SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITET

Machine Vision Based Traffic Surveillance Using Rotating Camera

A thesis submitted to the
School of Business and Social Sciences,
Department of Business Development and Technology
Aarhus University

In the partial fullfilment of requirements fo the degree of

DOCTOR OF PHILOSOPHY

Shivprasad Pandurang Patil

CTIF Global Capsule(CGC),
Department of Business Development and Technology,
School of Business and Social Sciences
Aarhus University, Herning, Denmark
2018

Machine Vision Based Traffic Surveillance Using Rotating Camera

A thesis submitted to the
School of Business and Social Sciences,
Department of Business Development and Technology
Aarhus University

In the partial fullfilment of requirements fo the degree of

DOCTOR OF PHILOSOPHY

2018

Shivprasad Pandurang Patil

CTIF Global Capsule (CGC),
Department of Business Development and Technology,
School of Business and Social Sciences
Aarhus University, Herning, Denmark.

PhD supervisor: Professor RAMJEE PRASAD,
CTIF Global Capsule (CGC), Department of Business Development and Technology, Aarhus University, Herning, Denmark.

PhD co-supervisor: Dr. Rajarshi Sanyal,
Belgacom International Carrier Services,
Brussels, Belgium.

PhD committee:

- Albena D. Mihovska, Associate Professor,
  Aarhus University, Denmark.
- Professor Mari Carmen Aguayo Torres,
  University of Malaga, Malaga, Spain.
- Professor May Huang,
  International Technological University, San Jose, USA.

PhD Series: Future Technologies for Business Ecosystem Innovation (FT4BI), Department of Business Development and Technology (BTECH), School of Business and Social Sciences, Aarhus University, Herning, Denmark.

CV

Shivprasad Patil received his Bachelor of Engineering (B.E.) degree in Electronics Engineering from University of Pune (India) in 1989. He worked as Development Engineer in Datar Switchgear Ltd, Nasik (M.S) for a year. From 1990 to 2006 he worked as lecturer, senior lecturer and Head of Department, Computer Engineering in K.K. Wagh Polytechnic, Nasik (India). He obtained his Master of Engineering (M.E.) degree in Electronics with specialization in Computer Technology from Shri. Ramanand Teerth Marathvada University (SRTMU), Nanded (India) in 2000. From 2006 to 2012, he worked with STES's Sinhgad Institute of Tecchnology, Lonavala (India) as Assistant Professor and Associate Professor. Since 2012 to till date he is working as a Professor in STES's NBN Sinhgad School of Engineering, Pune (India). He has over 28 years of experience in academia as well as in industry. His research interests include the areas of Image Processing, Computer Vision, multimedia data analysis and wireless multimedia communications.

DANSK RESUME

Shivprasad Patil modtog sin diplom Bachelor of Engineering (B.E.) i Electronics Engineering fra University of Pune (Indien) i 1989. Han arbejdede som udviklingsingeniør i Datar Switchgear Ltd, Nasik (M.S) i et år. Fra 1990 til 2006 arbejdede han som lektor, lektor og afdelingsleder, computerteknik i K.K. Wagh Polytechnic, Nasik (Indien). Han opnåede sin Master of Engineering (M.E.) grad i Electronics med speciale i Computer Technology fra Shri. Ramanand Teerth Marathvada University (SRTMU), Nanded (Indien) i 2000. Fra 2006 til 2012 arbejdede han med STES Sinhgad Institute of Technology, Lonavala (Indien) som adjunkt og lektor. Siden 2012 indtil dato arbejder han som professor i STES NBN Sinhgad School of Engineering, Pune (Indien). Han har over 28 års erfaring i den akademiske verden og i industrien. Hans forskningsinteresser omfatter områderne billedbehandling, computersyn, multimediedataanalyse og trådløs multimediekommunikation.

ENGLISH ABSTRACT

Traffic surveillance is considered as one of the indispensable aspects of smart city concept. Currently, in such kind of applications, rotating camera is preferred when comparing with the static camera. Rotating camera is preferred for the reason that, it minimizes the outlay during the transmission of information and possessions. In case of Video Surveillance (VS) systems with most favorable wireless smart city area network, some of the key areas such as transmission efficiency, lossless video data coding, data congestion, edge computing at transmission nodes can be considered. Thus, high-quality video streams are attained in spite of the transmission of the compressed information. Thus, the detection of moving objects and efficient streaming of video information has emerged as an important research topic. This research work utilizes several techniques for the effective detection of moving objects and streaming of video information using various motion estimation approaches. In this research work,quick detection of moving object along with accuracy and effective streaming of video information can be achieved with four proposed works. The first contribution includes a video surveillance system based on the statistical background subtraction model for identifying the moving objects in the case for a rotating camera. Here, the background model is evaluated in both spatial and temporal domain with respect to the distribution of each pixel in the background. The second contribution includes the analysis concerning the improved identification of objects with respect to the prediction of region in online video surveillance system using improved object identification based on Full Search Block Matching Algorithm (FS-BMA) approach. This approach leads to the prediction of foreground moving elements from the video sequences which are captured by a rotating sensor. The third contribution includes the consequence of using energy interpolated template coding for the identification of objects in case of compressing the video in traffic surveillance applications. Here, interpolation is done for the successive frames with respect to the time period instead of two successive frames. Due to decreased cost of computation, this approach becomes a good candidate for real time application. The fourth contribution includes a

diffusion of information through the wireless media and leads to the progressive streaming of video information for the traffic surveillance. Here, during the streaming of video, high quality of the data is maintained in spite of the compressed transmission of information. The experimental results show that our developed methods outperforms the existing approaches, when analyzed in terms of video quality and data throughput.

Lastly, conclusion of complete research work along with direction for future work is provided.

Keywords: Video Surveillance, Background Modelling, Adaptive Mixture of Gaussian, Least Mean Estimator, Motion Estimation, Rotating Camera, FS-BMA, Video Streaming, Error Resilience, Rate of Allocation, Video compression, Template coding, Energy Interpolation.

DANSK SUMMARY

Trafikovervågning betragtes som et af de uundværlige aspekter af smart city-koncept. I øjeblikket foretrækkes roterende kamera i sådanne applikationer, når man sammenligner med det statiske kamera. Roterende kamera er foretrukket af den grund, at det minimerer udlægget under overførsel af oplysninger og besiddelser. I tilfælde af videoovervågning (VS) -systemer med det mest gunstige trådløse smart city area-netværk kan nogle af de centrale områder som transmissionseffektivitet, tabsfri video data kodning, data overbelastning, kant computing ved overførsel knudepunkter overvejes. Således opnås højkvalitets videostrømme på trods af transmissionen af den komprimerede information. Påvisning af bevægelige objekter og effektiv streaming af videoinformation er således fremkommet som et vigtigt forskningsemne. Dette forskningsarbejde udnytter flere teknikker til effektiv påvisning af bevægelige genstande og streaming af videoinformation ved hjælp af forskellige bevægelsesestimeringsmetoder. I dette forskningsarbejde kan hurtig påvisning af bevægende objekt sammen med nøjagtighed og effektiv streaming af videoinformation opnås med fire foreslåede værker. Det første bidrag omfatter et videoovervågningssystem baseret på den statistiske baggrundsundertraktionsmodel til identifikation af de bevægelige objekter i tilfælde af et roterende kamera. Her evalueres baggrundsmodellen både i rumligt og temporalt domæne med hensyn til fordelingen af hver pixel i baggrunden. Det andet bidrag omfatter analysen vedrørende forbedret identifikation af objekter i forhold til forudsigelsen af regionen i online videoovervågningssystem ved hjælp af forbedret objektidentifikation baseret på FS-BMA-tilgang (Full Search Block Matching Algorithm). Denne fremgangsmåde fører til forudsigelse af forgrunds bevægende elementer fra videosekvenserne, som er fanget af en roterende sensor. Det tredje bidrag indbefatter konsekvensen af at bruge energiinterpoleret template-kodning til identifikation af objekter i tilfælde af komprimering af videoen i trafikovervågningsprogrammer. Her foretages interpolering for de efterfølgende rammer med hensyn til tidsperioden i stedet for to på hinanden følgende rammer. På grund af lavere omkostninger ved beregning bliver denne tilgang en god

kandidat til real-time ansøgning. Det fjerde bidrag omfatter en diffusion af information via det trådløse medie og fører til den progressive streaming af videoinformation til trafikovervågningen. Her under streaming af video opretholdes højkvaliteten af data på trods af komprimeret transmission af information. De eksperimentelle resultater viser, at vores udviklede metoder overgår de eksisterende tilgange, når de analyseres med hensyn til videokvalitet og datatransmission.
Endelig gives der konklusioner om komplette forskningsarbejde sammen med retningen for det fremtidige arbejde.

Nøgleord: Videoovervågning, Baggrundsmodellering, Adaptiv blanding af Gauss, Mindste estimat, Motion estimering, Roterende kamera, FS-BMA, Video streaming, Fejlfasthed, Tildelingshastighed, Video kompression, Template kodning, Energi interpolation.

ACKNOWLEDGEMENTS

नैनंछिन्दन्तिशस्त्राणिनैनंदहतिपावकः।

नचैनंक्लेदयन्त्यापोनशोषयतिमारुतः।।

nainan chhindanti shastraani nainan dahati paavakah.
na chainan kledayantyaapo na shoshayati maarutah.

Meaning: Weapons cannot cut, fire cannot burn, water cannot wet, and wind cannot dry this Atma.
Atma (energy) neither created, nor destroyed, only change from one soul (phase) to another soul (phase).

The Bhagavad Gita (2.23)

To begin with, I would like to remember the Almighty God for giving me strength, patience and endeavor to keep me moving on this journey of Ph.D.

Here I would like to express my acknowledgement to all those who have contributed in many way for the success of my PhD studies.

First and foremost, I would like to express my sincere gratitude to a great human, scientist and my SupervisorProfessor Ramjee Prasad. He is the pioneer of conceiving the platform of GISFI-PhD program for Indian students, to cherish their dream of PhD studies at Aalborg and Aarhus University. Throughout these years, he has been an inspirational force through his great supervisory role. I am and will remain, very grateful to him for giving me the opportunity to work with him and pursue my PhD studies.

It is a great honor and privilege for me to work with Dr. Rajarshi Sanyal, my co-supervisor. I whole-heartedly appreciate contributions of his time, suggestions and criticism in making my PhD research work technically inventive and fruitful. He has always encouraged me to perform better. I am grateful to him for supporting me in all aspect of my research work.

I would like to thank my former supervisor at Aalborg University, Associate Professor Zheng-Hua Tan. He has been very supportive of my work, especially at the time of problem definition. I appreciate contributions of his ideas and time in my research work.

My special thanks to Mrs. Jyoti Prasad, Mr. Rajiv Prasad and his family, for making our stay much comfortable with their love and support. Their motivation to follow our dreams and care is unforgettable. I am also thankful to Professor Mrs. Neelie Prasad for mentoring and supporting us through the path of our PhD studies.

Our PhD program was supported by Sinhgad Technical Education society (STES), Pune, India. I extend my sincere gratitude to, Hon. Founder President, Respected Prof. M.N. Navale, Hon. Founder Secretary Dr. Mrs. Sunanda Navale, Hon. Vice President (Admin) Mrs. Rachana Navale Ashtekar and Hon. Vice President (HR) Mr. Rohit Navale for the motivation, faith and great support. I would like to pay my respects and my compliments to Dr. A.V. Deshpande, Dr. S. S. Inamdar, Dr. Y.P. Reddy, Dr. S.D. Markande, Dr. M. S. Gaikwad and Dr. Rajesh Prasad for their faith on me and inexplicable support. I am also thankful to all my department colleagues at NBNSSOE, for helping me whenever and wherever possible.

Furthermore, I am extremely grateful to my family for their enduring support of me during my PhD journey. My parents, Pandurang and Lalita not only encouraged me but they have a persistent belief in me. I cannot forget the sacrifice of my wife,Indrayani and children’s, Chinmay and Jay. They extended their consistent support, they shower their love on me and standing by me through this long journey.I also thank my loving sisters for their blessings for this journey.

Last but not the least, I would like to thank all those who directly and indirectly involved in building this thesis and research work.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF ACRONYMS

- AVC : Advanced Video Coding
- AGMM : Adaptive Gaussian Mixture Model
- ANN : Artificial Neural Networks
- BEPN : Best-Effort Packet Networks
- BER : Bit Error Rate
- BES : Best Effort Support
- BG : Background
- BMA : Block Matching Algorithm
- BS : Background Subtraction
- BW : Bandwidth
- CA-AQM : Cross Layer Modeling
- CBVR : Content-Based Video Retrieval
- CDSSIM : Cumulative Distortion SSIM
- CLO : Cross Layer Optimization
- CNN : Convolution Neural Network
- CR : Compression Ratio
- CTU : Coding Tree Unit
- DCT : Discrete Cosine Transform
- DFC : Data Flow Control
- DMTC : Duel Metric Traffic control
- e2e : End to End
- EI-HIST : Energy Interpolated- HIST
- ERC : Error Resilience Coding
- ERVC : Error Resilient Video Coding
- FBC : Frame-Based Coding
- FC : Flow Control
- FDR : False Discovery Rate
- FG : Foreground
- FL-SSIM-RDO : Flow control SSIM-RDO streaming
- FNR : False Negative Rate
- FPGA : Field Programmable Gate Array
- FPR : False Positive Rate
- FSBMA : Full Search BMA
- GD : Gaussian distributions
- GMM : Gaussian Mixture Model
- HD : High definition
- HEVC : High Efficiency Video Coding
- HIST : Histogram energy based template matching
- HVS : Human visual system
- ITS : Intelligent Transportation System

- k-NN : k-Nearest Neighbor
- LM : Lagrange multiplier
- LMS : Least Mean Square
- LO : Lagrange Optimization
- MAC : Medium Access Control
- MAD : mean absolute difference
- MAN : metropolitan area network
- mMTC : Massive Machine Type Communications
- MB : Macro-block
- ME : Motion Estimation
- MoG : Mixture of Gaussians
- MPF : multi-path fading
- MSE : Mean Square Error
- MV : Motion vector
- NAL : Network Abstraction Layer
- NF : Neuro-Fuzzy
- OBC : Object-Based Coding
- OD : Object detection
- OF : Optical Flow
- PSNR : Peak signal to noise ratio
- QoS :Quality of Service
- RBM : Recurrent Block Matching
- RDO : Rate Distortion Optimization
- REM : Random Exponential Marking
- ROA : Rate of allocation
- ROC : Receiver Operation Characteristics
- SA : Statistical Approaches
- SAD : Sum of Absolute Diffrences
- SD : Service Data
- SI : Similarity Index
- SSE : Sum of Squared Errors
- SSIM : Structural Similarity Index
- SVM : Support Vector Machine
- TD : Temporal Differencing
- TMP : Template Match Prediction
- TL : Time Lapse
- TSS : Three-Step Search
- VC : Video Coding
- VCE : Video Compression Encoding
- VCL : Video Coding Layer
- VF : Video Frame
- VOP : Video Object Plane
- VS : Video Surveillance

CHAPTER 1. INTRODUCTION

1.1 INTRODUCTION

Surveillance refers to the close supervision or observation preserved over a group of people or a person. Visual Surveillance (VS) offers individuals, the chance to visualize the things happening in remote place; in addition, it facilitates observation of numerous remote places simultaneously [1]. VS systems have turned out to be an essential part of urban security supervision in recent years [2]. Monitoring of surveillance video demands continuous visual attention, where the brain cherry picks the constituents that would be investigated. The significant reduction in the cost of video sensors has promoted abundant use of VS systems[3]. Of let with the advent of Artificial Intelligence (AI) based systems, it is possible to detect suspicious objects, criminals, celebrities without any human intervention. This significantly helps in reducing human involvement in averting untoward incidents.

The VS system comprises of CCTV systems [5], using network of cameras. Witnessing the evolution of VS system over the time, we can categorise in three generations [6]. (1) The initial generation based on analogue CCTV technology. But this has some issues in data dissemination due to channel bandwidth and noise. (2) The next generation is based on digital video technology and networks, where problems with bandwidth restriction and channel noise are diminished. Thanks to the digital technology, the penetration of VS system have increased manifold e.g. railways, banks, supermarkets, airports and homes. (3) Third generation brings in new paradigms of VS technology. For example, with the advancement of network technology like MAN, mMTC it is possible to build an intricate city network with thousands of camera in a mesh and all centrally managed from the office location. Further with the help of AI based technologies, new features and functionalities like object or scene recognition, face recognition, vision-based motion control and alarming vision based mapping are realised [7].

The principal characteristic of any VS system is to compress massive quantities of recorded video efficiently and to enable the consequent operation. So the challenge is to manage the storage of data over the period of time. Hence any endeavour to compress the data to reduce the BW requirement and the storage requirement would be welcome[2]. Therefore, it is essential to find out new avenues in compression domain, like object detection and motion detection. This will also cut down the transmission overheads, thereby making suitable for real time applications[8].

Many of the implemented schemes related to VS depends on the investigation of visual features obtained from the temporal and/or spatial domain, and are generally dependent on the texture information, edge, or colour. Based on this hypothesis, the different tasks like moving object segmentation, action recognition, visual tracking, and OD, etc. can be actuated [9]. Many of the VS, particularly in susceptible locations such as banks and airports, are recorded in real time. Some more use cases are relevant, for example visual security in public transportation, monitoring vehicular traffic , large gathering or events [10].

It is also observed that the computational load of the systems based on some technologies can be a significant issue [11]. It is quite challenging to be deployed for real time supervision of a large-scale surveillance system. In order to overcome this difficulty, some research has focussed on computing video analytics in the compressed domain. The VS systems offer centralized monitoring, where bandwidth remains a primary concern. A good compression technology plays a pivotal role in optimizing bandwidth, where real time monitoring data is conveyed over a standard network protocol like TCP or UDP. [6].

Accordingly, in compressed video, motion data is embedded in the MV's, and are exploited in the motion compensation, and motion estimation process. Most of the accessible techniques for video object segmentation in the compressed area have been carried out in the MPEG domain. In addition, certain schemes utilize a mixture of MV's and DCT coefficients, whereas others exclusively utilize the data embedded in the MV's. Although a lot of compression standards and

VS systems exists, it is difficult to identify a committed system for archival of VS, which are beneficial for post investigation of occurrences and for comprehending the behaviors[3].

A lot of VS systems have been implemented for diverse scenarios. A distinctive VS system comprises of numerous modules for visual data processing. For illustration, single camera VS system involves foremost stages such as motion detection, BG modeling, event recognition/detection, and object tracking. All the modules form active research areas themselves [10].

VS have numerous security applications, including:

- Remote gate control
- Vandalism prevention
- Theft prevention
- Traffic control
- Perimeter protection
- Number plate recognition
- People counting
- Face recognition
- Boundary alarm

1.2 OBJECT DETECTION USING VIDEO COMPRESSION (VC)

In previous days, time lapse (TL) techniques have been deployed for video archiving in VS systems. It includes a larger space for storage, as the entire image is accumulated by FBC. As a novel method for finding a solution to this problem, OD-based coding algorithms have been implemented. On evaluating FBC with OBC, OBC can prefer to code significant FG objects like individuals with superior quality than the erstwhile segments of the scene [4]. Accordingly, in the second scheme [2], a technique related to the OBC approach is exploited for segmenting the objects. This scheme includes two procedures; MV analysis part and BG subtraction part. MV analysis is exploited to obtain moving objects for eliminating the false positive error owing to illumination variations, swaying leaves or branches, etc. In both techniques, FG objects are subjected to compression by means of an encoding scheme that is dependent on

DCT coding. BG subtraction is a characteristic scheme to detect FG objects by evaluating every new frame with an improved model of the scene BG in image sequence that is taken from a camera. Generally, motion compensation is necessary when deploying BG subtraction to a non-stationary BG. Actually, it is complicated to comprehend it to adequate pixel exactness[10].

Motion detection has attained significant consideration from the researchers. In numerous computer vision applications robust and real-time FG segmentation is a key issue [12]. The applications include OD [7], automated VS, vehicle-borne VS, and traffic surveillance network wherein cost effective sensors, such as rotating camera are employed for detecting small object. Motivation to bring rotating camera, which can cover the scene from 0 to 360 degrees, is number of stationary camera’s are replaced by single camera. This also reduces the cost of ownership. In such circumstances, BG subtraction cannot be employed directly. Motion compensation is necessary to recompense for the motion owing to the moving sensor. Subsequently, the BG is indexed perfectly and based on pixel level, FG can be detected. The fundamental postulations are that the motion representations have to be adequately precise and the constraints of the motion representation are precisely approximated. In addition, the sensing lenses are distortion-free. Actually, these postulations are complicated to realize [5]. In addition, these consume more time and inappropriate for applications in real time. Along with the estimation of the motion representation, the BG and the current image could not warp and record perfectly. This problem is moreover observed when exploiting the temporal difference approach.

The exploitation of BG modeling for detecting the moving object is found to be common in numerous applications. In the scene like VS, the BG model can be established by obtaining a BG image that doesn't comprise the stationary object, and such situations are hardly ever feasible. In certain circumstances, the BG is not accessible and/or there is a variation in illumination settings. Moreover, the object is removed or initiated from the scene. A lot of BG modeling techniques has been introduced, by taking into consideration the issues given above to formulate them more adaptive and robust. Even though the majority of these techniques exploit only a fixed camera, they offer a good initiating point for a rotating camera.

High Definition (HD) results in generating huge volume of videos which further require processing and analyzing. Two problems occur from these upcoming developments: (1) the accessible wireless network BW is not sufficient to transmit data to control stations; (2) Increased load is on researchers for data processing. A resolution to the initial issue is to carry out VCE prior to transmission, and as a result, it meets the necessity of actual channel BW in the environment of WSN. An additional resolution to the subsequent issue is to computerize the video recognition objects to facilitate improved and appropriate situational awareness and hence minimizes the workload of video-user. Nevertheless, precise situational awareness is practically unfeasible with the assurance of OD, which is not the present scope. Therefore, there were researches for evolving the present H.264 standards to make certain about the object recognition [2], however, it is not mature. The selections of coding constraints are usually engineered in the field of video quality evaluation. However traditional exploration in video quality evaluation is based on the utilization of subjective scores, which may make them inappropriate for video object recognition.

1.3 MOVING OBJECT DETECTION

In several wireless surveillance systems, camera sensors shares their video annotations to a central control station via wireless communication. Due to limited energy, BW and low computing power at the embedded camera, raw videos attained by cameras are generally pre-processed, encoded, and compressed before being distributed to the control station. An authoritative data centre at base station or central server can entirely exploit its excellent computing ability to carry out data fusion on videos from several cameras, generating a much improved comprehending of the VS than what is accessible from individual cameras. A characteristic automatic VS system comprises of five phases: OD, object classification, human identification, object tracking and understanding, and description of behaviors[5]. OD is the initial and important stage of the whole system, as identifying the object offers a focus of consideration for further operations, such as behavior analysis. Nevertheless, the unavoidable disruption of video quality occurring by compression considerably impacts the OD. For representing this, VS systems have to be modeled to enhance the computation of OD. Certain schemes exploited for detecting the moving objects are portrayed in the below sections.

1.3.1 TEMPORAL DIFFERENCING (TD)

It is a technique deployed for detecting the moving objects.In TD, areas that are moving are detected by considering the variations of pixel valuess in a video sequence of successive frames. The moving object is identified by obtaining the differentiation of image frames $t-1$ and t. TD is the foremost used technique for moving OD in case where there is a movement of the camera. Differing from static camera segmentation, in which the BG is constant; it will not be suitable to construct a BG model earlier for rotating camera due to unstable background. Hence in certain methods, the movement of the camera is approximated initially. This method is highly adaptive to dynamic changes in the scene as most recent frames are involved in the computation of the moving regions. Anyhow, it usually does not succeed in detecting entire significant pixels of certain kinds of moving objects. In addition, it erroneously detects the regions of trailing as moving object, if there remain any objects that are moving rapidly in the frames [5].

1.3.2 STATISTICAL APPROACHES (SA)

SA is exploited to prevail over the limitations of fundamental BG subtraction techniques. The BG subtraction mostly stimulates these statistical technique approaches for maintaining the data of the pixels which belong to the BG image. FG pixels are recognized by evaluating the statistics of all pixels with that of the BG model. This scheme is turning out to be more common owing to its consistency in scenes that include shadows, illumination variations and noise [11]. In addition, the statistical techniques that have been implemented portray an adaptive BG representation for tracking purpose. Accordingly, all pixels are individually modeled by a mixture of Gaussians that are updated by the received image data. With the intention of detecting if a pixel belongs to a BG or FG process, the Gaussian distribution of the mixture approach for the corresponding pixel is estimated.

1.3.3 OPTICAL FLOW (OF)

OF methods deploy the flow vectors of moving objects with respect to time to detect the moving regions in an image. In this scheme, the direction and velocity of each pixel should be calculated. It is an effectual method, however; the utilization of time is comparatively more.

BG motion approach stabilize the image of the BG plane that can be evaluated by means of optic flow. In addition, independent motion can be detected by this scheme as either in the form of flow in the direction of image gradient or by the residual flow that is not expected by the background plane motion. Accordingly, the technique can detect the MV in sequences from a BG and camera that were moving[12], nevertheless, the majority of the OF techniques are found to be complex and could not be deployed in real-time scenario, unless supported by particular hardware.

1.4 VIDEO TRANSMISSION

Video transmission remains as a significant media for entertainment and VS communications. [13]. The introduction of computers brought a revolution in the communication and compression of video [14]. Video Compression (VC) turns out to be a significant area of research, and it has facilitated several applications together with video broadcast. The popularity and development of the internet in mid1990's stimulated video transmission over best effort packed network (BEPN) [15]. Video transmission over BEPN is found to be complex by a several features together with time-varying and unknown BW, losses and delay, in addition to numerous other problems such as the fairly allocation of the network resources between several flows and the way to carry out one-to-many communication for renowned content efficiently [16][17].

There exists numerous varied video transmission and streaming applications that have extremely diverse operating characteristics or conditions. For instance, applications on video communication may be for multicast or broadcast communication or for point-to-point communication (For example, video conferencing or interactive videophone) [18][19]. Moreover the video channels may be dynamic or static and it may support a variable or constant bit rate transmission, and may sustain certain QoS measures or may offer only the best effort support (BES) [20]. The particular features of applications on video communication manipulate the model of the system powerfully [21][22].

1.5 EFFICIENT STREAMING OF VIDEOS DATA

Video streaming over WSN is persuasive for numerous significances, and several developing systems employ this technique [23]. For example, video streaming of entertainment clips and news. is extensively obtainable nowadays. For VS applications, cameras can be reasonably and flexibley set up, if connectionis offered by WSN [24]. A WLAN could connect a variety of audiovisual entertainment equipmentin at residence. While video streaming requires a steady flow of information and delivery of packets within a limit of latency, wireless radio networks find more difficulties to render high QoS and relability. It gets more challenging due to the conflict from other various nodes [25], in addition to intermittent interference from exterior radio sources like cordless phones or microwave ovens. For mobile nodes, shadowing and MPF may further raise the inconsistency in transmission error rate and link capacities. For such systems to convey the best end-to-end performance, reliable transport, wireless resource allocation, and VC have to be measured jointly, thus moving from the conventional layered system design to a cross-layered model [26].

The mixture of the rigid QoS requirement, unreliability of wireless links and the transmission of video over WSN are very demanding problem to address. For continuous video playback, the user has to decode and present a novel video frame at regular intervals (usually for every 33 msec) [27]. At the time when playback at the client side has begun, this entails rigid timing restraints on the VF transmission. If a VF is not entirely conveyed in time, the user may lose a portion or the total frame [28].

Normally, a small possibility of frame loss (play back starvation) is necessary for excellent apparent video quality. Moreover, the VF sizes (in byte) are extremely changeable and they are usually error prone at a high rate [29][30]. They establish a noteworthy number of bit errors that could deliver undecodable packet. The rigid timing parameters, on the other hand, permit only for restricted retransmissions. In addition, the wireless link errors are characteristically bursty and time-varying[31]. An error burst that might persevere for hundreds of msec could make the transmission temporarily impractical to the users, who are affected. These entire characteristics and needs make real-time streaming of video over WSN a very attractive domain of research [32].

Generally, there exist two approaches to deliver video over a packet switched network together with packet-oriented WSN's, (1) streaming, or (2) file download. By downloading the file, the whole video is downloaded to the terminal of the user before the commencement of playback. The video file is assessed with a consistent traditional transport protocol, like TCP [33]. The significance of file download is that it is comparatively easier and makes sure of an improved video quality [34]. This is owing to the WSN losses that are treated by the TCP protocol and the play–out could not instigate till the completion of video file download devoid of errors. The disadvantage of download is the increased response time, usually represented as service data (SD). The SD is the instance from when the client asks for the video till the commencement of playback. Particularly for small BW wireless links and huge video files, the SD can be extremely high [35].

In case of video streaming, playback starts prior to the whole file get downloaded to the terminal of user. In video streaming, normally only a small part of the video that ranges from a certain VF's to many frames (ranging from hundreds of msec to numerous sec or minutes) are downloaded prior to the commencement of streaming. The enduring section of the video is delivered to the client when the video playback is in progression [36]. A major trade-offs in video streaming is among the SD and the video quality, i.e., the lesser the amount of the video which is downloaded prior to the commencement of streaming, the more the uninterrupted video play back depends on the appropriate delivery of the remaining video over the unreliable WSN. The WSN further deteoriates the video quality because of low bit rate VF's and in certain cases, VF's are left out completely.

The limitation of video streaming relies on maintaining the quality deprivation to a level which is tolerable or noticeable while consuming the WSN resources powerfully (i.e., supporting as many synchronized streams as possible). Moreover, video streaming and file download with certain SD are appropriate only for pre-recorded video.

1.6 MAJOR CHALLENGES

Object detection can be employed vitally in computing the position of the object in consecutive frames in a video sequence [5]. Here, detecting the objects in a proper way can be considered to be a challenging task due to the variations in size, shape, location, and orientation of the objects. In object detection, several challenges have to be considered while operating a video detector; they are as follows,

- Illumination causes an impact on the emergence of BG and leads to false positive detections.
- It is very challenging to evaluate the BG when sensory camera is moving or rotating.
- During the evaluation of BG frame, the process mainly gets affected due to object occlusion..
- Difficulty in segmentation process due to the presence of BG clutter. Thus, it is impossible to represent a BG and divides the moving FG objects.
- Shadows transmitted from FG objects leads to the difficulties in processing with respect to BG subtraction. Hence, the overlapping shadows delays their partition and classification.
- BG subtraction techniques for VS have to deal with the signal that gets corrupted by various noises such as sensor noise and compression artifacts.
- In detecting the moving object, speed of the object plays a major role. Hence, if the movement of object is very slow, the uniform region conserved by the portion of the objects cannot be detected optimally.

