Multimodal Video Surveillance System

Submitted By

Vedang Shah

15MCEI26

DEPARTMENT OF COMPUTER ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2017

Multimodal Video Surveillance System

Major Project

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology in Computer Science and Engineering

(Information and Network Security)

Submitted By

Vedang Shah

(15MCEI26)

Guided By

Prof. Vishal Parikh

DEPARTMENT OF COMPUTER ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2017

Certificate

This is to certify that the major project entitled **"Multimodal Video Surveillance System"** submitted by **Vedang Shah (Roll No: 15MCEI26)**, towards the partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Engineering of Nirma University, Ahmedabad, is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this major project, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Prof. Vishal Parikh
Guide & Assistant Professor,
CE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Sharada Valiveti
Associate Professor,
Coordinator M.Tech - INS
Institute of Technology,
Nirma University, Ahmedabad

Dr. Sanjay Garg
Professor and Head,
CE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr Alka Mahajan
Director,
Institute of Technology,
Nirma University, Ahmedabad

Statement of Originality

I, **Vedang Shah**, Roll. No. **15MCEI26**, give undertaking that the Major Project entitled **"Multimodal Video Surveillance System**" submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Engineering** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made. It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

Signature of Student

Date:

Place:

Endorsed by

Prof. Vishal Parikh

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Prof. Vishal Parikh**, Assistant Professor, Computer Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. The appreciation and continual support he has imparted has been a great motivation to me in reaching a higher goal. His guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

It gives me an immense pleasure to thank **Dr. Sanjay Garg**, Hon'ble Head of Computer Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr. Alka Mahajan**, Hon'ble Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation she has extended throughout course of this work.

I would also thank the Institution, all faculty members of Computer Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

Finally, I would like to thank the almighty **GOD** and my **Family** for always being with me and blessing me with the strength and endurance required for the smooth completion of this project work. In any obstacle on my path, wholehearted prayer made divine atmosphere around me and boosted my soul with strength.

- Vedang Shah
15MCEI26

Abstract

Video Surveillance systems are used to monitor, observe and intercept the changes in activities, features and behavior of objects, people or places. A multimodal surveillance system incorporates a network of video cameras, acoustic sensors, pressure sensors, IR sensors and thermal sensors to capture the features of the entity under surveillance and send the recorded data to a base station for further processing. Multimodal surveillance systems are utilized to capture the required features and use them for pattern recognition, object identification, traffic management, object tracking, and so on. In the Indian scenario with the advent of the concept of 'Smart Cities', the need of an Intelligent Transportation System which is low-cost and limits power consumption is inevitable. The proposal is to develop an efficient camera placement algorithm for deciding placement of multiple video cameras at junctions and intersections in a multimodal surveillance system which will be capable of providing maximum coverage of the area under surveillance leading to complete elimination or reduction to a great extent the number of blind zones in a surveillance area, maximizing the view of subjects and minimizing occlusions in high vehicular traffic areas. Furthermore, the proposal is to develop a video summarization algorithm which can be used to create summaries of the videos captured in a multi-view surveillance system. Such a video summarization algorithm can be used further for object detection, motion tracking, traffic segmentation, etc. in a multi-view surveillance system. In addition to this, the use of video summarization algorithm can be extended in a multimodal surveillance system containing different types of sensors. The proposed algorithms have been designed while keeping into consideration the Indian transportation infrastructure scenario and will be able to reduce the cost of camera deployment, computational cost, power consumption; and provide efficient performance in a multi-view as well as multimodal surveillance system for an Intelligent Transportation System.

Abbreviations

AOV	Angle of View
CCD	Charge Coupled Device
CMOS	Complementary metal–oxide–semiconductor
DVR	Digital Video Recorder
FOV	Field of View
GOP	Group of Pictures
HD	High Definition
ITS	Intelligent Transportation System
NVR	Network Video Recorder
P/T/Z	Pan/Tilt/Zoom
SD	Standard Definition
WSN	Wireless Sensor Network

–

Contents

Certificate **iii**

Statement of Originality **iv**

Acknowledgements **v**

Abstract **vi**

Abbreviations **vii**

List of Tables **x**

List of Figures **xi**

1 Introduction **1**
1.1 Video Surveillance Systems . . . 1
1.1.1 Evolution of Video Surveillance Systems . . . 1
1.1.2 Applications of Video Surveillance Systems . . . 2
1.2 Multimodal Surveillance Systems . . . 3
1.2.1 Intelligent Transportation Systems . . . 3
1.2.2 Applications of a Multimodal Surveillance System in ITS . . . 3
1.3 The Indian Scenario . . . 4
1.3.1 Challenges . . . 5

2 Literature Review **7**
2.1 Domain of Survey . . . 7
2.1.1 Literature Survey . . . 7
2.2 Proposed System . . . 11

3 Cameras - Analog and Digital **12**
3.1 Types of Cameras . . . 12
3.1.1 Analog Video Cameras . . . 12
3.1.2 Digital Video Cameras . . . 13
3.1.3 IP-based Digital Video Cameras . . . 13
3.2 Commercially available Cameras . . . 14
3.2.1 Comparison of (Wired) Digital Video Cameras . . . 17
3.2.2 Comparison of IP Cameras . . . 18
3.3 Programmable Cameras . . . 21

4 Optimal Camera Placement in Multimodal Surveillance System **22**
4.1 Requirement of an Optimal Camera Placement Strategy 22
4.1.1 Camera System 23
4.1.2 An Optimal Camera Placement Algorithm 25
4.2 Tools and Simulators 30
4.2.1 OMNET++ Network Simulator 32

5 Video Summarization **37**
5.1 Introduction to Video Summarization 37
5.1.1 Types and Techniques of Video Summarization 38
5.1.2 Video Compression Picture Types and Group of Pictures 39
5.2 Video Summarization Algorithm 42
5.2.1 Need of Temporal Synchronization for Frames 44

6 Implementation and Results **48**
6.1 Simulation of Optimal Camera Placement Algorithm 48
6.1.1 Simulation Environment and Parameters 48
6.1.2 Results 49
6.2 Implementation of Video Summarization Algorithm 52
6.2.1 Dataset and Configuration 52
6.2.2 Demonstration of need of Temporal Synchronization for Frames . 54
6.2.3 Extensibility of the Video Summarization Algorithm 56

7 Conclusion and Future Scope **59**
7.1 Conclusion 59
7.2 Future Scope 60

Bibliography **61**

List of Tables

3.1 Difference between Analog Video Cameras and Digital(Wired and Wireless/IP) Video Cameras . . . 14
3.2 Comparison of (Wired) Digital Video Cameras . . . 18
3.3 Comparison of IP Cameras . . . 21

4.1 Comparison of Network Simulators . . . 31

6.1 Simulation Parameters for Optimal Camera Placement Algorithm . . . 48
6.2 Simulation results for Optimal Camera Placement Algorithm . . . 50
6.3 Experimental results for Video Summarization Algorithm . . . 53

List of Figures

4.1 Basic structure of a camera 24
4.2 FOV of a camera 25
4.3 Coverage of a Video Camera 26
4.4 Sample Surveillance Area with various possible Camera placements . . . 29
4.5 Eclipse-based Simulation IDE 33
4.6 C++ Source Code Editor with Code Reviewer 34
4.7 Content Assist for C++ Source Code Editor 35
4.8 Graphical NED Editor 35

5.1 GOP Structure with N=9 41
5.2 GOP Structure with N=15 41

6.1 Surveillance Area with various possible Camera placements 49
6.2 Optimal Camera Placement Design 51
6.3 Comparison of Frame Selection Techniques for Video Summarization Algorithm 54
6.4 Frames without Temporal Synchronization 55
6.5 Frames with Temporal Synchronization 56
6.6 Object Detection and Vehicle Tagging 57
6.7 Vehicle Tagging for Classification 58

Chapter 1

Introduction

This chapter provides a background about video surveillance systems and multimodal surveillance systems, in general and in the context of ITS along with their application areas.

1.1 Video Surveillance Systems

Video surveillance systems deal with monitoring, intercepting or observing activities, behavior, or any other changing information related to people, places or things. Video surveillance systems have evolved over three generations of surveillance systems namely,

- Analog Surveillance Systems
- Digital Surveillance Systems
- Smart or Intelligent Surveillance Systems

1.1.1 Evolution of Video Surveillance Systems

- Analog Surveillance Systems involved placement of analog video cameras in certain strategic or sensitive areas for monitoring. This led to the creation of hundreds of video tapes which were then examined manually by a security personnel.
- Digital Surveillance Systems involved the use of both Wired/Wireless IP Cameras, Closed-circuit television cameras, networked video recorders for monitoring. This led to the emergence of video-based intelligence software for enhanced surveillance and intelligence gathering.

- Smart Surveillance Systems involve automated pattern recognition and signal analysis software embedded in the hardware of the video sensors extract which "usable information" from video and sensor nodes automatically.

1.1.2 Applications of Video Surveillance Systems

- Indoor Surveillance and Monitoring
- Moving vehicle detection and classification
- Vehicle tracking systems
- Surveillance in Military Scenarios
- Public transport Security
- Scene recognition and Situation Awareness
- Human Motion Capture (i.e.) analyzing the movements of people or their posture and identifying the presence of elderly people, children, etc.
- Region categorization
- Surveillance of Car parking Areas
- Detection of unattended packages in public spaces
- Detection of stolen objects in a museum
- Monitor at-risk individuals, groups, and installations
- Automatic threat analysis of complex events
- Surveillance of Restricted Areas
- Vehicle number plate recognition
- Automatic detection of security-related events
- Object detection, recognition, and tracking
- Covert Surveillance of Sensitive Zones

- Gait Analysis
- Intrusion Detection and Motion Sensing Alarm Systems

1.2 Multimodal Surveillance Systems

A multimodal surveillance system normally consists of a wireless sensor network of video/image sensors, audio sensors, pressure sensors, thermal sensors and position sensors. Apart from these, some recent advances in sensor hardware also include an embedded data processing algorithm which is used to process the data captured by the sensor and send it to a base station. The main advantage of a multimodal surveillance system is its capability to,

- Integrate different kinds of sensors
- Process input from each sensor independently
- Provide stability in handling ambiguous or missing inputs
- Extract the features of the surveillance area from different viewing angles and hence adapt dynamically to the environment

1.2.1 Intelligent Transportation Systems

Intelligent transport systems apply data processing, communication protocols, and sensor technologies to vehicles (including cars, trucks, trains, aircraft and ships), transport infrastructure and transport users to increase the efficiency, effectiveness, overall performance, environmental performance, safety and resilience of the transport system.

1.2.2 Applications of a Multimodal Surveillance System in ITS

- Intersection control - Multimodal surveillance systems can be used in some of the most basic traffic management tasks like calculating the difference between timings of green signals among different traffic flows and deciding the total signal cycle at junctions and intersections.
- Incident detection - Multimodal surveillance systems can be used for identifying the locations of vehicle breakdown or accidents.

- Vehicle classification - Multimodal surveillance systems can be utilized in determining the type and proportion of vehicles that manoeuvre on a certain stretch of road. This information can in turn be used to choose appropriate road width and pavement materials while developing new transportation infrastructure.

- Object tracking - Multimodal surveillance systems can be utilized for tracking on-road objects (which includes both vehicles and human beings) and their movements across different paths.

- Historical traffic data - Data captured using multimodal surveillance systems and stored over a long period of time can be used for calibrating traffic signal cycles, planning new infrastructure, adding new public transport infrastructure, and so on.

- Individual vehicle management - Multimodal surveillance systems can be used to monitor parking places and obtain carbon footprint estimates of private vehicles.

- Public transport information - Multimodal surveillance systems can be used to track public transport services and predict the arrival of public transport vehicles which would reduce waiting times and delays.

- Accident handling - Multimodal surveillance systems can be used to provide quick emergency services after accidents by capturing various information from the surveillance area.

1.3 The Indian Scenario

The need of ITS in India is evident with the emergence of initiatives like Smart Cities and Digital India. While taking into consideration the current scenario of the transportation infrastructure in India and the application areas of multimodal surveillance systems; there are several factors leading to the need of such a system in India, and some of them are listed below:

- Initiative of Smart Cities in India

- Problem of traffic congestion, frequent accidents and traffic mismanagement

- Problem of improper lane driving

- Need of vehicle identification
- Problem of Over-speeding
- Need of low-cost solution
- Need of wireless solution

1.3.1 Challenges

There are several challenges, pitfalls and issues that need to be taken into consideration while designing a multimodal surveillance system for ITS in Indian transportation infrastructure [1] like,

- The complexity inherent in urban surveillance requires collaborative work from different sensors to enable certain capabilities such as detection of suspicious events despite the external factors affecting the scene.
- Multimodal surveillance systems are scalable but real dense surveillance networks are less connected and still not deployed. Thus access to numerous diverse multimodal sensors is a challenging task owing to the issue of limited bandwidths in the Indian scenario
- Vehicles on the road are typically in motion, introducing effects of relative motion. There is variability in the type, shape, color and size of vehicles encountered on the road.
- The on-road environment features present variations in illumination, background, and scene complexity. Man-made structures, complex shadowing, occlusions and ubiquitous visual clutter can introduce errors.
- Vehicles are also encountered in a variety of orientations, including overtaking traffic and cross traffic (no proper lane driving system in the Indian scenario).
- The on-road environment features, frequent and extensive scene clutter limit the full visibility of vehicles resulting in partially occluded vehicles.
- Furthermore, a vehicle detection system needs to operate at real-time speeds in order to provide the human or autonomous driver an advanced notice of critical situations

- Vehicle tracking aims to identify and measure dynamics and motion characteristics of moving vehicles; and predict and estimate the upcoming position of vehicles on the road from the captured frames. But, due to no proper lane driving it becomes a challenging task to track vehicles as it increases occlusions in the video.
- Requirement of less power consuming, less bandwidth hungry and economical solution.

Chapter 2

Literature Review

Multimodal surveillance systems in ITS have a wide application area and has been an emerging field of research. Many research studies have been carried out in this area and brief descriptions about some of them are provided in this chapter.

2.1 Domain of Survey

The area of multimodal surveillance system is very wide and since the focus mainly is to present novel methods for the Indian transportation scenario, the domain of survey carried out has been narrowed down to **Multimodal Video Surveillance for Intelligent Transportation System**.

2.1.1 Literature Survey

- Rhalem Zouaoui, et. al in [2] have proposed a multimodal system composed of two microphones and one camera integrated with on board video and audio analytics and fusion capabilities for surveillance. The system relies on the fusion of "unusual" audio events detection with target or object detection from the captured video sequences. The audio analysis consists of modeling the normal ambience and detecting deviation from the trained models during testing while the video analysis involves classification according to geometric shape and size. However even though the system succeeds in detecting robust 3D position of objects it employs only a single camera for surveillance which does not provide a robust multi-dimensional view of the object of interest.

- T. Wang, et. al. in [3] have presented a system for detecting and classifying moving vehicles. The system uses video sensors along with Laser Doppler Vibrometer (LDVs) – a kind of acoustic sensor for detecting the motion, appearance and acoustic features of the moving vehicles and later on using the data to classify them.

- M. Magno, et. al. in [4] have proposed a multimodal low power and low cost video surveillance system based on a CMOS video sensor and a PyroelectricInfraRed (PIR) sensor. In order to control the power consumption, the sensors do not transmit the full image but, some very limited amount of information such as number of objects, trajectory, position, size, etc. thus saving a large amount of energy in wireless transmission and extending the life of the batteries. However nothing is done from the point of view of data transmission and power consumption if the targeted object is not detected. In addition to this, this system is used only for detecting an abandoned or removed object from the perimeter under surveillance and hence there is no proper evidence of its usage in a large-scale, dynamically changing environment.

- H. Gupta, et. al. in [5] have designed a distributed visual surveillance system for military perimeter surveillance. The system is used to detect potential threats and create actionable intelligence to support expeditionary war fighting for the military base camp by using multimodal wireless sensor network. The system employs certain rule-based algorithms for detection of atomic actions from video. Some of the atomic actions that are automatically detected by the system are: a person appearing in a restricted area, tripwire crossing, a person disappearing from a protected perimeter, a person entering or exiting, leave behind action, loiters, take away action, etc. A geodetic coordinate system is used which provides metric information such as size, distance, speed, and heading of the detected objects for high level inference and inter-sensor object tracking.

- A. Prati, et. al. in [6] have proposed a PIR sensor based multimodal video surveillance system. In this system PIR sensors are used to bring down the cost of deployment of the surveillance systems and at the same time they are combined with vision systems for precisely detecting the speed and direction of the vehicles along with other complex events.

- R. Rios-Cabrera, et. al. in [7] have presented an efficient multi-camera vehicle identification, detection and tracking system inside a tunnel. In this system a network of non-overlapping video cameras are used to detect and track the vehicles inside a tunnel by creating a vehicle-fingerprint using the haar features of the vehicles despite poor illumination inside tunnel and low quality images.

- K. Lopatka, et. al. in [8] have proposed a system for detecting the traffic events which uses special acoustic sensors, pressure sensors and video sensors to record the occurrence of audio-visual events. A use-case of detection of collision of the two cars is demonstrated in this paper. The data collected by the multimodal sensors is sent to a computational cluster in real time for analysis of the traffic events. For this purpose a Real Time Streaming Protocol (RTSP) is used in the system.

