detection and Face Recognition for CCTV Surveillance System. IJCCS (Indones. J. Comput. Cybern. Syst.) 2018, 12, 107.

31. Wang, T.; Qiao, M.; Deng, Y.; Zhou, Y.; Wang, H.; Lyu, Q.; Snoussi, H. Abnormal event detection based on analysis of movement information of video sequence. Opt.-Int. J. Light Electron Opt. 2018, 152, 50–60

32. Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; Paluri, M. Learning Spatiotemporal Features with 3D Convolutional Networks. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 4489–4497.

33. Nehme, M.A.; Khoury, W.; Yameen, B.; Al-Alaoui, M.A., “Real time color based motion detection and tracking”, Proc. ISSPIT, 3rd IEEE International Symposium on Signal Processing and InformationTechnology, pp. 696 – 700.