1.7 OBJECTIVES

This work aims at the OD, video transmission and video streaming in the VS system. The objectives of the work can be explained here,

- To design detection technique for the detection of moving object and the BG frames in the VS system, which employs rotating camera as a sensor.
- To propose an approach for the enhancement of error free coding in VS system.
- To intend a technique based on VC to attain a high-quality video stream in spite of compressing the data during transmission.
- To establish an efficient coding technique to create an accurate template to attain enhancement in compression process.
- To develop an appropriate model to predict about the exact template and thereby reducing the processing time and overheads.

1.8 CONTRIBUTIONS OF THE THESIS

The contributions of this research work to perform OD, video transmission and video streaming in the VS system are enlisted as follows,

The first contribution of this work is the development of a statistical BG approach for the detection of moving object with respect to the motion compensation of the rotating camera. This technique can efficiently deal with both the outdoor and the cluttered scenes with high detection rate.

The second contribution of this work is the design of a coding approach in the VS for the elimination of noise that occurs during the detection process. Here, this technique can be found to me more accurate in the field of VS and thus high estimation probability can be attained. This permits its deployment in real time applications.

The third contribution of this work is the developement of a video streaming technique to control the flow of captured video data over the multi-hop network. In this technique, the video quality and the error resilience can be improved with respect to the high throughput. Due to this, said approach may be suitable for edge computing in smart city infrastructure.

The fourth contribution of this work is the design of an energy incorporated coding approach for the compression of video data for the traffic surveillance. This technique can be validated over the traffic surveillance data with high coding accuracy and less processing time.

1.9 ORGANIZATION OF THE THESIS

The organization of the research work explaining about the OD, video transmission and video streaming in the VS system is provided in this section.

Chapter 1 of the work provides the introduction to VS, OD using VC, moving OD, video transmission and video streaming in the VS system.

Chapter 2 explains the various literary works contributed towards the OD, video transmission and video streaming in the VS system.

Chapter 3 provides a brief explanation towards the basic moving OD techniques in the VS system. The experimental outcomes and the analysis of the proposed model are also explained here.

Chapter 4 explains the study and system modeling of FS-BMA for the detection of an object in the VS system with the improved rate. The experimental outcomes and the analysis of the proposed model are also explained here.

Chapter 5 explains the design and development of an OD approach in the VS system with respect to the template coding. The analysis includes both the algorithmic and the comparative analysis.

Chapter 6 provides a brief explanation towards the transmission of video data through the wireless channel in the VS system. The experimental outcomes and the analysis of the proposed model are also explained here.

Chapter 7 concludes the research work with the summary, research contributions, and the future work towards the detection of objects, transmission of video and streaming of video data in the VS system.

CHAPTER 2. LITERATURE REVIEW

2.1 INTRODUCTION

VS system can be manual, semi-automatic, or completely-automatic. Normally, the human operator is made accountable for scrutinizing the manual VS system. The complete mission is to examine the ocular information imminent from various dissimilar cameras (like static, rotating). It can be considered as a monotonous job. These systems can be problematic for outside and outdoor places as it is difficult to manage when there is massive proliferation of cameras [39]. A good example is VS in smart cities. Both the human operator and the AI assisted computer vision systems can manage the semi-automatic traffic surveillance system. Type of operation can be classified as, face recognition, motion detection and tracking, abnormality detection of patterns and classification and identification of the object [40, 41]. In computer vision, object tracking can be regarded as the most challenging task. The main intention of tracking in computer vision is to detect the object to be tracked and establish a model in a sequential frame series. Normally, every visual surveillance process commences with the identification of moving objects in the video streams [42].

2.2 CATEGORIZATION AND DESCRIPTION OF CONTEMPORARY RESEARCH

In this chapter, various approaches had been discussed for the traffic surveillance system.

1. Similarity measurement approaches
2. Video streaming approaches
3. Optimization based approaches
4. Learning-based approaches
5. Coding approaches
6. Motion estimation approaches

2.2.1 SIMILARITY MEASUREMENT APPROACHES

In 2017, Kourtis *et al.* [39] have proposed an improved video eminence measurement method intended for the next generation (5G) mobile configurations, targeting small cell deployment. This approach mainly depends on an improved handling of the SSIM, as a minimized reference metric and was made suitable for virtual network function (VNF). It mainly facilitates the in-service monitoring of the video quality delivered to the end user. A significant benefit that can be drawn from this is that the video eminence measurement was done at the edge of the network rather than user equipment itself, thereby saving considerable power consumption of device.

In 1997, Lu and Liou [40] have proposed an improved block search approach aiming to minimize computational overheads evaluating the movements. The TSS model was implemented for the evaluation of movement in case of the matched chunks. The system extensively employed in real time video applications. From the experimental analysis, it was noticed that, this approach attains better performance in terms of its efficiency and processing speed than standard approaches.

In 2009, Wang *et al.* [41] have offered a new similarity measurement approach depending on the neighborhood samples and label allocations. A graph dependent partly-supervised learning approach was implemented, which has been referred in several fields. On the other hand, the evaluation of the pair-wise similarity approach was not examined adequately because of its various critical characteristics. Usually, evaluation was done in terms of the resemblance between the two samples depending on the Euclidean distance between them. Here, the resemblance regarding these two samples was not associated to their Euclidean distance, but it was associated with the allocation of neighboring samples and labels. It was evident that this conventional distance based resemblance measurement approach may lead to the errors that were generated during the classification approach even for the simple sample sets. Generally, this type of resemblance based on the neighborhood between the two samples includes three features such as their distance, the dissimilarities in the allocation of the neighboring samples and the dissimilarities in the allocation of the neighboring labels. From the experimental outcomes, it was clear that this approach attains better similarity index when compared with the other traditional resemblance measurement approaches.

In 2014, Zhao *et al.* [42] have proposed an enhanced SSIM-based error-resilient RDO approach for improving the performance of transmitting the video series in the wireless channel. Initially, based on the SSE dependent RDO approach, based on Lagrange optimization method was combined along with the SSIM dependent RDO video coding in the error free surroundings. Moreover, the deformation in the SSIM dependent decoding of the end customer was evaluated at the encoder and it was incorporated in the RDO in order to include the deformation that were persuaded with the transmission in the encoding scheme. Furthermore, lagrange multiplier was obtained hypothetically for the optimization of the encoding scheme with respect to the assortment of the error flexible RDO approach. From the experimental outcomes it was clear that this approach attains excellent quality in case of transmitting the video series and better BER when compared with the other standard approaches.

In 2016, Sankisa *et al.* [43] have introduced two approaches for the analysis of the QoS of the video based on SSIM. This SSIM approach mainly utilizes both IDE and CDSSIM. In IDE approach, three sections of the frames were restructured iteratively, which was deployed for the integrations of three dissimilar losses in the packets. Moreover, the resultant deformations were also incorporated based on the probability in order to attain a complete expected deformation. In CDSSIM approach, a collective estimation scheme for the complete deformation was evaluated by adding the inter-frame possibilities. Moreover, this approach also includes NR based regression structure in order to identify the CSSIM template to get evade of the computational involvedness and was deployed for various real time applications. Here, both these two methods were estimated with respect to the distribution of the resources and packet prioritization.

2.2.2 VIDEO STREAMING APPROACHES

In 2018, Zhou *et al.* [44] have implemented description coding approach based on the transferred 3-dimensional set partitioning in hierarchical trees (SPIHT) scheme. This approach was established to produce variable autonomous descriptions in case of sub-streams depending on the condtion of the network. Moreover, an enhanced error avoidance safeguard scheme to significant components of bit stream has been offered. Furthermore, an efficient segmentation approach was established depending on the event of various dissimilar types of loss rate in the packet to improve the image resolution. From the simulation outcome, it was apparent that this scheme achieves better performance in terms of PSNR and ocular eminence when compared with the other predictable schemes.

In 2016, Wang *et al.* [45] have implemented an assessment approach depending on the eminence estimation and rate deformation in 3D videos. In this approach, first, the subjective eminence measurement testing on two databases that comprises of several asymmetric compressed stereoscopic 3D videos was transmitted with disproportionate transform quantization coding, their groupings, and numerous selections of post-processing approaches. Here, both the disproportionate stereoscopic video coding approaches and the proportionate coding approaches were compared together, and thus, it was validated with their probable enhancement in the coding gain. This approach permits for the calculation of coding attainment quantitatively depending on the variations of disproportionate video compression. From the experimental outcomes, it was obvious that this approach attains enhanced insight when compared with the other approaches.

In 2010, Xu *et al.* [46] have established a video eminence configuration with a supplementary amendment element to link the gap that occurs between the HVS and computed objective scores by machines.. At first, the video was depicted by various representing video series with huge entropy values. Next, several eminence expressions comprising of luminance, dissimilarity, configuration and spatiotemporal consistency were implemented in order to assess the eminence of the deformed video. For differentiating the spatiotemporal consistency, an improved descriptor known as rotation sensitive three dimensional consistency prototypes was formulated. Finally, the outcomes in the correction element stimulated by dissimilarity effects were enhanced. From the experimental outcomes, it was apparent that this approach attains better efficiency when compared with the other traditional approaches.

In 2012, Kim and Hwang [47] had implemented an enhanced approach for partitioning and extracting the moving substances in the video series. These moving objects in the video series were partitioned, and then, the VOPs were extracted. In case of the multiple VOPs in a scene, depending on the associated component analysis and efficiency related to the dislocation of VOPs in the consecutive frames was also examined. This approach mainly instigates with a vigorous dual edge map attained from the dissimilarity connecting two succeeding frames. The edge points present in the preceding frame was eliminated, and thus, the residual edge map and moving edge were deployed to extract the VOPs. From the experimental results, it was apparent that this approach achieves better outcomes than the other classical approaches.

In 2005, Lei and Georganas [48] had established an enhanced approach by investigating the constraints of buffer as well as the end-to-end impediment and thus, it explains about the situation that was to be followed by the buffer dependent transcoder, for example underflowing or overflowing of buffer. Moreover, the resource descriptions and variations in the scene of the pre-determined video series were also examined. Depending on the constrictions in the channel and the descriptions of the resource video series, an adaptive bit rate adaptation model was implemented in order to perform the operations of transcoding and thus, the pre-encoded video series was transmitted over the wireless channel. Here, by controlling the bits in the frames depending on the circumstances of the channels and buffer possession, the preliminary activated impediment of pre-encoded video series was minimized drastically.

In 2015, Xiang *et al.* [49] have introduced two substitute error resilient approaches for the transmission of multi-views in videos depending on the Wyner-Ziv coding approach. A light load based encoder with error resilient approach normally has no interactions connecting the cameras at the encoder, whereas the sequential redundancies can be investigated to produce the side information at the decoder. In this condition, it was not only vigorous to losses in the channels but also has autonomous encoders with small encoding involvedness. Moreover, an error-concealed based restructured frame was deployed at the receiver with respect to the side information in the WZ decoder. Thus, this approach mainly upholds the original multiple viewing sequences of bits in which, it was unchanged by basically totaling up WZ bits for the fortification. From the experimental outcomes, it was apparent that this approach attains better performance in terms of flexibility when compared with the other standard approaches.

In 2004, Mezaris *et al*. [50] have established a partitioning approach in the video entities. This approach mainly includes three phases such as, a preliminary partitioning approach which was done in the initial frame based on the information regarding the colour, motion and position using K-means approach, a sequential tracking approach using Bayes classifier with respect to the rule dependent dispensation approach was deployed for the relocation of transformed pixels to the existing areas by handling the resources based on the original areas, and a route dependent area reconciling practices which mainly utilizes the elongated phrase depending on the route concerning the areas, so as they were collected in accordance with the entities with dissimilar movements. From the experimental outcomes, it was obvious that this approach attains enhanced partitioning rate when compared with the other traditional approaches.

In 2014, Perkasa and Widyantoro [51] had developed a network for examining the traffic. Generally, traffic jam creates several severe predicaments, and thus it generates grief and severs to be the basis for the inadequacy of fuel utilization. Here, fraction of the elucidation to this predicament was considered to be as a network that can frequently detect the traffic obstruction intensity in a division of road. It mainly presupposes a stationary camera which was attached in the high location and thus facilitates it to observe the course of traffic in a division of road. The video series was examined by evaluating the density and speediness of the traffic depending on the movement of the vehicles. The arrangement of density and speediness of the traffic was deployed for the classification of the density level that occurs during traffic; moreover, it includes free flow, deliberated movement or overcrowding type. From the evaluation approach, it was obvious that this approach attains better accurateness and overcrowding detection rate than the other approaches.

In 2008, Lee and Chung [52] have introduced a novel approach depending on the cross-layer for the transmission of video series over the wireless networks. This type of intention mainly includes an adaptation approach depending on the rate in two layers such as physical layer and data link layer as well as the adaptation approach depending on the quality in the application layer. The adaptation approach depending on the rate was deployed to regulate the transmission rate of the information regarding the calculated received signal strength indicator at the transmitter side and thus, notify about the limitations within the rate to quality based adaptation approaches. Here, the adaptation approach mainly makes use of the limitations within the rate to control the quality of transmission of video series. From the experimental outcome, sit was apparent that this approach achieves better utilization quality rate when compared with the other standard approaches.

In 2014, Shao *et al.* [53] have introduced a new CBVR scheme for searching various human activities depending on the spatiotemporal localizations with the video series. This approach mainly includes several temporal localization parameters depending on the histogram related to the segments in time domain, and similarly, spatial localization depends on the histograms within a 2-D spatial network. Moreover, this CBVR approach mainly depends on the abovementioned localization, which was trailed by the consequent ranking approach and thus leads to the creation of elevated discriminative network, while taking less computation time than the other traditional approaches. From the experimental outcomes, it was clear that this approach attains improved localization rate when compared with the other basic CBVR approaches.

In 2004, Erdem *et al.* [54] have implemented various evaluative approaches to estimate the quantitaive performance of the partioning video substances as well as the tracking approaches without ground truth dependent partitioning maps. This approach mainly depends on the spatial dissimilarities of colour and movement along the periphery of the estimated VOPs as well as the dissimilarities that occur in the colour histogram of the existing entity plane and earlier one. They were exploited to confine the areas in both time and spatial domain depending on the quality of partitioning outcomes. Here, they were integrated together to capitulate a solitary based statistical determination to point out the righteousness of the periphery partitioning and tracking outcomes over a series. The influence of the projected routine was determined without ground truth map and has been established by several canonical correspondence based investigations with an additional series of ground truth (where information is available) on a video series. From the experimental outcomes, it was obvious that this approach achieves better partitioning rate when compared with the other standard approaches.

In 2010, Chen *et al.* [55] has implemented an analytical spatial harmonizing approach in case of inter-prediction depending on template matching. In addition to these environmental based restructured pixels, it leads to the creation of the templates depending on the analysis of the pattern identification based movement investigation which normally uses the various pixels. Moreover, a mode selection approach was established in order to investigate about the adaptively selected Pitch mapping approach at MB level. From the experimental outcome, it was clear that this approach attains better performance in terms of low BER when compared with the other traditional approaches.

In 2005, Laptev [56] had established a conception of spatial interest points into the spatiotemporal domain, and it demonstrates about the consequential characteristics which were often replicated by appealing events that was deployed for a compact illustration of video information as well as for the analysis of spatiotemporal incidents.In order to identify the spatiotemporal incidents, a suggestion was made for the construction of Harris and Forstner interest point operators, and thus, the confined configurations in the spatiotemporal domain were also identified, where, the values of each image have momentous confined changes in the spatiotemporal domain. Here, the spatiotemporal coverage of the identified incidents was evaluated with respect to the exploitation of the regularized spatiotemporal Laplacian operator depending on their extents. In case of denoting the identified incidents, confined, spatiotemporal, scale-invariant approaches were also estimated, and thus classification was done for each incident with respect to its descriptor. From the experimental analysis, it was clear that this approach attains better performance in case of identifying several features in the scenes with enhanced rate when compared with the other approaches.

In 1997, Davis and Bobick [57] had implemented a novel view dependent technique which was used to denote and identify several actions within the image series. The source of this representation was regarded to be as a temporal metric, wherein a motionless vector image was considered to be as function of the motion features at the consequent spatial position in an image series.Normally, two modules were deployed to represent the power with respect to the metrics. Here, the first value denotes the binary value in which, it describes about the occurrence of the movement and similarly, the second value was considered to be as the task denoting the frequency of the movements in the image series. Finally, an identification approach was suggested in order to map both the spatial and temporal characteristics depending on the movements in an image. From the experimental outcomes, it was obvious that this approach achieves better partitioning and classification rate when compared with the other conventional approaches.

In 2003, Chalidabhongse *et al.* [58] have formulated an estimation approach known as perturbation detection rate which was deployed for the measurement of performance with respect to the background subtraction approaches. This approach has several benefits when compared with the investigation of ROC. Particularly, this type of approach does not need any kind of foreground distribution. This approach was generally deployed to measure the sensitivity of a BGS approach for the identification of the small disparity objects aligned with the various background conditions.

2.2.3 OPTIMIZATION BASED APPROACHES

In 2008, Maddalena and Pestrosino [59] had suggested an approach depending on the self association in the course of ANN, which has been extensively exploited in human image processing configurations and more usually in cognitive disciplines. This technique has been able to deal with various prospects comprising of several movable backgrounds, steady enlightenment dissimilarities, and concealment. Moreover, it mainly does not involve any bootstrapping boundaries but, it exploits the background scheme concerning the transmitted shadows by stirring objects, and attains enhanced identification for several dissimilar varieties of videos which were captured using motionless cameras. From the experimental outcomes, it was obvious that this approach achieves enhanced identification rate and speed when compared with the other modelling approaches.

In 2014, Evangelio *et al.* [60] have established an investigation regarding some of the appropriate GMM techniques and thus revise about their essential postulations and intend assessments. Here, GMM classifiers depending on the pixels were regarded to be like the most significant preference during the identification of the change in the video based domain. In this approach, the configurations were enhanced with respect to the variance controlling approach and the integration of region analysis based feedback. From the experimental outcomes, it was obvious that this approach attains better performance in terms of identification rate when compared with the other standard approaches.

In 2013, Huang *et al.* [61] have established an approach based on the correlation of video coding and the GMM classifier. The typical GMM classifier mainly depends on the arithmetical information of every pixel, and thus, it tends to change depending on the illumination variations. Before evaluating each and every pixel in the videos, it should be deciphered into unprocessed videos. Here, both the MV’s and the intra mode were deployed to locate the foreground comprehensive chunk and then it appends the overhead flag in the compressed video to specify it. In the deciphering process, the probable foreground regions were deciphered, and the moving objects were identified in these regions. This approach mainly deploys two datasets in which, both datasets were investigated with inimitable and changing lighting circumstances. From the investigational outcomes, it was obvious that this approach achieves enhanced detection rate when compared with the other classical approaches.

In 2016, Chen *et al.* [62] have developed an orientation scheme for high competence video coding. In this coding approach, three chief methodological assistances were formulated. In first contribution, the background reference was created progressively by revising the chunk instead of renewing the picture. This revision formulates the approach which was free of bit rate burst and was made more appropriate for real time applications and thus can produce high quality background location even with intricate foreground. In second contribution, a scheme to choose the background CTUs depending on both temporal and spatial smoothness was implemented. In third contribution, a scheme to choose a particular background CTUs with coding characteristics were implemented based on the motion of the whole picture, which effectively follows the GoP-level finest routine during the creation of CTU-level decisions. This background location was formulated into HEVC and thus founds to have better efficiency in terms of coding and decoding involvedness.

In 2015, Sriharsha and Rao [63] had proposed an approach regarding establishing a moving object using a motionless digital camera and correlating it in uninterrupted video series.. In the first phase of testing, both the background subtraction and series dissimilarities approaches were deployed for the identification of objects, and thus, the movement was evaluated by correlating the centroid of the moving object in each dissimilar video series. Mobility based foreground areas were tracked and assumed to be as one of the main decisive needs for the surveillance configurations. In the second phase of testing, similar approaches were selected for identifying the objects, but the movement of each tracking objects was evaluated by Kalman filtering. On the other hand, the most excellent approximation was prepared by integrating the prediction knowledge and amendment methods that were included as a component for the creation of Kalman filter. Consequently, kernel dependent tracking phenomenon based on the mean shift presumption was formulated for tracking a particular entity in terms of prejudiced occlusion. Depending on the spatial masking with an isotropic kernel, the histogram based objective illustrations were standardized. The masking persuades spatially flat resemblance task that is appropriate for inclination dependent optimization. While considering the metric attained from the Bhattacharyya Coefficient, resemblance measurement was deployed, and consequently, mean shift approach was deployed for the execution of the optimization approach. For enhancing the effectiveness of the tracking process, an object tracking approach based on the Kalman filter was amalgamated with the mean shift scheme. Here, first, the configuration version of Kalman filter was created, and thus, the interior of the object was expected to be deployed in mean shift scheme in order to locate the target in the frame.

In 2005, Lee *et al.* [64] have proposed an efficient approach to enhance the convergence rate without the compromising of GMM permanence. For the representation of non-stationary sequential distributions of pixels in the video, several adaptive Gaussian mixtures were deployed. However, a frequent predicament for this scheme was considered to be matching among mock-up union swiftness and permanence. This was attained by reinstating the comprehensive, motionless maintenance features with an adaptive erudition rate deliberated for each Gaussian at every frame. Considerable enhancements were revealed on both real and unreal video series. From the simulation outcomes, it was apparent that this approach was integrated with the statistical framework in order to attain better enhancement in segmentation when compared with the other classical approaches.

In 2009, Xiang *et al.* [65] have implemented an improved 2D layered multiple description coding which was exploited for the broadcasting of error-resilient video transmission over the unpredictable system. Here, this approach was deployed to distribute the multiple depictions of series of sub-bits based on the 2-Dimensional scalable series of bits related to the system pathways with unequal loss rates. In order to reduce the end to end distortion specified in the entire rate resources and possibilities regarding the packet loss, the resources and the path charges were optimally distributed depending on the hierarchical sub-levels of the scalable series of bits. Here, the conservative Lagrangian multiplier scheme was avoided to resolve several predicaments due to computational cost. Hence, for resolving the rate distortion based optimization predicament, Genetic algorithm was utilized. From the simulation outcome, it was seen that this approach attains better performance when compared with the other standard approaches.

In 2013, Mukherjee *et al.* [66] have proposed two kinds of enhancements such as an improved distance measure depending on the local support weight and gradient of histograms to make available the distinct cluster values and exploitation of the conception regarding the background level to divide the foreground appropriately. This approach mainly utilizes number of clusters which was deployed for the simplification procedures. The benefits of this approach involve inherent exploitation of association of pixels through the distance measure with the slightest adaptation to the conservative GMM approach and efficient elimination of background noise in the course of the utilization of the conception of the background level without the implication of the post-processing steps. From the experimental outcomes, it was apparent that this approach attains better accuracy when compared with the other standard approaches.

In 2012, Chen *et al.* [67] have established an improved approach for the evaluation of the end to end deformation depending on the quantization after encoding and arbitrary broadcast inaccuracies due to broadcasting the video frames in the video communication systems. This approach principally fluctuates from the imperative conventional approaches with several filtering schemes. For instance, an interpolation that occurs in the sub-pixel motion compensation as executed in the video coding sequences. The evaluation of deformations for both pixels and its sub-pixels with respect to the filtering schemes mainly necessitates the estimation of the arbitrary values in terms of the second moment of a biased averaging process. Here, it does not demands the likelihood distributions for the estimation of the arbitrary values in terms of the second moment of a biased averaging process.

In 2016, Shen *et al.* [68] have introduced a precise and computationally proficient background subtraction approach for embedded camera network. Here, a baseline description was implemented depending on the utilization of luminance and then it was expanded for employing the colour information. The primary design of this approach was to exploit arbitrary projection matrix for minimizing the dimesionality of the information keeping significant information of data. Depending on the numerous datasets, the accurateness of this background subtraction approach is analogous to that of the conventional background subtraction approaches. Furthermore, it is demonstrated that, the computational efficiency is independent of embedded platforms. The authentic functioning illustrates that this approach was constantly enhanced and was several times more rapid when compared with the other standard approaches.

In 2013, Maddalena and Petrosino [69] had established a structure to partition the motionless foreground substances aligned with poignant foreground substances in particular inspection series received from the motionless cameras. The repetitive detection of several objects that are abondened in a video series is an appealing area of computer vision. Some of the illustrations such as stolen stuff in the airports, railway stations, and irregularly parked vehicles were considered to be as the momentous problems. Here, an approach based on the image sequence was attained through learning in a self-organized neural network changes in the image sequences. It was observed as the trajectories depending on the pixels with respect to time period and was implemented within the model dependent structure. From the experimental outcomes, it was apparent that this approach attains better accuracy rate when compared with the other classical approaches.

In 2010, Bhaskar *et al.* [70] have formulated an extensive clustering based BS approach with an assortment of established symmetric alpha stable allocations. In order to identify the moving substances in the video series, background subtraction scheme was regarded to be as the most effective approach. An undemanding BS approach mainly includes the construction of a template regarding the background, and thus it tends to remove the areas of the foreground substances, for the motionless camera and thus, there subsist no activities in the background. Depending on the log moment approach, an online self-adaptive scheme for model parameters was made accessible. From the experimental results, it was apparent that this approach attains enhanced identification rate with respect to the information from the motionless and moving video cameras when compared with the other traditional approaches.

In 2005, Liu and Zheng [71] have implemented an enhanced partitioning and tracking approach in terms of extracting the object. When compared with the conventional techniques, this approach mainly originates the separation of the video object from the background as a categorization problem. Here, each frame was alienated into diminutive chunks. Subsequent to the physical partitioning done in the first frame, the chunks present in this first frame were deployed as the training samples for the classifier with respect to the background objects. Moreover, an improved tool known as Si-learning was exploited to guide the classifier which has better performance than the traditional SVM classifiers in linearly with non-distinguishable conditions. To covenant with outsized and multifaceted substances, a multilayer approach assembled with a hyper-plane tree was implemented. Each and every node in the tree denotes a hyper-plane, which is responsible for classification of the training samples. Here, several hyper-planes were made indispensable to categorize the complete deposits. Depending on the tracking stage, the centriod pixels which present in each and every chunk within a consecutive frame were categorized with respect to the hyper-plane series from the core node to the leaf node of the tree based hyper-plane, and thus the chunks with each class were detected consequently. All the chunks with entities thus generates entity of concern in which, the periphery was regrettably assumed to be in the form of stairs with respect to the consequences of the chunks. This method iteratively chooses a few revealing pixels in case of the inspection of class labels, and thus, minimizes the improbability regarding the authentic periphery of the entity.

In 1999, Stauffer and Grimson [72] had implemented a modelling approach depending on the concoction of Gaussians and online estimation to renew this model. Here, the allotments regarding the concoction of Gaussian with respect to GMM were then estimated to establish the expected outcome from the background model. Each pixel was categorized depending on the allocation of the Gaussian concoction which normally denotes the efficient and accurate segment in the background model. From the simulation results, it was noticeable that, this outperforms with better reliability while dealing with lighting changes and recurring movement due to clutter , than the other approaches.

In 2008, Bouwmans *et al.* [73] have established an assessment for the inventive classification approach with several enhancements. Moreover, several techniques were also discussed in case of the consequences regarding the minimization of the computational time. Initially, an improved MoG approach was repeated and examined with respect to the issues that occur in the video series. Here, several enhancements were classified in terms of the policies which were deployed to enhance the innovative MoG depending on the crucial circumstances that are claimed to be handled.

2.2.4 LEARNING-BASED APPROACHES

In 2012, Zhu *et al.* [74] have established a new recursive Bayesian learning dependent approach for the proficient and precise segmentation of video with respect to the dynamic background. Here in this approach, pixels in each frame can be described as the layered normal distributions which lead to dissimilar contents in the background images with respect to the scene. The layers were associated with a confident term, and thus only the layers were deployed to gratify the specified assurance and thus, it has been restructured through the evaluation of recursive Bayesian learning. This leads to the formulation in which the erudition of movement regarding the background to be more precise and proficient. Finally, a local texture correspondence scheme was also established in order to fill the vacancies and thus eliminates the incomplete false foreground areas. From the simulation outcomes, it was observed that this approach attains better improvements in case of partitioning the background from the scenes when compared with the other approaches.

In 2013, Wang *et al.* [75] have implemented an enhanced scheme to identify human activities across cameras through recostuctable paths. Here, each activity was represented as a collection of visual expressions depending on the spatiotemporal characteristics. Even though demonstration of activity was susceptible to several variations in the scrutiny, the re-makeable pathway was made capable to interpret the activity descriptors of one camera to another camera. In the learning of the paths, a dictionary was considered to be more erudite beneath each sight to renovate the activity descriptors into a sparsely demonstrated space, and a linear mapping function was concurrently cultured to overpass the semantic gap connecting the source and target spaces, such that each domain configuration can be entirely discovered. Along the re-makeable paths, an unidentified activity from the end inspection was accurately restructured into any source observation, and hence the SVM classifiers trained in source observations were capable to discriminate this unidentified activity from target observation.

In 2013, Zhang *et al.* [76] have established a statistical scheme for the exponentially weighted moving average (EWMA) dependent background modeling approache. This background modeling scheme was deployed to renew the features depending on EWMA with predetermined learning rates.. This scheme normally describes a new manner to investigate the changes that occur in the pixel intensities in video sequences and thus constructs an intensity point movement likelihood map, which was considered to be as a recursively renewed 2 D lookup table for recovering adaptive learning rates. From the experimental outcomes, it was apparent that this approach attains enhanced adaptive rate when compared with other dissimilar approaches.

In 2010, Cheng *et al.* [77] have established an outline for the classification of human activities and localization in the video series depending on the structured learning of confined spatiotemporal characteristics. Various local patches were deployed to represent the human activities. In this approach, a discriminative hierarchical Bayesian classifier (DHBC) approach was employed to choose several interest points depending on the spatiotemporal characteristics which were made beneficial for each and every movement. Those concise characteristics were then passed to a SVM with protrusion of PCA which was deployed for the classification assignment. In the meantime, the localization depending on the human activities was performed based on the dynamic conditional random fields (DCRF) established to integrate the spatiotemporal organizational constraints of several super-pixels which were attained from these characteristics. In the video series the super-pixels mainly defined on the information regarding the contour and activities with respect to the consequent characteristic areas. From the simulation outcome, it was obvious that this approach attains enhanced effectiveness and robustness with respect to the identification of human activities when compared with the other standard approaches.

In 2013, Kazemian and Ouazzane [78] have presented NF relevance based approach to the transmission of MPEG-4 video series in IEEE802.15.4 ZigBee wireless standards. Normally, ZigBee can function within the frequency range of 2.4GHz with respect to the information rate of 250kb/s, and thus impedes with the other wireless appliances such as WiFi and Bluetooth, which were operating with the similar frequency band. The variable bit rate (VBR) video has various requirements such as high bandwidth which may lead to the loss in the information and delay with respect to its time instant with an inadequate information rate due to the elevated changes in the bit rate. Subsequently, in the ZigBee channel, it was approximately unworkable for the VBR video which was to be transmitted. This approach was implemented in order to investigate both the input and output in case of accumulated information which was unconstrained with traffic adaptable buffer. Here, the input of the buffer was regulated by a NF approach which was deployed to guarantee about the amendable traffic buffer which was not flooded and starved with the video information. Similarly, the output of the amendable traffic buffer was examined by a second order NF approach which was deployed to make confirmation regarding the departure rate based on the situations of the traffic in ZigBee. From the experimental outcomes, it was obvious that this approach achieves enhanced quality with the pictures when compared with the other traditional approaches.