- Y. K. Wang, et. al. in [9] have proposed a large scale video surveillance system for wide area monitoring which has capability of monitoring and tracking a moving object in a widely open area using an embedded component on the camera for detailed visualization of objects on a 2D/3D interface. In addition to this, it is also capable of detecting illegal parking and identifies the driver's face from the illegal parking event.

- A. van den Hengel, et. al. in [10] have proposed a genetic algorithm for automatic placement of multiple surveillance cameras which is used to optimize the coverage of cameras in large-scale surveillance systems and at the same find overlapping views between cameras if necessary.

- E. Yildiz, et. al. in [11] have presented a bi-level algorithm to determine an optimal camera placement with maximum angular coverage for a WSN of homogeneous and heterogeneous cameras.

- J. Zhao, et. al. in [12] have presented two binary integer programming (BIP) algorithms for finding optimal camera placement and network configuration. Moreover they have extended the proposed framework to include visual tagging of subjects in the surveillance environments.

- L. Liu, et. al. in [13] have presented a Multi-Modal Particle Filte technique to track vehicles from different views (frontal, rear and side view). In addition to this they have also discussed a technique for occlusion handling in surveillance systems.

- S. Denman, et. al. in [14] have presented a system for automatic monitoring and tracking of vehicles in real time using optical flow modules and motion detection from videos captures by four video cameras.

- K. Wang, et. al. in [15] have proposed an effective foreground object detection technique for surveillance systems by estimating the conditional probability densities for both the foreground and background objects using feature extraction techniques and temporal video filtering.

- R. Zheng, et. al. in [16] have proposed a key frame selection technique based on motion-feature based approach in which motion information for each key frame from the traffic surveillance video stream is computed in a GPU based system and key frames with motion information greater than their neighbors are selected. By implementing GPU based processing capabilities, the authors have shown a significant increase in the accuracy and processing speed of the algorithm.

- R. Panda, et. al. in [17] have proposed a novel sparse representative selection method for summarizing multi-view videos, that is videos captured from multiple cameras. They have used inter-view and intra-view similarities between the feature descriptors of each view for modelling multi-view correlations.

- S. K. Kuanar, et. al. in [18] have proposed a bipartite matching method for multi-view correlation of features like visual bag of words, texture, color, etc. and extracting frames for summarization of multi-view videos. In this method the authors have used Optimum-Path Forest algorithm for clustering the intra-view dependencies and removing intra-view redundancies.

- S. Liu, et. al. in [19] have proposed a unique method for visualizing object trajectories in multi-camera videos and creating video summaries of suspicious movements in a building.

2.2 Proposed System

On the basis of the domain of survey carried out, and keeping into consideration the Indian transportation infrastructure scenario; the proposal is to develop an efficient camera placement algorithm in a multimodal surveillance system capable of providing maximum coverage of the area under surveillance, maximizing the view of subjects, minimizing occlusions in high vehicular traffic areas and completely eliminating or reducing to a great extent the number of blind zones in a surveillance area. This camera placement algorithm can be used to find optimal placements for multiple-cameras in a multimodal surveillance system for deployment at intersections, junctions and crossroads. Furthermore, the proposal is to develop a video summarization algorithm which can be used to create summaries of videos captured in a multi-view surveillance system. Such a video summarization algorithm can be used further for object detection, motion tracking, traffic segmentation, etc. The application of the proposed video summarization algorithm can also be extended for use in a multimodal surveillance system containing different kinds of sensors.

Chapter 3

Cameras - Analog and Digital

Over the years the technology used in surveillance systems for recording and storing the videos have changed drastically. From analog video cameras to the latest IP-based digital video cameras, and use of video tapes to DVRs and now cloud storage; the type of cameras used in multimodal surveillance systems have a profound impact on the output of such systems. Different applications require cameras with application-specific intrinsic and extrinsic parameters. With respect to the scope of the proposed system, the first step was to decide the kind of camera that would be suitable to the applicability of the proposed system. This chapter provides a brief overview of how technological advances have led to the evolution of cameras and a comprehensive comparison of various types of cameras available commercially based on different intrinsic and extrinsic camera parameters.

3.1 Types of Cameras

3.1.1 Analog Video Cameras

Analog video cameras involve use of VCR to store videos and they use wired connections to transmit the video from source to the base station. Analog video cameras can stream the video to a video monitor through which the perimeter under surveillance can be monitored by a security personnel. Analog video cameras do not have P/T/Z capabilities, due to this they have a limited FOV. Analog video camera systems involve human intervention for changing the video tapes manually. It also requires a security personnel to take decisions (like raising an emergency alarm) based on the information visible on the video monitor. Analog video cameras do not have the functionality of providing a clear, visible video output in low light conditions. Analog video cameras provide a SD resolution video output. Analog video camera systems are cost effective as compared to Digital video camera systems. Also these cameras require less bandwidth to transmit the videos to the base station.

3.1.2 Digital Video Cameras

Digital Video Cameras provide HD resolution video output. Digital Video Cameras provide low-light video capturing features by using integrated IR Sensors. Digital Video Cameras use DVR which converts the analog signal to digital signal and stores the video on an attached hard drive or memory card for later retrieval. Digital Video Cameras also provide P/T/Z facility which increase the angular FOV to a great extent. Moreover certain digital video cameras provide 360° rotation thus providing a wider camera coverage. DVR systems have inbuilt intelligence software that can detect motion and objects in the scene. Thus, with digital video cameras, the decisions are taken by the system itself based on the intelligence gathered from the streaming video. The most advanced digital video cameras also provide a two-way audio communication mechanism.

3.1.3 IP-based Digital Video Cameras

IP-based digital video cameras, also called networked digital video cameras are the most technologically advanced version of digital video cameras. Each IP-based camera has an IP address associated with it along with an in-built wireless transmitter which is capable of broadcasting the videos captured over a wireless network to a base station and also remotely stream the videos live through an app/web based interface. IP-based cameras use a Network Video Recorder (NVR) for digital signal processing, recording and compressing the captured videos. IP-based digital cameras have an in-built P/T/Z functionality which can be controlled remotely. They also have video analytics capability which can be used to process the video signals and mine intelligence from it in real time. IP-based cameras also provide excellent low light video imaging capabilities. IP-based cameras provide encryption of the captured videos during streaming and later on inside the storage where they are stored. These cameras require a large bandwidth to stream the high resolution videos over the network in real time. Also these cameras are costlier than the analog cameras due to the additional features embedded in them.

Parameter	Analog Video Cameras	Digital(Wired and Wireless/IP) Video Cameras
Resolution	SD	HD upto 1080p
Low-light Imaging	Not Possible	Possible with IR Sensors
Angular FOV Coverage	Narrow	Wide
P/T/Z	Not Available	Available, endless 360° coverage
Storage	Video Tape	Hard drive attached to DVR, Network Attached Storage via NVR
Video Analytics Capability	Not Embedded in the Camera	Embedded in the Camera
Video Transmission Medium	Wired	Wired/Wireless
Remotely Controlled	Not Possible	Possible
Bandwidth requirement for streaming the captured videos	Low	High
Cost	Low	High

Table 3.1: Difference between Analog Video Cameras and Digital(Wired and Wireless/IP) Video Cameras

3.2 Commercially available Cameras

There are various surveillance cameras available commercially in the market ranging from simple digital video cameras with minimal functionalities to high-end IP-based cameras with real time cloud storage capabilities. Various parameters are used for comparing these cameras and brief descriptions for primary parameters are provided below:

- Image Sensor: The image sensor operates like a retina for the video camera. It converts the visible image into an electrical (analog/digital) signal by using the light sensitive photosites present on them which record the information seen through the lens. There are two types of sensor chips namely CCD and CMOS, and the size of the sensor chip determines the quality of the image produced. Higher the size of the chip, more will be the photosites and better would be the image quality.

- Lens (Focal Length): The main function of the lens is to focus the illuminating object onto the sensor. The focal length of a lens is used to determine the angular FOV covered by the camera. Less the focal length, larger the FOV and wider is the coverage of the perimeter under surveillance. Also there are mainly two kinds of lens; one with a fixed focal length and another with a variable focal length (also called vari-focal lens); the later one is used to provide zoom-in and zoom-out capability to the video camera.

- Sensitivity (Illumination): The sensitivity of a camera is used to determine its performance in low-light conditions. That is it is a measure of determining the minimum level of light required to get an acceptable image and is measured in a unit called lux (lx). It is defined by minimum illumination for color and B/W images. However sensitivity is dependent on many other factors like color temperature, reflection ratio, etc.

- Pan/Tilt/Zoom: In a video camera, P/T/Z is used to determine the horizontal and vertical movement of a camera and the zoom-in/zoom-out capability of the lens. This feature provides a wider coverage of the area under surveillance. This mechanism is used in cameras used for moving object detection, specific object detection, face detection, etc.

- Signal-to-Noise (S/N) Ratio: The S/N ratio is used to compare the level of desired signal to the level of background noise. The S/N ratio [20], [21], [22] is measured in decibels (dB). This ratio is used to determine how much noise will be superimposed over a picture signal inside the camera. Every time a video signal is processed there is some form of noise that is introduced due to the presence of electronic components (resistors, transistors and capacitors) inside the device. Higher the ratio, lesser is the presence of noise in the picture and better the picture quality.

- Frame rate: Frame rate is a measure of the number of frames of images captured in one second by the video camera. While lower frame rates can lead to disrupted and less sharp images, higher frame rate leads to smoother image. Although higher frame rate also increases the size of the video file, it also requires more bandwidth to stream the video in real time and more storage to store the video for later retrieval.

- Resolution: It defines the number of pixels per inch represented in horizontal and vertical field that can be recorded in a single frame and processed by the camera. Higher the number of pixels, higher is the resolution leading to sharper and finer images. With high resolution, it is possible to zoom in on the image without losing the detail of the image.

- Compression: The process of transmitting a video over a network in a way that the original quality of the video is maintained as well as the total size of the video file being transmitted is optimized so as to reduce the total data that the system needs to process is called compression.

3.2.1 Comparison of (Wired) Digital Video Cameras

The table 3.2 below presents a detailed comparison of various wired digital video cameras available commercially.

Product/ Parameter	**iBall Guard B8062SW**	**Samsung SDC-73-40BCN**	**Panasonic WV-CW594A**	**Panasonic WV-CW304L**	**Sony SS-CCB564R**	**SMTSEC SP19E**
Image Sensor	1/3" DIS	1/3" CMOS	1/4" CCD	1/3" CCD	1/3" Ex-View HAD CCD II	1/4" Sony CCD Effio DSP
Resolution	800 TVL	720 TVL (976 x 494 pixels)	650 TVL (976 x 494 pixels)	650 TVL (976 x 494 pixels)	700 TVL (976 x 494 pixels)	480 TVL (795 x 596 pixels)
S/N Ratio	52 dB Minimum	46 dB minimum	52 dB minimum	52 dB minimum	55 dB minimum	60 dB minimum
Day/Night Light Sensitivity (ICR)	Yes	Yes	Yes	Yes	Yes	Yes
Minimum Illumination Color	-	7 lx	0.5 lx	0.15 lx	0.06 lx	0.01 lx
Minimum Illumination B/W	-	0 lx	0.04 lx	0 lx	0 lx	0.01 lx
IR LED	Yes	Yes	No	Yes	Yes	Yes
IR Range	25 m	25 m	-	20 m	30 m	120 m
Lens (Focal Length)	6 mm	3.8 mm	3.3 mm	2.0 mm	2.8 mm to 10.5 mm	3.9 mm
Angle of View	60°H	75°H	60°H	90°H	101.8°to 27.4°H	70°H
Pan/Tilt	-	-	360° endless Panning Range, 400°/s (Preset Mode) Tilting Speed	-	-	360° endless Horizontal rotation, 180° Vertical auto reversal, 200°/s Tilting Speed
Zoom	-	-	-	Upto 10x Digital Zoom	Upto 10x Digital Zoom	Upto 12x Digital Zoom

Table 3.2 continued						
Parameter /Product	**iBall Guard B8062SW**	**Samsung SDC-7340BCN**	**Panasonic WV-CW594A**	**Panasonic WV-CW304L**	**Sony SS-CCB564R**	**SMTSEC SP19E**
Signal Mode	PAL	NTSC/ PAL	NTSC/ PAL	PAL	NTSC	NTSC/ PAL
Video Output	1.0Vp-p composite output(75 ohm/BNC)	1.0 V [p-p] /NTSC/ PAL composite 75	1.0 V [p-p] /NTSC/ PAL composite 75 /BNC	1.0 V [p-p] /PAL composite75 0	1.0 V [p-p] /NTSC composite75 0	1.0Vp-pcomposite output(75 ohm/BNC)
Water Resistant	Yes, IP66	Yes, IP66	Yes, IP66	Yes, IP66	Yes, IP66	-
Power Source	12V DC	12V DC	24V AC, 12V DC	24V AC, 12V DC	24V AC, 12V DC	12V DC
Cost (Approximate Conversions - at the time of writing)	USD 39.99 (Rs. 2800)	USD 79.99 (Rs. 5500)	USD 259.99 (Rs. 17500)	USD 369.99 (Rs. 25000)	USD 449.99 (Rs. 34000)	USD 299.99 (Rs. 20500)
Source	[23]	[24],[25]	[26],[27]	[28],[29][30]	[31],[32]	[33]

Table 3.2: Comparison of (Wired) Digital Video Cameras

3.2.2 Comparison of IP Cameras

The below table 3.3 presents a detailed comparison between some of the most prominent wireless digital video cameras available commercially for video surveillance.

Product/ Parameter	**YI Dome Camera**	**Belkin NetCam HD+**	**Amcrest ProHD IP2M-841B**	**Samsung Smart-Cam HD Pro SNH-P6410BN**	**D-Link DCS-2310L**	**AXIS Q8665-E**
Image Sensor	1/4” I Megapixel CMOS	1/3.2” CMOS	1/2.7” 2 Megapixel progressive scan CMOS	1/2.8” 2M CMOS Sony, FHD	1/4” 1 Megapixel progressive scan CMOS	1/2.9” Progressive Scan RGB CMOS
Resolution	1280 x 720 pixels	1280 x 720 pixels	1920 x 1080 pixels	1920 x 1080 pixels	1280 x 720 pixels	1920 x 1080 pixels
Minimum Illumination Color	-	-	0.1 lx	0.3 lx	1 lx	0.5 lx

Table 3.3 Comparison of IP Cameras continued						
Product/ Parameter	**YI Dome Camera**	**Belkin NetCam HD+**	**Amcrest ProHD IP2M-841B**	**Samsung Smart-Cam HD Pro SNH-P6410BN**	**D-Link DCS-2310L**	**AXIS Q8665-E**
Minimum Illumination B/W	-	-	0 lx	0 lx	0.5 lx	0.04 lx
Lens (Focal Length)	-	3.37 mm	4.0 mm	2.8 mm	3.45 mm	4.7 mm
Field of View	112°(Wide Angle View)	95°/57°/-76° (D/V/H)	90° (Wide Angle View)	128°/ 62°/ 111° (D/V/H)	70°/36°/ 60° (D/V/H)	59°(H)
Day/Night Filter	Yes	Yes	Yes	Yes	Yes	Yes
IR Vision	Yes	Yes	Yes	Yes	Yes	Yes
IR Range	3 m	8 m	9 m	10 m	5 m	-
Pan/Tilt	Upto 345° Horizontal Panning and upto 115°Vertical Tilting	-	Yes (Remote Viewer Support)	Yes (Viewer Support)	Yes	Pan Speed of 0.02°/s to 100°/s – Endless 360° Viewing, Tilt Speed of 0.02°/s to 40°/s
Zoom	-	-	Yes	Upto 10x Digital Zoom	Upto 10x Digital Zoom	Upto 12x Digital Zoom
Frame rate	20 fps	25 fps	30 fps	30 fps	30 fps	25-30 fps
Video Compression Format	H.264	H.264	H.264H, H.264B, H.264, MJPEG	H.264, MJPEG	H.264, MJPEG	H.264, MJPEG
Audio Compression Format	-	G.711, PCM	G.711MU, G711A, ACC	G.711, G.726	AAC, G.711	-
In-built Memory Storage	Yes, upto 32 GB	No (Cloud Storage)	Yes, upto 64 GB	Yes, upto 64 GB	Yes, upto 32 GB	-
Inbuilt Video Analytics Capability	Motion Detection	Motion Detection	Motion Detection	WDR, Motion and Audio Detection	Motion and Audio Detection	Motion Detection

Table 3.3 Comparison of IP Cameras continued

Product/ Parameter	**YI Dome Camera**	**Belkin NetCam HD+**	**Amcrest ProHD IP2M-841B**	**Samsung Smart-Cam HD Pro SNH-P6410BN**	**D-Link DCS-2310L**	**AXIS Q8665-E**
Wireless Transmission	Wi-Fi 802.11b/g/n	Wi-Fi 802.11b/g/n	WiFi (802.11-b/g)	Wi-Fi 802.11 a/b/g/n (Dual-band)	Wi-Fi 802.11 a/b/g/n	-
Protocol Support	-	No	IPv4, IPv6, HTTP, HTTPS, TCP/IP, UDP, UPnP, ICMP, IGMP, RTSP, RTP, SMTP, NTP, DHCP, DNS, PPPOE, DDNS, FTP, IP Filter, QoS	TCP/IP, DHCP, SMTP, DNS, RTSP, RTCP, RTP, HTTP, TCP, UDP, STUN, TURN, XMPP, SIP, uPNP, SNTP, IPv4, ICMP, Bonjour	IPv6, IPv4, ARP, TCP /IP,UDP, ICMP, DHCP client,NTP client(D-Link), DNS client, DDNS client(D-Link), SMTP client, FTP client, HTTP/ HTTPS, Samba client, PPPoE,UPnP, UPnP port forwarding, RTP/RTSP /RTCP,IP filtering, QoS, CoS, DSCP, Multicast,IGMP, ONVIF compliant, Bonjour, SNMP v1, v2c, v3	-

Table 3.3 Comparison of IP Cameras continued						
Product/ Parameter	**YI Dome Camera**	**Belkin NetCam HD+**	**Amcrest ProHD IP2M-841B**	**Samsung Smart-Cam HD Pro SNH-P6410BN**	**D-Link DCS-2310L**	**AXIS Q8665-E**
Power Consumption	-	-	Max 7.5 W	Max. 11 W	Max. 5.5 W	-
Cost (at the time of writing)	Rs. 9501	Rs. 12849	Rs. 13958	Rs. 17297	Rs. 19021	Rs. 300000
Source	[34],[35], [36]	[37],[38], [39]	[40],[41], [42]	[43],[44], [39]	[45],[46]	[47],[48]

Table 3.3: Comparison of IP Cameras

3.3 Programmable Cameras

There are several IP cameras available commercially which can be used for real-time streaming however taking into consideration the cost factor of such cameras, another alternate is to customize a camera according to the requirements. There are several camera modules available which can be mounted over a motherboard (like Raspberry Pi Camera Module) and can be programmed accordingly to accomplish the requirements. However, comparison of such programmable cameras is not provided since it comes outside the scope of the proposed system.

