

**Automated Video Summarization for Suspicious  
Event Detection By Making Pipeline**

**Habib University**

**Project Mid Progress Report**

**Syed Muhammad Hussain | Azeem Haider |  
Affan Habib**

**April 30, 2023**

**Contents**

<b>1</b>	<b>Abstract</b>	<b>2</b>
<b>2</b>	<b>Introduction</b>	<b>2</b>
<b>3</b>	<b>Literature Review</b>	<b>3</b>
3.1	Subsection 2.1 . . . . .	3
3.2	Subsection 2.2 . . . . .	3
<b>4</b>	<b>Methodology</b>	<b>3</b>
4.1	Video Summarization . . . . .	4
<b>5</b>	<b>Conclusion</b>	<b>4</b>

# 1 Abstract

With the widespread use of surveillance cameras in public places, there is an increasing need for automated methods to quickly analyze video footage and detect suspicious events. In this paper, we propose a novel approach for automated video summarization for suspicious event detection by using multi-modal. Our proposed method first extracts visual features from the input video as an event then classifies these events and generates a summary of the events.

To evaluate the proposed approach, we conduct experiments on a publicly available dataset of surveillance videos. The results demonstrate that our method outperforms existing state-of-the-art techniques in terms of both summarization quality and suspicious event detection accuracy. Our approach can be useful for a variety of real-world applications, such as security surveillance, public safety, and law enforcement.

# 2 Introduction

Surveillance through CCTV cameras has been in the works since the 1940s. Surveillance cameras have proven extremely useful in stopping crime due to the fear of getting caught. However, the sheer amount of data these cameras generate is overwhelming and requires much manual labor to analyze. This is where computer vision comes in. Computer vision is the science of making computers see and understand the world. It is a field that has been proliferating in the past few years.

With the advent of computer vision technology and algorithms, this manual labor has changed and has been made easier for humanity. Suspicious activity is now tracked through computer vision algorithms that notify the stakeholders when suspicious activity is recognized. Detecting and recognizing suspicious activities to stop them in time is a challenging task. However, with the help of computer vision, this task has become easier by instilling fear in the hearts of the culprits and making catching the culprits easier.

With the crime index of Pakistan being among the highest in the world, especially Karachi standing at 55.0 crime index, suspicious activity detection models have become the need of the hour in the country. Especially completely automated detection models that detect and recognize suspicious activities without human intervention. Summarizing 24-hour CCTV footage is another crucial task for the models to carry out so individuals do not have to skim through the footage to find suspicious activity.

While many models have been proposed in the past, they have yet to be able to achieve the desired results. This is due to various reasons, such as the need for a proper dataset, lack of proper training, and lack of proper testing. In this paper,

we propose a novel approach for automated video summarization for suspicious event detection by using multi-modal. Our proposed method first extracts visual features from the input video as an event then classifies these events and generates a summary of the events through the activity labels. We have conducted experiments on a publicly available dataset of surveillance videos.

Automated security surveillance has become a need of the hour for the people of Pakistan and Karachi. Our model can be a real game changer since it can be used to detect suspicious activities in the city. In addition, the model can also be used to summarize 24-hour CCTV footage so that individuals do not have to skim through entire footage to find suspicious activity.

### 3 Literature Review

This is section 2.

#### 3.1 Subsection 2.1

This is subsection 2.1.

#### 3.2 Subsection 2.2

This is subsection 2.2.

### 4 Methodology

The objective of this research paper is to present a methodology for efficiently detecting abnormal activity in surveillance videos using a two-stage process. In the first stage, we use OpenCV to summarize long surveillance videos into unique and important events. In the second stage, we use two different approaches, ConvLSTM and LRCN, to detect abnormality in the summarized events based on a training dataset.

The research design consists of the following steps:

- **Data Collection:** The data used in this research was collected from a surveillance system installed in a public area. The data consisted of long surveillance videos of the area.
- **Data Preprocessing:** The collected data was preprocessed to ensure that it was in a suitable format for analysis. This involved converting the video data into frames, which were then analyzed using OpenCV to detect and summarize unique and important events.
- **First Stage Model:** A first-stage model was developed using OpenCV to summarize long surveillance videos into unique and important events. The

model uses a combination of image processing techniques, such as background subtraction, object detection, and tracking, to identify important events by frame to frame checking process.

- **Second Stage Models:** Two different models were developed for the second stage: ConvLSTM and LRCN. The ConvLSTM model is a combination of convolutional neural networks and LSTM networks that can learn both spatial and temporal features of the data. The LRCN model is a combination of CNNs and Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks, which can effectively capture the temporal dependencies in the data. Both models were trained using a training dataset that included labeled data with normal and abnormal events.
- **Evaluation:** The efficiency and effectiveness of the proposed method will be evaluated using several metrics such as detection rate, false-positive rate, and accuracy. The evaluation will be performed on a separate dataset to ensure the generalizability of the proposed method.

#### 4.1 Video Summarization

### 5 Conclusion

This is the conclusion section.