2.2.5 CODING APPROACHES

In 2014, Abdelali *et al.* [79] have suggested an approach depending on the identification of the moving object and its tracking behavior regarding the video series based on the characteristics of the colours. In this scheme, both the likelihood product kernels were regarded as a resemblance mesures, and it was combined with the integral images in order to calculate the histograms of all probable areas of objects which were tracked with respect to the data series. The main aim of this approach was to correlate the objects in successive video outlines. The correlation was considered to be more complicated depending on the rapid movement of the objects as compared with the frame rate. A different condition which augments the involvedness of the difficulty was considered depending on the tracking regarding the variations in the objects orientation over the time. From the experimental outcomes, it was apparent that this scheme achieves enhanced exactness regarding tracking when compared with the other traditional models.

In 2011, Schmidt and Rose [80] have investigated a source channel coding for error resilient video steaming depending on the redundant encoding technique. In this approach, the end to end distortion with respect to the encoded comprehensive chunk in the course of the expansion of the optimal pixel which was approximated in a repeated manner to include several superfluous diffusions. Moreover, three encoding approaches were also created with dissimilar gain-complexity tradeoffs. This approach was considered to be more common and could be executed on top of hybrid video codec. From the simulation outcome, it was clear that this approach attains better performance in terms of gain when compared with the other traditional error flexible encoding approaches.

In 2008, Wang *et al.* [81] have introduced three improved approaches such as running average, norm, MoG, which was exploited for the modelling of background from the compressed video series, and a dual phase partitioning scheme depending on this background representations. This approach mainly deploys coefficients based on DCT, of the chunk in order to demonstrate about the background. Moreover, it adapt the background by renewing the coefficients of

DCT. This partitioning approach was made to haul out the foreground items based on the accurateness in the pixel. Here, initially, an innovative background subtraction approach in the DCT field was subjugated in order to detect the areas of the chunks completely or moderately engaged by the foreground objects, and then pixels from these foreground chunks were categorized depending on the spatial domain. From the experimental outcomes, it was obvious that this approach attains better accuracy rate in terms of partitioning when compared with the other conventional approaches.

In 2000, Robinson and Shu [82] had developed an approach for difference-image residues in the video coding. Here, a structured spatial pattern was deployed for mapping the residue pixel standards into a quadtree configuration, which is then implied in importance order with the SPIHT approach. Thus, the classical zero tree coding (ZTC) approach based on the wavelet coefficients were substituted by the untransformed residue pixel standards. Moreover, an improved pattern based ZTC approach as well as the wavelet based ZTC was deployed to compress the codes in the errorless channels when compared with the DCT approach. Similarly, in the noisy channels, pattern-based ZTC was exploited to create flexibility in the error, thus permit it for the diffusion of the deposited data without any error control overheads. From the outcome, it was noticeable that, this scheme includes improved suppression rate than the other standard approaches.

2.2.6 MOTION ESTIMATION APPROACHES

In 2012, Chen *et al.* [83] have developed a hierarchical approach depending on the segmented area and pixel descriptors for video background subtraction. Here, an enhanced hierarchical approach depending on the background scheme was established with respect to the segmented background images. First, a mean shift approach has been deployed in order to partition the background images into various regions. Next, a hierarchical approach comprised of both area and pixel schemes were generated. The scheme based on the area was considered to be as the most significant type of approach known as accurate GMM which was attained depending on the histogram of a particular area. Similarly, the pixel scheme depends on the dissimilarities that occur during the co-occurrence of an image defined based on the histogram of oriented gradients of concerning pixels in each area. Benifits that occurs in the segmentation of background images leads to both area and pixel schemes depending on the dissimilar areas which were exploited to set various dissimilar features. Here, the pixel descriptors were estimated from the adjacent pixels in the similar entities.

In 2014, Ghahremani and Mousavinia [84] had presented an enhanced Adaptive Energy model based predictive Motion Estimation (AEME) scheme to assess an active resemblance scheme connecting the blocks and it was compared with the energy histograms. Block matching approaches were frequently deployed to evaluate the movement. Among these approaches, the predictive block matching approaches attempts to estimate the position of the finest identical chunk before the exploration of its significant synchronization. Finally, an adaptive two action search approach was established to evaluate the movement of chunk. From the simulation outcomes, it was obvious that this approach achieves better accuracy when compared with the other standard approaches.

In 2010, Li *et al.* [85] have implemented coordination for involuntarily identifying and examining composite participant activities in moving background sports video series, aspiring at action-dependent sports videos offering kinematic capacities for instructor support and performance enhancement. Normally, this configuration operates in a coarse-to-fine manner. In the central granularity point, the activity categories were identified to maintain activity-dependent video repossession and indexing. In the end of the fine granularity point, the decisive kinematic constraints of participant activities were attained for sports professional's guidance principles. On the other hand, the composite and active background of sports videos and the involvedness of participant activities convey extensive intricacy to the repeated examination. To accomplish such kind of task, robust approaches comprising global motion estimation alongwith adaptive outliers filtering, partitioning of objects depending on the creation of adaptive background, and repeated tracking of human bodies were formulated.

In 2008, Kamolrat *et al.* [86] have implemented a technique regarding the video coding. Here, its particular characteristics of the intensity based channel are exploited to compress the information with respect to the intensity. Enhancing the optimization based rate deformation in case of inter-frame calculation; Binary Partition Tree (BPT) was implemented to facilitate the adaptive segmentation of the intensity frames. From the simulation outcome, it was observed that this approach attains better enhancement in terms of segmenting the intensity based information when compared with the other approaches.

In 2009, McHugh *et al.* [87] have implemented an approach based on the foreground adaptive background subtraction with respect to the adjustment of several threshold values to modify the information regarding the video depending on the statistical approaches. The most flourishing background subtraction approaches pertain several likelihood phenomenon to deal with the background intensity developed with respect to the instant, non-parametric and mixture of Gaussian schemes. Based on the identification threshold selection, it includes involvedness in modelling robust background subtraction approaches. Additionally, other than a nonparametric background approach, a foreground approach was implemented depending on the small spatial neighborhood to enhance the discrimination sensitivity. Moreover, a Markov scheme was applied to vary the labels to enhance the spatial consistency of the identification process

In 2016, Bernal *et al.* [88] have proposed two different schemes to enhance the effectiveness in motion estimation of video sequences. First, an exceedingly competent model-independent approach was implemented that estimates the path and extent of activity regarding the objects in the scene and thus, calculates the best possible exploration path and vicinity position for activity vectors. Next, a model-dependent approach was implemented to find out the prevailing spatiotemporal characteristics of the activity based approaches which were confined in the video all the way through the statistical schemes and facilitated the minimized explorations depending on the created approaches. From the experimental substantiation, it was obvious that this approach achieves better detection rate and extent of neighborhoods when compared with the other conventional activity-based assessment approaches.

In 2015, Muthuswamy and Rajan [89] have implemented an approach to identify the prominent video objects with respect to the particle filters, which were directed by spatiotemporal prominent records and colour characteristics with the capacity to rapidly recover from fake identifications. This approach for producing both spatial and activity prominent records normally depends on evaluating the confined characteristics with respect to the prevailing characteristics in the frame. Moreover, for spatial prominent records, both the hue and the saturation characteristics were deployed. It was seen this approach achieves better activity prominent identification rate when compared to the other state of the art approaches.

In 2011, Zhang *et al.* [90] have implemented a multiple viewing approache for the segment of the foreground objects comprising of an assemblage of populace into entity based individual substances, and track them in the sequence of video. Intensity and occlusion information reconstructed from the scenes regarding the multiple viewing was incorporated into the identification of the object, segmenting the object and the tracking phenomenon. Here, the adaptive background penalty with occlusion reasoning was projected to disconnect the foreground areas from the background in the preliminary frame. Multiple indications were utilized to fragment the entity based human substances from the assemblage. To disseminate the partitioning in the course of video, each object area was autonomously followed by motion compensation and unceranity refinement, and the occlusion depending on the motion was attempted as conversion with respect to the level. From the experimental outcomes, it was apparent that this approach attains better performance in terms of effectiveness when compared with the other state of art approaches.

In 2008, Zhao *et al.* [91] have offered an unequal error protection approach known as an adapted Perceived Motion Energy (PME) scheme for wireless H.264 video transmission. Here, the unequal protection of error on the transmission of video was extensively deployed to contest with bit errors in the wireless channel.

Nevertheless, contemporary unequal protection model models with respect to the heuristic phenomenon as well as the distinctiveness of human visual system were not taken into description. Depending on the susceptible features related to the video activities of human eyes, this enhanced approach was considered for performing the encoding process with respect to the characteristics of H.26 4/AVC. In this approach, the bit streams in the video were partitioned into various eminent layers, and thus, the asymmetrical error fortification was intended to defend the transmission of video it steams over the wireless channels. From the experimental outcomes, it was obvious that this approach attains better performance in terms of enhanced quality in transmitting the video when compared with the other standard approaches.

In 2010, Han *et al.* [92] have presented an approach depending on the single frame interpolation and multiple frame interpolation. In this approach, the representation in terms of the attributes regarding the activity in the video sequence was investigated. Subsequently, the representations depending on the activities were customized to minimize the calculation and system complexity. Finally, Kalman filtering approach was exploited to interpolate the image vigorously to achieve high resolution.

In 2013, Lijun and Kaiqi [93] have presented a video dependent crowd density estimation approach and prediction networks for the applications related to the extensive locale surveillance. In monocular visual images, the Accurate Mosaic Image Difference approach was exploited for the extraction of crowded regions with asymmetrical movement. Here, the number of individuals and swiftness of a crowd can be effectively approximated by this network depending on the compactness of crowded regions. Based on the multiple camera networks, the calculations of density of crowd were attained, quite a few minutes prior.

In 2002, Mikolajczyk and Schmid [94] had implemented an affine invariant interest points. This approach includes three suggestions such as; first, it exploits a second-moment matrix which was estimated in a particular direction which was again deployed for the regularization of a particular area in this approach. Next, the magnitude of the neighboring configuration was specified by several confined extrema of regularized derivatives with respect to the magnitude. Finally, an affine adapted based Harris detector was used to establish the position of interest points. Here, for the initialization process, a multi-scale version of detector was deployed. In case of identification and mapping an image, series of affine invariant points were considered. Also an affine based conversion approach was also correlated with this approach. From the experimental analysis, it was apparent that this approach attains better identification rate in case of various deformations in the invariant affine points as well as the conversion rate when compared with the other standard identification approaches.

In 2005, Dollar *et al.* [95] have implemented an undeviating 3D matching part was frequently deployed with respect to the 2D interest point detectors which were insufficient, and thus an unusual approach was employed. For securing these interest points, an identification approach depending on the spatiotemporal characteristics was deployed with a better rate.

In 2009, Seshadrinathan and Bovik [96] have developed an approach based on the video eminence indicator which was referred as MOVIE indicator which was deployed to incorporate both the temporal and the spatial characteristics regarding the distortion consideration. In this approach, movement plays an imperative responsibility in the human perception of videos and thus, it experiences from various objects that have to be compacted with the erroneousness in the illustration of movement in the test video compared to the oriented video. This approach unambiguously exploits information with respect to the movements from the oriented video and estimates the eminence of the assessment video depending on the movement in the oriented videos. From the experimental analysis, it was clear that this approach attains better performance in terms of the objects present in the video with better rate when compared with the other standard approaches.

In 1994, Koller *et al.* [97] have developed an approach for examining the traffic-related prospects, which is an essential component of Intelligent Vehicle Highway Systems (IVHS). The information regarding the traffic scenes was deployed to optimize the flow of traffic throughout the hectic periods and thus detect the delayed vehicles and accidents. Moreover, it assists in the creation of assessments in terms of an independent vehicle regulator. Various enhancements in this technology with respect to the machine vision based visualization and elevated point emblematic interpretation were exploited to implement a network based on the comprehensive, consistent examination of traffic scenes. The machine vision based approach network mainly utilizes a shape tracker and an affine movement approach depending on the Kalman filters to acquire the routes of vehicle over a traffic scene in an image series. The symbolic analysis constituent mainly deploys a dynamic belief network to create presumptions regarding the traffic measures including the variations in the path of the vehicle and stalls. Here, the key assignments were conferred depending on the visualization and analytic mechanisms, as well as their incorporation into an operational model.

2.3 RESEARCH GAPS AND ISSUES

There have been a lot of attainments in the research area of VS techniques in case of traffic, though there were still some issues that desire to be addressed for this technology. The expedition for the enhanced traffic information includes an improving the dependence in case of the traffic surveillance and thus has resulted in a requirement for enhanced identification of vehicles, but, due to the elevated outlays and security threats, there arises various issues in the traffic surveillance and thus have to be engaged in the exploration towards the in-persistent detection techniques.

In smartcity concept, wherein ITS is a vital component, video based detection system is the core of all.

The rapidly diminishing outlay in case of the image attainment procedures and the accessibility depending on the inexpensive, as well as the authoritative central processing units, have generated various concerns in the exploration of the computer vision approaches for the supervision and controlling of traffic purposes. The supervision of crossroads pretences several difficulties in terms of highways, which are associated to the decidedly changeable configuration of the crossroads, and also the existence of the numerous flows of the vehicles depending on the turning movements and the assorted traffic ranges leads to the impediment of the vehicles at the traffic signals. Moreover, detailed classification and occlusion based supervision approaches are necessary.

Further, there are millions of cameas are installed for various surveillance reasons, and incrase is presumed in days to come. In this scenario, it is challenging to send video data from cameras to control server. Therefore, it is essential to have ‘machine learning on the edge’. Camera need to do some intelligent local processing and send ‘data of interest’, which is small in amount, to the server or cloud in real time.

Hence, understanding the activities of objects in a scene by the use of video is both a challenging scientific problem and a very fertile domain with many promising applications. Thus, it draws attention of several researchers, institutions and commercial companies.

2.4 SUMMARY

In this chapter, various existing methods were formulated for the object detection in traffic surveillance technology. This chapter surveys a number of existing methods for the classification of traffic surveillance system based on coding, similarity measurement, optimization based, learning based, motion estimation and video streaming techniques. Several object detection techniques in traffic surveillance system were deployed in all the works to enhance the performance of the traffic surveillance coding, similarity measurement, optimization, learning, motion estimation and video streaming approaches. Thus in traffic surveillance system, object detection was consistently the eventual objective. Here, the approaches used for the object detection and classification in traffic surveillance system were clearly described, the performance of different object detection and classification techniques based on video object segmentation, ANN, GMM, spatiotemporal correlations and interest point matching approach were analyzed, and thus the benefits and drawbacks of these techniques were also described. From the existing works, the approaches for the detection of objects in traffic surveillance can be classified into six categories, namely similarity measurements, video streaming, optimization based, learning based, VS coding and motion estimation. From this insight, it can be concluded that there is a scope for further improvements in all classified categories to make them more efficient and accurate. This may support the VS system for their real time application, which is a backbone of any smart city infrastructure.

CHAPTER 3. MOVING OBJECT DETECTION IN SURVEILLANCE

3.1 INTRODUCTION

In this chapter, a novel model for BS of remote scene monitored by a camera that is static or rotating is presented. Accordingly, Adaptive Gaussian Mixture Model (AGMM) framework is exploited to approximate the BG design. The allocation of every BG pixel is spatially and temporally modeled. Depending on the arithmetical representation, a pixel in the present frame is categorized as belonging to the BG or FG. This enhanced BG model achieves better results in detecting moving object. Also, the implemented approach can efficiently manage the outdoor scenes. Sample videos from real surveillance system were ingresses in the BG model to detect moving object, particularly car as an object. For purpose of training the model, a real time video sequence obtained from static and rotating camera are injected in to the system. Due to major differences in the operating parameters between static and rotating camera, the model derived for static camera cannot directly fit in rotating camera use case. Hence, it has to be enhanced to be made compatible with the rotating camera ecosystem.

3.2 PROPOSED MODEL

3.2.1 BACKGROUND MODELLING

The implemented scheme exploits the AGMM formulation which has been introduced by Kaew and Bowden [98]. Accordingly, the suggested scheme is more precise and could be trained rapidly. In fact, AGMM permits multimodal BG modeling, and continuous updating of the background with respect to varying condition of shadow and lighting. Initialisation of building BG model with AGMM can be described as follows:

In this approach, value of every pixel in the given frame is computed from Eq.(3.1 and 3.2). This can be termed as model developement by a mixture of H Gaussian distributions. This can be further elaborated as follows:

The pixel distribution is symbolized by a mixture of H Gaussians as given by Eq. (3.1).

$$g(I_t = y) = \sum_{j=1}^{k} f_{j,t} \eta(y, \mu_{j,t}, \sigma_{j,t}) \tag{3.1}$$

where, $f_{j,t}$ indicates the weight parameter of the j^{th} Gaussian component, $\eta(y, \mu_{j,t}, \sigma_{j,t})$ denotes the normal distribution of j^{th} Gaussian component with standard deviation $\sigma_{j,t}$ and intensity mean $\mu_{j,t}$ and $g(I_t = y)$ is the probability of a pixel has a value of y at time t.

In general, the value of H alters from three to five on the basis of obtainable storage and computational power. The H distributions are well-organized on the basis of the fitness value $\frac{f_k}{\sigma_k}$. Initially, M distributions are deployed as a model of the background scene, in which, M is approximated as given in Eq. (3.2), where T denotes the threshold for the minimum fraction of the BG model.

$$M = \arg_m \min\left(\sum_{j=1}^{m} f_j > T\right) \tag{3.2}$$

Moreover, BG subtraction is carried out by pointing out a FG pixel that is present away from any of the M distributions by higher than 2.5 standard deviations. The updating of the constraints that are matched is made with the subsequent updated forms as given by Eq. (3.3), Eq. (3.4) and Eq. (3.5).

$$\hat{f}_j^{t+1} = (1-\alpha)\hat{f}_j^{t} + \alpha \hat{X}(\omega_j / y_{t+1}) \tag{3.3}$$

$$\hat{\mu}_j^{t+1} = (1-\alpha)\hat{\mu}_j^{t} + \beta y_{t+1} \tag{3.4}$$

$$\hat{\sigma}_j^{t+1} = (1-\alpha)\hat{\sigma}_j^{t} + \beta\left(y_{t+1} - \hat{\mu}_j^{t+1}\right)\left(y_{t+1} - \hat{\mu}_j^{t+1}\right)^T \tag{3.5}$$

Where,

$$\beta = \alpha\eta\left(y_{t+1}, \hat{\mu}_j^{t}, \hat{\sigma}_j^{t}\right)$$

$$\hat{X}\left(\omega_j / y_{t+1}\right) = \begin{cases} 1, \text{if } \omega_j \text{is first Gaussiancomponent} \\ 0, \text{if not.} \end{cases}$$

$$\frac{1}{\alpha} = \text{Time constant}$$

If no distributions equal the pixel value, then a distribution with high variance, low weight, and current value as its mean are replaced in the place of least significant component of the mixture representation.

3.2.2 DETECTING FOREGROUND OBJECTS

The objective is to set out a highly performant system of OD. Accordingly, in this chapter, a perspective transform [17] is applied from four relative points from every two consecutive frames. With this, we achieve shifting of an object from one coordinate of the frame to the coordinate of the frame translation matrix. This is necessary for the compensation of camera motion [5, 115].This transformed frame is taken away from the preceding frame to obtain the detected moving objects. In the subsequent phase, an arithmetical BG model is constructed. The corresponding BG model depends on Bowden's and Kaew Trakulpong algorithm. At first, the BG model for N frames is built up that includes one entire rotary motion of the camera. Moreover, the value of N is evaluated, from cameras fps. According to arithmetical representation (Eq.3.2), a pixel in the present frame is categorized as belonging to the BG or FG with respect to respective BG model. Explained approach is illustrated in Fig. 3.1.

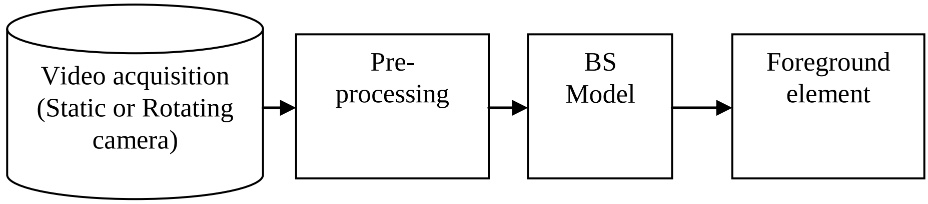


Figure 3-1:Architecture for Foreground Detection.

3.3 EXPERIMENTS BASED ON PROPOSED MODEL

3.3.1 PROCEDURE

This project requires real-time investigation of the video stream for detection of objects. For experimental setup, the free accessible Open CV-library that was implemented in C and C++ code has been exploited. The sequence revealed here are 352x272 images. In addition, an adaptive combination of five Gaussian components was exploited.

3.3.2 EVALUATION METRICS

The performance measures such as accuracy, sensitivity, specificity, precision, FPR, FNR and FDR are evaluated for the proposed AGMM and conventional model. The definitions and formulations of the measures are described below.

Accuracy: It is defined as "weighted arithmetic mean of Precision and Inverse Precision (weighted by Bias) as well as a weighted arithmetic mean of Recall and Inverse Recall (weighted by Prevalence)". In Eq. (3.6), TP indicates the true positive, TN denotes true negative, FP signifies false positive and FN implies false negative.

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \tag{3.6}$$

Sensitivity: It is defined as "the study of how uncertainty in the output of a model can be attributed to different sources of uncertainty in the model input". Eq. (3.7) reveals the formulation of sensitivity.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{3.7}$$

Specificity: It is defined as "the ability of a test to preciously identify foreground pixels which are true positive."

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3.8}$$

Precision: It is defined as "the probability that a (randomly selected) retrieved document is relevant".

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.9}$$

FPR: It is calculated as "the ratio between the number of negative events wrongly categorized as positive (false positives) and the total number of actual negative events (regardless of classification)".

$$\text{FPR} = \frac{FP}{FP + TN} \tag{3.10}$$

FNR: It is defined as "the proportion of positives which yield negative test outcomes with the test, i.e., the conditional probability of a negative test result given that the condition being looked for is present".

$$\text{FNR} = \frac{FN}{TP + FN} \tag{3.11}$$

FDR: It is defined as "the expected proportion of rejected hypotheses that are mistakenly accepted".

$$\text{FDR} = \frac{FP}{TP + FP} \tag{3.12}$$

Here, accuracy, sensitivity, specificity and precision are considered as positive measures. Increase in accuracy refers to the increase in the better performance of the proposed model. Accordingly, increase in sensitivity refers to understand how the proposed algorithm matches with information provided by direct observation without wrongly identifying the foreground. Increase in specificity refers to the detection of negative proportions that are correctly identified. These metrics i.e. accuracy, sensitivity, specificity and precision have to be high for improved performance of the system. The FPR, FNR and FDR are considered as the negative performance measures. FPR usually refers to the expectancy of the false positive ratio. FNR refers to the conditional probability of a negative test result given that the condition being looked for is present. FDR refers to conceptualizing the rate of type I errors in null hypothesis testing when conducting multiple comparisons. These measures, i.e. FPR, FNR and FDR have to be low for better performance of the system.

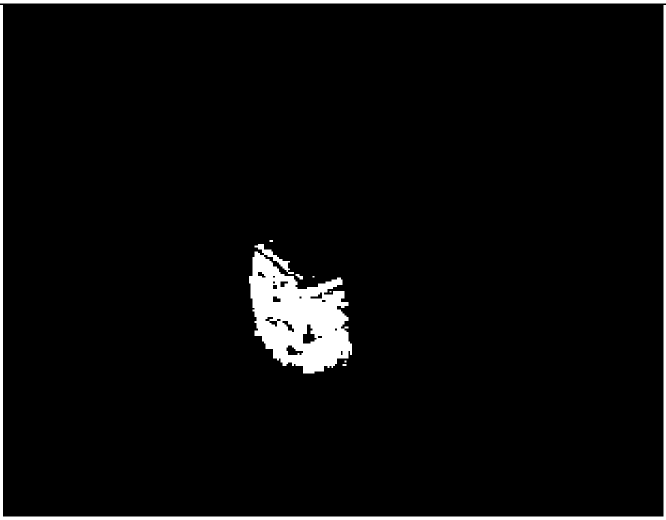
3.3.3 COMPARISON TECHNIQUES

The proposed Adaptive Gaussian Mixture Model (AGMM) was compared with conventional adaptive statistical model and the enhanced outcomes of the suggested scheme were proved from the simulation results. Since the conventional Model [108] relies on the statistical information of each pixel, it fails to detect moving objects accurately in certain datasets. On the other hand, the implemented AGMM algorithm detects the moving objects accurately in all datasets.Moreover, the implemented AGMM approach is robust in opposition to illumination variations and it is also robust against noise factors. In addition, the implemented AGMM presents a novel and practical choice for intelligent video surveillance systems using static cameras and rotating cameras and the results were attained.

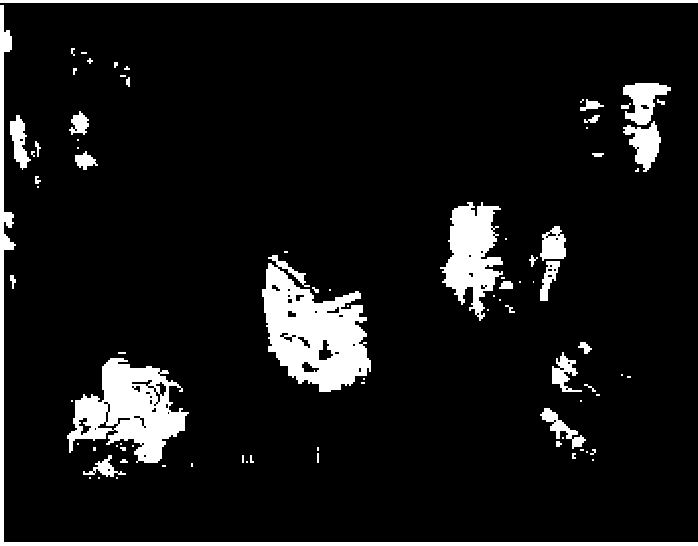
3.3.4 STATIC CAMERA

Fig. 3.2 demonstrates the FG segmentation outcomes with of static camera by deploying perspective transform and adaptive statistical mixture representation.

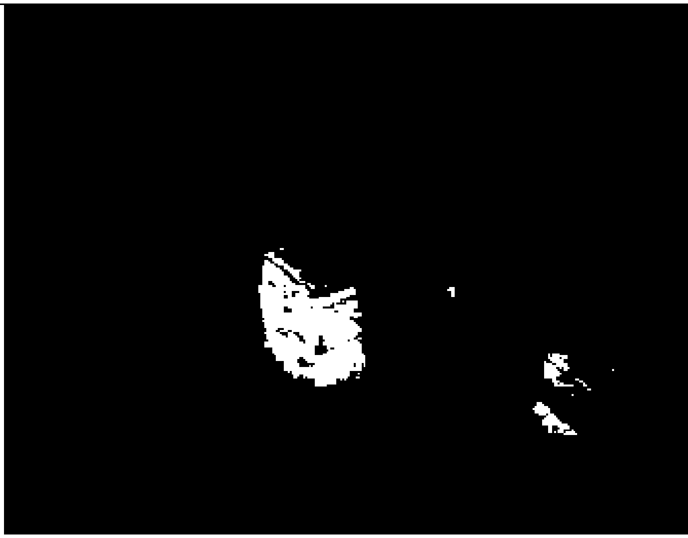
(a)



(b)



(c)



(d)

Figure 3-2: Sample image (a) Original image (b) Ground truth image (c) Conventional image (d) Proposed image (Static camera)

3.3.5 PERFORMANCE ANALYSIS (STATIC CAMERA)

The performance measures of the proposed methodology in terms of positive measures such as accuracy, sensitivity, specificity, and precision is demnonstrated by Fig. 3.3 for static camera. From Fig. 3.3, the adopted scheme is 5% better in terms of accuracy, 5% better in terms of specificity and 33.3% better in terms of precision when distinguished with the conventional techniques.

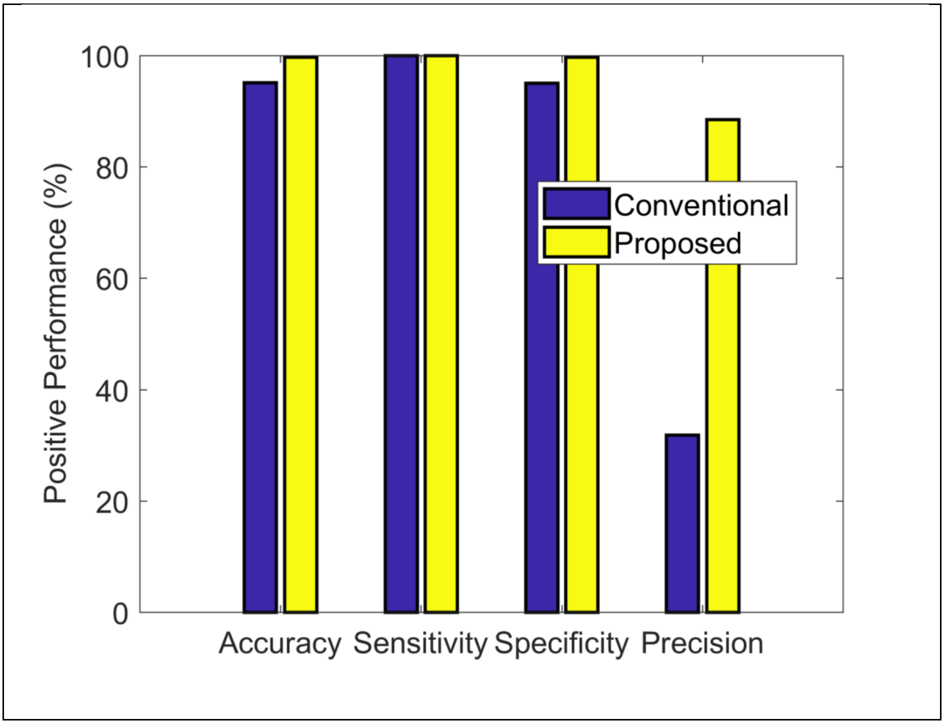


Figure 3-3: Performance analysis of the proposed scheme and conventional scheme in terms of positive measures for static camera

Likewise, the performance of the suggested methodology with respect to the negative measures is specified by Fig. 3.4, in which the FPR of the implemented method is 80% superior when compared with the traditional model and FDR of the suggested scheme is 82.35% superior when distinguished with the conventional approach. Lower negative performance (FPR, FNR and FDR) is reflection of better performance. At the same time, higher positive performance (accuracy, sensitivity, specificity and precision) is desirable as far as performance of the system is concerned. Thus the enhanced outcomes of the proposed scheme have been confirmed by the simulation results.