Chapter 4

Optimal Camera Placement in Multimodal Surveillance System

While designing a multimodal surveillance system, it is paramount that the placement of cameras is done in such a way that maximum area of surveillance is covered and at the same time the purpose of deployment of such a system is fulfilled. As the system should be such that it is capable of dealing with multimedia vector data of audio and video obtained from the camera, for modeling and understanding the complexity of a surveillance system, it requires knowledge and competencies of various areas from networking to data management. The modeling of multimedia data in a multimodal surveillance system requires a network simulator for simulating the complex process based on real values before arriving upon a real-world solution. Thus a network simulator is required to validate the design and behavior of the network of cameras, model the network topology and analyze the output data for performance evaluation. This chapter provides a brief comparison of some of the open-source and commercially available network simulators. Later in this chapter, an optimal camera placement algorithm is discussed which can be used to decide the placements of multiple cameras while designing a multimodal surveillance system.

4.1 Requirement of an Optimal Camera Placement Strategy

The placement of cameras in a multimodal surveillance system has a substantial impact on its performance, however the placement of cameras depends on the objective of the system which needs to be served. In the context of an ITS, there are several factors which need to be considered while deciding the placement of cameras like:

- Maximizing the coverage of cameras with respect to the area of surveillance
- Maximizing the view of subjects in the video
- Minimizing the amount of occlusions in high vehicular traffic areas
- Generating high resolution videos from multiple cameras so as to track the objects in motion from multiple views
- Minimizing the overall cost of the system and maximizing the performance of the system
- Completely eliminating or reducing to a great level the number of blind zones in a permieter under surveillance

While the goal of many optimal camera placement strategies has been to minimize the overlapping views; with respect to the objective of the proposed system, overlapping views were necessary so as to track the complete path of motion of the subjects under surveillance from multiple views, maximize their visibility and maximize the degree of coverage of the surveillance perimeter. Many large-scale multimodal surveillance systems have used human experts for camera selection and placement, however such a technique is not capable to effectively design a system while considering the multitude of factors listed above. Also a straightforward method to deploy the video cameras would be to deploy them uniformly around the surveillance area. However, in real-world deployment scenarios, such a method of uniform placement is not practical since the placement of cameras is restricted by many constraints like costs, availability, visibility, applicability, feasibility, and other factors which have been listed above. This study has investigated the effect of all the factors listed above and an optimal camera placement strategy has been designed which satisfies all these factors.

4.1.1 Camera System

In order to understand the impetus behind the proposed optimal camera placement strategy, it is important to understand the basics of a camera system. As shown in figure 4.1, below are some technical terms corresponding to the camera system:

- Focal Length: It is the distance between the sensor of the camera and the optical centre of the lens when the lens is focused on a subject lying at infinity (means at an unknown distance).

- Angle of View (AOV): It is the angle between the two farthest points on the area of the subject projected on the sensor by the lens.

- Field of View (FOV): It is another manner of representing the AOV but expressed as a measurement of the area of the subject rather than the angle.

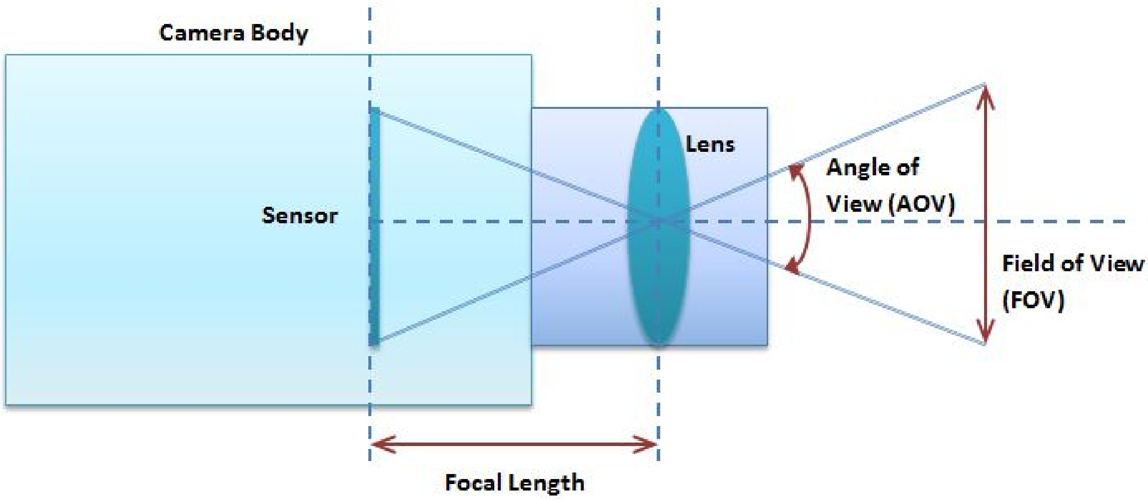

Figure 4.1: Basic structure of a camera

In a three-dimensional (3-D) space, the FOV of a camera can be defined by its placement, and its Horizontal Angle of View and Vertical Angle of View. Figure 4.2 highlights a cross-section of the FOV projected in the three-dimensional space by a camera while viewing a subject. However, it is worth noting that the FOV of a camera depends on the size of the camera sensor and the focal length of the lens; and shorter the focal length, larger is the FOV for a camera.

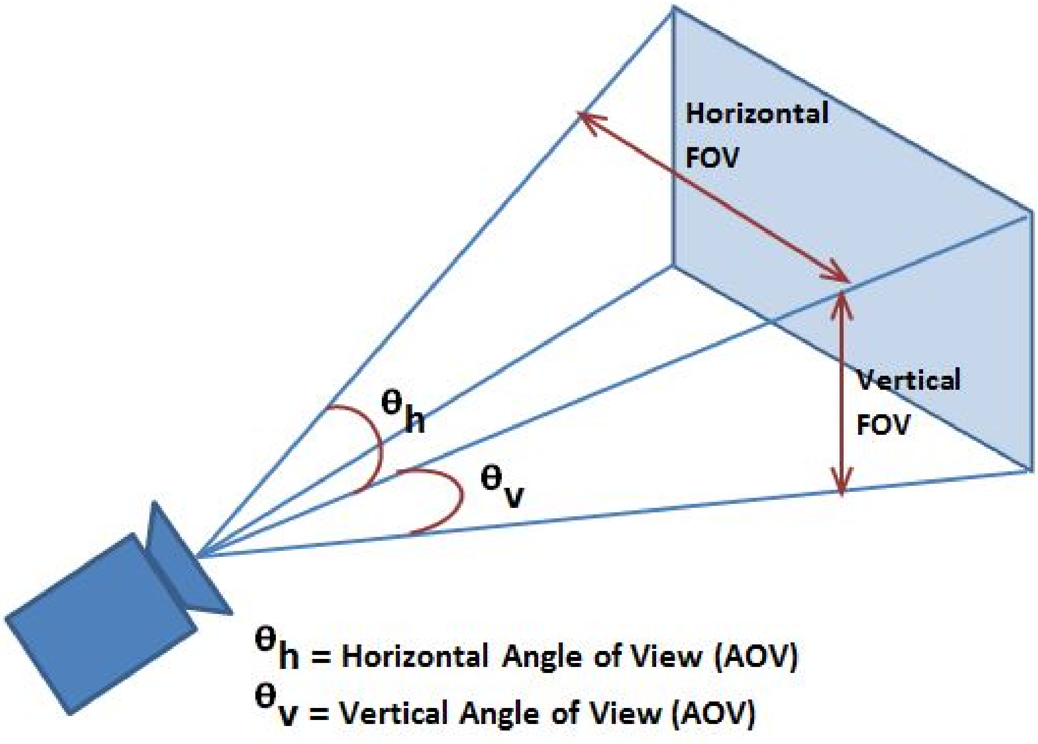

Figure 4.2: FOV of a camera

4.1.2 An Optimal Camera Placement Algorithm

In this section, the algorithm developed to achieve an optimal camera placement for a multimodal surveillance system in an ITS is discussed. For this algorithm there were several assumptions which have been considered and the same have been listed below:

- The cameras deployed in the system were fully calibrated.
- All the cameras shared the same intrinsic and extrinsic parameters like FOV, AOV, Focal length, Resolution, Lens aperture, Sensor size, Height at which the cameras were mounted, Mounting angle of the cameras, GOP size, Encoding method, Video codec, Audio codec and frame rate.
- All the cameras had fixed focal length and zooming feature was not available in them.
- The surveillance area modelled was Rectangular in shape and the maximum depth of the area was same as the height of the camera mounted in the system.

As shown in figure 4.2, in a three-dimensional space, the coverage area of a camera can be modelled as a rectangular pyramid whose apex is directly at the centre of the base. The figure 4.3 gives the coverage of a video camera C in three-dimensional space. With reference to the figure 4.3, point V is the position of the video camera V(x,y,z) and point G indicates the centre of gravity for the video camera V. The four points A,B,C and D are the extreme points in the FOV of V and can be computed using horizontal AOV, vertical AOV and position of the video camera. These points also form the base plane of the rectangular pyramid. Point X is an arbitrary point present in the FOV of video camera V which is to be observed using V.

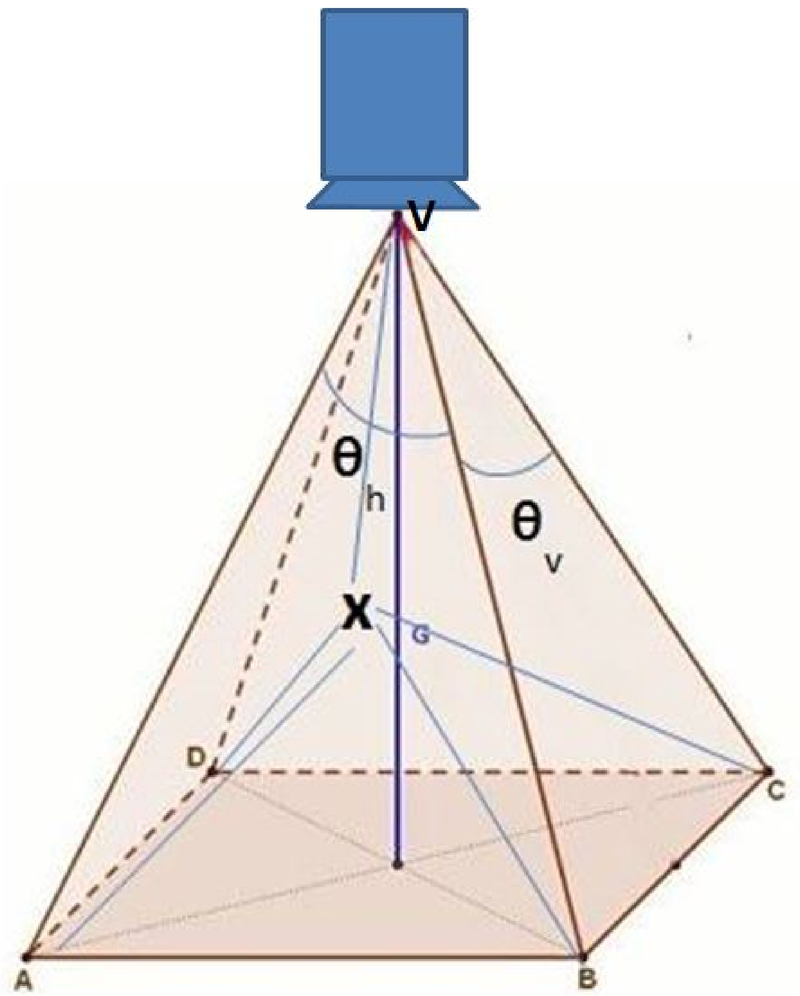

Figure 4.3: Coverage of a Video Camera

The volume of the rectangular pyramid formed by the points {V,A,B,C,D} (that is the volume of the coverage area of the camera C) is given by the equation 4.1,

$$V_C = \frac{l * b * h}{3} \tag{4.1}$$

where h is height of apex from the base

Now since X is an arbitrary point inside the FOV of V, it forms four tetrahedrons with the four sides of the pyramid and point V as the apex. Volume of each such tetrahedron is given by the equation 4.2,

$$V_i = \frac{\sqrt{2} * Area_{base} * h}{12} \tag{4.2}$$

where h is height of apex from the base and i=1 to 4

Consider V_{total} as the total volume computed by adding volumes of all the four tetrahedrons and pyramid created with X as the apex. V_{total} is given by the equation 4.3,

$$V^X_{total} = V_{base} + \sum V_i \tag{4.3}$$

for all i= 1 to 4

In equation 4.3, V_{base} is the volume of pyramid with point X as apex and points A,B,C,D as the base plane. The equation 4.4 is used to test the presence of a point X within the FOV of a video camera C whose coverage area can be modelled as a rectangular pyramid of volume V, and returns true or false accordingly.

$$FOV(C, X) = \begin{pmatrix} true, if \ V_C = V^X_{total} \\ false \end{pmatrix} \tag{4.4}$$

Algorithm 1 Optimal Camera Placement

Result: $C_i(x, y, z)$

Require: Surveillance area P of size LxBxH

Cameras C_i, where i = 1 to 4

Initialization

Find midpoint M_j for each side of the region LxB of P and divide the region LxB of P into four equal regions R_i; where i = 1 to 4

for $i = 1$ to 4 **do**

  For region R_i, midpoints M_j on the adjacent edges of R_i and camera C_i, find $C_i(x, y, z)$ using **function** FOV(C_i,M_j) such that $C_i(x, y, z)$ lies inside $R_i(x, y, z)$

end for

function: FOV(C_i,M_j)

  Input: $C_i(x, y, z), M_j$

  Output: TRUE/FALSE

  $result$ = FALSE

  For camera C_i find volume of its coverage area V_i

  For point M_j find individually volume of four tetrahedrons $V_n^{M_j}$ formed by M_j with $C_i(x, y, z)$ as apex and each of the four sides of the coverage area of $C_i(x, y, z)$, where n =1 to 4

  Find volume $V_{base}^{M_j}$ of the pyramid formed with M_j as apex and the base plane of the coverage area of $C_i(x, y, z)$ as the base plane of the pyramid

  Find the total volumes of all the tetrahedrons and pyramid created with M_j as: $V_{total}^{M_j} = V_{base}^{M_j} + \sum V_n^{M_j}$, where $n = 1$ to 4

  if $V_{total} == V_i$

    $result = TRUE$

  end if

  return $result$

end function

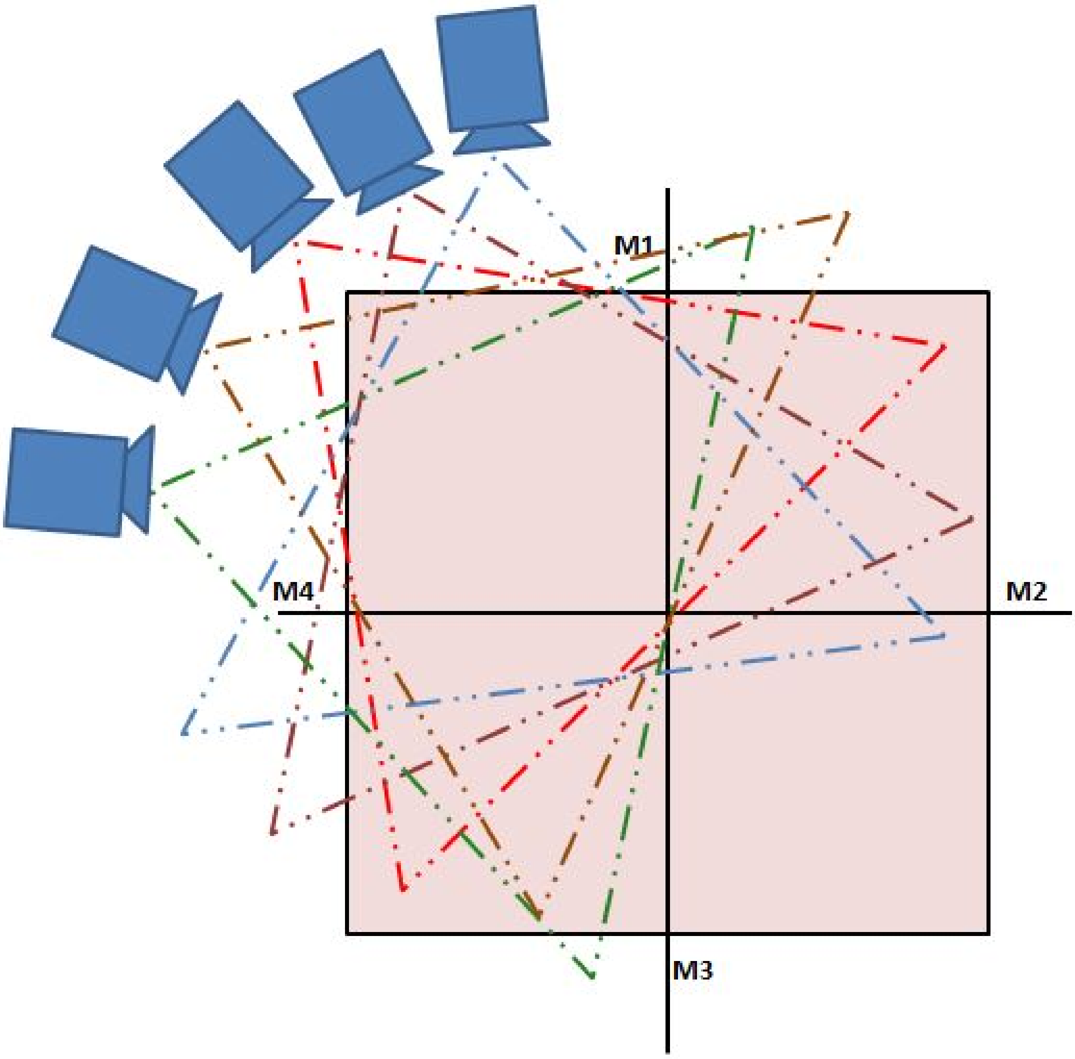

Figure 4.4: Sample Surveillance Area with various possible Camera placements

Now consider the surveillance area as shown in the figure 4.4. The first step is to divide the surveillance area into four equal parts and find the midpoint for each side of the surveillance area. In the figure 4.4, points M1, M2, M3 and M4 are the midpoints of the surveillance area. The next step is to place a camera at the region between any two adjacent sides such that the FOV of that video camera would cover the midpoints of any two adjacent sides so as to cover the maximum part of that region. For a multimodal surveillance system, each video camera should fulfill the equation 4.4 for all the midpoints of the sides of the surveillance perimeter and the position of the video cameras thus obtained are the best placements that provide the best coverage of the whole region under surveillance. The Algorithm 1: Optimal Camera Placement depicts systematic steps for calculating optimal camera placement in a multimodal surveillance system. Also note that this algorithm was designed keeping in mind the assumptions listed earlier at the beginning of the section 4.1.2. This algorithm can be used to decide the placement of multiple cameras at intersections, junctions and crossroads and achieve the best possible coverage of the surveillance area.

4.2 Tools and Simulators

The validity of the algorithm 1: Optimal Camera Placement was checked using a network simulator. At the time of writing this text and designing the algorithm 1: Optimal Camera Placement, wide range of network simulators were available in the market; some of them were open source while some were commercial. The open source network simulators covered here include OMNET++, NS-2, J-Sim while commercially available simulators covered in this text include Riverbed Modeller and QualNet. These simulators were compared mainly on the basis of programming language used by them, protocol support, their extendability and their features. The table 4.1 presents a brief comparison of some of these network simulators.

Simulator	Type	Language	Features	Visualization	Extension
OMNET++	Open source [49]	C++, NEtwork Description(NED)	Component-based, modular and open architecture discrete event simulator framework GUI for simulation execution Graphical output vector plotting tool	Yes	Castalia, MiXiM, INET, etc.
NS-2	Open source [50]	C++, OTcl	Object-oriented, discrete event driven network simulator Event scheduler Support for simulation of TCP, routing, and multicast protocols over wired and wireless networks. It uses C++ and OTcl languages.	Yes	Mannasim
J-Sim	Open source [51]	Java, Tcl	Component-based, compositional simulation environment Real-time process driven simulation Partial realization of network emulation	No	INET, Diffserv
Riverbed Modeller	Commercial [52]	C, C++	Discrete event simulator Self-contained constructive simulation	Yes	-
QualNet	Commercial [53]	C++	High-fidelity modeling discrete event simulator Multi-protocol support	Yes	-

Table 4.1: Comparison of Network Simulators

4.2.1 OMNET++ Network Simulator

For the purpose of simulating the optimal camera placement algorithm 1 in a multimodal video surveillance system, OMNET++ network simulator [49] was used. While the results of the simulation have been presented in table 6.2, this section provides a general overview and features of OMNET++ Network Simulator. OMNET++ is a discrete event simulation engine. It follows a generic modular, component-based architecture consisting of simple modules that can communicate with each other using messaging connections called gates. Compound modules are formed by grouping the simple modules. It is also possible to make an architecture of hierarchically nested modules. While all the simple modules are written using the simulation class library in C++; a network topology containing various simple modules, their parameters and gates is defined using a NEtwork Definition (NED) language. The main features of OMNET++ are-

- Rich Eclipse-based Simulation IDE
- Flexible Module Parameters
- Communication links and gates between modules
- Domain-specific functionality
- Extensive GUI Support
- Command-line Interface
- Support for Ad-hoc Networks, Sensors Networks, Wireless Networks
- Rich extensions for Network emulation, Database Integration, Real-time Simulation
- Graphical NED Editor
- Predefined simulation models for modeling and simulating state-of-the-art network protocols and models

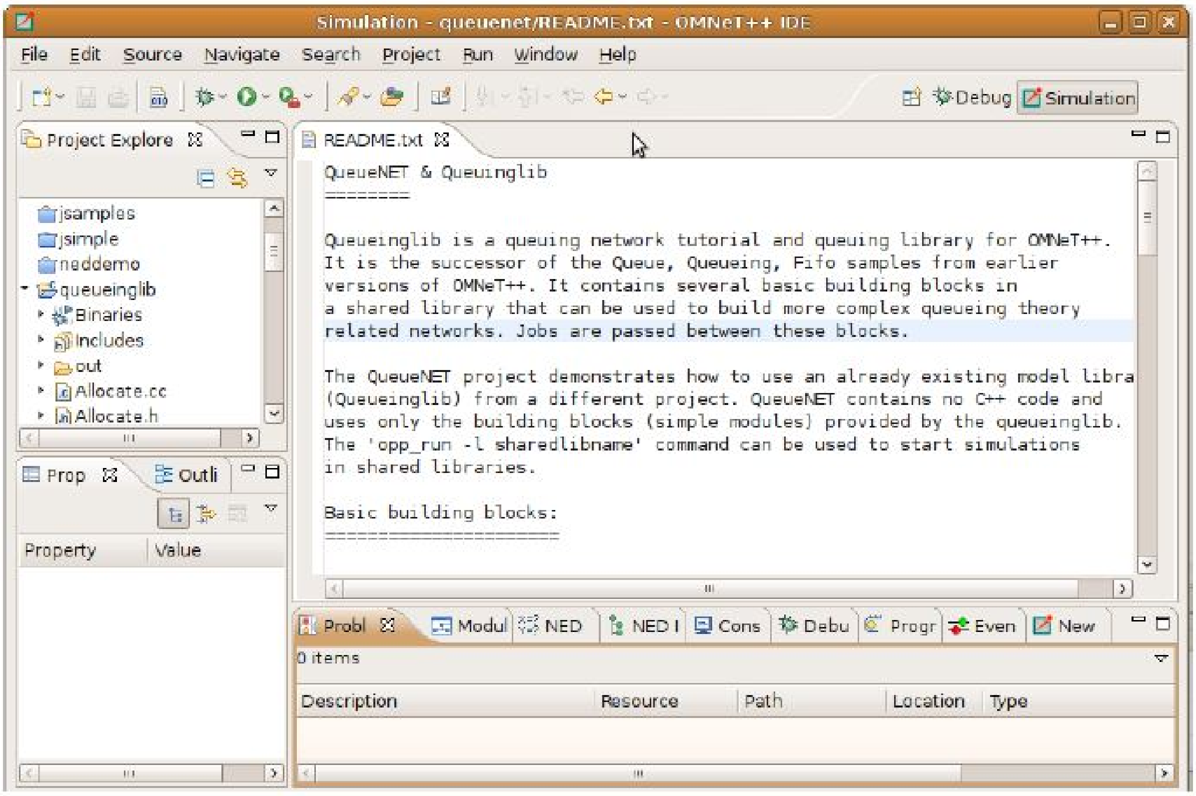

Figure 4.5: Eclipse-based Simulation IDE

The above figure 4.5 is a screenshot of the rich Eclipse-based simulation IDE provided in OMNET++.

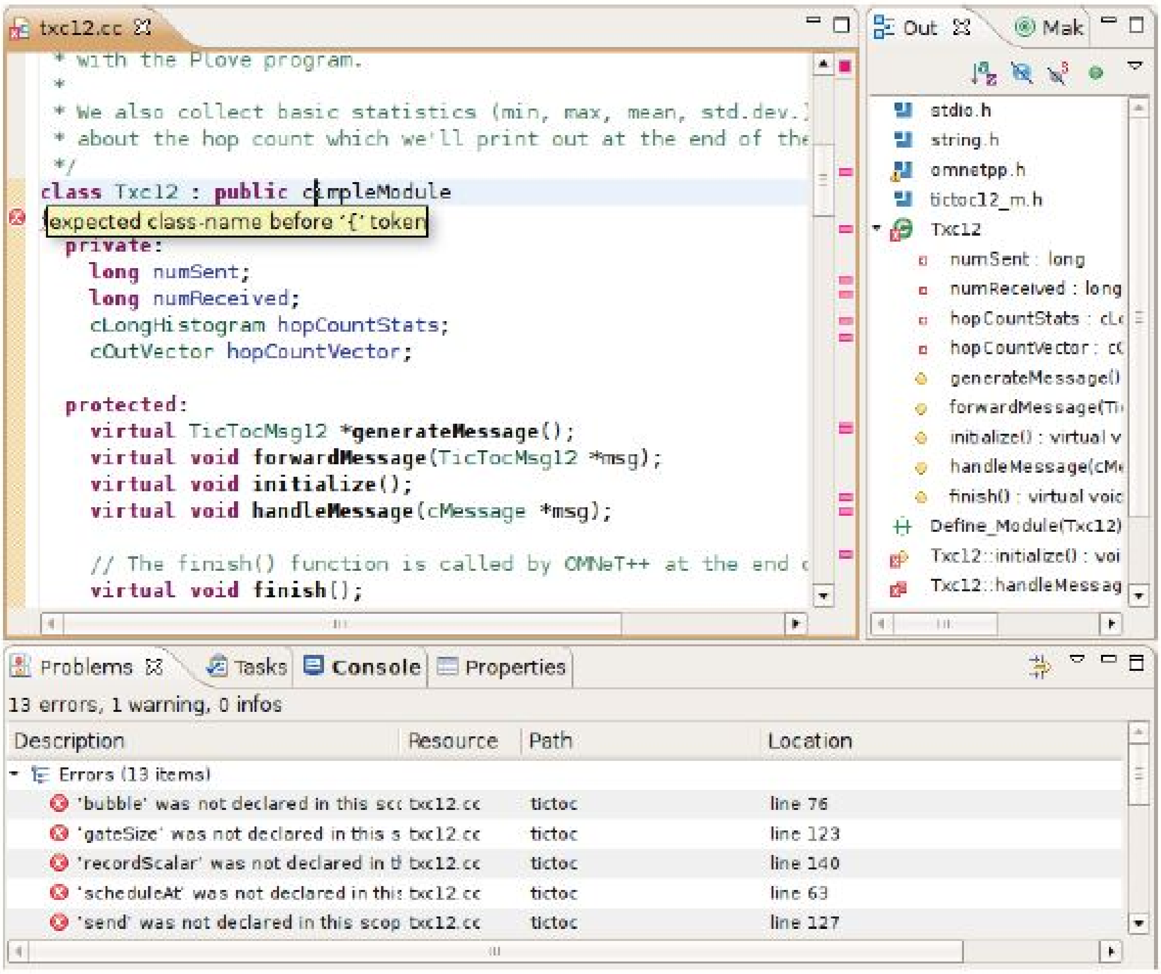

Figure 4.6: C++ Source Code Editor with Code Reviewer

Similarly, OMNET++ simulation IDE also provides a code editor supporting C++ along with code reviewing functionality as shown in figure 4.6 above. The figure 4.7 is a screenshot of the content assist for C++ source code editor provided in OMNET++ simulation IDE. OMNET++ uses NED for defining the description of the network models and topology as described earlier, and figure 4.8 is a screenshot of the graphical NED editor provided in OMNET++.