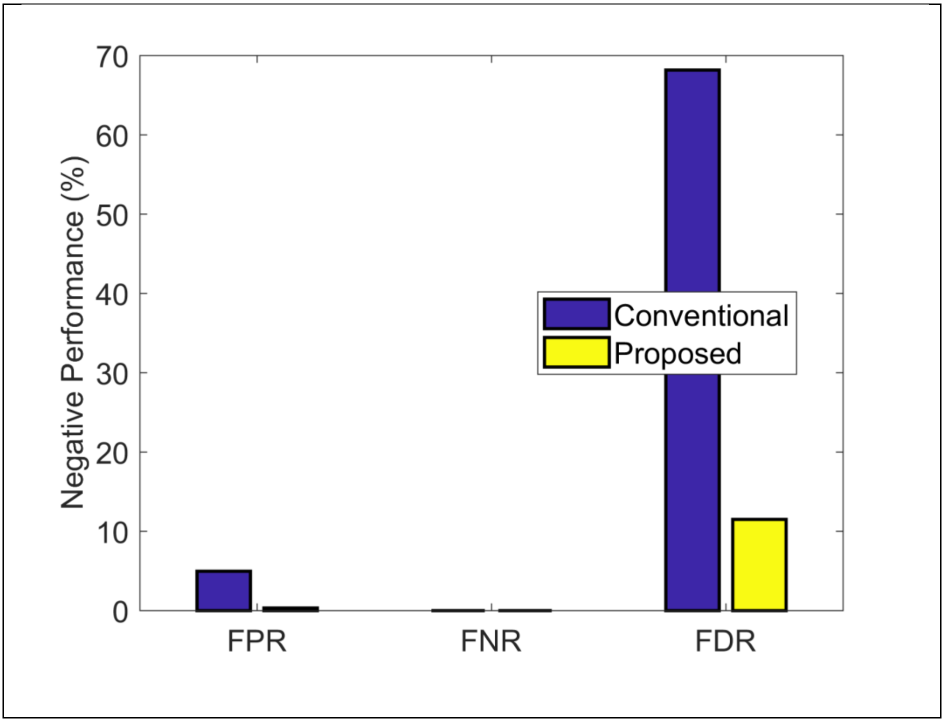
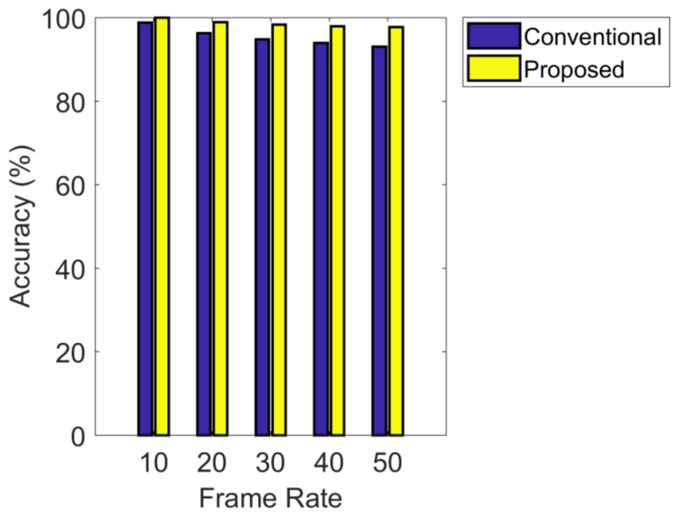
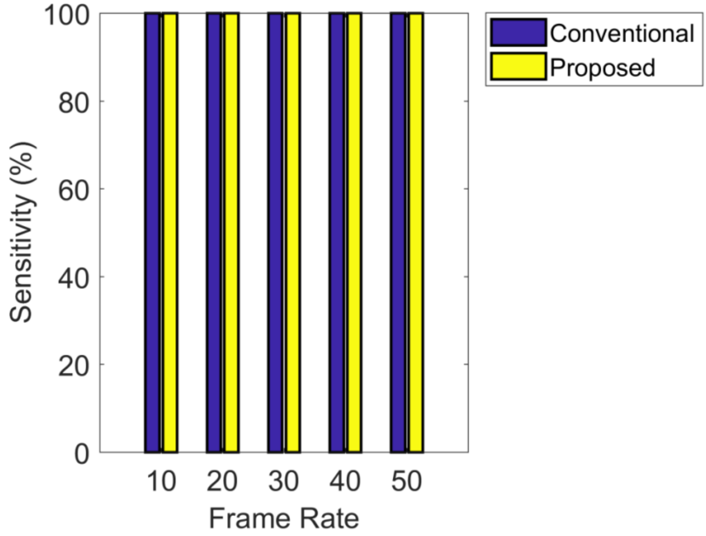


Figure 3-4: Performance analysis of the proposed scheme and conventional scheme in terms of negative measures for static camera

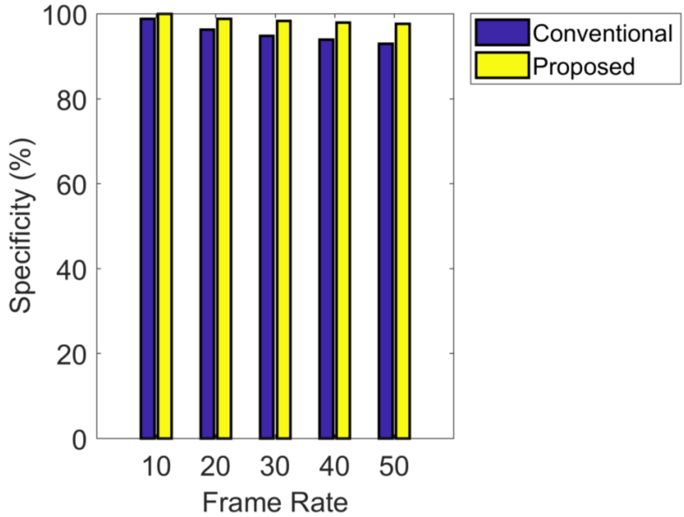
The accuracy of the suggested scheme for varying frame rates such as 10, 20, 30, 40 and 50 can be obtained from Fig. 3.5(a), where the proposed method is 2% better for 10^{th} frame rate, 4.21% better for 20^{th} frame rate, 2.08% better for 30^{th} frame rate, 2.10% better for 40^{th} frame rate and 2.10% better for 50^{th} frame rate. Similarly, from Fig. 3.5(c), the specificity of the introduced scheme can be attained which is 2% , 3.15%, 1.03%, 2.10% and 2.10% superior for 10^{th}, 20^{th}, 30^{th}, 40^{th} and 50^{th} frame rates, respectively. Also from Fig. 3.5(d), the presented scheme in terms of precision is 38% superior to conventional scheme for 10^{th} frame rate and 52.63% superior to conventional scheme for 20^{th} frame rate. In addition, from Fig. 3.5(e), the FPR of the suggested model can be obtained, which is 99%, 73.68%, 20% and 64.28% better than conventional model for 10th, 20th, 30th and 40th frame rates correspondingly. Also, FDR of the suggested scheme is superior to the traditional algorithm by 62%, 39.34%, 28.57%, 21.62% and 15.38% for 10th, 20th, 30th, 40th and 50th frame rates correspondingly. Thus the enhanced performance measures of the static camera can be attained from the execution results.



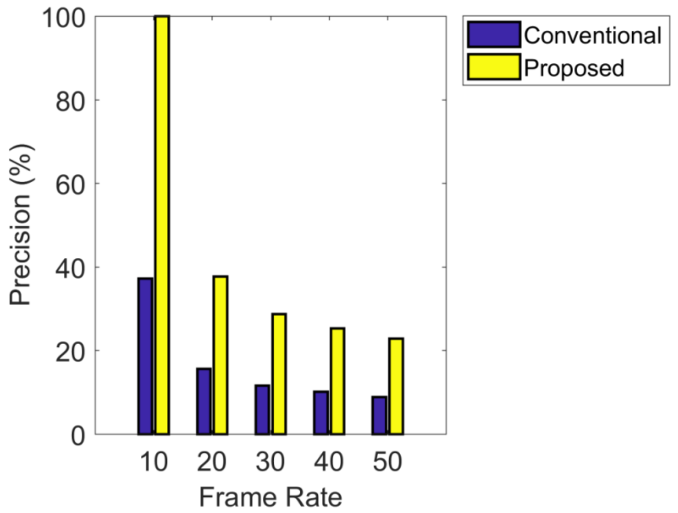
(a)


100
80
60
40
20
0
Sensitivity (%)
10
20
30
40
50
Frame Rate
Conventional
Proposed

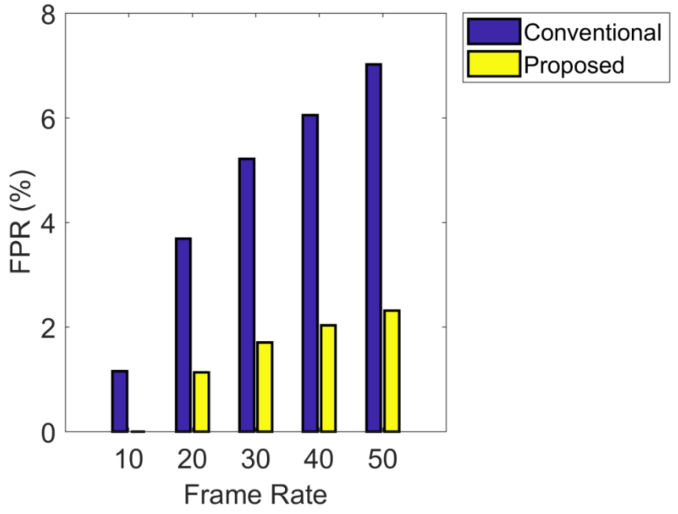
(b)


100
80
60
40
20
0
Specificity (%)
10
20
30
40
50
Frame Rate
Conventional
Proposed

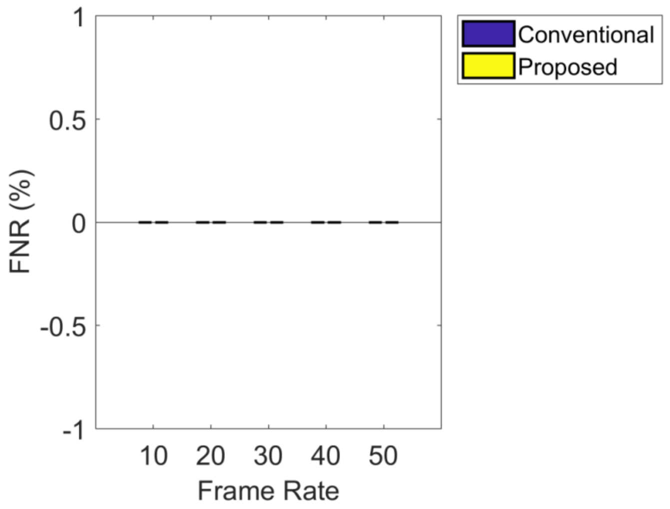
(c)


100
80
60
40
20
0
Precision (%)
10
20
30
40
50
Frame Rate
Conventional
Proposed

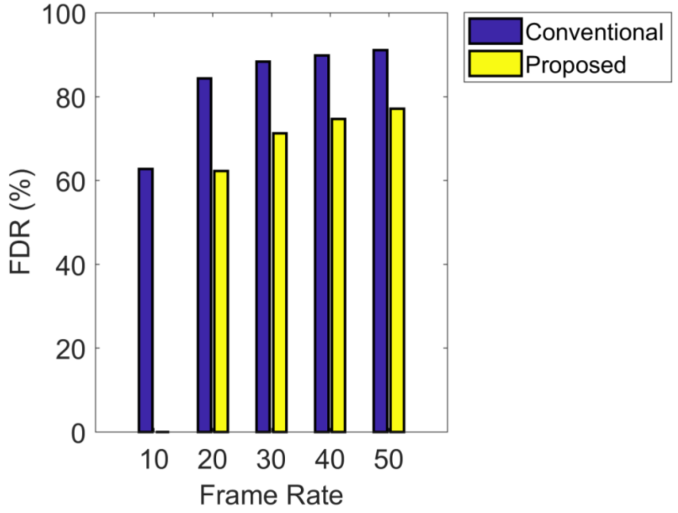
(d)


8
6
4
2
0
FPR (%)
10
20
30
40
50
Frame Rate
Conventional
Proposed

(e)



(f)



(g)

Figure 3-5: Experimental analysis of the proposed approach for static camera in terms of (a) Accuracy (b) Sensitivity (c) Specificity (d) Precision (e) FPR (f) FNR (g) FDR

3.3.6 ROTATING CAMERA

The MV detection has been a major challenge in image processing and gained a lot of attention in the recent years. However, if the camera is non-stationary, it becomes difficult task since the image motion is due to combined effects of object motion and camera motion. Therefore, the algorithms helpful for stationary camera cannot be employed when camera is rotating.

Fig. 3.6 demonstrates the FG segmentation outcomes with motion compensation by deploying perspective transform and adaptive statistical mixture representation for rotating camera.

(a)

(b)

(c)

(d)

Figure 3-6: Sample image (a) Original image (b) Ground truth image (c) Conventional image (d) Proposed image (Rotating camera)

3.3.7 PERFORMANCE ANALYSIS (ROTATING CAMERA)

The performance measures of the proposed methodology in terms of positive measures such as accuracy, sensitivity, specificity, and precision is exposed by Fig. 3.7. From Fig. 3.7, the implemented scheme with respect to accuracy and specificity is 17.34% better than conventional approach and 90% superior to conventional scheme in terms of precision.

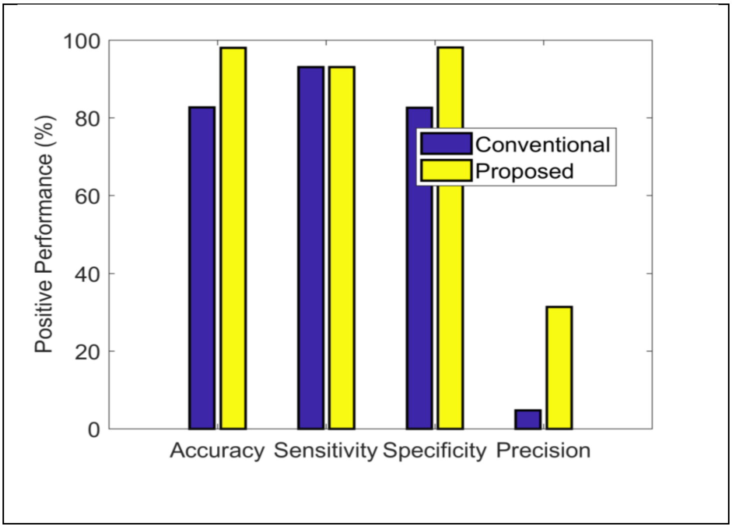


Figure 3-7: Positive measures of the proposed model for Rotating camera

Similarly, the performance of the implemented methodology in terms of negative measures is given by Fig. 3.8, where the FPR of the suggested scheme is 89.47% better than conventional approach, and on considering FDR, the adopted scheme is 37.14% superior to the traditional scheme. Thus the enhanced computation of the proposed approach has been validated successfully.

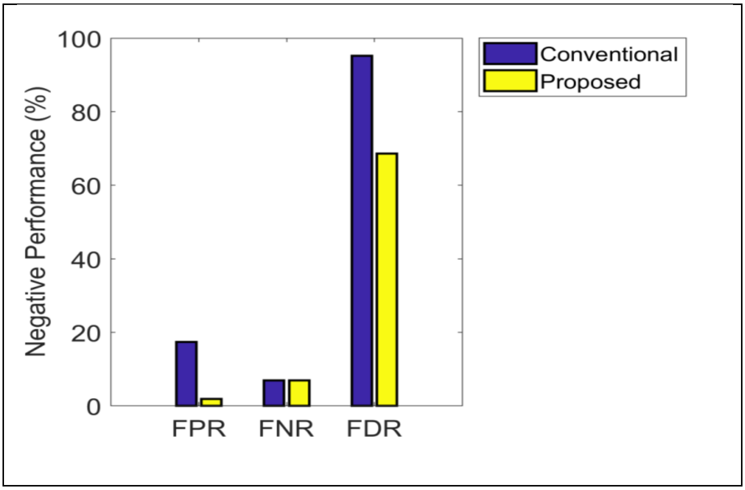
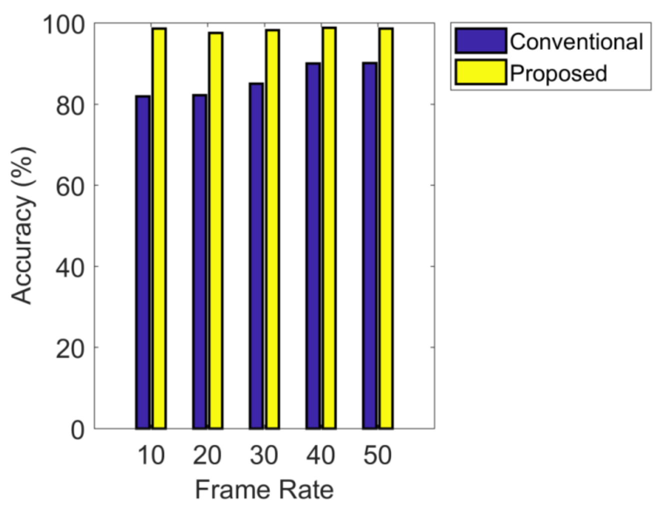
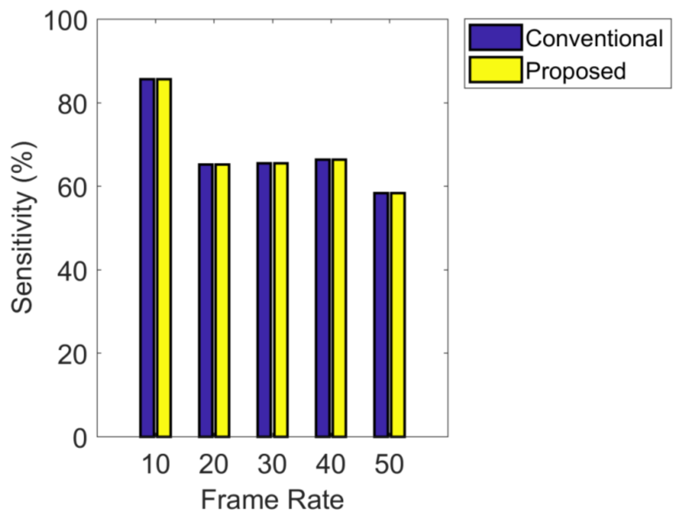


Figure 3-8: Negative measures of the proposed model for Rotating camera

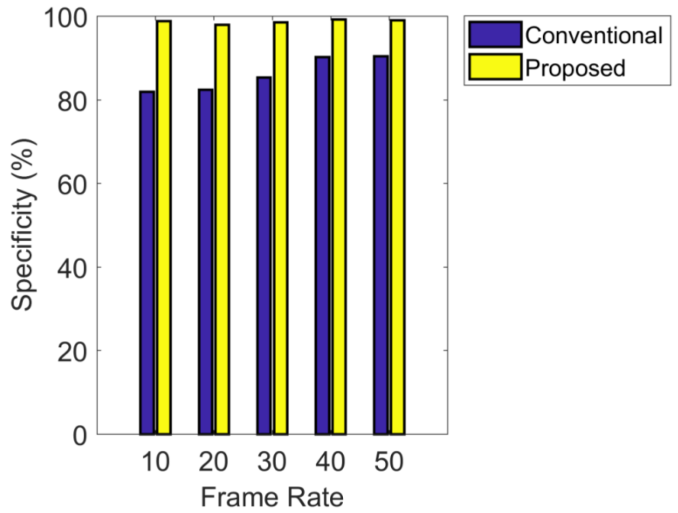
The overall performance measures of the proposed scheme with respect to variation in frame rate (10, 20, 30, 40 and 50) are given by Fig. 3.9. The analysis was scrutinizedin terms of positive measures such as accuracy, sensitivity, specificity and precision and negative measures such as FPR, FNR and FDR. From Fig. 3.9 (a), the proposed methodology for frame rate 10 is 19.19% better, frame rate 20 is 17.53% better, for frame rate 30 is 15.31% better, for frame rate 40 is 9.09% better and frame rate 50 are 9.09% better than the conventional approach. Similarly, from Fig. 3.9(b), the implemented scheme in terms of sensitivity can be attained, in which the suggested and conventional approaches are found to have the similar values for all the considered frame rates. In addition, from Fig. 3.9(c), the analysis of adopted scheme in terms of specificity was attained, which is 19.19% superior for frame rate 10 , 17.53% superior for frame rate 20, 15.31% superior for frame rate 30, 9.09% superior for frame rate 40 and 9.09% superior for frame rate 50 when compared with the conventional approach. Moreover, from Fig. 3.9(d), the analysis of presented scheme with respect to precision was achieved, which is 88.88% better for frame rate 10 , 84% better for frame rate 20, 85.29% better for frame rate 30, 87.5% better for frame rate 40 and 82.5% better for frame rate 50 when distinguished with the traditional scheme. From Fig. 3.9(e), the investigation of offered scheme with respect to FPR was achieved, which is 88.88% better for frame rate 10 , 85.71% improved for frame rate 20, 85.33% enhanced for frame rate 30, 90% superior for frame rate 40 and 81.82% improved for frame rate 50 when distinguished with the conventional design. The FNR of the suggested scheme can be attained from Fig. 3.9 (f), which is found to be similar for conventional and adopted techniques for all the frame rates. Finally, from Fig. 3.9(g), the analysis of implemented scheme with respect to FDR was achieved, which is 72.73% better for frame rate 10 , 28% better for frame rate 20, 41.17% better for frame rate 30, 42.11% better for frame rate 40 and 52.38% better for frame rate 50 when compared with the traditional scheme.


100
80
60
40
20
0
Accuracy (%)
10
20
30
40
50
Frame Rate
Conventional
Proposed

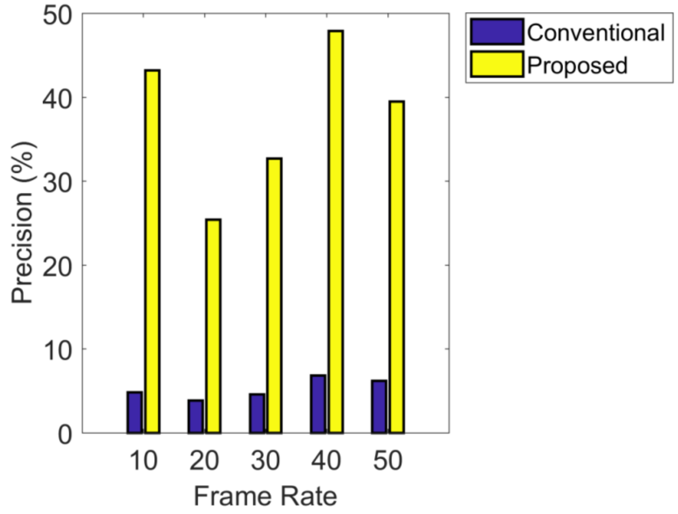
(a)


100
80
60
40
20
0
Sensitivity (%)
10
20
30
40
50
Frame Rate
Conventional
Proposed

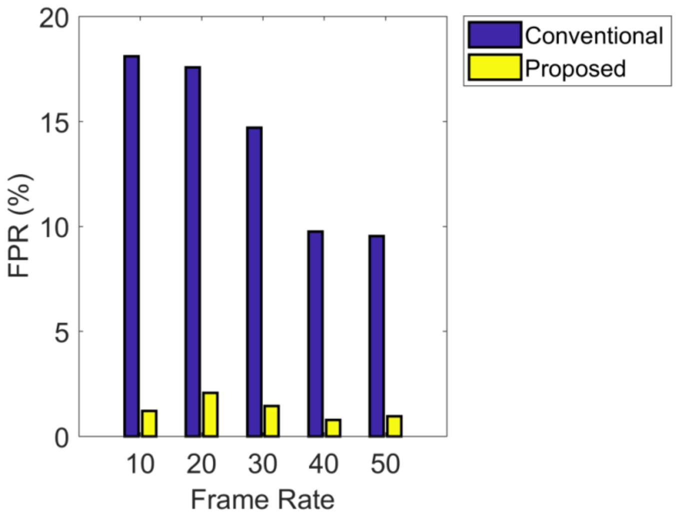
(b)


100
80
60
40
20
0
Specificity (%)
10
20
30
40
50
Frame Rate
Conventional
Proposed

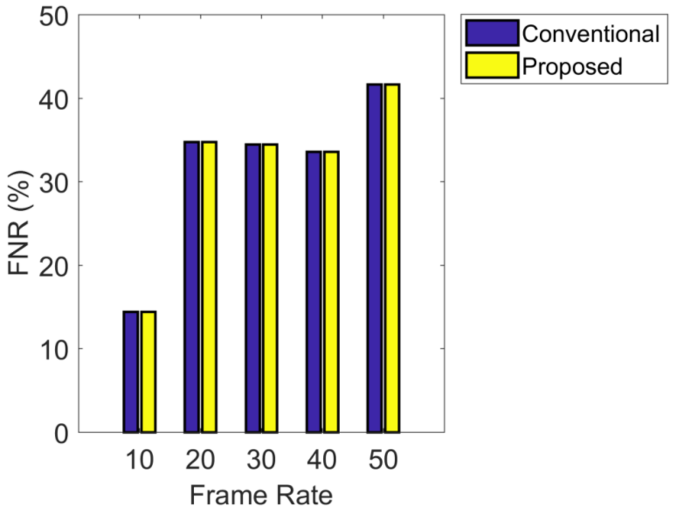
(c)


50
40
30
20
10
0
Precision (%)
10
20
30
40
50
Frame Rate
Conventional
Proposed

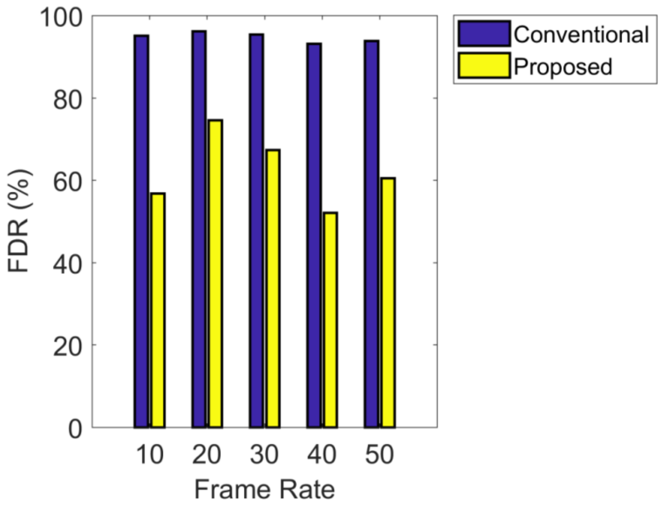
(d)


Conventional
Proposed
FPR (%)
Frame Rate
0
5
10
15
20
10
20
30
40
50

(e)


Conventional
Proposed
FNR (%)
Frame Rate
0
10
20
30
40
50
10
20
30
40
50

(f)



(g)

Figure 3-9: Experimental analysis of the proposed approach for Rotating camera in terms of (a) Accuracy (b) Sensitivity (c) Specificity (d) Precision (e) FPR (f) FNR (g) FDR

It is desirable to have lower negative performance (FPR, FNR and FDR) for better performance. At the same time, higher positive performance (accuracy, sensitivity, specificity and precision) is desirable as far as performance of the system is concerned

3.4 SUMMARY

In this chapter, a statistical background subtraction technique had been implemented for detection of moving object that considers motion of rotating camera. In conventional approach, the results of static camera in motion detection were found to be enhanced while the results of rotating camera were not found to be encouraging. Our proposed model, for static and rotating camera has been proven to be the better option than the state of the art. This can be inferred from the key performance indicators derived from the experimental results. Even though, the proposed model has better outcomes, it is evident from Fig. 3.2, Fig. 3.3 and Fig. 3.6, Fig. 3.7, the motion detection of rotating camera is complex issue to address. The dataset deployed was a video clip, captured from static as well as rotating camera. The experimental results are compared with frame differencing scheme, and it was proved that the results were more precise than conventional ones. There is still scope for further improvements in terms of accurate moving object detection for the sequences arriving from rotating camera.

CHAPTER 4. ENHANCED OBJECT DETECTION BASED ON FULL SEARCH BLOCK MATCHING ALGORITHM (FS-BMA)

4.1 INTRODUCTION

There is a modern trend to execute VC systems for video surveillance. A variety of coding methods had been implemented to improve evaluation accuracy in development of VC. As the traditional coding techniques have certain limitations, while optimizing image processing for moving bodies when the BG is characteristically motionless. The issue becomes more critical in case of rotating cameras. The limitation in this case is to separate the dynamic objects when the BG rotates at a predetermined velocity. This chapter provides a technique for the enhancement of error-free coding in VS by deploying the rotating cameras. Moreover, a least mean estimator scheme is exploited and the intermittent entire search motion estimator logic is described for the calculation of FG moving constituents from the video sequence that arrives from a rotating sensor. The advantages which are obtained from this technique provide more precise recognition of the definite moving object, thus, minimizing the data redundancy by eradicating the unnecessary BG data. Also, this edge processing reduces the burden of transmission of this intelligence to central control center.

4.2 PROPOSED METHOD

4.2.1 DE-NOISING USING LMS ALGORITHM

The LMS algorithm is an adaptive technique that exploits a gradient-dependent approach of steepest decent. The LMS technique utilizes the approximate of the gradient vector from the accessible information. LMS integrates an iterative course of action, which makes consecutive improvements to the weight vector in the direction of the gradient vector, which is negative that ultimately leads to the minimum MSE [99]. When distinguished with other various schemes, the LMS algorithm is comparatively uncomplicated; it does not necessitate the correlation function computation nor does it necessitate matrix inversions. Thus from the technique of steepest descent, the weight vector formulation is specified as in Eq. (4.1), where, μ denotes the step-size constraint and it manages the convergence features of the LMS approach, $e^2(m)$ points out the MSE among the output $y(m)$ and the required output, as specified by Eq. (4.2).

$$U(m+1) = U(m) + \mu\left[-\nabla\left(E\left\{e^2(m)\right\}\right)\right] \tag{4.1}$$

$$e^2(m) = \left[d(m) - w(m)x^T(m)\right]^2 \tag{4.2}$$

The gradient vector in the abovementioned weight update formulation can be evaluated as in Eq. (4.3), where, R signifies an autocorrelation of input signal $x(m)$and r indicates a cross-correlationamong the required input and response.

$$\nabla\left(E\left\{e^2(m)\right\}\right) = 2RU(m) - 2r \tag{4.3}$$

In the technique of steepest descent, the major crisis is the calculation concerned in discovering the values of r and R matrices in real time. The LMS approach eases this crisis by deploying instantaneous values as given by Eq. (4.4) and Eq. (4.5). As a result, the weight update formulation can be specified by the subsequent Eq. (4.6).

$$R = x(m)x^T(m) \tag{4.4}$$

$$r = d(m)x(m) \tag{4.5}$$

$$U(m+1) = U(m) + \mu x(m)\left[d(m) - x^T(m)U(m)\right] \tag{4.6}$$

$$= U(m) + \mu x(m)e(m)$$

The LMS approach is instigated with a random value $U(0)$ for the weight vector at $m = 0$. The consecutive modifications of the weight vector ultimately show the way to the minimum value of the MSE. Thus, the LMS approach can be briefed in the subsequent Eq. (4.7), Eq. (4.8) and Eq. (4.9).

$$y(m) = U^T x(m) \tag{4.7}$$

$$e(m) = d(m) - y(m) \tag{4.8}$$

$$U(m+1) = U(m) + \mu x(m) e(m) \tag{4.9}$$

This estimated weight offers an optimal value for noise removal. Over this de-noised video sample, a novel ME scheme is implemented. This scheme is an expansion to the FS-BMA formulation.

4.2.2 MOTION PREDICTION

The ME and compensation method has been extensively exploited in video compression owing to its potential of minimizing the temporal redundancies among the frames. Majority of the techniques introduced for ME until now are block-dependent methods, said to be the BMA. According to this method, the present frame is split into a predetermined size of blocks, and subsequently, each block is evaluated with candidate blocks in a reference frame which is present in the exploration areas [40][101] [100] [106]. The extensively utilized techniques for the BMA are the FSBMA that scrutinizes the entire candidate blocks contained by the exploration area in the reference frame to acquire a MV. The MV can be described as the displacement among the blocks in the present frame and the most excellent matched block in the reference frame in vertical and horizontal directions. The ME approach is carried out with an inconsistent size of search area based on block varieties changing from an 8x8 block to the entire frame. The video sequences for VC applications at low bit-rate like, video-conferencing and videophone have various restrictive motion descriptions. A block in a particular region in the preceding frame can belong to the similar region at that location in the current frame; in other words, a block in the BG province may stretch out in the BG area in the present frame.

The varying block demonstrates the percentage of the dissimilarity from the BG to the active region or vice versa. The erstwhile labels indicate that the block types are similar in consecutive frames. Moreover, in the entire video sequences, the proportion of BG blocks in the succeeding frames is extremely high. The varying blocks engage only 30% below, signifying that the motion field of every block is extremely high in the succeeding frames for the other blocks. In addition, the model of distribution is much identical devoid of consideration to video sequences. It is revealed that the temporal correlation among the succeeding frames is extremely high, specifically, if a block in the preceding frame belongs to active regions or BG regions, the block that is positioned in the identical position in the present frame may be categorized as an active moving block or BG block, correspondingly, with a strong probability. The fundamental concept behind block matching is represented in the Fig. 4.1, in which the displacement for a block $(L \times L)$ in frame K is described by taking into consideration a window of size $[(L+2w)\times(L+2w)]$ in frame $K+1$ for discovering the position of the best-matching block of the equivalent size. The search is generally restricted to $(L+2w)^2$ region, which is said to be the search window. Here, K indicates the present frame and $K+1$ denotes the search frame.

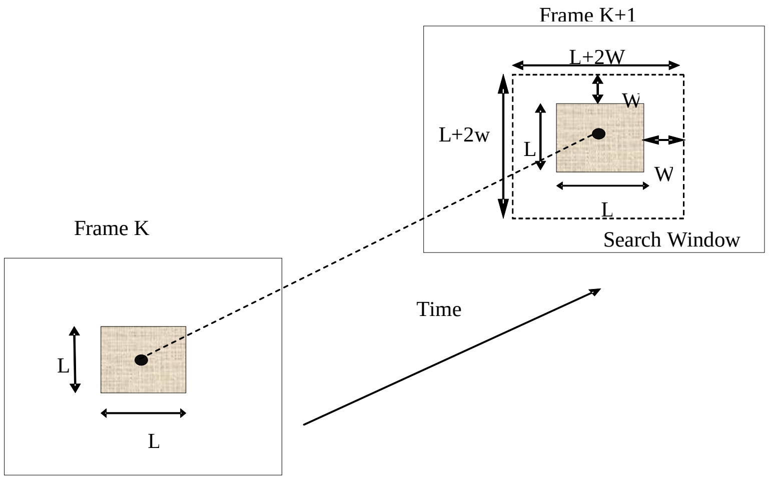


Figure 4-1: Architecture for matching approach

Block matching approaches vary in

- Criteria of matching
- Plan of search
- Block size determination

4.2.3 MATCHING CRITERIA

The matching of the blocks can be enumerated based on a variety of criteria, where the less expensive and well-renowned is MAD, specified by Eq. (4.10). Another measure for MSE is specifiedby Eq. (4.11), where, L denotes the side of the block, and $s_{i,j}$ and $p_{i,j}$ points out the pixels being distinguished in the block from the current frame and the block from the search frame, correspondingly. In general, MSE is not exploited, as it is complicated to comprehend the square function in hardware.

$$MAD = \frac{1}{L^2}\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\left|p_{i,j} - s_{i,j}\right| \tag{4.10}$$

$$MSE = \frac{1}{L^2}\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\left(p_{i,j} - s_{i,j}\right)^2 \tag{4.11}$$

4.2.4 BLOCK SIZE DETERMINATION

The assortment of a suitable block is necessary for any block-dependent ME approach. There are contradictory necessities on the size of the search blocks. If the blocks are much undersized, a match may be introduced among blocks including identical gray level patterns that are not related in the sense of motion. Also, if the blocks are excessively large, then the real MV’s may diverge within a block; contravening the postulations of a single MV for each block. Accordingly, the block size for the implemented model is determined by carrying out continuous testing, captivating a different mixture of frame sizes with diverse frame skips.

4.2.5 RECURRENT ESTIMATION LOGIC

The tracer identifies this preferably and fragments the region over the entire frames, and not just the frames, where it moved. Usually, this phase outlines the computational bottleneck of the entire algorithm. The recurrent exploration of an overlapped pixel is revealed by Fig. 4.2.

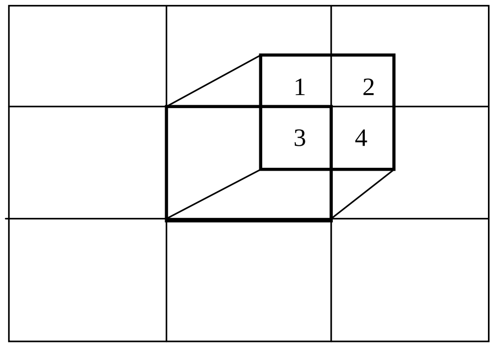


Figure 4-2:Recurrent exploration of an overlapped pixel

Tracing MV's provides itself to a recursive solution in nature. All the blocks with non-zero MV's in every frame correspond to a "seed" call to the tracing operation. A moving block will, generally, transform into an area equivalent to four blocks as shown by Fig. 4.3. The tracing approach commences with a seed call. This seed block could move into a lot of four erstwhile blocks, and all of these blocks are called by the tracing function recursively. The intention of the tracing function is just to recognize the suitable moving pixels depending on the block regions and MV's, and subsequently to seed further calls to it. Motion tracing has an uncomplicated elucidation only in a single direction, which is temporal. In other words, tracing has to be carried out in both the reverse and forward temporal directions for obtaining the most excellent segmentation outcomes.

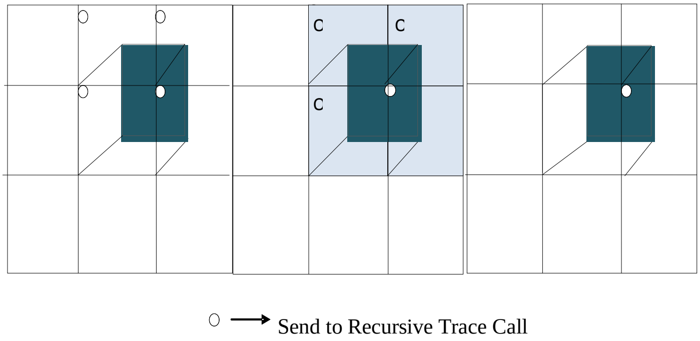


Figure 4-3:Process of exploring frames by means of R-FSBMA

On considering moving block, only the pixels matching to that moving block are correlated with motion, however, the entire four regions interrupted by the block are seeded to the consecutive tracing call. It is considered as the most precise scheme; however, it is also the most troublesome in computation. The subsequent approach is to seed the entire four blocks in addition and to treat the entire pixels contained by the four seeded blocks as having moved instead of the definite moving pixels. This estimation simplifies the tracing approach significantly, and it moreover raises the effectiveness of the algorithm considerably, since a block which is seeded to the tracing function does not require to be ever seeded again.

A last approach is to notice the entire moving pixels as in the common case; however, it seeds only the block matching to the highest overlap. If there are similar overlaps, then several blocks are seeded. Even though this disparity only approximates the tracing crisis, it can be much quicker as each trace call generally seeds only one recursive call instead of four. In the most common case, the tracing approach functions slow. For enhanced speed, MV's are calculated not among every frame, but among every n frames and tracing is performed on this little set of MV's.

4.3 EXPERIMENTAL ANALYSIS

4.3.1 SIMULATION OBSERVATION

To examine the introduced work, a video sequence is comprehended, in which a set of video frames is chosen,and the tracing technique is deployed. The attained outcomes are as shown below. The video file is captivated at a higher location at the centre of a cross road, and the sensor is rotated for 360 degrees to captivate the traffic images. The video sequence demonstrates the movement of the vehicle and other static portions in the surrounding area. The video sample is captivated at 25fps, with a resolution of 272x 352. Moreover, the extracted video frames from the video file are revealed by Fig. 4.4.

Figure 4-4:Extracted Video frames from the video file

A set of successive frames is extracted from the captured video sequence. Further, they are used for processing. The extracted frames after LMS filtration are illustrated in figure 4.5.

Figure 4-5: De-noised sample after LMS filtration

It is required to eliminate the noises so as to achieve higher accuracy in the estimation of moving objects. To achieve this, a conventional adaptive LMS filter is applied to denoise the affected sample. The obtained result for such filtration is given in Fig.4.6 and Fig. 4.7. It is observed that a higher visual quality is achieved with this approach.

Figure 4-6: Extracted sample after mean filtration

Figure 4-7: Extracted sample after median filtration

Over the filtered sample, a full search block matching algorithm is applied to compute the moving element. It is observedthat as the camera is rotating, the BG objects will change their corresponding position for each frame. Hence, such components are also detected as moving elements in predicted video frames. Moreover, the samples that are extracted from noise images are revealed by Fig. 4.8.

Figure 4-8: Extracted sample of noised image

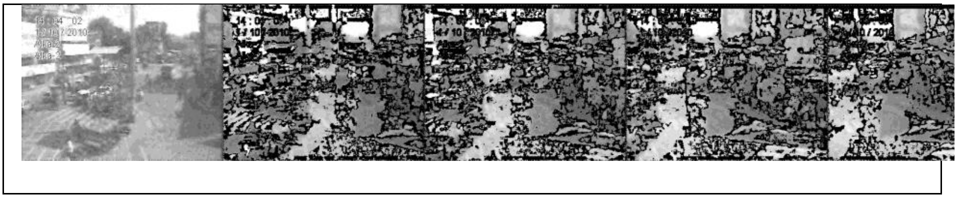


Figure 4-9: Predicted motion elements of FSBMA scheme

In proposed Recurrent FSBMA approach, due to the successive computation of Motion vector in both inter and intra frames, the elimination of a BG element is possible as depicted by Fig.

4.9. Hence, this approach detects the moving elements more accurately than the FSBMA approach, which is exposed in Fig 4.10.

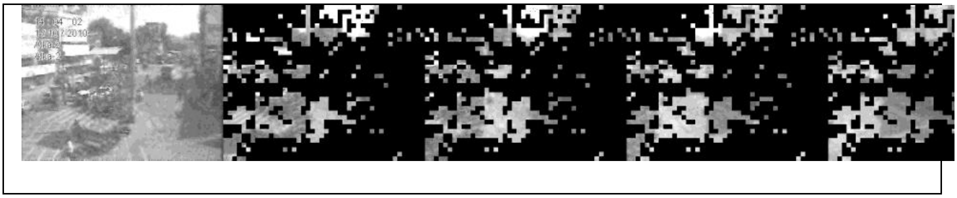
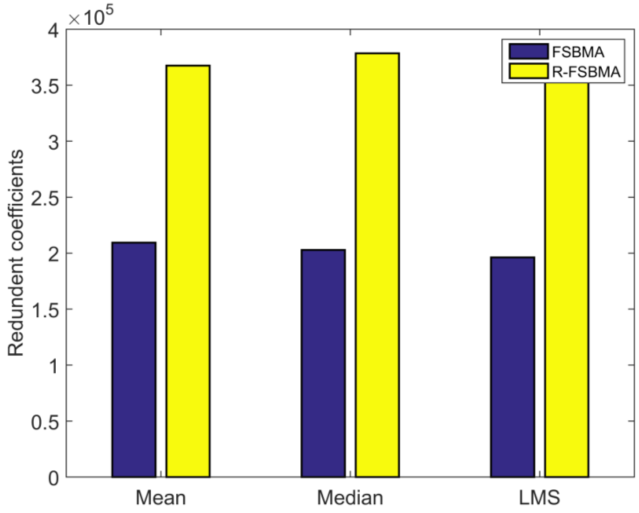


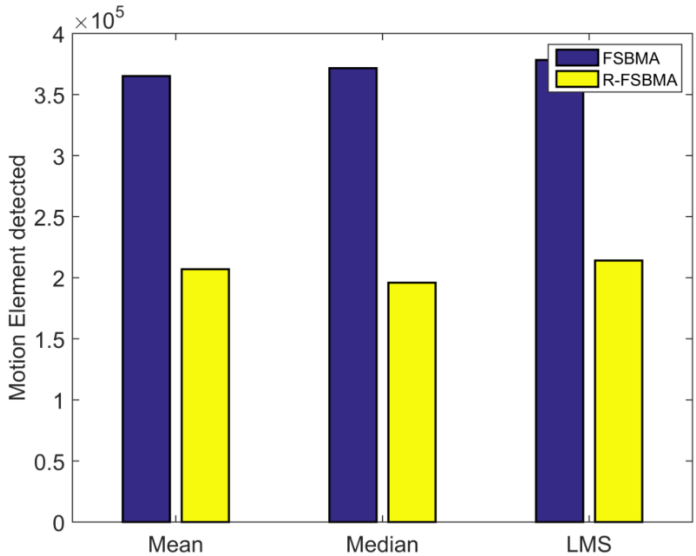
Figure 4-10: Predicted motion elements of R-FSBMA scheme.

4.3.2 FILTER COMPARISON

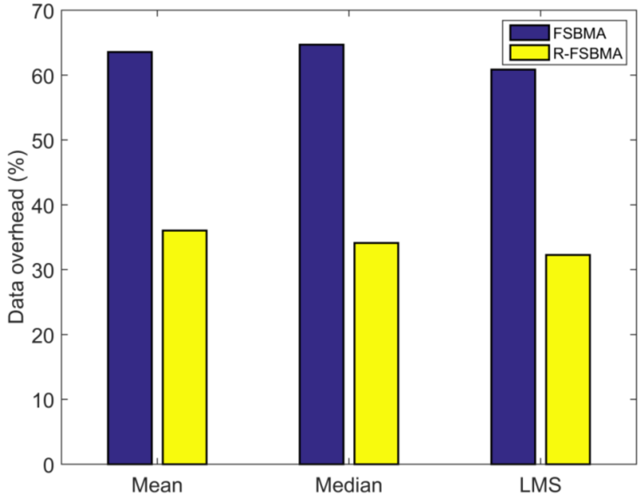
The experimental results for the proposed R-FSBMA approach in terms of filter comparison are given by Fig. 4.11. From Fig. 4.11 (a), the redundant coefficients of the proposed scheme for mean filter, median filter, and LMS filter is 42.1%, 47.37% and 47.37% better than FSBMA schemes. In addition, from Fig. 4.11 (b), the motion element detection for mean filter, median filter, and LMS filter is 47.37%, 48.65% and 39.47% superior to FSBMA techniques. Also, from Fig. 4.11 (c), the data overhead of the suggested scheme for mean filter, median filter, and LMS filter is 46.15%, 87.71% and 45% enhanced than FSBMA model. Thus the filter comparison for the proposed R-FSBMA technique has been substantiated successfully.


×10^5
4
3.5
3
2.5
2
1.5
1
0.5
0
Redundent coefficients
FSBMA
R-FSBMA
Mean
Median
LMS

(a)


×10^5
4
3.5
3
2.5
2
1.5
1
0.5
0
Motion Element detected
FSBMA
R-FSBMA
Mean
Median
LMS

(b)

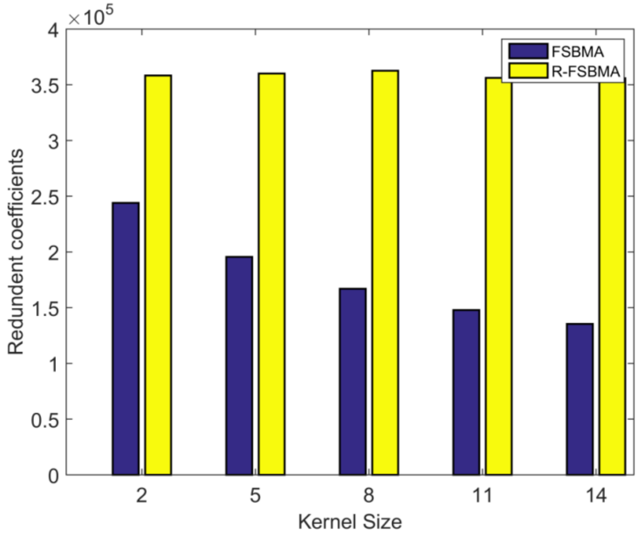


(c)

Figure 4-11:Filter comparison for the proposed and conventional schemes for (a) Redundant Coefficients (b) Motion element detected (c) Data overheads

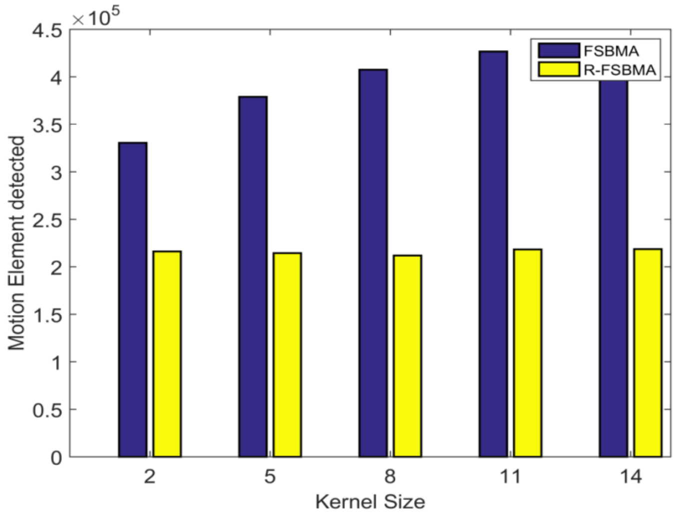
4.3.3 KERNEL SIZE VARIATION

The experimental results for the proposed R-FSBMA approach in terms of kernal size variation are given by Fig. 4.12. From Fig. 4.12 (a), the redundant coefficients of the proposed scheme for 2nd kernel size is 3.14% better than FSBMA scheme. In addition, from Fig. 4.12 (b), the motion element detection for 5th kernel size is 47.37% superior to FSBMA technique. Also, the data overhead of the suggested scheme at 8th kernel size is 45.71% enhanced than FSBMA algorithm. Thus the kernel size variation for the proposed R-FSBMA performance has been confirmed effectively.



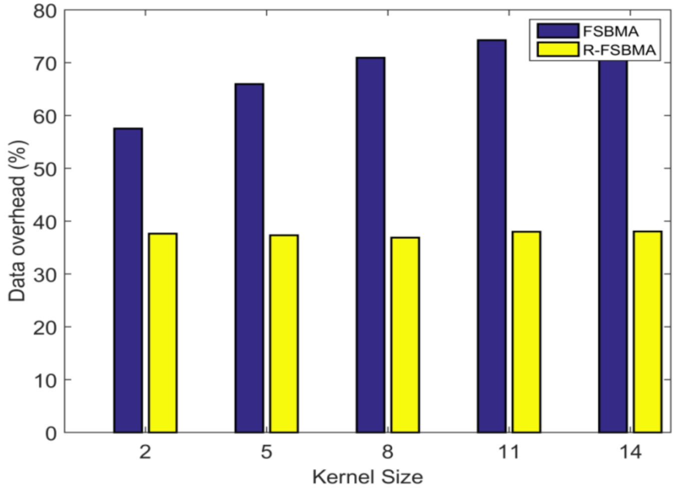
×10^5
4
3.5
3
2.5
2
1.5
1
0.5
0
Redundent coefficients
FSBMA
R-FSBMA
2
5
8
11
14
Kernel Size

(a)



×10^5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Motion Element detected
FSBMA
R-FSBMA
2
5
8
11
14
Kernel Size

(b)



(c)

Figure 4-12: Kernel size variation for the proposed and conventional schemes for (a) Redundant Coefficients (b) Motion element detected (c) Data overheads

4.4 SUMMARY

This chapter has presented a novel coding scheme for VS. The integration of innovative coding approach for denoising by means of the MSE estimator results in advanced estimation probability. This denoising scheme was a dynamic design and therefore, was appropriate for the entire kind of system interface. The application of recurrent FSBMA logic result in detection improvement of a moving object in a video sequence, produced from a rotating camera. From the outcome, it could be concluded that the implemented work depending on the persistent block matching scheme was found to be more effectual and precise in the field of VS, by deploying rotating sensors. It is significant to point out that, very less data for 'information of interest' can be transmitted to central server through this edge processing. Further, the suggested approach has major significance in the metropolitan surveillance system with wireless or wired rotating camera execution, where processing resource optimization and BW were the foremost challenges.

CHAPTER 5. TEMPLATE CODING BASED OBJECT DETECTION

5.1 INTRODUCTION

In video coding, deployment of temporal correlation among frames is a significant step for the minimization of redundant data in succeeding video frames. On the other hand, dynamic nature of video content establishes complexity in discovering temporal correlation. In this chapter, a novel template coding technique to compress the video data for traffic surveillance is implemented that deals with above mentioned complicatedness. In this work, the traditional technique of the template coding, in which two consecutive frames are measured, is enhanced by a 'dynamic model' of the template. The dynamic property of template selection is attained by means of an energy interpolation of consecutive frame data over certain duration, instead of only two consecutive frame data. Moreover, a coherent histogram representation is introduced to construct accurate template to accomplish development in compression. The implemented proficient template matching technique predicts accurate template, thus reducing the processing time and overheads in processing. This makes it a strong candidate for an edge computing. The attained simulation outcome exposes that, the implemented approach provides precise template localization, thus enhancing the accurateness in coding and the coding speed, when compared over traditional template-based compression models. This enables the use of this approach in real time traffic surveillance applications, which is a foremost requirement of a smart city infrastructure.

5.2 PROPOSED METHOD

5.2.1 TEMPLATE BASED CODING

In a variety of video compression schemes, template-based coding is favored for video compression owing to its reduced computational complication and thus offering the fast processing. Among various techniques, an ANN was offered in [103]. This technique offers a data portioning depending on hierarchical extension and k-mean clustering to develop a template block. The technique is practiced on overlapped block size and obtains a residual error which is dependent on the direct evaluation of the information of two consecutive frames. In [104] a TMP was described to build up compression framework. The TMP technique obtain the self similarity content of the two consecutive frame details and obtain the frame correlation dependent on the distance measure of neighbor and current pixels. This representation exploits the static pixel information to set up a template match. Accordingly in this technique, motion details are not deployed. To develop the TMP in [102], an energy interpolation of static frame content by means of histogram is introduced. This technique describes a temporal and spatial localization of template that depends on histogram mapping. In the procedure of temporal coding, a group of histogram describing the temporal template for diverse slices of the frames is introduced. The established histogram correlation E_d is classified into a set of temporal matching histogram, E_c where $E_c \in E_d$ .The classification of templates thus paves the way to quicker localization of template design in the exploration process of motion element in a specified video sample E_s. In the exploration process, the intersection of histogram template for temporal localization is exploited. The intersection of the histogram is defined by Eq. (5.1) in case of such coding, where E_N denotes the normalized histogram for the video observation.

$$I = (E_s, E_d) = \sum_{i=1}^{m} \left(\frac{\min\left(\mathrm{E}_\mathrm{s}^\mathrm{i} - \mathrm{E}_\mathrm{d}^\mathrm{i}\right)}{\mathrm{E}_\mathrm{N}^\mathrm{i}} \right) \tag{5.1}$$

The most important assumption in this representation is that a high value for $I(E_s, E_d)$ is predictive that time frame d contains part of frame s. Accordingly, it is to be noticed that template classification is performed by means of a temporal template, which is described in dataset, and the dynamic property of the video content is disregarded. Here, dynamicity is defined as the nonlinear dissimilarities in video contents owing to vehicle flow, which is unpredictable. It is significant to observe that in these dynamic contents, maximum energy contents are available and if it is not taken into account, it may pave the way to erroneous template definition. Moreover in a variety of video samples the disparities are non redundant, specifically, the foreground and background moving, or foreground static-background moving, foreground and background static moving and in several application where the cameras are moved, in which a false motion is noticed owing to movement of the camera. In such circumstances, the template differentiation requires to be dynamic to accomplish higher accurateness and rapid processing.

5.2.2 ENERGY INTERPOLATED TEMPLATE CODING

In traditional energy dependent template technique, two consecutive frames are taken into consideration for processing function. It is obvious that the accurateness of this technique will not be sufficient if video contents are varying dynamically. As a result, it is necessary to encompass a dynamic template representation to prevail over the problems occurring from distortions in video data. Accordingly, in this chapter, a dynamic template representation depending on the video content is obtained through few consecutive frames histogram correlation. A multi-frame inter correlative frame error is calculated depending on the recurrent frame histogram correlation. In this scheme, for a specified video sample, every frame energies is calculated, and an energy set is described as given in Eq. (5.2), where, E_i indicates the energy histogram for all frames of a video file and N corresponds to the number of frames.

$$E_i(m) = \left[E(mN), E_i(mN-1), \ldots . E_i(mN)\right] \tag{5.2}$$

To calculate the frame error (F) for the specified two frame data, in which the energy interpolates are evaluated to derive energy difference as given by Eq. (5.3)

$$F_{i,E}(m) = E_{i,t}(m) - E_{i,t+1}(m) \tag{5.3}$$

The set of frame correlation on energy plane is evaluated, and an optimal value is selected from the attained frame errors for the gratifying condition of $\min(F_{i,E}(m))$. For obtaining the frame interpolation template, a reference energy histogram is attained from a consecutive frame data observed for certain duration. In addition, a histogram normalization course is derived by a weight parameter for optimizing the template estimation as given by Eq. (5.4), where $V(K) = \lfloor V_0(m), V(m) \ldots V_{M-1}(m)^T \rfloor$ denotes a set of frame weight described for every frame.

$$E_i(K) = E_i(K)V(K) \tag{5.4}$$

The values are arbitrarily initialized and proliferated to a minimal error value in a repeated manner. The $F_{i,E}(m)$ is subsequently updated as in Eq. (5.5)

$$F_{i,E}(m) = E_{i,t}(m) - E_{i,}(m)V(m) \tag{5.5}$$

An initial error value of frame error $F_{i,E,0}$ is recorded,and the frame error $F_{i,E}(m)$ is updated for the entire consecutive frames. For all the iteration, the weight values are updated depending on $F_{i,E,0}$ and the histogram energy as shown in Eq. (5.6).

$$V(m+1) = V(m) + \mu \Sigma_{i=0}^{N-1} \frac{\mathrm{E_i^T}(\mathrm{m})}{\|\mathrm{E_i}(m)\|^2} F_{i,E,0}(m) \tag{5.6}$$

The weight value is modified with a step size μ that is exploited to manage the weight value to a steady update instead of an arbitrary update. The deviation in the bin disparity of the histogram is subsequently incorporated over a period of 0 to N as given by Eq. (5.7), in which incorporating the approximation, over ‘ N ’ observation phase accumulates the evaluation of ‘ N ’ inter-frame errors.

$$D\left(E_{i,N}\right)=\int_{0}^{N}\mu\sum_{i=0}^{N-1}\left(2E\right)\left[\frac{E_{i,n}(m)\tilde{\nabla}F_{i,N}(m)}{\left\|E_{i,n}(m)\right\|^{2}}\right]-\mu E\left[\frac{e_{i,N,E}^{2}(m)}{\left\|E_{i,n}(m)\right\|^{2}}\right] \tag{5.7}$$

All the frames with minimum estimate error are subsequently chosen as the selected histogram bin, and an intersection bin is then obtained from Eq. (5.1). This template selection results in the assortment of template region for any motion constraint with reduced correlative $F_{i,E,0}$. The implemented approach is depicted in Fig. 5.1.

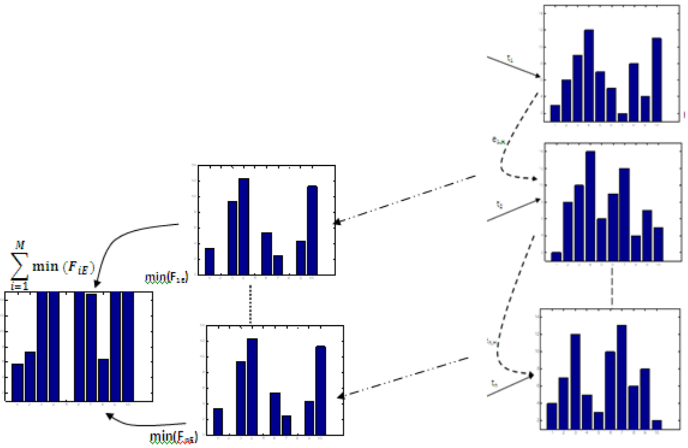


Figure 5-1:Energy correlative template selection scheme

The k-NN representation is exploited for the interpolation of the frame back in decoder unit. A signature is deployed with motion constraint for the entire template regions, and the noticeable template region is interpolated by the duplication of reference template to redevelop the video frame. To authenticate the suggested approach, an executionis performed for the introduced system in evaluation to traditional energy approach and template-dependent compression scheme.

5.3 EXPERIMENTAL RESULTS

5.3.1 PROCEDURE

The suggested system was simulated in MATLAB 8.1 and verified for applications in real time traffic surveillance. The daytimetest sample was captivated from city of Pune (India) traffic flow by means of a rotating capturing camera mounted over a junction point. The video samples were captured at 25fps. A frame sample of captivated video is revealed in Fig. 5.2.

Figure 5-2: Captured sample of a traffic surveillance camera

5.3.2 PROCESSED OUTPUT

For operating the captivated video sample, a frame reading is performed at a frame time skip of ten frames to attain motion information that is dominant. The extorted frames for processing are revealed in Fig 5.3.

Figure 5-3: Extracted frames for processing

The frames are chosen to comprise three different motion constraints, in which the foreground is noticed to be moving that includes certain moving vehicles, the background is stationary that reflects the tree, sign boards, and a false motion design owing to camera motion is obtained which provides a motion effect to stationary roads, footpath and static environment. A template extraction to this captivated frame is performed by means of TMP [104] and HIST [102] that are distinguished with the suggested EI-HIST.