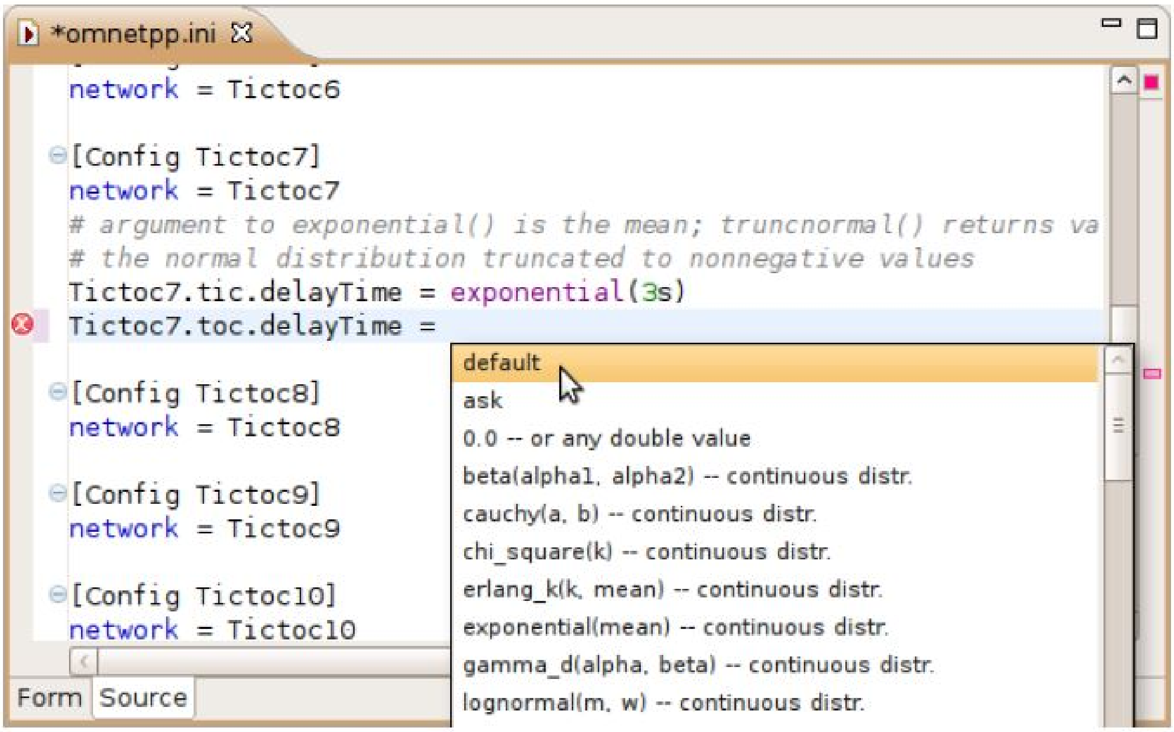

Figure 4.7: Content Assist for C++ Source Code Editor

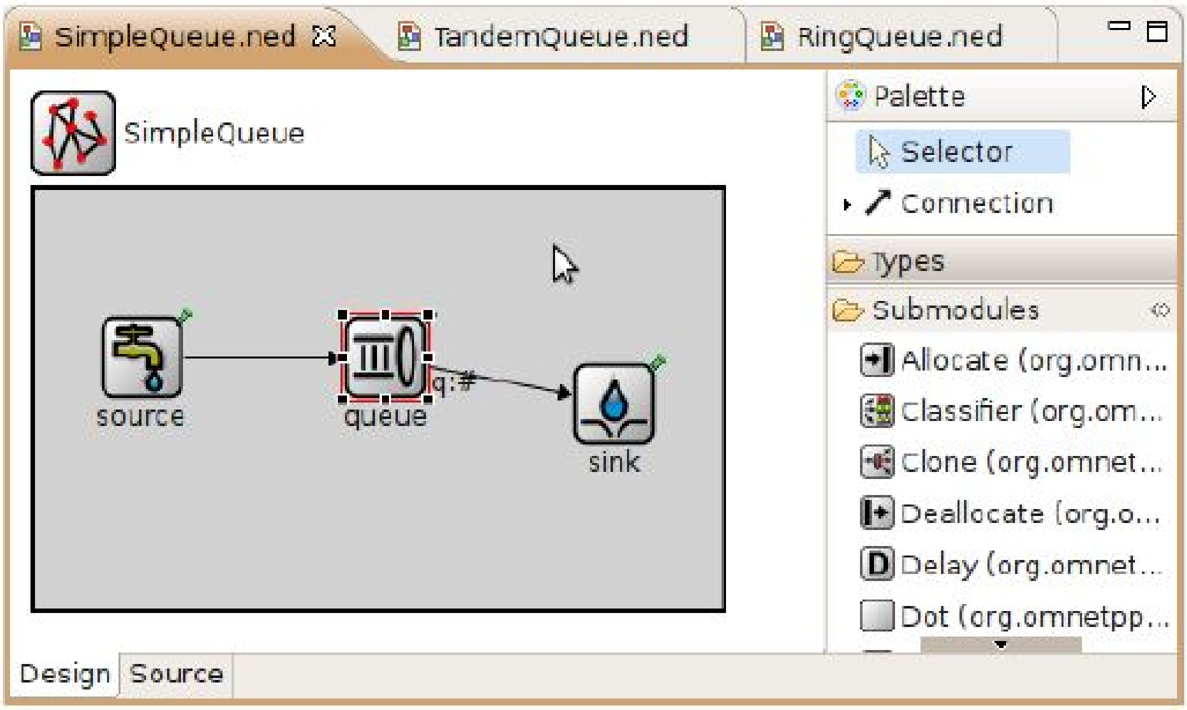

Figure 4.8: Graphical NED Editor

There are several extension frameworks available on top of OMNET++ like INET Framework, OverSim, SimuLTE, Castalia, etc.; but for simulating WSNs, networks of low-powered embedded devices and Body Area Networks, Castalia framework [54] is more suitable than other extension frameworks as it has been designed for modeling of distributed algorithms for WSNs under realistic communication conditions. For the purpose of simulating the proposed optimal camera placement algorithm and checking its validity, Castalia extension framework is used on top of OMNET++.

Chapter 5

Video Summarization

In this chapter, an overview about video summarization, its requirement and techniques used to perform video summarization are discussed. In the later part of this chapter, the algorithm developed for summarizing videos with respect to the purview of the proposed system is discussed.

5.1 Introduction to Video Summarization

Recent advances in technology have led to an increase in the amount of the data generated by the systems. This applies even to the field of surveillance where multiple cameras keep on recording the activities happening in surveillance perimeters. These video streams are voluminous and often unstructured. Although they contain important information captured according to the purpose of such surveillance systems, they do contain redundant data that would span along long duration of time and it is a painful task to extract some important video sequence from such large sized and lengthy video streams. Inside the control rooms and base stations of large-scale multimodal surveillance systems, often there are human experts who operate such systems and monitor the videos to detect some abnormal events or grab some important contextual information. But, monitoring such lengthy and large-sized video streams, and extracting important information from them is often a tedious task and less user-friendly. A video summary is a short version of a video stream which contains all the important and necessary information extracted from the original video. Video summarization is a process of creating shorter versions of original video streams, by retaining the relevant scenes and events and deleting redundant information from the original video streams. Video summarization has a wide applicability and in the area of ITS it has many applications, some of which are listed below:

- During night time, since there is less traffic; video summaries can be used to extract only important information from the video streams of night time surveillance.
- Video summaries can be used to track vehicles from the original video streams and create a summarized video for each particular vehicle's motion.
- Video summarization can be used to summarize video streams generated over a long period of time, and later on use these summaries to plan new infrastructure, calibrate traffic signals, and so on.
- Video summaries of traffic data can be used for crime scene investigation when there is a lot of data to go through in less time.
- Video summaries can be used to identify and track people from the original video streams and create a summarized video for each particular person's motion.
- Video summaries of traffic data can also be created for the purpose of vehicle classification.
- Video summaries can be used for traffic segmentation by removing the redundant data.

5.1.1 Types and Techniques of Video Summarization

There are mainly two kinds of techniques used for Video Summarization, namely-

- Key-frame Extraction
- Video skimming

5.1.1.1 Key-frame Extraction

A shot in a video is represented by a continuous sequence of captured images. Each shot consists of a key-frame which can be used to represented the whole shot individually since it contains enough information to characterize the content of the whole shot. In this technique, key-frames are extracted from a video stream at fixed intervals which collectively represent the whole video. The key-frames may be extracted from each shot using different techniques like simple temporal-based key-frame extraction, histogram-based key-frame extraction, pixel-to-pixel comparison based key-frame extraction, etc. Then these key-frames are fused together in the two-dimensional space and the summary thus created is called a static video summary or static video storyboard. Using this technique, the most important frames can be selected which provide accurate description of the information needed to be extracted from the original video.

5.1.1.2 Video Skimming

In this technique, short video clips called video skims are extracted from the original video streams. These video skims are joined together in a sequence by applying different effects like wipe-out, dissolve, fade, etc. and played. Video skimming has an added advantage that motion elements and audio both can be incorporated very easily in the video summary. The video summary created using this technique is called a dynamic video summary or a moving storyboard. Video skimming can be carried out by rather a lengthy process for example by using semantic analysis of motion models, singular value decomposition (SVD), etc. Using video skimming techniques, it becomes easier to detect moving objects, analyze the video, track objects, detect events, extract motion-based information, etc. from individual clips of the original video stream.

5.1.2 Video Compression Picture Types and Group of Pictures

As discussed in section 3.2, a video is encoded and compressed before transmitting it from a video camera source to a base station or any other intermediate receiver. In order to compress a video, the video encoder or decoder should have information about the frames of the video. A video is composed of frames and each frame is in turn composed of two fields. A field represents one of the many still images or pictures which are displayed in sequential order to create a motion of images on the screen. There are three types of video frames [55],[56],[57],[58] which are used during video compression.

- **I-frame:** An I-frame or an Intra-coded picture frame (also known as Key Reference frame) stands on its own and does not depend on any other type of frame to describe itself. It carries with itself all the information that is required for decoding, and it also serves as a starting or reference point for other frames. Since I-frames carry the information of a fully specified image, it requires relatively more space but at the same time since it is coded all by itself it is relatively quicker to decode.

- **P-frame:** A P-frame or a Predicted picture frame (also known as Delta frame) is coded with forward prediction using the reference from previous I-frame or P-frame or both in addition to the information from the preceding I-frame or P-frame related to changes in image. For example, in a scene where an object is moving along a stationary background, the information about the background pixels remains the same and hence the P-frame needs to be encoded only with the information related to the object's movement. P-frames are relatively slower to decode but they occupy relatively lesser space than I-frames.

- **B-frame:** A B-frame or a Bi-directionally predicted picture frame is coded with forward prediction as well as backward prediction from the previous and next I-frame and P-frame respectively. It requires more information from the surrounding I-frame and P-frame for prediction leading to more encoding/decoding memory but relatively less space to store the information of the frame after prediction.

All the frames are organized in a sequence of groups of pictures (GOP) in video streams. The GOP structure is used to define the order in which the inter- and intra-frames are arranged. The GOP structure always begins with an I-frame followed by a number of P and B-frames. The size of GOP structure is often dynamic in nature. GOP structure is defined by two numbers M and N, for example if M=3 and N=8 then M defines the distance between two anchor frames and N defines the distance between two full images (i.e.) I-frames. The figures 5.1 and 5.2 show GOP structures with varying sizes. There are mainly two types of video compression formats used for digital video surveillance cameras namely, H.264 and MJPEG(Motion JPEG) [59]. A digital video encoded or compressed using MJPEG consists of a sequence of individual JPEG images in which each image will have the same guaranteed quality. In MJPEG, since each frame is independent of each other, the quality of the video is not affected in case if one of the frames is skipped during transmission. In contrast, H.264 uses lossless coding technique which results in upto 50% reduction of the size of the video without compromising the image quality. This means H.264 needs lesser bandwidth to transmit the video. Due to these benefits, the most commonly used standard for compression [60] is H.264 or MPEG-4 Part 10, Advanced Video Coding (MPEG-4 AVC).

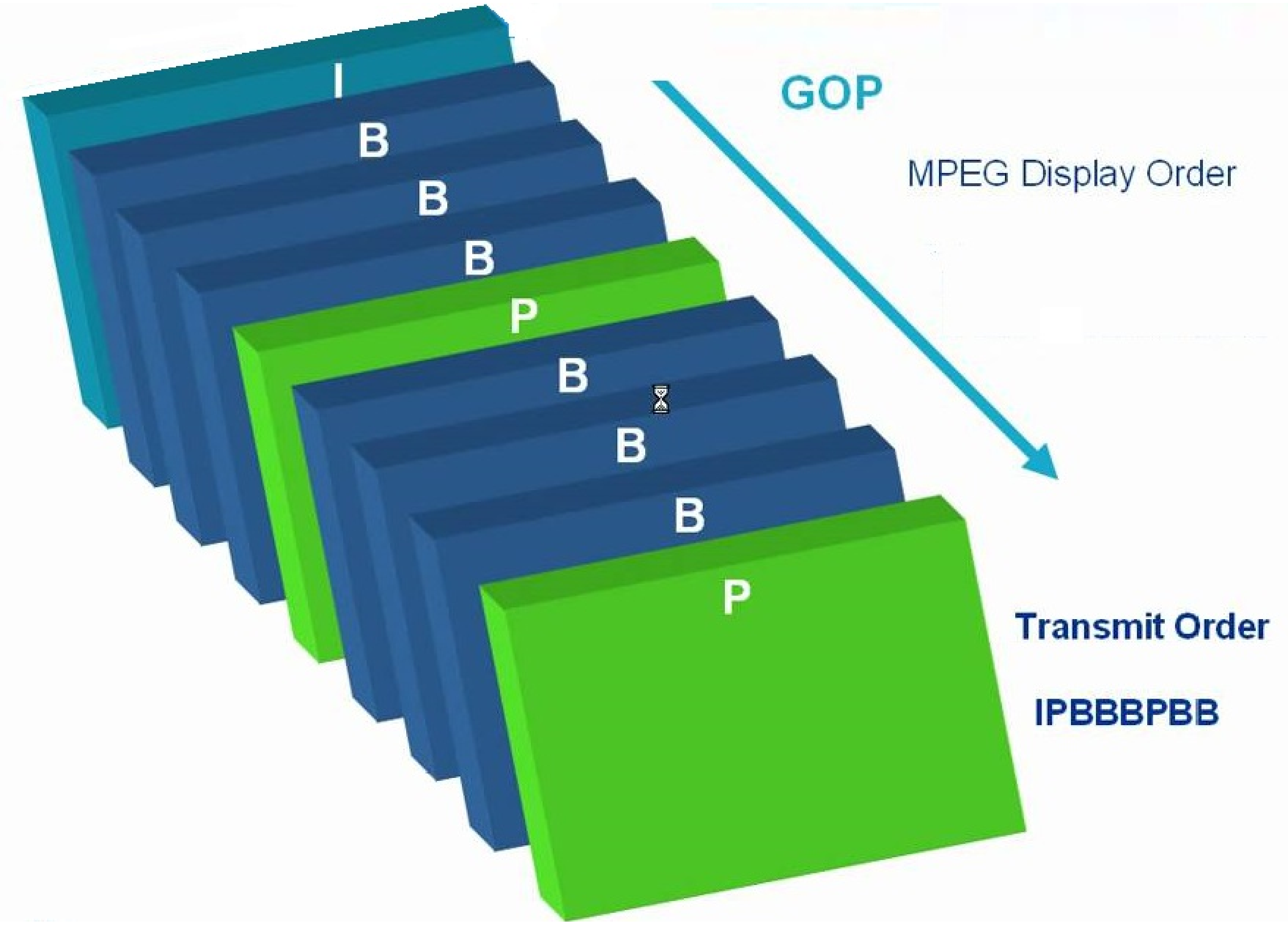

Figure 5.1: GOP Structure with N=9
(Source: https://i.ytimg.com/vi/P7abyWT4dss/maxresdefault.jpg)

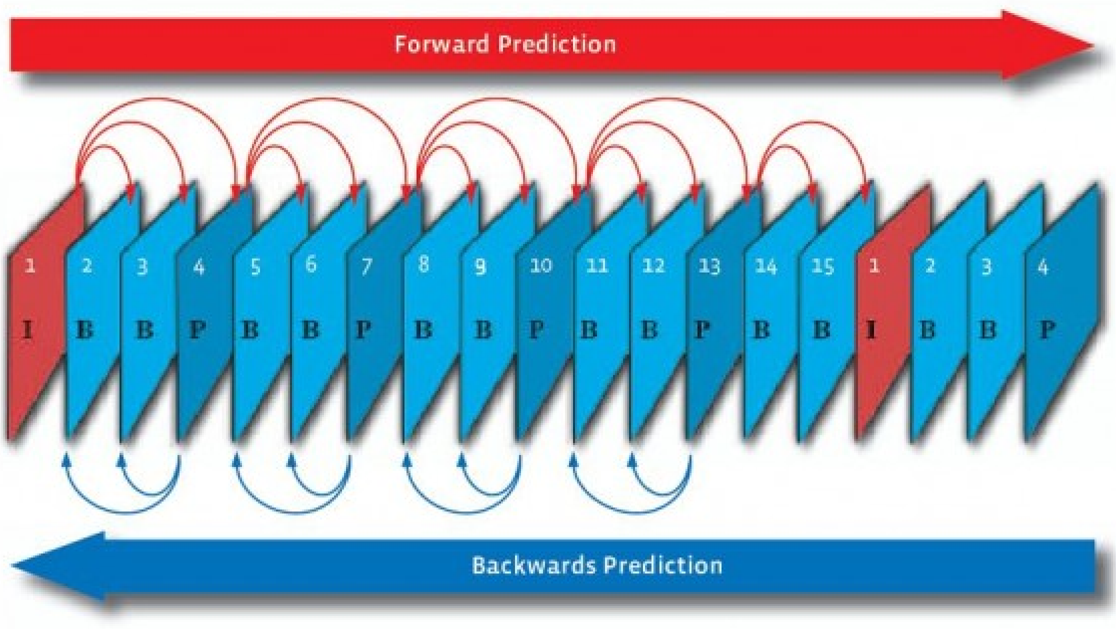

Figure 5.2: GOP Structure with N=15
(Source: http://cfile3.uf.tistory.com/image/232D443656D6B1FE2415F4)

5.2 Video Summarization Algorithm

In this section a video summarization algorithm for a multi-view surveillance system is discussed. For this algorithm there were several assumptions which have been considered and the same have been listed below:

- The cameras deployed in the system were fully calibrated, which also implies that all the cameras were synchronized temporally (time-based synchronization).
- All the cameras shared the same intrinsic and extrinsic parameters like FOV, AOV, Focal length, Resolution, Lens aperture, Sensor size, Height at which the cameras were mounted, Mounting angle of the cameras, GOP size, Encoding method, Video codec, Audio codec and frame rate.
- All the cameras had fixed focal length and zooming feature was not available in them.
- All the cameras were configured to record videos using a GOP structure of size greater than or equal to 8 and the video was encoded using the H.264 or MPEG-4 Part 10, Advanced Video Coding (MPEG-4 AVC) standard.
- All the cameras had overlapping views of the surveillance area.
- The video streams captured from multiple cameras were archived in a base station (common receiver) where the process of video summarization was performed.

In this research, a unique approach for video summarization is taken by merging the concepts of both the techniques of video summarization namely key-frame extraction and video skimming. Thus a hybrid approach incorporating both the techniques of video summarization is presented. Below are the mentioned steps of the algorithm to create summarized videos from multiple cameras in a multi-view surveillance system:

1. **Number of Video Cameras (Input Sources):** Give the total number of video streams to be given as input for summarization.
2. **Frame Extraction:** Extract I, B and P-frames from all the video streams and store them independently by numbering them sequentially according to their type and GOP index corresponding to each video stream.

(a) **Temporal Synchronization:** All the frames (I,B and P) extracted from each video stream should be synchronized based on the time at which they were captured. For this purpose, map the frame capture timestamp of each frame with the corresponding type of frame for that particular video stream. This step has to be performed simultaneously along with Frame Extraction.

3. **Frame Selection Technique:** For the purpose of summarization, only some frames have to be retained while the remaining ones have to be dropped. To fulfill this requirement, apply any one of the three frame selection techniques (mentioned below) to each GOP of each video stream and store the selected frames along with their frame capture timestamp information.

4. **Generate a Video Summary:** For each video stream, join the selected frames based on the Frame Selection Technique employed and generate a video using a pre-defined encoding method and GOP size. The resulting video generated using this technique is the summarized video of the original video stream.