The template adopted for compression model in terms of TMP template matching is obtained depending on the assessment of consecutive frames. The attained template for mapping is exposed in Fig. 5.4.

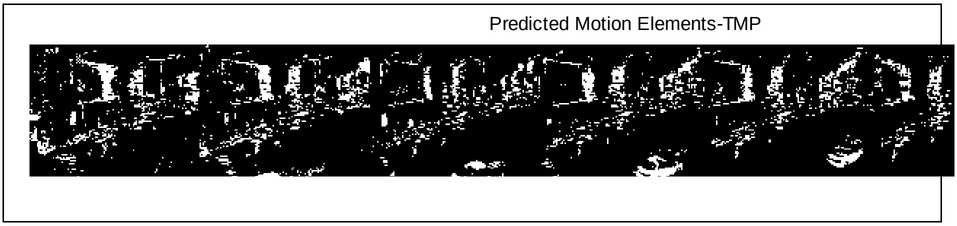


Figure 5-4:TMP dependent template coefficient [104]

For the calculation of template more efficiently, energy prediction representation by means of histogram feature as sketched out in [102] is introduced. The attained template coefficient by means of HIST is revealed in Fig. 5.5.

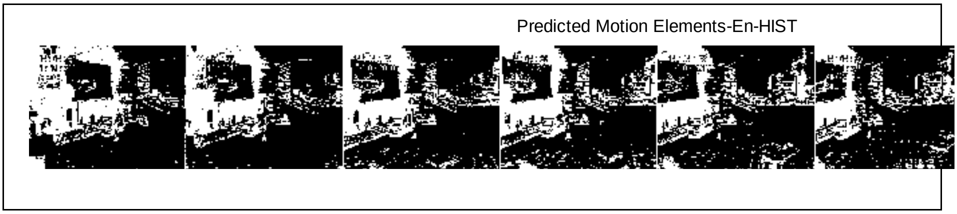


Figure 5-5:Template derived by deploying Histogram mapping [102]

The attained template for video compression by means of the suggested EI-HIST is revealed in Fig 5.6. The finest coarseness in the assortment of moving constituent depending on the interpolation scheme could be noticed. The finer calculation results in lower coefficients, thus attaining better compression.

Figure 5-6: Template derived from EI-HIST

The processing accurateness and computational overhead are calculatedwith respect to accuracy in recovered frame data and the time consumed for encoding and decoding of the introduced system. The resultant frame by the interpolation by means of TMP, HIST, and EI-HIST is revealed in Fig. 5.7, Fig. 5.8 and Fig. 5.9 correspondingly.

Figure 5-7: Recovered frame by deploying TMP technique

Figure 5-8: Recovered frame by deploying HIST technique

Figure 5-9: Recovered frame by deploying EI-HIST technique

5.3.3 PERFORMANCE ANALYS

The computation of the established system is calculatedwith respect to the actual motion element detected. The coefficient calculation for the traditional technique and suggested technique is given in Table 5.1.

Table 5.1: Motion coefficients stated by the three established techniques

Methods / **Parameter value**	**TMP [104]**	**HIST [102]**	**EI-HIST**
Original Sample Size	765952	765952	765952
Motion Element detected	168120	205202	542884
Redundant coefficients	597832	560750	223068
Data Overhead	56.94%	46.65%	17.63%

The accurateness of these three established techniques is distinguished by the attained PSNR, where MSE is given in Eq. (5.8 and 5.9) in which denotes the original and frames and I' points out the interpolated frames

$$PSNR = 10 \log \left(\frac{[\max(I)]^2}{MSE} \right) \tag{5.8}$$

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left(I - I' \right)^2 \tag{5.9}$$

The assessment of the PSNR attained for the established technique is revealed in Fig. 5.10. The PSNR for the energy interpolation part is noticed to be superior owing to spectral domain processing instead of time domain processing. On considering the TMP technique, coding part stores further redundant data that causes error in decoding process that consequence in lower PSNR value.

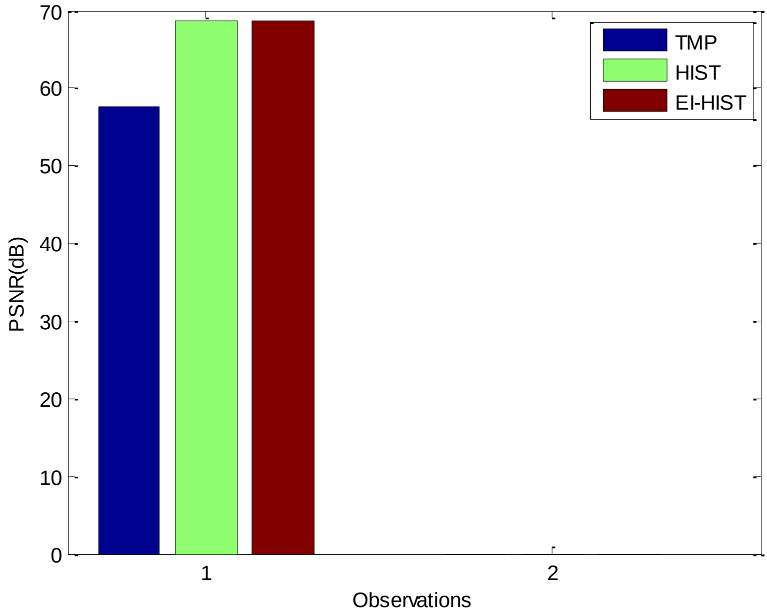


Figure 5-10:PSNR evaluation for the introduced scheme

The template model has an effect on the performance time depending on the obtained template region. The precise region recognition results in advanced interpolation accurateness and low calculation duration. The evaluated time delay for processing that is processed over Intel(R) core i5 CPU at 2.3GHz processor, is revealed in Fig. 5.11. An enhancement of 0.3 Sec and 0.8 sec is noticed in assessment to the traditional TMP and Histogram interpolation schemes correspondingly.

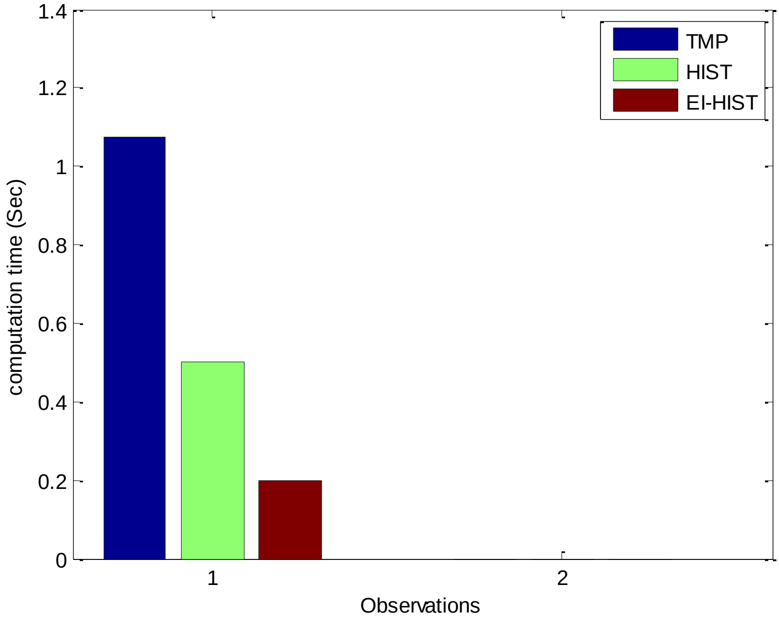


Figure 5-11: Computation time plot for the three introduced schemes

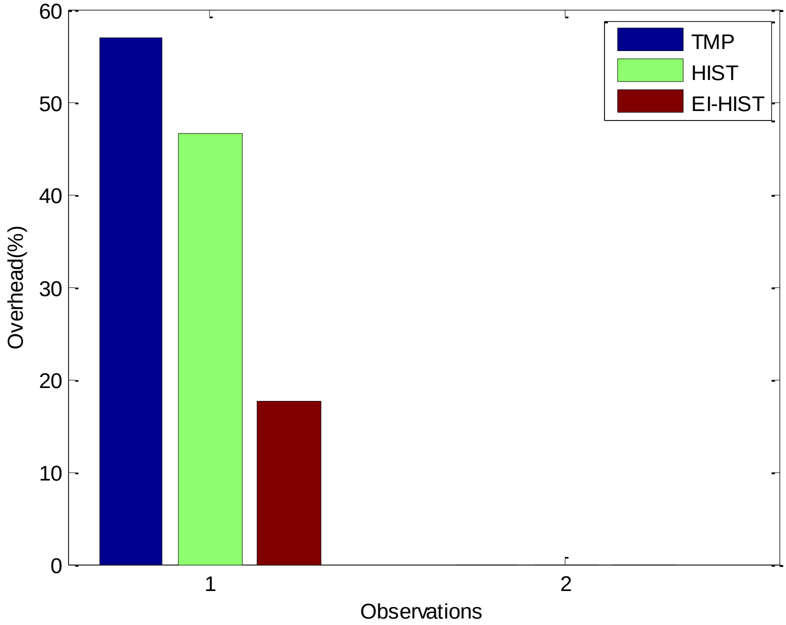


Figure 5-12: Overhead annotations of the introduced schemes

The noticed overhead in processing data for the three schemes is accessible in Fig 5.12. The eradication of data redundancy causes reduction of data overhead in the evaluation process. It is found to be minimized by 29% and 39% in assessment to HIST and TMP techniques correspondingly.

5.3.4 COMPARATIVE ANALYSIS

Here, computation time, data overhead, motion element detected, MSE, PSNR, redundant coefficients and SSIM are computedin terms of various observations.

Accordingly, in the first observation, eight frames are taken into account, in second observation, 10 frames are considered, in third observation, 15 frames are taken into account, and in fourth observation, 25 frames are focused. From Fig. 5.13, the computational time of the proposed EI-HIST at 1st observation is 82.85% better than TMP and 70% better than HIST approaches. Similarly, from Fig. 5.14, the data overhead at 1st observation is 67.27% superior to TMP and 60% superior to HIST scheme. Also, from Fig. 5.15, the motion element detection at 1st observation is 67.24% better than TMP and 68.51% better than HIST techniques. Moreover, from Fig. 5.16, the error analysis of proposed scheme at 2nd observation is 71.42% superior to TMP and HIST models. From Fig. 5.17, the PSNR of the implemented scheme is 14.71% better than TMP approach. Also, from Fig. 5.18, the redundant coefficient of the suggested scheme at 3rd observation is 78.57% superior to TMP and 12% superior to HIST methods. Also, from Fig. 5.19, the SSIM of suggested technique at 4th observation is 60% superior to TMP model. Thus the enhancement of the proposed EI-HIST in terms of frames observation has been substantiated successfully.

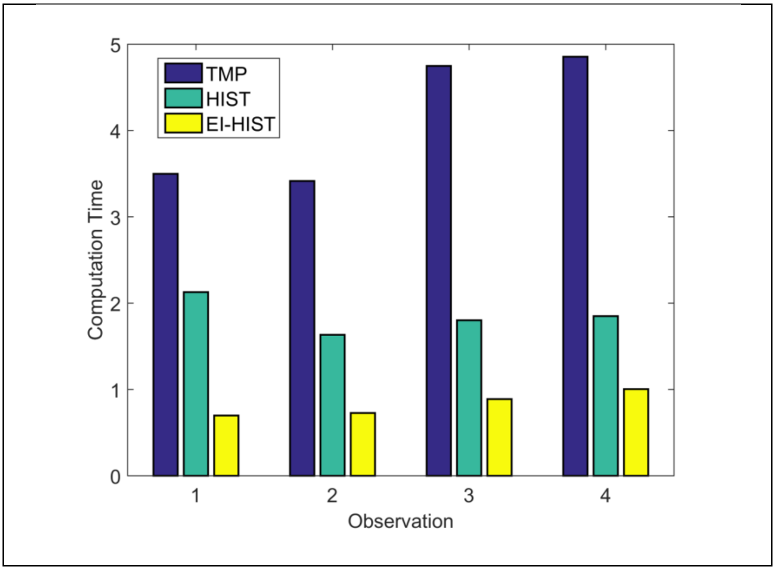


Figure 5-13: Computation analysis of the suggested and traditional schemes

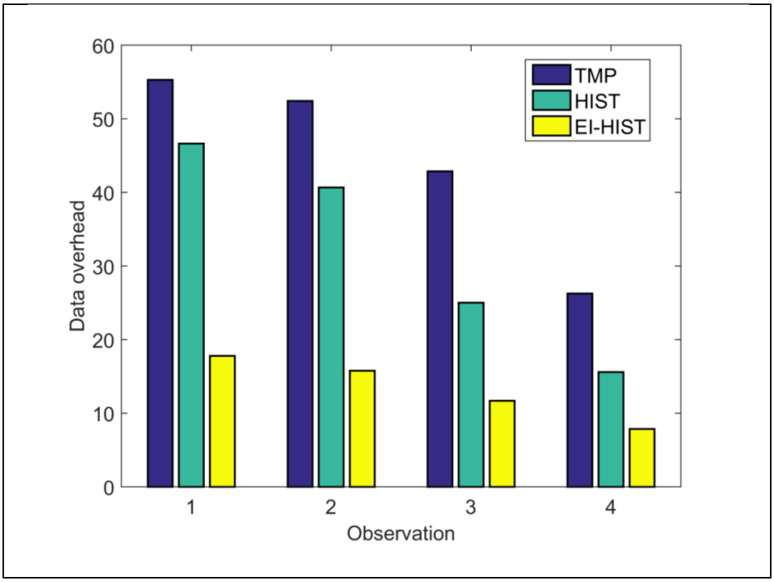


Figure 5-14: Data overhead analysis of the suggested and traditional schemes

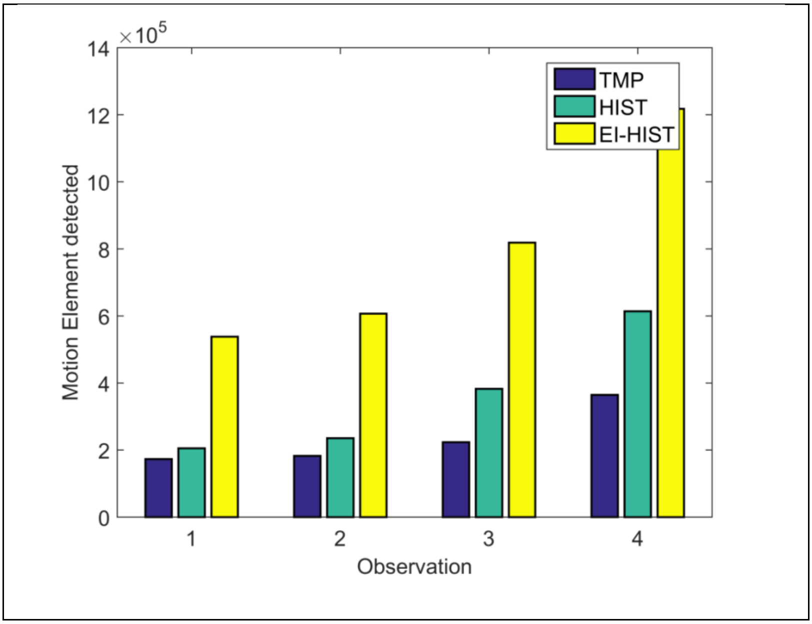


Figure 5-15: Motion element analysis of the suggested and traditional schemes

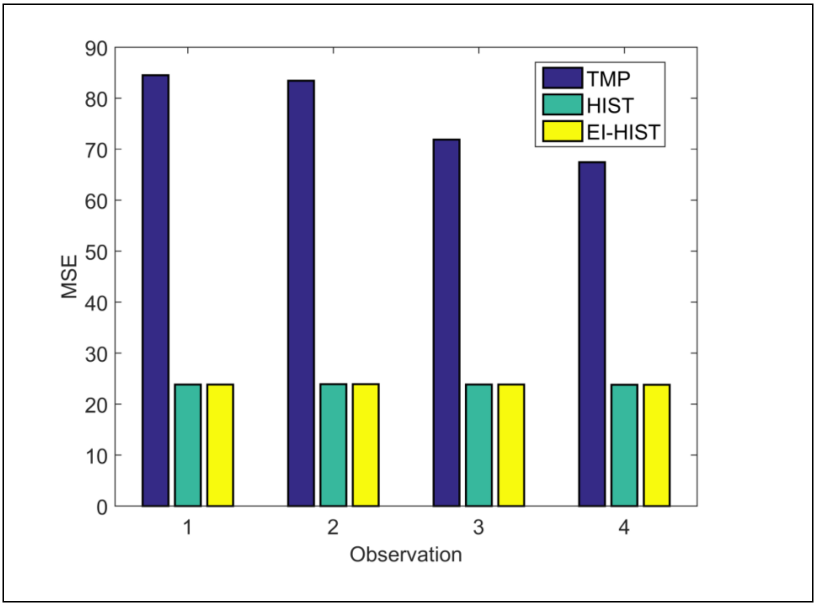


Figure 5-16: Error analysis of the suggested and traditional schemes

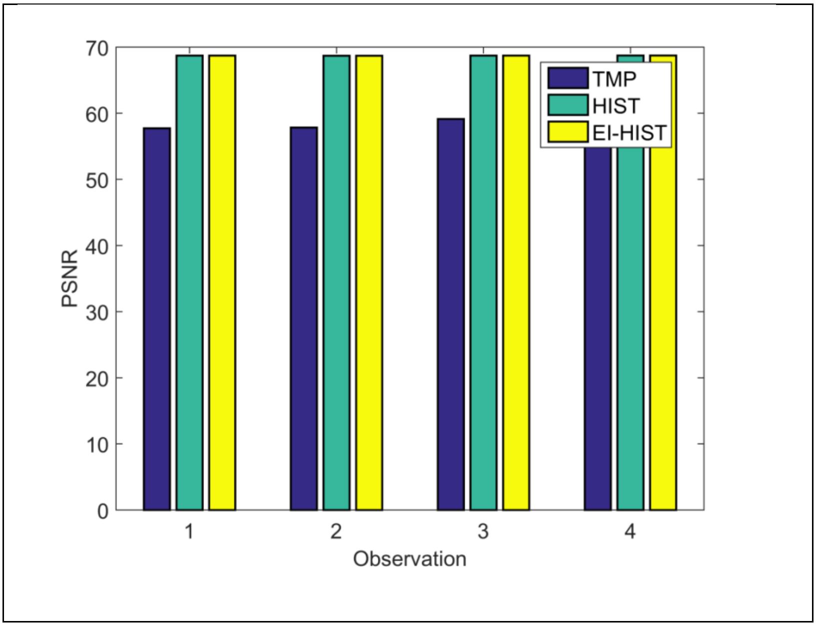


Figure 5-17: PSNR analysis of the suggested and traditional schemes

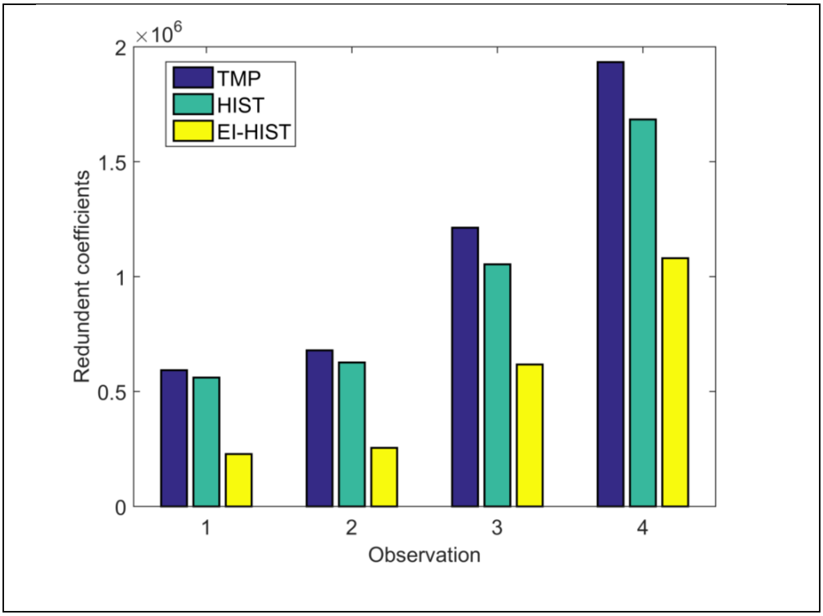


Figure 5-18: Redundant co efficiency analysis of the suggested and traditional schemes

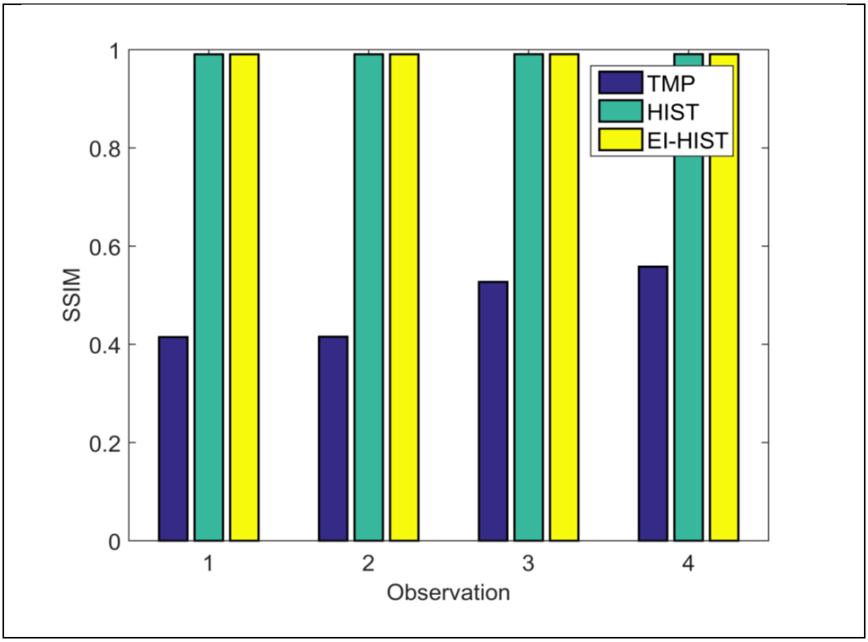


Figure 5-19: SSIM analysis of the suggested and traditional schemes

5.4 SUMMARY

The effectiveness of template-dependent video compression is based on the precise description of template region. Since the erroneous interpretation raises the delay factor and processing overhead, an oversampled template reduces the accuracy of interpolation. In this chapter, an energy interpolated template coding depending on inter correlative histogram was implemented. The traditional representation of template match prediction and histogram based coding was distinguished with the implemented energy interpolated histogram coding. The simulation outcome attained for the suggested approach analyzed over a traffic surveillance data demonstrated a superior coding precision together with enhancement in coding speed owing to precise template derivation. Thereby, allowing this approach to be a strong contender for real time surveillance applications.

CHAPTER 6. DATA DIFFUSION THROUGH WIRELESS MEDIA

6.1 INTRODUCTION

Nowadays, traffic surveillance remains as a fundamental attribute in smart city conception, where rotating camera is favoured over static cameras. Inspiration regarding this replacement is for minimizing the expenditure of overall cost of ownership and data transmission. For modeling an optimal wireless smart city area network in the case of VS systems, some important areas should be focused, i.e., transmission efficiency, edge computing at transmission nodes, data congestion and so on. The final intention is to accomplish the received video streams with high quality regardless of data transmission that is compressed. Certain investigational procedures in this field are significant. For instance, SSIM based RDO in wireless environments is an effectual approach for improving the video quality. On the other hand, existing scheme does not focus on the congestion of network, which inspired in presenting a novel video streaming technique dependent on a new packet data queue management system. It is dependent on a SSIM technique that integrates the ROA approaches as a function. The investigational outcomes reveal that the suggested DMTC can accomplish improved data throughput and increased video quality.

6.1.1 SSIM-RDO VIDEO STREAMING

For ERC, the exploitation of SSIM-RDO in video streaming is extremely common in recent times. The motivation behind this is SSIM performs better than conventional techniques, such as MSE and PSNR, that have established to be reliable with human observation. Accordingly, the implemented algorithm exploits the theory of SSIM-RDO that is essential to identify the contents of the same. In [42], the SSIM-RDO dependent ERC method for H.264/AVC is suggested. To develop the wireless video streaming computation, an arithmetical association was obtained via the LO scheme to attain the reduced distortion. In addition, the SSIM is deployed as an distortion measure, and LM with reduced complexity for SSIM-dependent RDO for error-free coding is obtained primarily. The SSIM- dependent decoding for minimizing the distortion is established in encoder to manipulate ERVC. It is valuable to note that, in SSIM dependent technique, the LM is hypothetically obtained from SSE based LO procedure, thus minimizing arithmetical complication.

6.1.2 SSIM-DEPENDENT RDO FORMULATION DEPENDING ON SSE-BASED RDO

In video processing, the desired encoding mode could be described by attaining the best trade-off among the attained video quality and quantity of coding bits. This problem could be designed [42] as in Eq. (6.1), that points out that the video encoder has to reduce the perceptible distortion ‘ D ’ with the quantity of encoding bits amount ‘V ’,subsequent to the parameters of bits amount ‘V_c’ by choosing the suitable encoding mode ‘ m ’ [42].

$$\min_{\{m\}}\{D\} \text{subject to} V \leq V_c \tag{6.1}$$

Here, the LO scheme is exploited as specified in Eq. (6.2), to formulate the objective.

$$\min_{\{m\}}\{j\} \text{with } j = D + \lambda V \tag{6.2}$$

where, ‘ j ’ indicates LO cost and ‘ λ ’ denotes the LM for RDO

In general, in LO, the distortion measures, such as SAD and SSE, are metrics of video quality. Moreover, the LM is exploited to adjust the SAD or SSE dependent distortion and the quantity of coding bits. Accordingly, in this chapter, SSIM is deployed to determine the distortion unit. The SSIM index is evaluated to calculate the relativeness of contrast, local luminance, and structure among a distorted image and an original image independently. The SSIM index is computed in windows for the two images with varied block sizes. For two images windows of block y and x, the local SSIM index of the two images is specified as specified in Eq. (6.3), where, ‘ σ_{xy}

',denotes the cross-correlation, 'σ_x' indicates the standard deviation, and 'μ_x' points out the mean among the two image windows.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + G_1)(2\sigma_{xy} + G_2)}{(\mu_x^2 + \mu_y^2 + G_1)(\sigma_x^2 + \sigma_y^2 + G_2)} \tag{6.3}$$

The 'G_1' and 'G_2' are employed to sustain the constancy when the variances and means are close to 0. The LO method from [42] could be designed as in Eq. (6.4), when the coding distortion is determined by means of SSIM-based distortion, where, D_{SSIM} indicates the SSIM-based distortion and "λ_{SSIM}" denotes the LM for the SSIM-dependent RDO.

$$\min_{\{m\}}\{j\} \ with \ j = D_{SSIM} + \lambda_{SSIM}.V \tag{6.4}$$

As the distortion is notedin terms of the SSIM metric, 'λ_{SSIM}' must be selected appropriately to acquire the optimal tradeoff among the SSIM-based interruption and the coding bits quantity. Therefore, the core issue for SSIM-based RDO is to find out the SSIM-based LM 'λ_{SSIM}'.The LM 'λ_{SSIM}' for SSIM based RDO can be designed by just scaling the 'λ_{SSE}' for a fixed scaling factor 'f'. As a result, the SSIM-dependent LM can be attained as shown by Eq. (6.5).

$$\lambda_{SSIM} = \frac{-D_{SSIM}}{V} = \frac{\left(\frac{D_{SSE}}{f}\right)}{V} = \frac{1}{f}.\frac{D_{SSE}}{V} = \frac{\lambda_{SSE}}{f} \tag{6.5}$$

Therefore, for the SSIM-based LO development, it could be designed by only leveling the existing SSE-based LO model [105] with a fixed parameter 'f' as specified by Eq. (6.6).

$$\min_{\{m\}} \left\{\frac{\lambda_{SSE}}{f}\right\} with \ \frac{\lambda_{SSE}}{f} = \frac{D_{SSE}}{f} + \frac{\lambda_{SSE}}{f}.V \tag{6.6}$$

6.1.3 SSIM-BASED ERVC

According to this scheme, for offering the network optimality, the NAL and the VCL are modeled for H.264/AVC in terms of VC criterion. The VCL functions for the compression action while; the NAL is executed to present the resources of accurate distribution at the network level. Generally, the transmission channel is erroneous and time-varying for wireless communication. Moreover, an independent channel representation is exploited for the reduction of an error throughout the propagation of signal. With the knowledge of the BER, the packet loss probability 'ρ' of the transmission channel for a NAL unit including 'L' bits are linked as in Eq. (6.7).

$$\rho = 1 - (1 - ber)^L \tag{6.7}$$

Throughout the encoding procedure, the video streams are separated into frame slices symbolized as $s_{n,m}$. For the m^{th} slice, the BER is denoted by $ber_{n,m}$ in the n^{th} frame, which is the packet loss rate for slice and is the BER channel for the transmission of the m^{th} slice of the n^{th} frame. The LM 'λ_{SSIM}' was attuned to attain the objective of the ERC to a least value. The LM was established and progressed depending on the "distortion metric", D_{SSIM}. As the distortion evaluation is carried out at the final stage of encoder, a module in the encoder is integrated, to model the process of decoding with the assistance of acknowledgement notification that informs the encoder if the transmitted packet is delivered or undelivered to the receiver. While encoding the n frame, for an acknowledgement notification of 'nr' frame received by the encoder, the encoding data is accumulated and the integrated decoding unit decodes the 'nr' frame and obtains the decoded frames, which are expected from $nr+1$ to the $n-1$ frame. When the expectations of the decoded reference frames or the decoded reference frames are provided, the pixel values were attained. Accordingly, the decoding distortion that is expected is obtained by the Eq. (6.8), where, $b_{n,m,k}$ points out the original MB,

$b_{n,m,k}^{e_c}$ denotes the hidden error MB with packet loss and $b_{n,m,k}^{n_l}$ signifies the decoded MB devoid of packet loss.

$$E\{DSSIM_{n,m,k}\} = 1 - \rho_{n,m}.SSIM\left(b_{n,m,k}, b_{n,m,k}^{e_c}\right) - \left(1 - \rho_{n,m}\right)SSIM\left(b_{n,m,k}, b_{n,m,k}^{n_l}\right) \tag{6.8}$$

The appropriate modification of the LM dependent on the distortion measure is calculated as given in Eq. (6.9).

$$\lambda_{SSIM}{}' = \frac{-D_{SSIM}(V)}{V} = \frac{\left(1 - \rho_{n,m}.SSIM\left(b_{n,m,k}, b_{n,m,k}^{e_c}\right) - \left(1 - \rho_{n,m}\right)SSIM\left(b_{n,m,k}, b_{n,m,k}^{n_l}\right)\right)}{V} = \frac{\left(\rho_{n,m}.SSIM\left(b_{n,m,k}, b_{n,m,k}^{e_c}\right)\right)}{V_{n,m,k}} + \frac{\left(1 - \rho_{n,m}\right)SSIM\left(b_{n,m,k}, b_{n,m,k}^{n_l}\right)}{V_{n,m,k}} \tag{6.9}$$

Approximately, it is specified as given in Eq. (6.10), where' λ_{SSIM} ' points out the LM for the RDO in the error-free surroundings.

$$\frac{\left(\rho_{n,m}.SSIM\left(b_{n,m,k}, b_{n,m,k}^{e_c}\right)\right)}{\partial V_{n,m,k}} = \frac{\left(1 - \rho_{n,m}\right)SSIM\left(b_{n,m,k}, b_{n,m,k}^{n_l}\right)}{V_{n,m,k}} \approx \lambda_{SSIM} \tag{6.10}$$

When the LM, ' λ_{SSIM}'' is attuned to be less than ' λ_{SSIM} ' the ERC- RDO could choose additional intra-coded macroblocks to hold back the error propagation. Eq. (6.10) points out that' λ_{SSIM}'' is adjusted in an adaptive manner to be less than ' λ_{SSIM} ' with the varied rates of packet loss to endorse the robustness of error in the video streaming process.

6.2 FL-SSIM-RDO APPROACH

It is observed that, for error resilience coding, the SSIM-RDO technique is found to be simpler and effectual for video streaming. Accordingly, on considering video streaming over a wireless channel for traffic surveillance, it is to be made certain that the visual quality and the transmission rate requires to be good for improved monitoring. In traffic surveillance applications, owing to remote capturing, the assigned resources will be restricted. In such a situation, appropriate intermediate node support and exploitation of resources is extremely necessitated to attain improved performance. On the other hand, devoid of the control of data flow this resilience may not consequence in effectual visualization at the monitoring end owing to an acquired latency problem. Therefore with this error resilience, a rate allocation is essential in order to attain the objective of an increased throughput with improved visualization. In the SSIM-RDO scheme, the SSIM-based video coding considered the SSIM as distortion metric among the received and original videos. The SSIM has presented the structural similarity among the recovered and the original videos. If the SSIM is increased, the quality of service or vice versa will also be increased. Though error resilience coding is necessary to develop the visualization, congestion in the channel would corrupt its performance. Therefore, the access controlling is further required with error controlling. Thus, both the objectives are obtained altogether by the implemented scheme.

In this chapter, all the intermediate nodes are taken into account to be a router. Thus, every node practices heavy traffic, which may cause congestion at the particular node. For prevailing over this problem, queue management is deployed. The CLO of video stream traffic at the router level was implemented in [107]. The technique of coding was established at NAL, in which the queue dependent congestion control subsequent to the relative QoS, and AQMis mapped for scheduling the traffic flow rate. The CLO scheme is said to be the CA-AQM approach on an evaluated queue length, and it obtains the dropping probability $d(t)$ or packet Enqueue depending on the data traffic that has been received. At VCL, the video source is blocked into segments and transmitted to NAL for rate distribution. The technique calculates the $d(t)$ as given by Eq. (6.11), where, $\min_{th}$ indicates the least queue threshold and $\max_{th}$ denotes the highest queue limit.

$$d(t) = \begin{cases} 0; & q(t) < \min_{th} \\ 1; & q(t) > \max_{th} \\ \max_p \times \dfrac{q(t) - \min_{th}}{\max_{th} - \min_{th}}; & otherwise \end{cases} \tag{6.11}$$

According to the approach of CA-AQM, the drop probability is adapted as given by Eq. (6.12), where, $p(t)$ denotes the cost in time t and ϕ indicates the stable value 1.001 which is described as REM.

$$d(t) = 1 - \phi^{-p(t)} \tag{6.12}$$

The price varies from time to time on the basis of the average queue length, input rate, and output rate of the queue. The CA-AQM model governs the traffic flow by dropping or accommodating the video packets depending on the significance of packet to drop and the probability index. The cost is increased if the input rate goes beyond the output rate, or else, it is decreased. The recommended controlling approach of CA-AQMis sketched in Fig. 6.1, where, $U(i)$ points out the importance index of i packet in the queue.

Even though the above mentioned techniques of CA-AQM and SSIM-RDO schemes are introduced as rate control to video quality estimation, they were deficit in providing optimal rate controlling depending on the channel distortion level. As pointed out earlier, in [42], the 'network property' is ignored, while in [107] the 'error factor' is ignored. Therefore, it is essential to include an integrated approach of data rate controlling along with error control in video monitoring for improved quality visualization. With this intention in this work, a flow control design depending on enhanced error metric and queue management is obtained as given below.

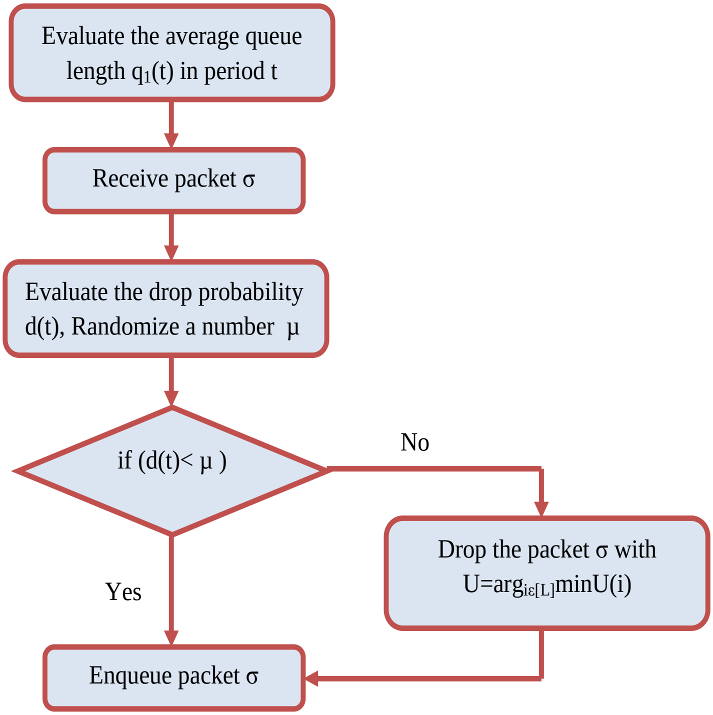


Figure 6-1:Flow diagram for CA-AQM

6.2.1 FLOW CONTROL BASED ON CONGESTION LEVEL

The queue management method as outlined in [107], it is noticed that the level of congestion is managed at two levels and $d(t)$ is subsequently described as of one or zero as specified in Eq.(6.11). It is noticed that traffic flow below $\min_{th}$ is regarded as a non-"congestive zone" and higher than $\max_{th}$ is regarded as a "congestive zone". The region lying among $\min_{th}$ and $\max_{th}$ limitsare considered as a "random zone", in which the packets were arbitrarily dropped or enqueued depending on $d(t)$.

However, in contemplation of traffic flow and error resilience, a flow control depending on video streaming is implemented, which is referred as "FL-SSIM-RDO". The implemented approach is described as; under the restraint of node congestion, deploy queue management with the ROA as portrayed in Eq. (6.13), where, $V_{alloc}(t)$ denotes the ROA, Δt indicates the incremental data rate, $Q_{current}$ implies the present length of queue, $Q_{\min}$ specifies lowest limit of queue, $Q_{\max}$ defines the highest limit of queue and $V(t)$ points out the full rate.

$$V_{alloc}(t) = \begin{cases} V(t) + \Delta t & if\ Q_{current} < Q_{\min} \\ V(t) + (\Delta t - d(t)) & if\ Q_{\min} < Q_{current} < Q_{\max} \\ V(t) - \dfrac{V(t)}{d(t)} & if\ Q_{current} \geq Q_{\max} \end{cases} \tag{6.13}$$

It can be noticed that from Eq. (6.13), the assigned data rate is changing with reference to the congestion level of node. In case, if the $Q_{current}$ is at the least point, then, the data would be assigned with an augmentation of 'Δt'.If $Q_{current}$ lies among the maximum and minimum levels, the assigned data rate will be based on the dropping probability depending on Eq. (6.11). Likewise, if the current queue length goes beyond the maximum queue length, that signifies the level of congestion, the assigned data rate will differ on the basis of the dropping probability, that is, the assigned data rate would be minimum.

Based on Eq. (6.13), the data rate assigned is changing in common. In addition, it has an effect on the modifications of the LM given in Eq. (6.9), and could be updated as shown by Eq. (6.14).

$$\lambda_{SSIM-FL}{}' = \frac{-D_{SSIM}(V_{alloc}(t))}{V_{alloc}(t)} = \frac{\left(1-\rho_{n,m}.SSIM\left(b_{n,m,k}, b_{n,m,k}^{e_c}\right) - \left(1-\rho_{n,m}\right)SSIM\left(b_{n,m,k}, b_{n,m,k}^{n_l}\right)\right)}{V_{alloc(n,m,k)}(t)} =$$
$$\frac{\left(\rho_{n,m}.SSIM\left(b_{n,m,k}, b_{n,m,k}^{e_c}\right)\right)}{V_{alloc(n,m,k)}(t)} + \frac{\left(1-\rho_{n,m}\right)SSIM\left(b_{n,m,k}, b_{n,m,k}^{n_l}\right)}{V_{alloc(n,m,k)}(t)} \tag{6.14}$$

Eq. (6.14) could also be indicated as given by Eq. (6.15).

$$\frac{\left(\rho_{n,m}.SSIM\left(b_{n,m,k}, b^{e_c}_{n,m,k}\right)\right)}{V_{alloc(n,m,k)}(t)} = \frac{\left(\left(1-\rho_{n,m}\right)SSIM\left(b_{n,m,k}, b^{n_l}_{n,m,k}\right)\right)}{V_{alloc(n,m,k)}(t)} \approx \lambda_{SSIM} \quad (6.15)$$

From Eq. (6.15), it can be noticed that the LM is based on the rate assigned for a specific n^{th} slice in the m^{th} frame. As a result, by adjusting the assigned data rate, λ_{SSIM} is also attuned, and it offers a proficient congestion free and error resilient coding. From the Eq. (6.15), the SSIM-dependent LM in Eq. (6.5) can be updated as in Eq. (6.16).

$$\lambda_{SSIM-FL} = \frac{-D_{SSIM}}{V_{alloc}(t)} \begin{cases} \dfrac{-D_{SSIM}}{V(t)+\Delta t} & if\ Q_{current} < Q_{\min} \\ \dfrac{-D_{SSIM}}{V(t)+\left(\Delta t - d(t)\right)} & if\ Q_{\min} < Q_{current} < Q_{\max} \\ \dfrac{-D_{SSIM}}{V(t)-\dfrac{V(t)}{d(t)}} & if\ Q_{current} \geq Q_{\max} \end{cases} \quad (6.16)$$

From Eq. (6.16), it is obvious that the modification of the LM in LO approach is based on the assigned data rate. For the evaluated distortion D_{SSIM} and for the evaluated LM, the LO scheme (6.2), can be updated as in Eq. (6.17), where, 'D_{SSIM} indicates the SSIM-dependent distortion and '$\lambda_{SSIM-FL}$' indicates the LM for the FL-SSIM-dependent RDO.

$$\min_{\{m\}}\{j\}\ with\ \ j = D_{SSIM} + \lambda_{SSIM-FL}.V \quad (6.17)$$

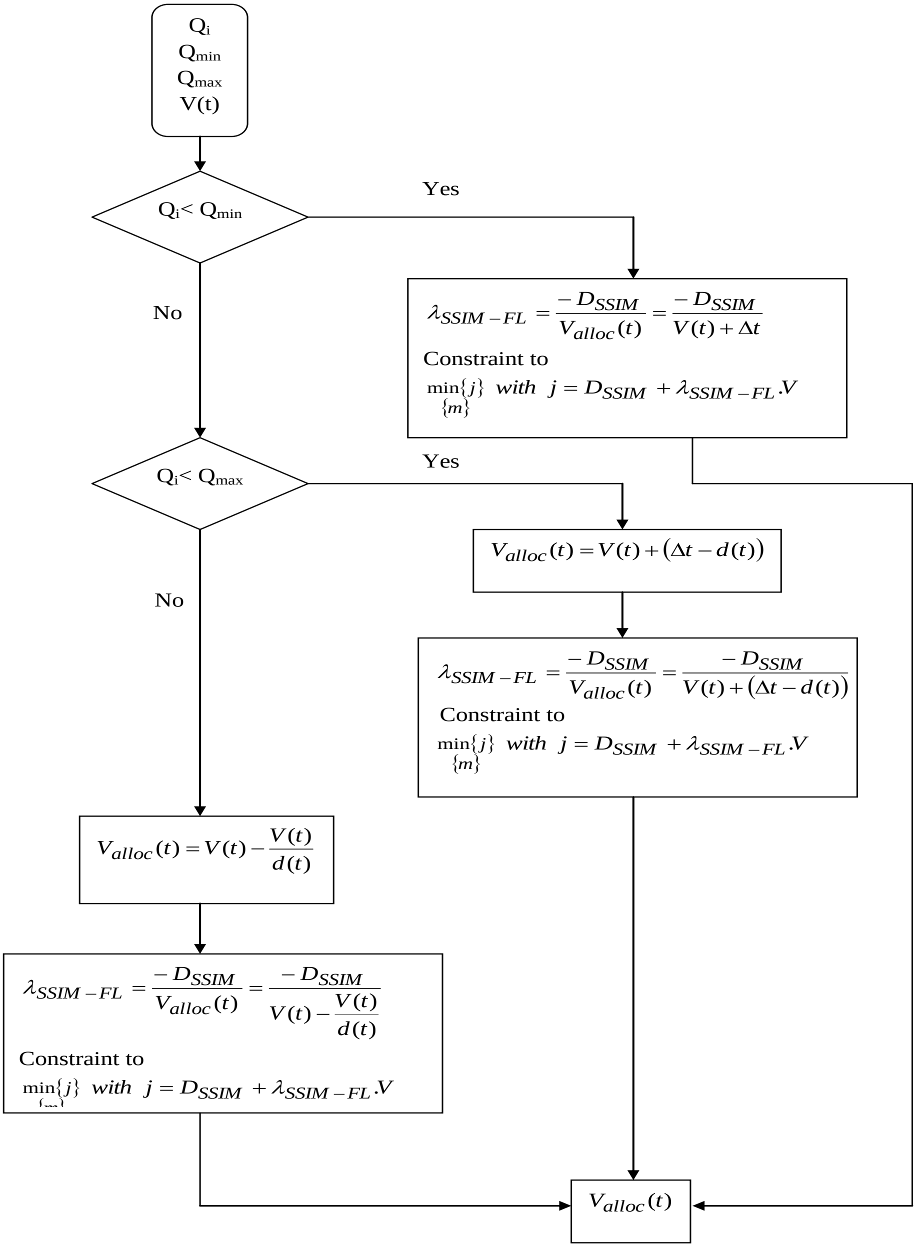


Figure 6-2: Flowchart of FL-SSIM-RDO Algorithm

6.3 DMTC APPROACH

From the queue management method, it is noted that the congestion level is controlled at two levels and $d(t)$ is illustrated as of 'one' or 'zero' as specified by Eq. (6.11). In addition, it is noticed; traffic flow beneath $\min_{th}$ is measured as a non-congestive zone and beyond $\max_{th}$ is measured as a congestive zone. The area between these two limits is obtained as a random zone, in which the packets have arbitrarily been enqueued or dropped depending on $d(t)$ as shown by Eq. (6.13). To initiate the nonlinear distortion variations owing to dynamic channel circumstances, subsequent feasible cases emerge as second monitoring constraint.

Case 1: Under Invariant Channel Condition

In the transmission system, in which the channel is time-invariant, a fixed scaling factor could be described. Based on Eq. (6.13), the data rate allocated is found to be dynamic. This further influences on the adjustment of the LM as portrayed by Eq. (6.8). The LM in this caseis described by Eq. (6.18), where E is specified as in Eq. (6.19).

$$\lambda'_{SSIM-DMTC} = \frac{E}{V_{alloc(n,m,k)}(t)} \tag{6.18}$$

$$Where, E = \left(1-\rho_{n,m}.SSIM\left(b_{n,m,k}, b^{e_c}_{n,m,k}\right) - \left(1-\rho_{n,m}\right)SSIM\left(b_{n,m,k}, b^{n_l}_{n,m,k}\right)\right)$$

$$= \frac{\left(\rho_{n,m}.SSIM\left(b_{n,m,k}, b^{e_c}_{n,m,k}\right)\right)}{V_{alloc(n,m,k)}(t)} + \frac{\left(1-\rho_{n,m}\right)SSIM\left(b_{n,m,k}, b^{n_l}_{n,m,k}\right)}{V_{alloc(n,m,k)}(t)} \tag{6.19}$$

It can also be addressed as shown in Eq. (6.15).

Now, LM can be modified as;

$$\lambda_{SSIM-DMTC} = \begin{cases} \dfrac{-D_{SSIM}}{V(t)+\Delta t} & if\ Q_{current} < Q_{\min} \\ \dfrac{-D_{SSIM}}{V(t)+(\Delta t - d(t))} & if\ Q_{\min} < Q_{current} < Q_{\max} \\ \dfrac{-D_{SSIM}}{V(t)-\dfrac{V(t)}{d(t)}} & if\ Q_{current} \geq Q_{\max} \end{cases} \tag{6.20}$$

For the measured distortion D_{SSIM} and for the measured LM $\lambda_{SSIM-DMTC}$, the LO scheme can be modified as given by Eq. (6.21), where, ‘ $\lambda_{SSIM-DMTC}$ ’ indicates the LM.

$$\min_{\{m\}}\{j\}\ with\ j = D_{SSIM} + \lambda_{SSIM-DMTC}.V \tag{6.21}$$

Case 2: Under Variant Channel Condition

In this condition noise effects are very dynamic. Therefore, SI measure for channel distortion estimation is not much effective [43] and hence, conventional SSIM technique also not effective. Therefore, said approach has to be modified to a cumulative distortion SSIM as given in Eq. (6.22)

$$CDSSIM = 1 - SSIM \tag{6.22}$$

In the dynamic interference state, the integrated distortion inference is exploited for the LO function, portrayed by the assessment of network estimate. For the reduction of distortion, an optimization regression approach that reduces the I/O dependent residual is extracted. The regression coefficient attained by the summation of absolute value is further illustrated as [43] given in Eq. (6.23), in which ‘ n ’ indicates the entire amount of blocks in group of blocks. The expression W denotes the RC vector, λ indicates the regularization factorand W_0 denotes the intercept.

$$\min_{W_0,W}\left\{\frac{1}{2n}\sum_{i=1}^{n}\left(CDSSIM - W_0 - W^T I_i\right)^2 + \lambda\left\|W\right\|\right\} \tag{6.23}$$

The regression coefficient is then defined by replacing the Lagrange regularize factor as given by Eq. (6.24), in which λ'_{SSIM} denotes the regularizing factor by means of SI measure at the assigned transmission rate, and CDSSIM signifies the distortion calculated over an observation period.

$$\min_{W_0,W}\left\{\frac{1}{2n}\sum_{i=1}^{n}\left(CDSSIM - W_0 - W^T I_i\right)^2 + \lambda_{SSIM'}\left\|W\right\|\right\} \tag{6.24}$$

The adopted approach has the concept of duel metric observations for reducing the distortion, in which the SI measure is deployed as a measuring constraint for ROA by means of W_0, W and λ'_{SSIM} to optimize CDSSIM. Therefore, the duel metric optimization attains the distortion minimization and data rate allocation beneath dynamic noise condition. The ROA beneath variant channel condition is based on the cumulative error function. It is portrayed by the optimization of regression constraint, in which the reduction of cumulative distortion error owing to SI is performed. The Lagrange regulator and the CDSSIM constraints are noted for ROA. The described rate allocation is subsequently described as specified by Eq. (6.25).

$$\lambda_{SSIM-DMTC} = \begin{cases} \dfrac{-CDSSIM}{V(t)+\Delta t} & if \;\; Q_{current} < Q_{\min} \Rightarrow \min\limits_{W_0,W} \left\{ \dfrac{1}{2n} \sum\limits_{i=1}^{n} \left(CDSSIM - W_0 - W^T I_i \right)^2 + \lambda_{SSIM'} \|W\| \right\} \\ \dfrac{-CDSSIM}{V(t)-\Delta t} & if \;\; Q_{current} < Q_{\min} \not\Rightarrow \min\limits_{W_0,W} \left\{ \dfrac{1}{2n} \sum\limits_{i=1}^{n} \left(CDSSIM - W_0 - W^T I_i \right)^2 + \lambda_{SSIM'} \|W\| \right\} \\ \dfrac{-CDSSIM}{V(t)+\left(\Delta t - d(t)\right)} & if \;\; Q_{\min} < Q_{current} < Q_{\max} \Rightarrow \min\limits_{W_0,W} \left\{ \dfrac{1}{2n} \sum\limits_{i=1}^{n} \left(CDSSIM - W_0 - W^T I_i \right)^2 + \lambda_{SSIM'} \|W\| \right\} \\ \dfrac{-CDSSIM}{V(t)-\left(\Delta t - d(t)\right)} & if \;\; Q_{\min} < Q_{current} < Q_{\max} \not\Rightarrow \min\limits_{W_0,W} \left\{ \dfrac{1}{2n} \sum\limits_{i=1}^{n} \left(CDSSIM - W_0 - W^T I_i \right)^2 + \lambda_{SSIM'} \|W\| \right\} \\ \dfrac{-CDSSIM}{V(t)-\dfrac{V(t)}{d(t)}} & if \;\; Q_{current} \geq Q_{\max} \Rightarrow \min\limits_{W_0,W} \left\{ \dfrac{1}{2n} \sum\limits_{i=1}^{n} \left(CDSSIM - W_0 - W^T I_i \right)^2 + \lambda_{SSIM'} \|W\| \right\} \\ 0 & if \;\; Q_{current} \geq Q_{\max} \not\Rightarrow \min\limits_{W_0,W} \left\{ \dfrac{1}{2n} \sum\limits_{i=1}^{n} \left(CDSSIM - W_0 - W^T I_i \right)^2 + \lambda_{SSIM'} \|W\| \right\} \end{cases} \tag{6.25}$$

Accordingly, the allocation issueis portrayed as a LO function λ'_{SSIM}, that is a function of allocation data rate with respect to SI measure. The variant is calculated as a factor of cumulative distortion measure ($CDSSIM$), in the dynamic conditions which has to be optimized for rate allocation. If the minimization cost function is fulfilled, the ROA is raised by a factor of Δt under the constraint of minimum threshold. Under similar condition, if the regression model could not attain an optimization value, the allocation rate is minimized to attain the convergence of reduced distortion. Beneath the intermediate region, the data are dropped in an arbitrary approach depending on $d(t)$ and the allocation is managed and subjected to the reduction of regression error. The identical process is done with the maximum bound limit for two observing cases. Moreover, the data traffic is completely blocked under the state of convergence not meeting to the reduction measure. The dual monitoring factor results in maximum accurateness and increased throughput beneath dynamic channel condition. Accordingly, the cumulative distortion results in reduction of distortion in channel variant condition. As the distortion is noticed in terms of the SSIM measure, the assigned rate is optimized with regard to node congestion in addition to the distortion probability in video coding. The implemented approach for DMTC-SSIM-RDO is specified in Fig. 6.3.

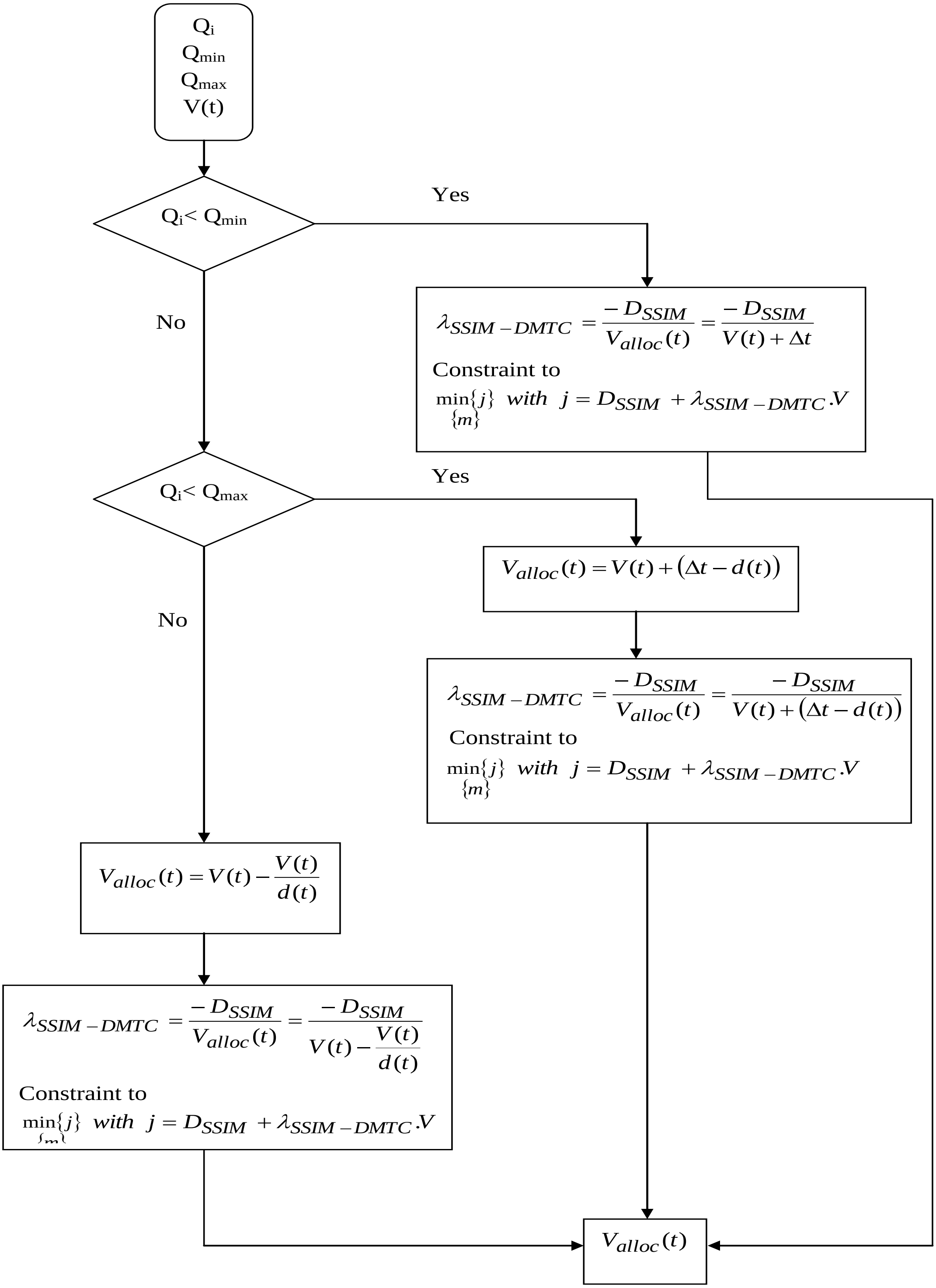


Figure 6-4:Communication model for traffic surveillance

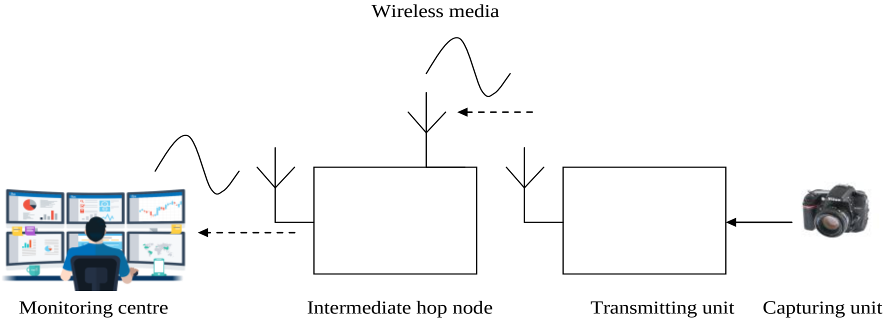


Figure 6-5:Operational data flow for traffic surveillance

At VCL, the captivated video is simulated for compression. The components for motion are derived by means of a RBM model. In addition, the derived motion vectors are compressed by means of a stream out to NAL and entropy encoder. At every node, the NAL calculates the present congestion level and evaluates the rate of transmission that is allocated in contemplation with the error constraint as concise previously. To estimate the implemented scheme, an objective and subjective investigation of the introduced approach are performed based on the SSIM dependent ROA technique. The performance of the implemented approach is calculatedwith regard to the SSIM, e2e delay, throughput, assigned data rate and the node overhead.

A network layout with a captivating node, a monitoring center, and two intermediate hop nodes is established as demonstrated in Fig. 6.6. The Fig. 6.7 demonstrates the sequence that is captivated from a traffic junction. The captivating unit is established at the prevailing traffic light poles with a revolution of 360^0 orientation, and the video is captivated from a high resolution camera with a 272 x 352 pixel resolution at a frame rate of 25fps.For video sample processing, the captivated video sequence is taken as frames that are obtained at a bounce of five frames to simplify the performance overhead. The obtained frame deployed for processing is revealed in Fig. 6.8.Recovery of frames is doneby means of traditional SSIM-RDO, devoid of DFC as shown in Fig. 6.9. Recovery of frames is madeby means of SSIM-RDO together with DFC is given by Fig. 6.10 and Fig. 6.11.

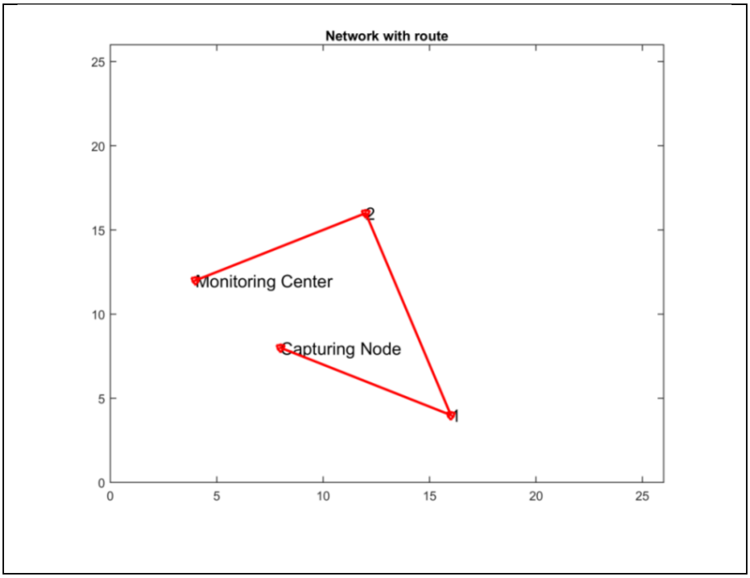


Figure 6-6: Network model deployed for execution

The network factors exploited for the modeling of communication is described in Table 6.1.