As mentioned in the above step 3, three different kinds of Frame Selection Techniques are developed which are mentioned below.

5.2.0.1 Video Summariazation using Even Frame Selection Technique

In this technique, for each video stream, all the I-frames extracted from the video stream are retained. However, for each GOP only even numbered P-frames and even-numbered B-frames are retained and the remaining P-frames and B-frames for each GOP are discarded. Later on before creating a video summary, for each GOP its corresponding selected I-frame, P-frame(s) and B-frame(s) are joined together; and the frames are synchronized according to their temporal information while joining so as to maintain the flow and order of events from original video in the summarized video. The final list of synchronized frames selected using this technique are encoded using a pre-defined GOP size, pixel information, frame rate and codec in order to generate a video summary.

5.2.0.2 Video Summarization using Odd Frame Selection Technique

In this technique, for each video stream, all the I-frames extracted from the video stream are retained. However, for each GOP only odd numbered P-frames and odd-numbered B-frames are retained and the remaining P-frames and B-frames for each GOP are discarded.

Later on before creating a video summary, for each GOP its corresponding selected I-frame, P-frame(s) and B-frame(s) are joined together; and the frames are synchronized according to their temporal information while joining so as to maintain the flow and order of events from original video in the summarized video. The final list of synchronized frames selected using this technique are encoded using a pre-defined GOP size, pixel information, frame rate and codec in order to generate a video summary.

5.2.0.3 Video Summarization using Randomized Frame Selection Technique

In this technique, for each video stream, all the I-frames extracted from the video stream are retained. However, for each GOP a randomization method is applied for the selection of P and B-frames which is as follows:

- For the first GOP, only two P frames are selected randomly and rest all are dropped.
- For the second GOP, two P frames and two B frames are selected randomly, and rest all are dropped.
- For the third GOP, four B frames are selected randomly and rest all are dropped.
- For the fourth GOP, only four randomly chosen B frames and all even numbered P frames are selected, and rest all the frames are dropped.
- For the fifth GOP, only four randomly chosen B frames and all odd numbered P frames are selected, and rest all the frames are dropped.

In this manner this randomization sequence is continued for the next sequence of GOPs till all the GOPs have been covered. Later on before creating a video summary, for each GOP its corresponding selected I-frame, P-frames and B-frames are joined together; and the frames are synchronized according to their temporal information while joining so as to maintain the flow and order of events from original video in the summarized video. The final list of synchronized frames selected using this technique are encoded using a pre-defined GOP size, pixel information, frame rate and codec in order to generate a video summary.

5.2.1 Need of Temporal Synchronization for Frames

As mentioned in earlier sections, while generating the video summary, all the selected frames were synchronized based on their frame capture timestamp information mapped with them. The purpose of applying temporal frame synchronization mechanism was due to the reason that multiple cameras with overlapping views required to be synchronized so as to maintain the alignment of events in video while creating summaries. When the output of the three or more video cameras are taken to create a single video summary for applications like object detection, object tracking, motion tracking, vehicle classification, event detection, etc. then it becomes necessary that the contextual information is not distorted and jitter in the video summary is reduced. Sudden changes in scenes from multiple cameras may also cause variations in spatial information and the scenes in the summarized video may appear unordered due to temporal misalignment. Thus frame-by-frame timing information is maintained and temporal synchronization using frame capture timestamp information was applied while creating video summaries.

Algorithm 2 Video Summarization

Result: Summarized Video

Require: Video Stream

Initialization

Determine the total number of GOP ($totalNumberOfGOP$) present in the Video Stream

for GOP_{index} from 1 to $totalNumberOfGOP$ **do**

  $i - frameList$.add(I-frame extracted from GOP at index GOP_{index})

  $p - frameList$.add(P-frame(s) extracted from GOP at index GOP_{index})

  $b - frameList$.add(B-frame(s) extracted from GOP at index GOP_{index})

end for

if $frameSelectionTechnique$ = Even **then**

  for GOP_{index} from 1 to $totalNumberOfGOP$ **do**

    $selected - p - frameList$.add(Even Numbered P-frames from $p - frameList$ corresponding to the GOP at index GOP_{index})

    $selected - b - frameList$.add(Even Numbered B-frames from $b - frameList$ corresponding to the GOP at index GOP_{index})

  end for

else if $frameSelectionTechnique$ = Odd **then**

  for GOP_{index} from 1 to $totalNumberOfGOP$ **do**

    $selected - p - frameList$.add(Odd Numbered P-frames from $p - frameList$ corresponding to the GOP at index GOP_{index})

    $selected - b - frameList$.add(Odd Numbered B-frames from $b - frameList$ corresponding to the GOP at index GOP_{index})

  end for

else if $frameSelectionTechnique$ = Randomized **then**

  $GOP_{seqCount} = 1$

  for GOP_{index} from 1 to $totalNumberOfGOP$ **do**

    if $GOP_{seqCount} = 1$ **then**

      $selected - p - frameList$.add(Two P-frames chosen randomly from $p - frameList$ corresponding to the GOP at index GOP_{index})

      $GOP_{seqCount}$ ++

Algorithm 2 Video Summarization (continued)

else if $GOP_{seqCount} = 2$ **then**

$selected-p-frameList$.add(Two P-frames chosen randomly from $p-frameList$ corresponding to the GOP at index GOP_{index})

$selected-b-frameList$.add(Two B-frames chosen randomly from $b-frameList$ corresponding to the GOP at index GOP_{index})

$GOP_{seqCount}$ ++

else if $GOP_{seqCount} = 3$ **then**

$selected-b-frameList$.add(Four B-frames chosen randomly from $b-frameList$ corresponding to the GOP at index GOP_{index})

$GOP_{seqCount}$ ++

else if $GOP_{seqCount} = 4$ **then**

$selected-p-frameList$.add(All Even Numbered P-frames from $p-frameList$ corresponding to the GOP at index GOP_{index})

$selected-b-frameList$.add(Four B-frames chosen randomly from $b-frameList$ corresponding to the GOP at index GOP_{index})

$GOP_{seqCount}$ ++

else if $GOP_{seqCount} = 5$ **then**

$selected-p-frameList$.add(All Odd Numbered P-frames from $p-frameList$ corresponding to the GOP at index GOP_{index})

$selected-b-frameList$.add(Four B-frames chosen randomly from $b-frameList$ corresponding to the GOP at index GOP_{index})

$GOP_{seqCount} = 1$

end if

end for

end if

Merge all the I,P,B-frames selected using any of the frame selection technique and apply temporal synchronization to it : $finalFramesList = temporalSynchronization(merge(i-frameList, selected-p-frameList, selected-b-frameList))$

Generate a summarized video : $generateVideo(finalFramesList)$

Chapter 6

Implementation and Results

In this chapter, the implementation details and experimental results of the algorithms discussed in previous chapters is discussed.

6.1 Simulation of Optimal Camera Placement Algorithm

6.1.1 Simulation Environment and Parameters

The optimal camera placement algorithm 1 discussed previously was simulated in OMNET++ Network Simulator. The following table 6.1 lists the simulation parameters that were considered while checking the results and validity of the algorithm. Also all the assumptions mentioned in 4.1.2 were taken into consideration.

Parameter	**Value**
Simulation Time	300 s
Surveillance Area	25m x 20m x Depth of Surveillance Area (mentioned below)
Depth of Surveillance Area	10m to 15m
Number of Video Cameras	4
Focal Length	4.0 mm
AOV of each Camera	90° to 120°
Camera Deployment (Coordinates in three-dimensional space)	Random

Table 6.1: Simulation Parameters for Optimal Camera Placement Algorithm

The network topology was configured in a way that each video camera would be placed randomly inside or on the edges of the one-fourth part of the surveillance area as shown in the figure 6.1 below since the surveillance area is divided into four equal parts. Also as mentioned in the algorithm, each camera needed to have the two midpoints of adjacent sides in their FOV to have the best possible coverage of the surveillance area.

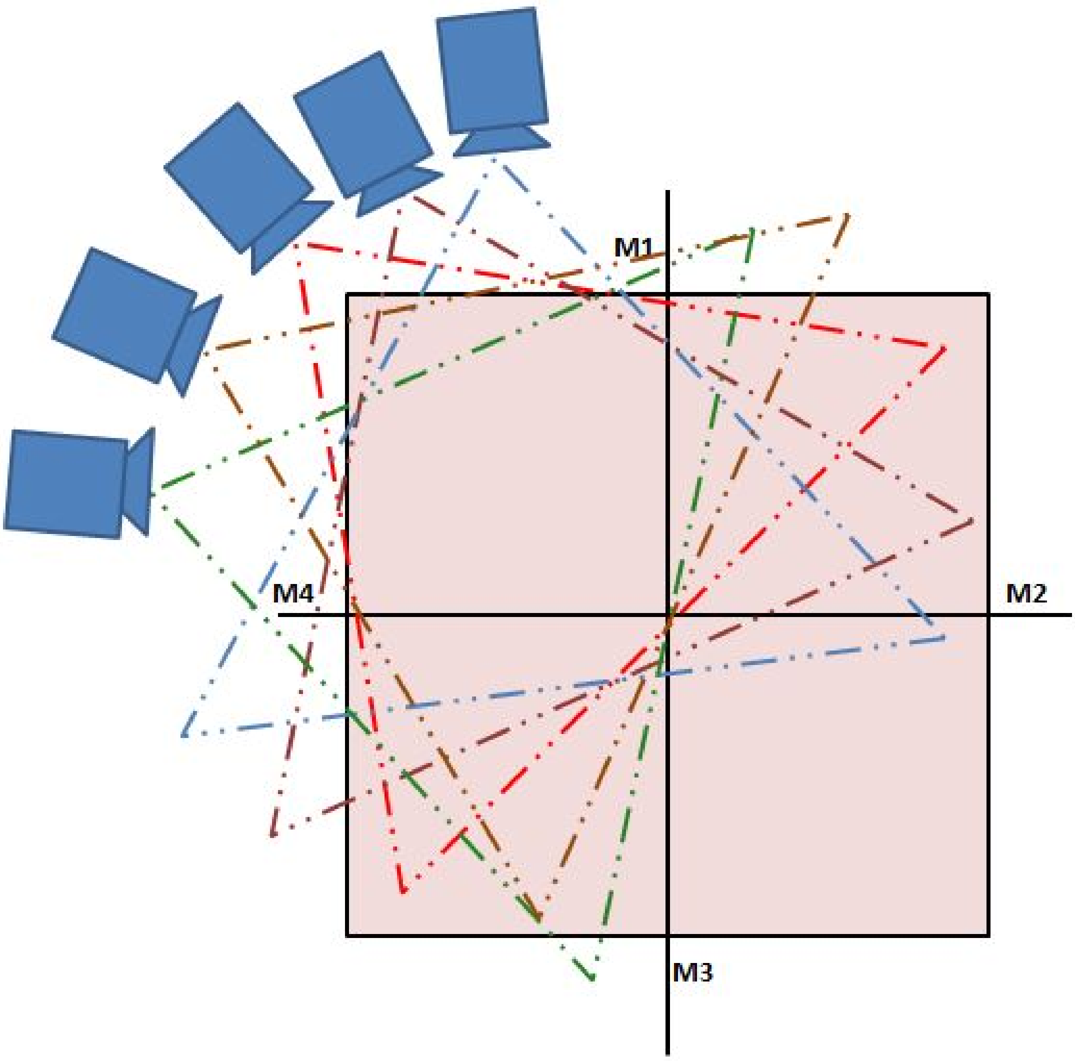

Figure 6.1: Surveillance Area with various possible Camera placements

6.1.2 Results

During the simulation the cameras were placed randomly inside each one-fourth part of the surveillance area, on the sides of each one-fourth part of the surveillance area, near the midpoints and on the vertex joining the two edges. At any moment of time, the AOV and Depth of Field of each camera was fixed and identical, though it was selected randomly from the range specified in table 6.1.

Angle of View (degree)	**Depth of Field (m)**	**Camera Placement**	**Area Covered (%)**
90	10	Random-inside the one-fourth part of Surveillance Area	29
90	11	Random-inside the one-fourth part of Surveillance Area	34
100	12	Random-inside the one-fourth part of Surveillance Area	45
100	13	Random-inside the one-fourth part of Surveillance Area	50
105	14	Random-inside the one-fourth part of Surveillance Area	48
100	15	Random-inside the one-fourth part of Surveillance Area	53
90	10	On the sides of the one-fourth part of Surveillance Area	43
95	11	On the sides of the one-fourth part of Surveillance Area	50
110	12	On the sides of the one-fourth part of Surveillance Area	61
100	13	On the sides of the one-fourth part of Surveillance Area	70
105	14	On the sides of the one-fourth part of Surveillance Area	73
120	15	On the sides of the one-fourth part of Surveillance Area	74
95	10	Near the midpoints	23
110	11	Near the midpoints	28
120	12	Near the midpoints	39
100	13	Near the midpoints	38
100	14	Near the midpoints	27
95	15	Near the midpoints	24
95	10	At the vertices joining the two edges	94
110	11	At the vertices joining the two edges	97
105	12	At the vertices joining the two edges	95
120	13	At the vertices joining the two edges	98
105	14	At the vertices joining the two edges	95
120	15	At the vertices joining the two edges	97

Table 6.2: Simulation results for Optimal Camera Placement Algorithm

The results of the simulation are presented in table 6.2. From the results, it can be inferred that the best placement for all the cameras in a multimodal surveillance system is near the vertex joining any two edges of the surveillance area. It can also be derived that higher depth of field and wider angle of view produces better region coverage for a surveillance camera leaving less than 5% of area uncovered. However, since all the cameras have overlapping views, it is possible to achieve better coverage of the whole area leaving very less to zero blind spots. The figure 6.2, shows the best placement for video cameras in a multimodal surveillance system from the results derived using the optimal camera placement algorithm discussed in section 4.1.2.

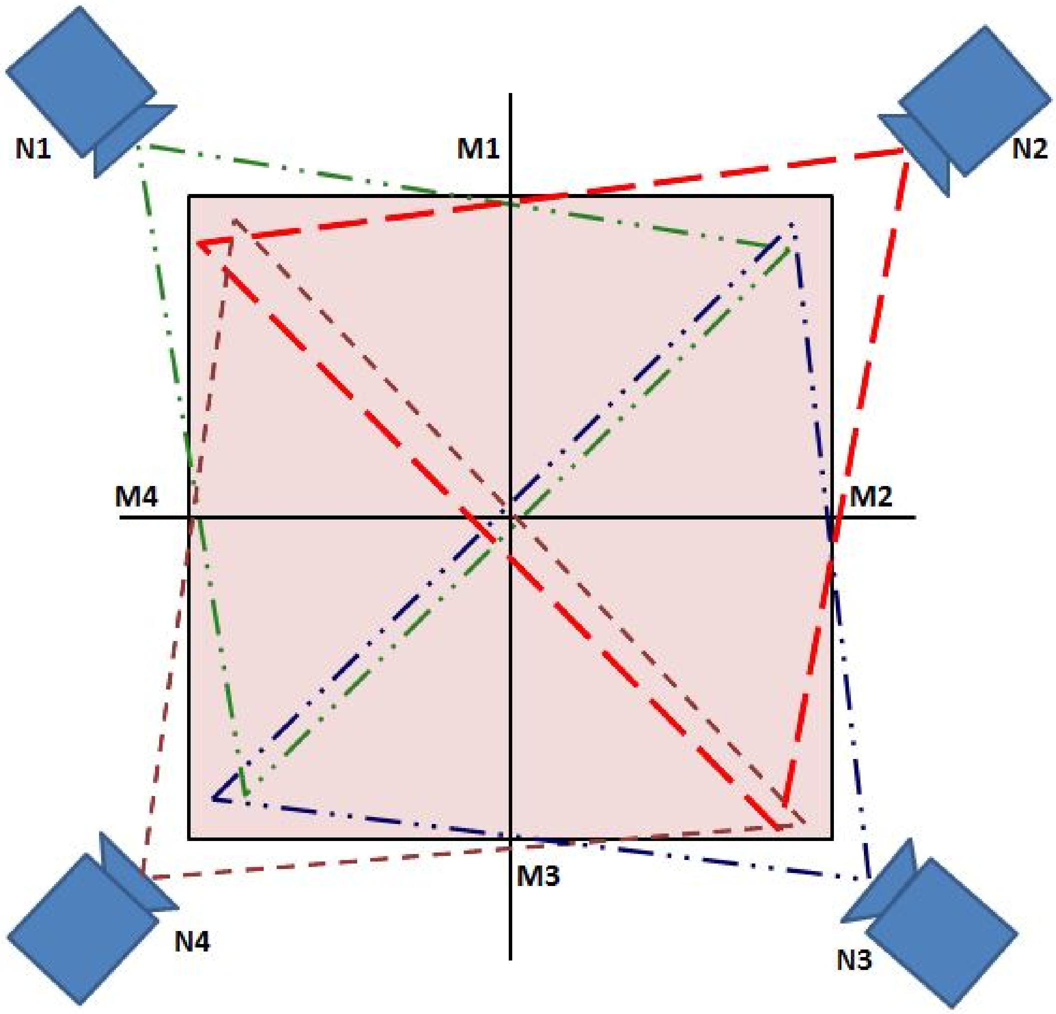

Figure 6.2: Optimal Camera Placement Design

6.2 Implementation of Video Summarization Algorithm

6.2.1 Dataset and Configuration

For the evaluation of the Video Summarization algorithm using three frame selection techniques discussed in section 5.2, the Ko-PER Intersection Dataset[61] was used which comprised of the following information:

- highly accurate reference trajectories of cars
- raw laser scanner measurements
- undistorted monochrome camera images

For evaluating the performance of the three techniques of frame selection for video summarization discussed in section 5.2, the assumptions listed in the section 5.2 were considered. From the dataset, four videos were generated of varying GOP size N=8,16,24 and 30 with a constant frame rate of 30 fps and were encoded using the H.264 or MPEG-4 Part 10, Advanced Video Coding (MPEG-4 AVC) standard. These different videos were used to check the performance of each algorithm individually. The table 6.3 presents the results of algorithm wrt each frame selection technique. From the table, it can be observed that the execution time of the randomized frame selection technique for video summarization is the fastest as compared to even frame selection technique and odd frame selection technique. Even though the duration of input video is very small, but for large scale multi-view surveillance systems generating large-sized and lengthy videos, the execution times do matter a lot. In such a scenario the proposed video summarization algorithm using randomized frame extraction would prove to be very efficient in terms of computations and memory utilization. However, the other two techniques for video summarization namely even frame extraction technique and odd frame extraction technique can also be used in multi-view surveillance systems. It is worth noting that in all the three proposed techniques for video summarization, all the key-frames (I-frames) have been retained and none of them are dropped which is very useful since each key-frame independently can be used to describe the information of the whole GOP leading to zero information loss.

Video Summarization Algorithm)	Input Video GOP Size	Duration of Input Video (s)	Execution Time of Algorithm (s)	Duration of Summarized Video (s)	Reduction of Summarized Video(%)
Even Frame Selection Technique	8	60	50	37	60
Even Frame Selection Technique	16	60	51	37	60
Even Frame Selection Technique	24	60	52	37	60
Even Frame Selection Technique	30	60	58	37	60
Odd Frame Selection Technique	8	60	49	37	60
Odd Frame Selection Technique	16	60	50	37	60
Odd Frame Selection Technique	24	60	51	37	60
Odd Frame Selection Technique	30	60	58	37	60
Randomized Frame Selection Technique	8	60	46	37	60
Randomized Frame Selection Technique	16	60	48	37	60
Randomized Frame Selection Technique	24	60	49	37	60
Randomized Frame Selection Technique	30	60	57	37	60

Table 6.3: Experimental results for Video Summarization Algorithm

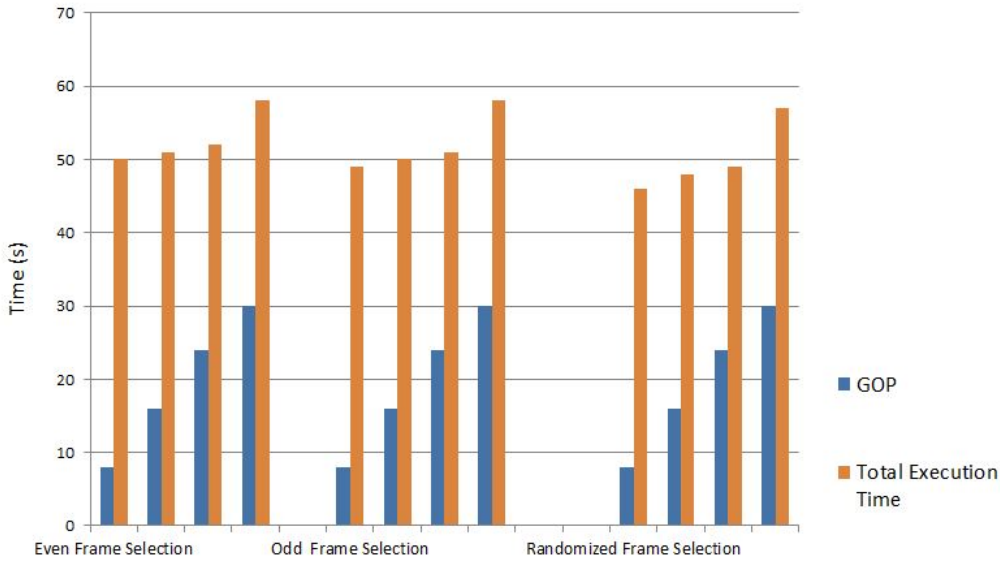

Figure 6.3: Comparison of Frame Selection Techniques for Video Summarization Algorithm

The graph presented in figure 6.3 shows the comparison of the total execution times of the proposed three frame selection techniques for video summarization. There are a couple of more observations derived from the results presented in table 6.2 and figure 6.3 with respect to the three proposed techniques for video summarization:

- As the GOP of the input video stream increases, the execution times of all the three techniques. Thus with increase in GOP size, the execution time of the algorithms increases.
- Using any of the technique of frame selection for video summarization, the duration of the final summarized video is same since no key-frames (I-frames) have been dropped in the process and hence all the three techniques are suitable to create video summaries without the loss of important contextual information from the video.

6.2.2 Demonstration of need of Temporal Synchronization for Frames

The figure 6.4 shows the difference between the order of frames selected when temporal synchronization algorithm using frame capture timestamp information was not applied while generating video summaries. As shown in the figure the order of the frames Frame-206, Frame-207 and Frame-208 is not maintained and the scene in Frame-206 should be

Figure 6.4: Frames without Temporal Synchronization

selected after Frame-209. To avoid such distortion of order, temporal synchronization was applied to the selected frames before generating the video summaries so that all the events in the generated video summary are aligned with the occurrence of the same events in the original video. As depicted in figure 6.5, the order of occurrence of events is maintained when temporal synchronization using frame capture timestamp information was applied to the selected frames before generating video summaries. Hence it becomes necessary to apply temporal synchronization methods to frames while generating video summaries so as to preserve the order of occurrence of events wrt to the original video.

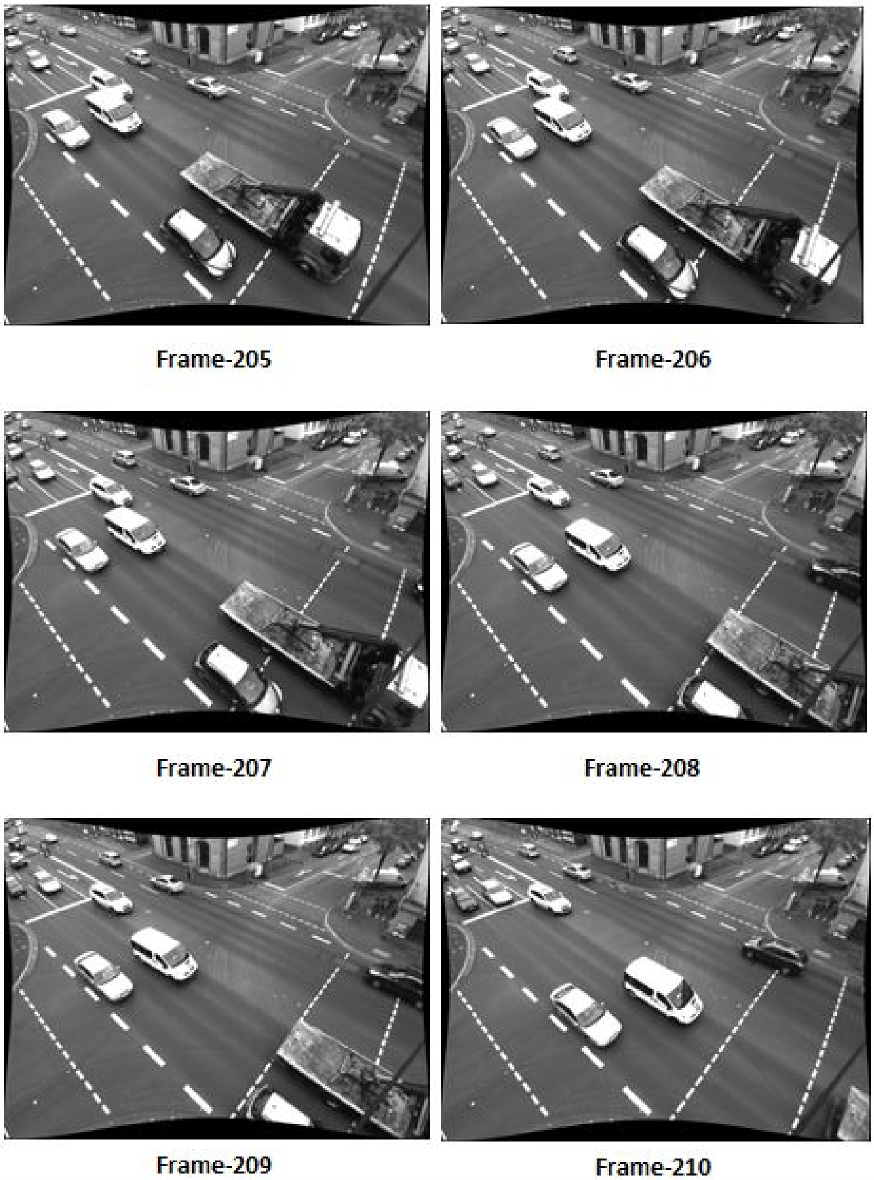

Figure 6.5: Frames with Temporal Synchronization

6.2.3 Extensibility of the Video Summarization Algorithm

In order to prove the extensibility of the proposed video summarization algorithm, an object detection and tagging algorithm using haar feature-based cascade classifiers was implemented on the videos created from the dataset. The haar-based cascade classifier was trained with a lot of positive and negative images of cars obtained from a publicly available dataset [62]. The video was fed to the object detection algorithm and cars were detected and tagged according to their haar features. For the purpose of both training and detection, Opensource Computer Vision (OpenCV) [63] library was used which comes with pre-built object detection functions. The figures 6.6 and 6.7 show the cars (while in motion) detected using the algorithm. Now using this information, the proposed video summarization algorithm can be extended in the context of multi-view surveillance systems where the tagged cars can be further tracked using multiple cameras and a summarized video by merging information obtained from all the cameras can be created using any of the proposed techniques of frame selection and applying temporal synchronization to frames. Moreover, object detection is useful in the context of multi-view surveillance systems especially in the Indian transportation infrastructure scenario since improper lane driving is followed by the drivers at many places and such a method can help in tracking the traffic rules offenders. Furthermore, a user or human expert can be asked to give an input to locate a particular car or object from the video using the information of vehicle tags obtained through object detection and tagging algorithm discussed above; and a summary of the video containing the motion and activity of only that particular selected car or object can be created by merging frames from multiple cameras and applying the proposed frame selection techniques and frame capture timestamp based temporal synchronization.

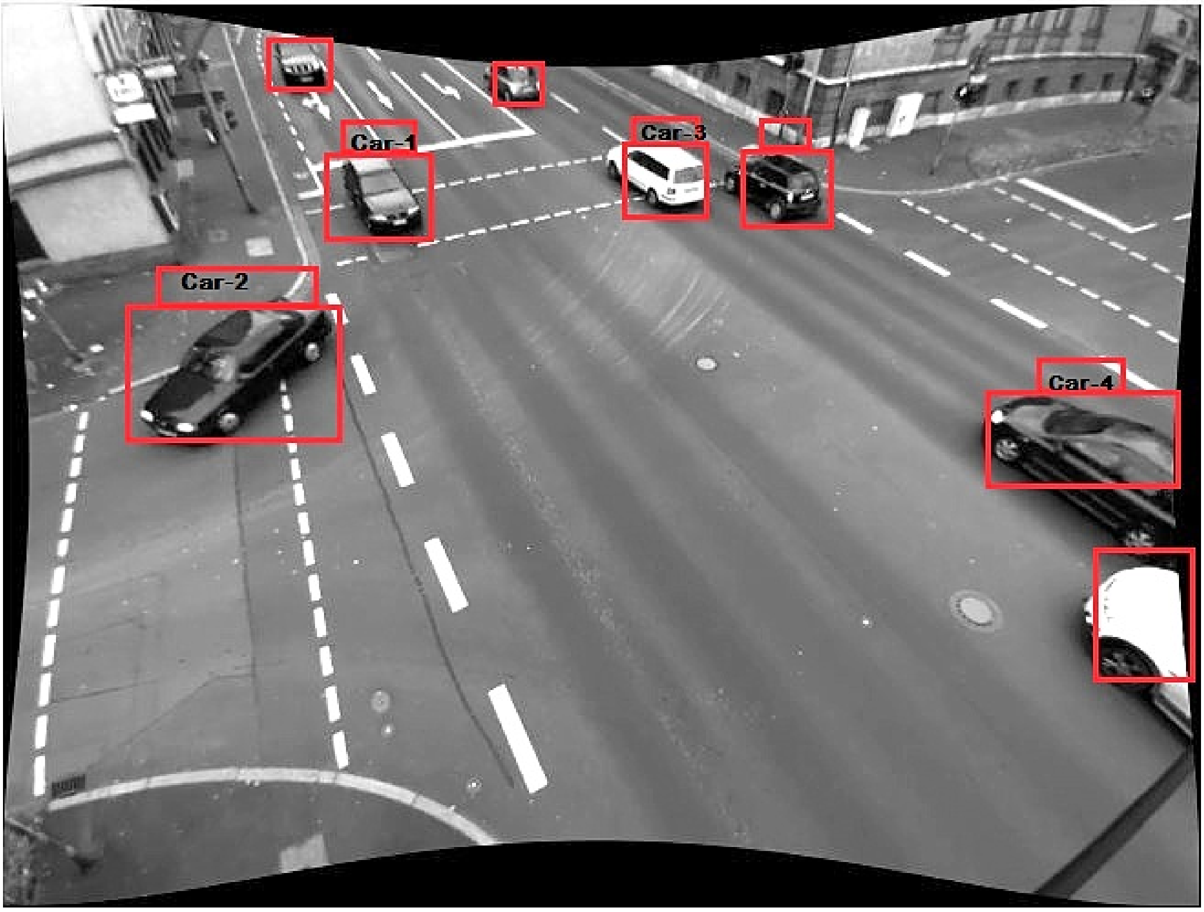

Figure 6.6: Object Detection and Vehicle Tagging

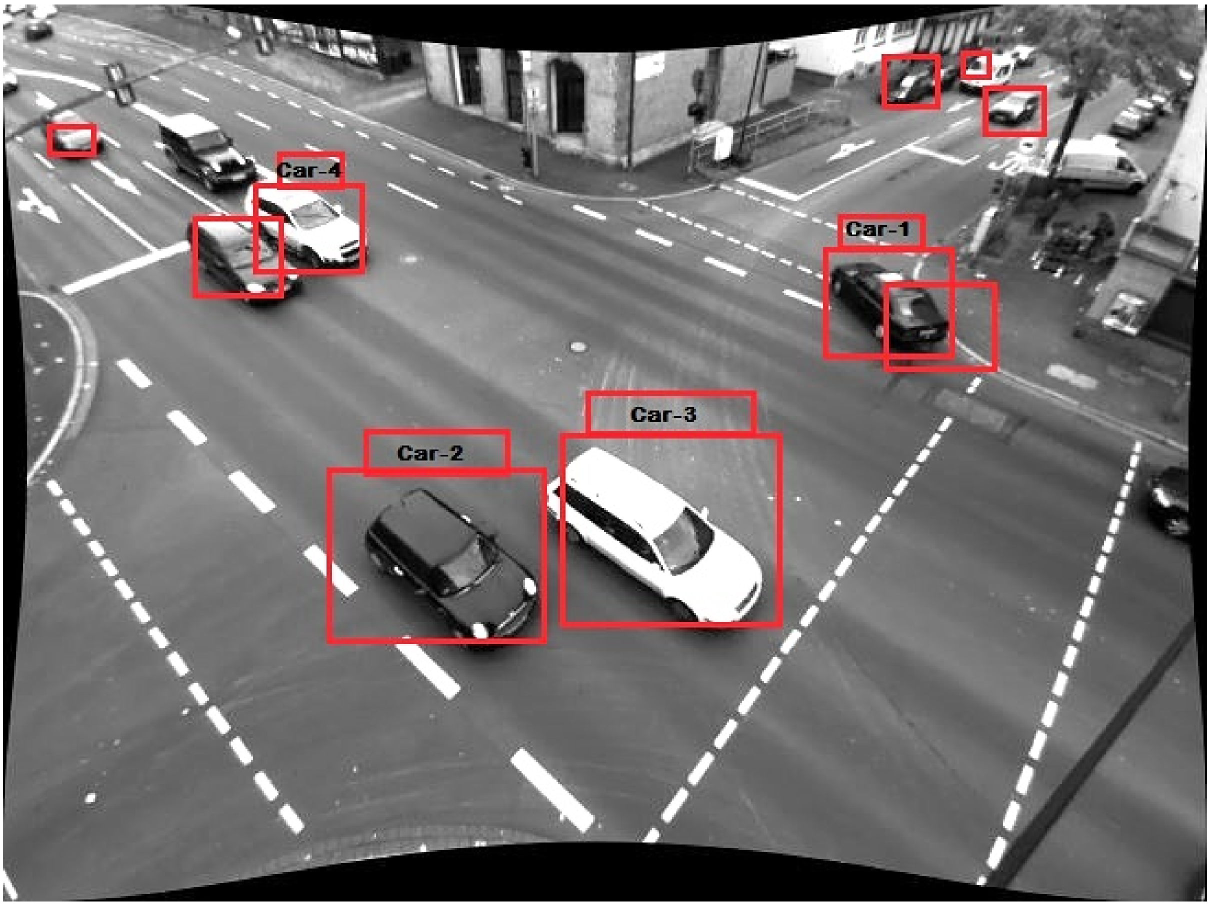

Figure 6.7: Vehicle Tagging for Classification

In additon to this, this multi-view model can be extended by adding PIR sensors, acoustic sensors, thermal sensors to create a an advanced multimodal surveillance system. For example, a thermal imaging sensor can be attached along with the video cameras and when an event like an explosion or accident is detected by the thermal imaging sensor (due to generation of large amount of heat energy in the surveillance area), summaries of traffic surveillance videos before and after the event using information from multiple cameras and thermal imaging sensor can be generated using the proposed video summarization algorithm.

Chapter 7

Conclusion and Future Scope

7.1 Conclusion

Smart multimodal surveillance architectures are possible only if they are based on devices or sensors with an adequate trade-off between cost, power consumption and processing capabilities. There are several challenges in the Indian transportation infrastructure scenario which need to be resolved while deploying an effective surveillance system for intelligent traffic and transportation management and, a multimodal surveillance system of low-cost, low-power consuming cameras can provide the best solution for it.

The proposed optimal camera placement algorithm can be used for deciding the placement of multiple cameras at intersections, junctions and cross-roads without compromising the coverage area of the deployed video cameras and cost of deployment. This optimal camera placement algorithm is capable of providing maximum coverage of the area under surveillance leading to - complete elimination or reduction to a great extent the number of blind zones in a surveillance area, maximizing the view of subjects and minimizing occlusions in high vehicular traffic areas.

In addition to this, a video summarization algorithm using three different techniques of frame selection for multi-view surveillance systems is presented which can be used to create summaries of large-sized, lengthy video streams of traffic surveillance data and, at the same time reduce the computational processing for creating the video summaries. These summaries can be later used for performing traffic data analysis and accomplishing various objectives like face-detection, licence-plate recognition, crime-scene investigation, accident-investigation, event detection, and many more. The proposed algorithm can be used to create video summaries with zero information loss since all the key-frames from the original videos are retained while generating video summaries.

Collectively, both the proposed algorithms will be able to reduce the cost of camera deployment, computational cost, power consumption and, provide efficient performance in a multi-view as well as multimodal surveillance system for an Intelligent Transportation System while keeping into consideration the Indian transportation infrastructure scenario.

7.2 Future Scope

As demonstrated, the proposed video surveillance algorithm can be extended for application in multimodal surveillance systems. By integrating information from multiple video cameras and multiple sensors like thermal imaging sensors, pressure sensors, acoustic sensors, etc. in a multimodal surveillance system, the proposed video summarization algorithm can be used to generate video summaries using frames from multiple cameras by applying the proposed frame selection techniques and frame capture timestamp information based temporal synchronization.

Bibliography

[1] S. Sivaraman and M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1773–1795, Dec. 2013.

[2] S. A. F. C. H. B. S. J. D. S. T. L. Rhalem Zouaoui, Romaric Audigier, "Embedded security system for multi-modal surveillance in a railway carriage," in *Proc. of SPIE, 2016*, Jan. 2016.

[3] T. Wang and Z. Zhu, "Multimodal and multi-task audio-visual vehicle detection and classification," in *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, Sep. 2012, pp. 440–446.

[4] M. Magno, F. Tombari, D. Brunelli, L. D. Stefano, and L. Benini, "Multimodal abandoned/removed object detection for low power video surveillance systems," in *Advanced Video and Signal Based Surveillance, 2009. AVSS '09. Sixth IEEE International Conference on*, Sep. 2009, pp. 188–193.

[5] H. Gupta, L. Yu, A. Hakeem, T. E. Choe, and N. Haering, "Multimodal complex event detection framework for wide area surveillance," in *CVPR 2011 WORKSHOPS*, Jun. 2011, pp. 47–54.

[6] A. Prati, R. Vezzani, L. Benini, E. Farella, and P. Zappi, "An integrated multi-modal sensor network for video surveillance," in *Proceedings of the Third ACM International Workshop on Video Surveillance &Amp; Sensor Networks*, 2005, pp. 95–102.

[7] R. Rios-Cabrera, T. Tuytelaars, and L. V. Gool, "Efficient multi-camera vehicle detection, tracking, and identification in a tunnel surveillance application," *Computer Vision and Image Understanding*, vol. 116, pp. 742 – 753, 2012.

[8] K. Lopatka, J. Kotus, M. Szczodrak, P. Marcinkowski, A. Korzeniewski, and A. Czyzewski, "Multimodal audio-visual recognition of traffic events," in *2011 22nd International Workshop on Database and Expert Systems Applications*, Aug. 2011, pp. 376–380.

[9] Y. K. Wang, C. T. Fan, and C. R. Huang, "A large scale video surveillance system with heterogeneous information fusion and visualization for wide area monitoring," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on*, Jul. 2012, pp. 178–181.

[10] A. van den Hengel, R. Hill, B. Ward, A. Cichowski, H. Detmold, C. Madden, A. Dick, and J. Bastian, "Automatic camera placement for large scale surveillance networks," in *2009 Workshop on Applications of Computer Vision (WACV)*, Dec. 2009, pp. 1–6.

[11] E. Yildiz, K. Akkaya, E. Sisikoglu, and M. Y. Sir, "Optimal camera placement for providing angular coverage in wireless video sensor networks," *IEEE Transactions on Computers*, vol. 63, no. 7, pp. 1812–1825, Jul. 2014.

[12] J. Zhao, S. C. Cheung, and T. Nguyen, "Optimal camera network configurations for visual tagging," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 4, pp. 464–479, Aug. 2008.

[13] L. Liu, J. Xing, and H. Ai, "Multi-view vehicle detection and tracking in crossroads," in *The First Asian Conference on Pattern Recognition*, Nov. 2011, pp. 608–612.

[14] S. Denman, C. Fookes, J. Cook, C. Davoren, A. Mamic, G. Farquharson, D. Chen, B. Chen, and S. Sridharan, "Multi-view intelligent vehicle surveillance system," in *2006 IEEE International Conference on Video and Signal Based Surveillance*, Nov. 2006, pp. 26–26.

[15] K. Wang, Y. Liu, C. Gou, and F. Y. Wang, "A multi-view learning approach to foreground detection for traffic surveillance applications," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 4144–4158, Jun. 2016.

[16] R. Zheng, C. Yao, H. Jin, L. Zhu, Q. Zhang, and W. Deng, "Parallel key frame extraction for surveillance video service in a smart city," vol. 10, pp. 1–8, Aug. 2015. [Online]. Available: https://doi.org/10.1371/journal.pone.0135694

[17] R. Panda, A. Dasy, and A. K. Roy-Chowdhury, "Video summarization in a multi-view camera network," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, Dec. 2016, pp. 2971–2976.

[18] S. K. Kuanar, K. B. Ranga, and A. S. Chowdhury, "Multi-view video summarization using bipartite matching constrained optimum-path forest clustering," *IEEE Transactions on Multimedia*, vol. 17, no. 8, pp. 1166–1173, Aug. 2015.

[19] S. Liu and S. Lai, "Schematic visualization of object trajectories across multiple cameras for indoor surveillances," in *2009 Fifth International Conference on Image and Graphics*, Sep. 2009, pp. 406–411.

[20] (2016) S/n ratios demystified. (Date last accessed 10-December-2016). [Online]. Available: http://www.provideocoalition.com/s_n_ratios_demystified/

[21] (2016) Why is s/n ratio important to cctv. (Date last accessed 10-December-2016). [Online]. Available: https://www.fmsystems-inc.com/sn-ratio-cctv/

[22] (2016) Signal to noise ratio. (Date last accessed 10-December-2016). [Online]. Available: http://www.cctv-information.co.uk/i/Signal_to_Noise_Ratio

[23] (2016) iball ib-b8062sw 800tvl bullet ir camera. (Date last accessed 10-December-2016). [Online]. Available: http://www.iball.co.in/Product/800TVL-Bullet-IR-Camera--6mm-Lens--/11282

[24] (2016) Samsung sdc-7340bc weatherproof night vision camera. (Date last accessed 10-December-2016). [Online]. Available: https://www.samsungsv.com/Product/Detail/85/Samsung-SDC-7340BC-Weatherproof-Night-Vision-Camera

[25] (2016) Sdc-7340bc weather-resistant night vision camera — samsung us. (Date last accessed 10-December-2016). [Online]. Available: http://www.samsung.com/us/smart-home/security/security-systems/sdc-7340bc-weather-resistant-night-vision-camera-sdc-7340bc/

[26] (2016) Wv-cw594a — specifications — panasonic security system. (Date last accessed 10-December-2016). [Online]. Available: http://security.panasonic.com/products/wv-cw594a/spec/

[27] (2016) Wv-cw590 series — specifications — panasonic security system. (Date last accessed 10-December-2016). [Online]. Available: ftp://ftp.panasonic.com/videosurveillance/wvcw594/wv-cw594_specsheet.pdf

[28] (2016) Wv-cw304l / specifications — panasonic security system. (Date last accessed 10-December-2016). [Online]. Available: http://security.panasonic.com/products/wv-cw304l/spec/

[29] (2016) Wv-cw300l series fixed cameras panasonic. (Date last accessed 10-December-2016). [Online]. Available: http://www.panasonic.com/in/business/security-systems/analog-cameras/fixed-cameras/wv-cw300l-series.html

[30] (2016) Wv-cw300l security system. (Date last accessed 10-December-2016). [Online]. Available: http://www.panasonic.com/my/business/security-system/analog-products/fixed-cameras/wv-cw304l.html

[31] (2016) Sony product detail page ssccb564r. (Date last accessed 10-December-2016). [Online]. Available: https://pro.sony.com/bbsc/ssr/cat-securitycameras/cat-cctv/product-SSCCB564R/

[32] (2016) Ssccb564r features. (Date last accessed 10-December-2016). [Online]. Available: http://pro.sony.co.in/pro/product/video-security-analogue-cameras-fixed/ssc-cb564r/

[33] (2016) Smtsec sc-sp19. (Date last accessed 10-December-2016). [Online]. Available: http://www.smtsec.com/pro.aspx?id=256

[34] (2016) Yi dome camera. (Date last accessed 10-December-2016). [Online]. Available: http://www.yitechnology.com/yi-dome-camera

[35] (2016) Amazon.com — yi dome camera. (Date last accessed 10-December-2016). [Online]. Available: https://www.amazon.com/gp/product/B01CW4BG4K/ref=s9_top_hd_bw_bxklz_g421_i2?pf_rd_m=ATVPDKIKX0DER&pf_rd_s=merchandised-search-2&pf_rd_r=KJRP22NM2MQFX5X5D6TH&pf_rd_r=KJRP22NM2MQFX5X5D6TH&pf_rd_t=101&pf_rd_p=7405d5e1-a244-46a1-8af9-928a5a8ab59c&pf_rd_p=7405d5e1-a244-46a1-8af9-928a5a8ab59c&pf_rd_i=14241151

[36] (2016) Amazon.in — yi dome camera. (Date last accessed 10-December-2016). [Online]. Available: http://www.amazon.in/YI-93002-Dome-Camera-White/dp/B01CW4BG4K/ref=sr_1_1?ie=UTF8&qid=1480903946&sr=8-1&keywords=YI+Dome+Camera+Pan%2FTilt%2FZoom+Wireless+IP+Security+Surveillance+System

[37] (2016) Belkin netcam hd+ wi-fi camera with glass lens and night vision. (Date last accessed 10-December-2016). [Online]. Available: http://www.belkin.com/us/p/P-F7D7606/

[38] (2016) Amazon.in — belkin netcam hd+ wi-fi camera with glass lens and night vision. (Date last accessed 10-December-2016). [Online]. Available: http://www.amazon.in/Belkin-enabled-WeMo-Infrared-Cut-off/dp/B00KNM763E/ref=sr_1_1?ie=UTF8&qid=1480942576&sr=8-1&keywords=Belkin+NetCam+HD%2B

[39] (2016) Belkin netcam hd+ wi-fi camera review. (Date last accessed 10-December-2016). [Online]. Available: https://www.cnet.com/products/belkin-netcam-hd-plus-wi-fi-camera/review/

[40] (2016) Wifi security camera — amcrest. (Date last accessed 10-December-2016). [Online]. Available: https://amcrest.com/amcrest-1080p-wifi-video-security-ip-camera-pt.html

[41] (2016) Amazon.com — amcrest ip2m-841. (Date last accessed 10-December-2016). [Online]. Available: https://www.amazon.com/Amcrest-IP2M-841-1920TVL-Wireless-Camera/dp/B0145OQTPG

[42] (2016) Amazon.in — amcrest ip2m-841. (Date last accessed 10-December-2016). [Online]. Available: http://www.amazon.in/Amcrest-ProHD-Wireless-Security-Camera/dp/B0145OQXCK

[43] (2016) Amazon.in — samsung smartcam hd pro. (Date last accessed 10-December-2016). [Online]. Available: http://www.amazon.in/Samsung-Smartcam-Pro-Notification-SNH-P6410BN/dp/B00J38NVHE

[44] (2016) Samsung smartcam hd pro. (Date last accessed 10-December-2016). [Online]. Available: http://www.samsungsv.com/Download/SNH-P6410BN-User-Manual.pdf

[45] (2016) Dlink dcs-2310l. (Date last accessed 10-December-2016). [Online]. Available: http://www.dlink.com/uk/en/home-solutions/view/network-cameras/dcs-2310l-outdoor-hd-poe-day-night-cloud-camera

[46] (2016) Amazon.in — dcs-2310l. (Date last accessed 10-December-2016). [Online]. Available: http://www.amazon.in/D-Link-Outdoor-Surveillance-mydlink-Enabled-DCS-2310L/dp/B0092KZA0E/ref=sr_1_4?ie=UTF8&qid=1481001501&sr=8-4&keywords=Dlink+DCS-2310L

[47] (2016) Axis q8665-e ptz network camera — axis communications. (Date last accessed 10-December-2016). [Online]. Available: http://www.axis.com/in/en/products/axis-q8665-e

[48] (2016) Axis q8665-e ptz network camera. (Date last accessed 10-December-2016). [Online]. Available: http://www.axis.com/files/datasheet/ds_q8665e_61646_en_1512.pdf

[49] (2016) Omnet++ discrete event simulator. (Date last accessed 10-December-2016). [Online]. Available: https://omnetpp.org/

[50] (2016) The network simulator - ns-2. (Date last accessed 10-December-2016). [Online]. Available: http://www.isi.edu/nsnam/ns/

[51] (2016) J-sim official. (Date last accessed 10-December-2016). [Online]. Available: https://sites.google.com/site/jsimofficial/

[52] (2016) Riverbed modeler. (Date last accessed 10-December-2016). [Online]. Available: https://www.riverbed.com/in/products/steelcentral/steelcentral-riverbed-modeler.html

[53] (2016) Qualnet network simulator. (Date last accessed 10-December-2016). [Online]. Available: http://web.scalable-networks.com/qualnet-network-simulator

[54] (2016) Castalia wireless sensor network simulator. (Date last accessed 10-December-2016). [Online]. Available: https://castalia.forge.nicta.com.au/index.php/en/

[55] (2016) Frames, fields, pictures (i, p, b). (Date last accessed 10-December-2016). [Online]. Available: http://www.bretl.com/mpeghtml/pixtypes.HTM

[56] (2016) Frame types. (Date last accessed 10-December-2016). [Online]. Available: https://wiki.multimedia.cx/index.php/Frame_Types

[57] (2016) I-p-b frames explained. (Date last accessed 10-December-2016). [Online]. Available: http://forum.digital-digest.com/f20/i-p-b-frames-explained-9785.html

[58] (2016) Video compression picture types. (Date last accessed 10-December-2016). [Online]. Available: https://en.wikipedia.org/wiki/Video_compression_picture_types

[59] (2016) Video compression. (Date last accessed 10-December-2016). [Online]. Available: http://www.axis.com/th/en/learning/web-articles/technical-guide-to-network-video/compression-formats

[60] (2017) H.264/mpeg-4 avc. (Date last accessed 14-May-2017). [Online]. Available: https://en.wikipedia.org/wiki/H.264/MPEG-4_AVC

[61] J. Krause, M. Stark, J. Deng, and L. Fei-Fei, "The ko-per intersection laserscanner and video dataset," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Oct. 2014, pp. 1900–1901.

[62] E. Strigel, D. Meissner, F. Seeliger, B. Wilking, and K. Dietmayer, "3d object representations for fine-grained categorization," in *4th IEEE Workshop on 3D Representation and Recognition, at ICCV 2013 (3dRR-13)*, Dec. 2013.

[63] (2017) Opencv. (Date last accessed 16-May-2017). [Online]. Available: http://opencv.org/