Table 6.1: Network Constraints

Network Constraints	Values
Node placement	Static
Transmission range	40 units
MAC protocol	IEEE 802.11
Number of nodes	4
Network area	25 x 25
Qmin	0.15xM
Memory size / node, (M)	3M
Qmax	0.75xM
Initial blockage probability	0.1

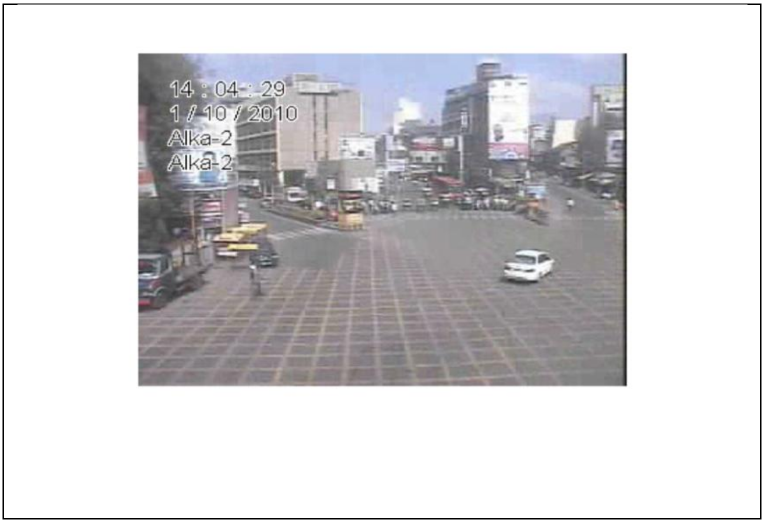


Figure 6-7: Captivated video data surveillance

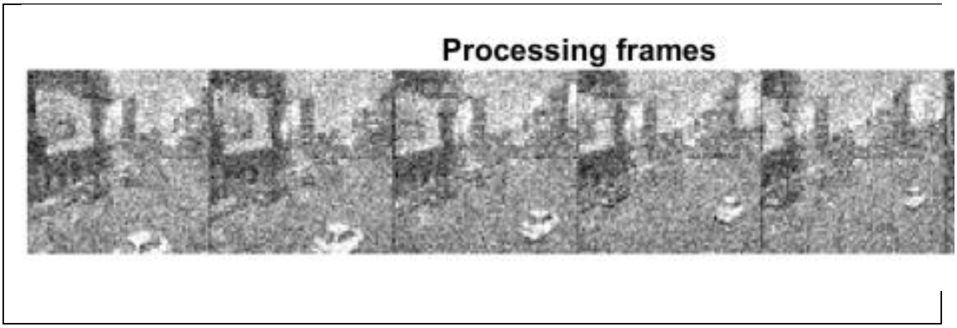


Figure 6-8: Processing frames for the captivated video sequence

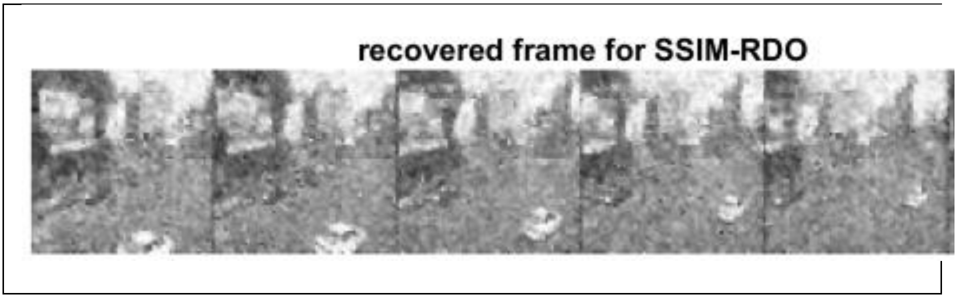


Figure 6-9: Recovered frame by means of SSIM-RDO model

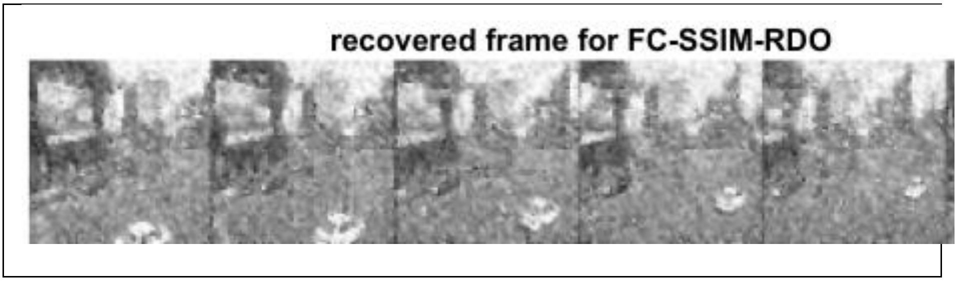


Figure 6-10: Recovered frame by means of FC model

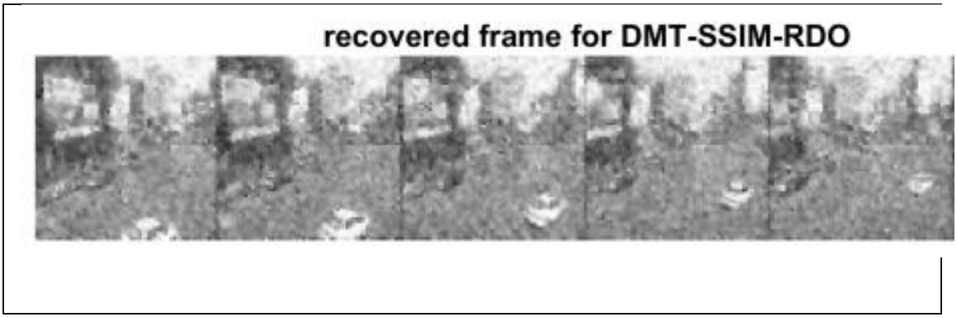


Figure 6-11: Recovered frame by means of DMTC model

Fig. 6.12, Fig. 6.13, Fig. 6.14 to Fig. 6.15 demonstrate quality metrics of data transmission, that is, throughput, route overhead, allocated data rate and e2e delay of SSIM-RDO, flow control and DMTC scheme for non-variant channel conditions. According to this execution, ideal channel is considered, that is almost without noise. The figures demonstrates that the computation of DMTC is improved than prevailing schemes of SSIM-RDO technique owing to the DFC system of DMTC.

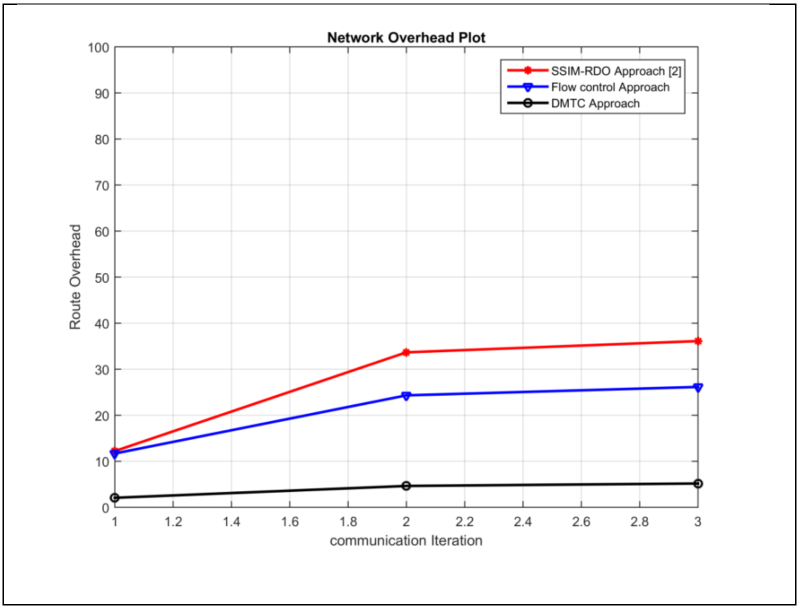


Figure 6-12: Network overhead plot

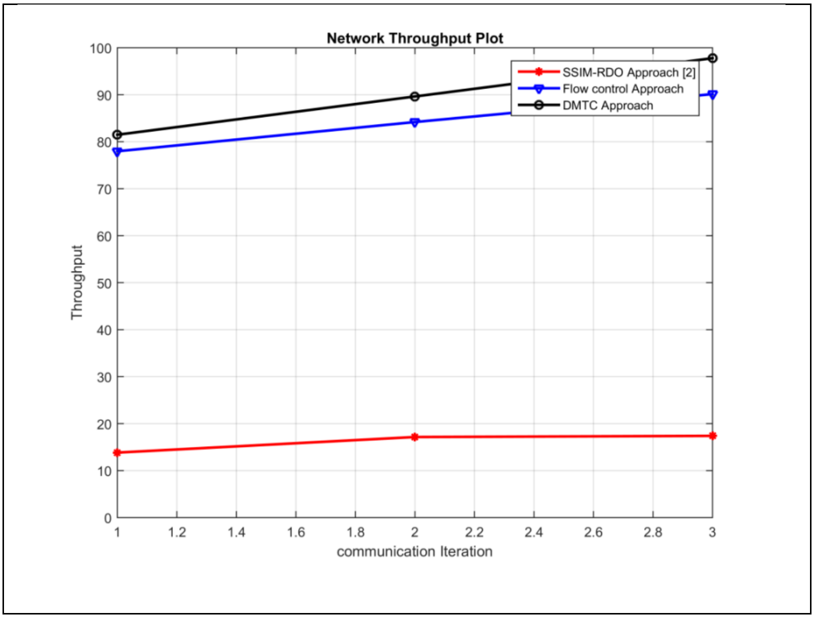


Figure 6-13: Throughput plot for the suggested model

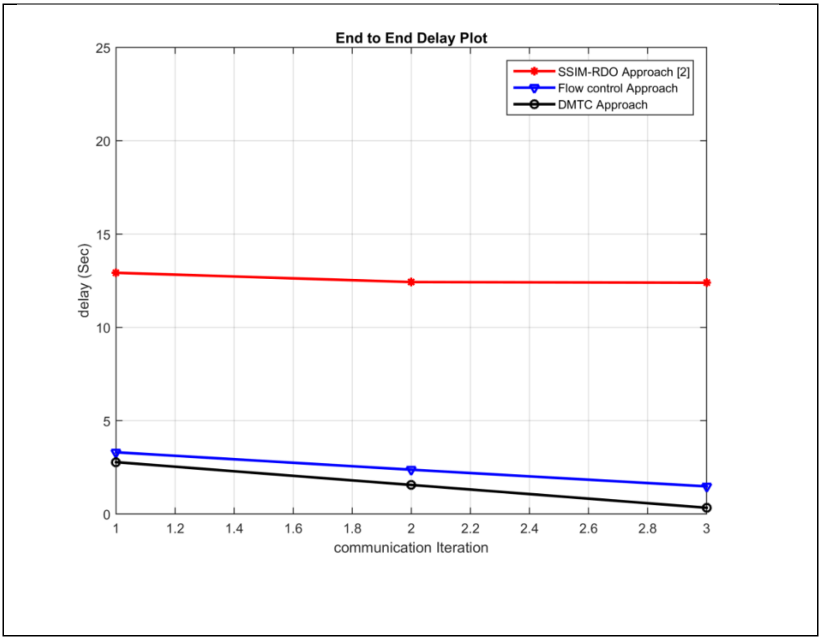


Figure 6-14: e2e delay for introduced scheme

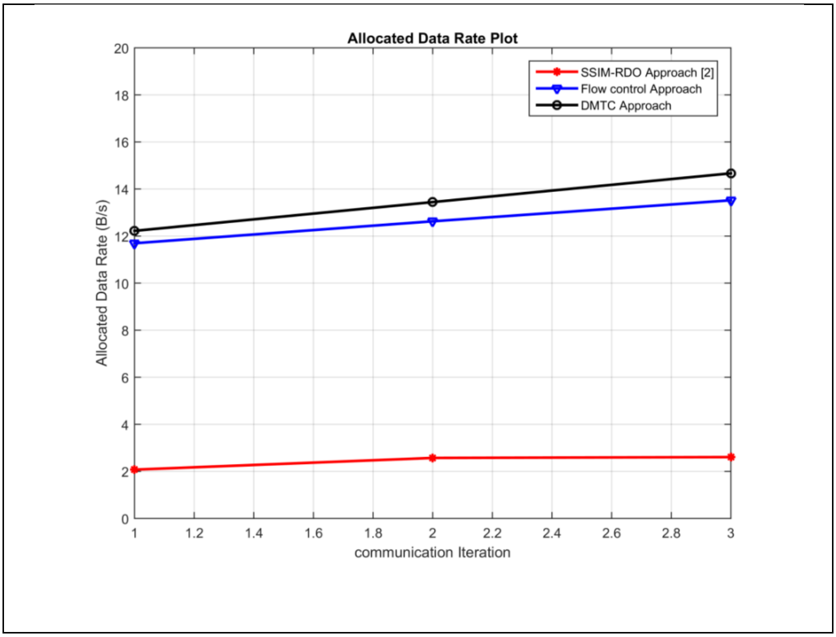


Figure 6-15: Assigned data rate plot for introduced scheme

6.4.1 EXAMINATION UNDER DIVERSE CHANNEL CONDITIONS

Case 1 Variance = 0.1 (Figs. 6.16, 6.17, 6.18, 6.19)

Fig. 6.20, Fig. 6.21, Fig. 6.22 and Fig. 6.23 demonstrate throughput, route overhead, allocated data rate and e2e delay of SSIM-RDO, FC without DMTC and with DMTC at variance = 0.1.

Accordingly, in this execution, channel noise level at 10% is considered. The graph demonstrates that the computation of DMTC is improved than conventional model of SSIM-RDO due to the effectual DFC in DMTC scheme. There are certain enhancements in DMTC from its corresponding item i.e. FC devoid of noise evaluation.

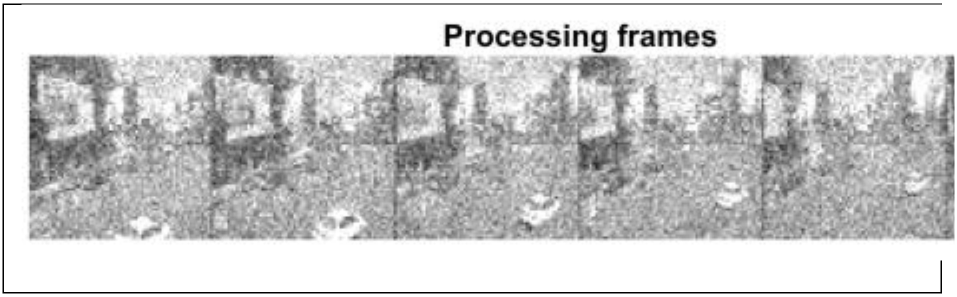


Figure 6-16: Noised sample

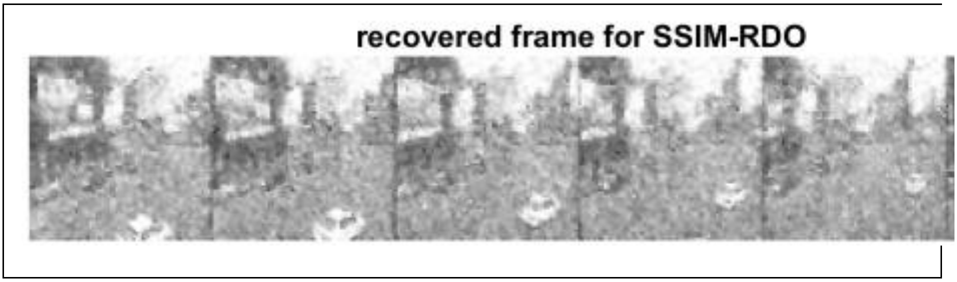


Figure 6-17: Recovered sample by SSIM model

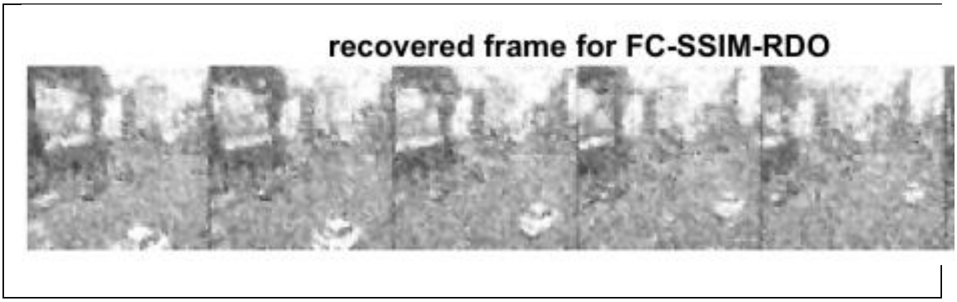


Figure 6-18: Recovered sample by means of FC model

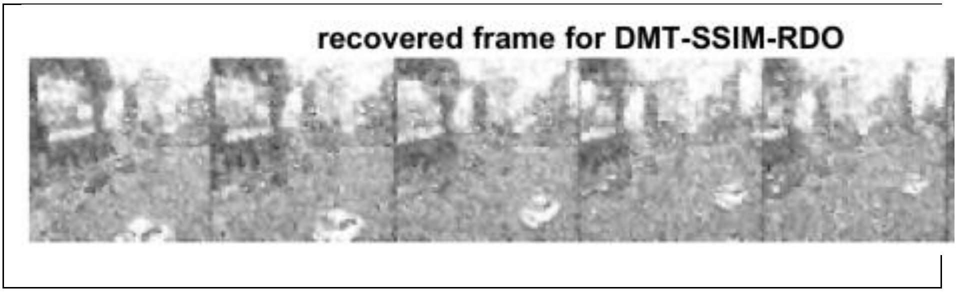


Figure 6-19: Recovered sample by means of DMTC model

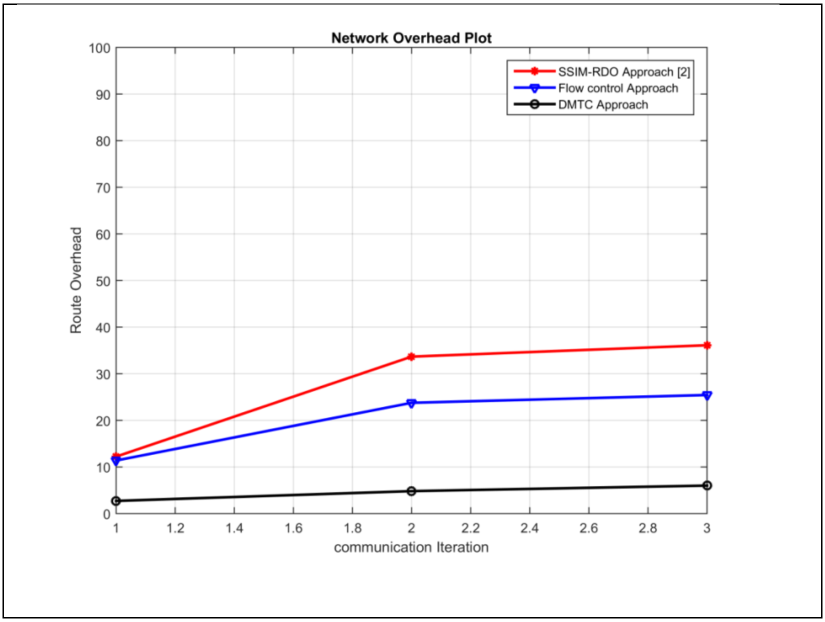


Figure 6-20: Route overhead plot

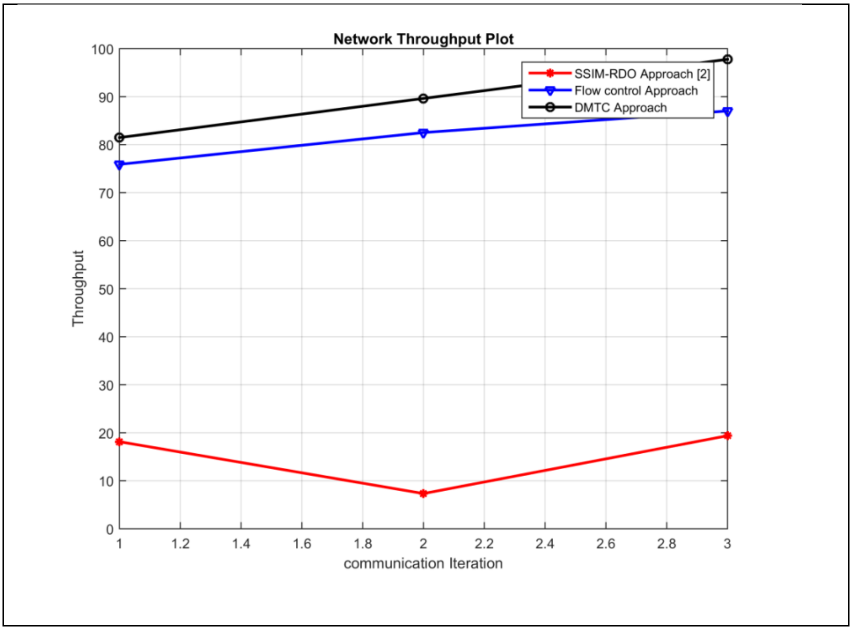


Figure 6-21:Network throughput plot

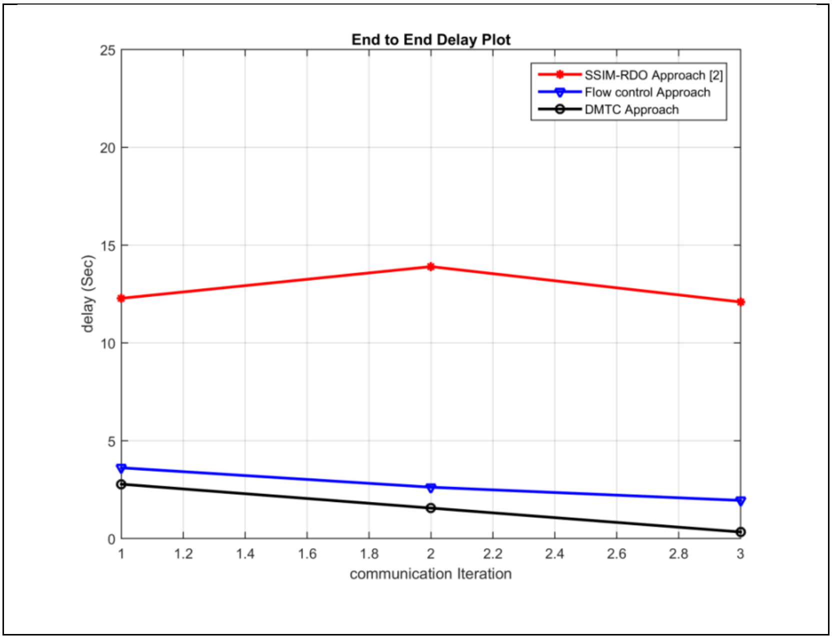


Figure 6-22:e2e delay plot

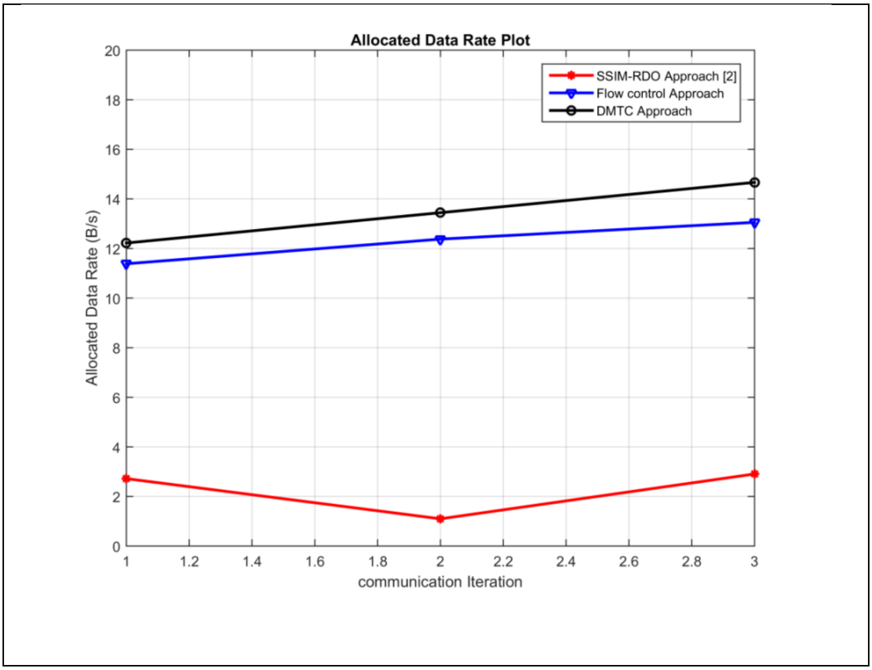


Figure 6-23: Assigned data rate plot

Case 2 Variance = 0.3 (Figs. 6.24, 6.25, 6.26, 6.27)

Fig. 6.28, 6.29, 6.30 and 6.31 show throughput, route overhead, allocated data rate and e2e delay of SSIM-RDO, FC (DMTC without channel noise) and DMTC scheme at variance = 0.3. According to this execution, channel noise level at 30% was measured. The graph reveals that the computation of DMTC is improved than conventional scheme of SSIM-RDO owing to its efficient DFC in DMTC mechanism. In addition, there exists some enhancement in DMTC from its corresponding item, i.e. FC exclusive of noise contemplation. In general, a raise in channel noise has an effect on aforesaid quality metrics. However, it is obvious from the graphs, DMTC scheme doesn't allow it to reduce, instead it is more or less stable. As a result, the suggested scheme is evaluated as an enhancement in the performance.

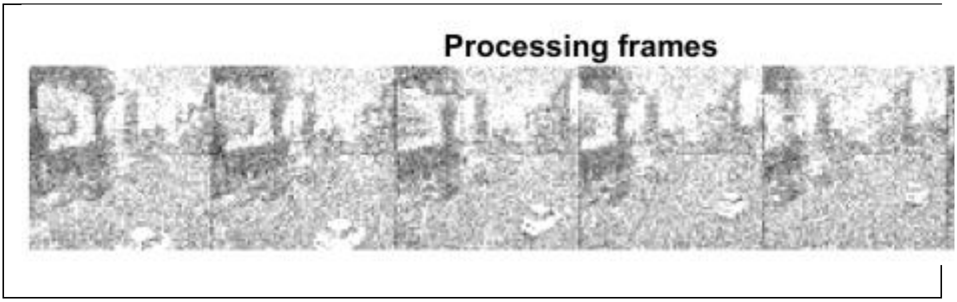


Figure 6-24:Noised sample

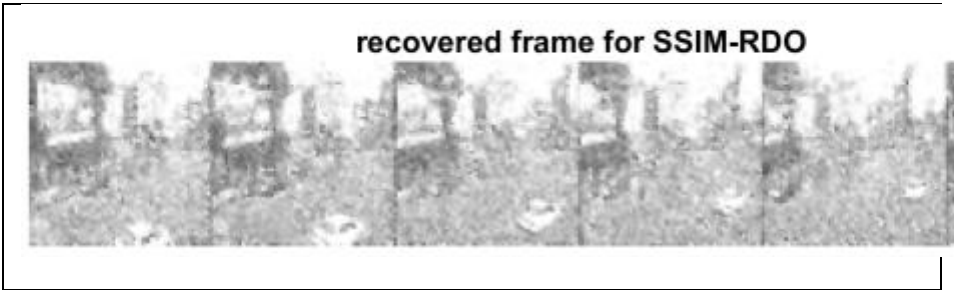


Figure 6-25:Recovered sample by means of SSIM model

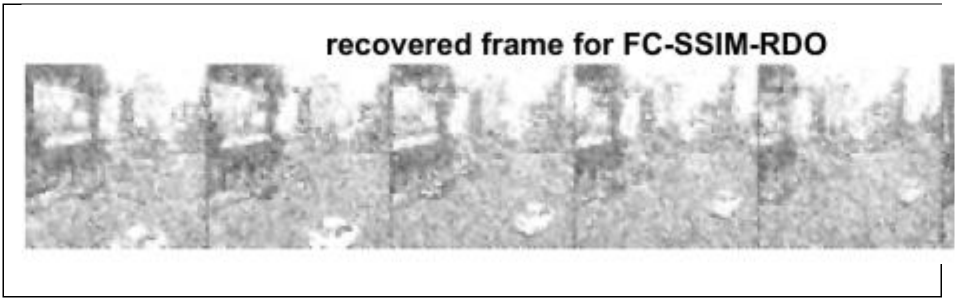


Figure 6-26:Recovered sample by means of FC model

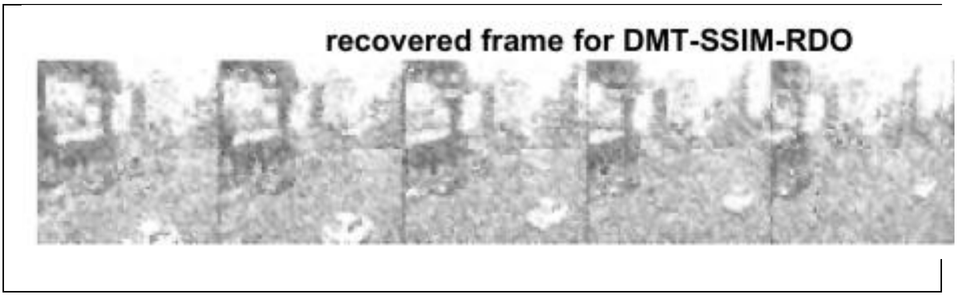


Figure 6-27: Recovered sample by means of DMTC model

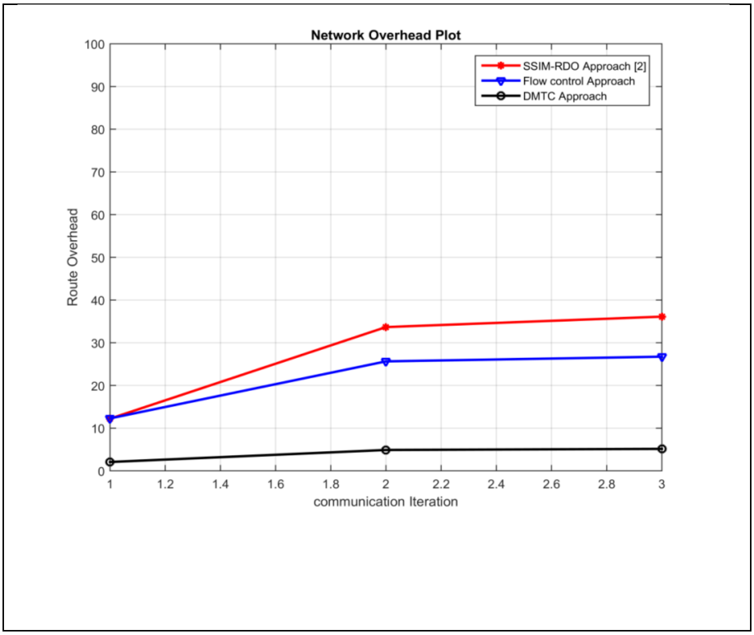


Figure 6-28: Route overhead plot

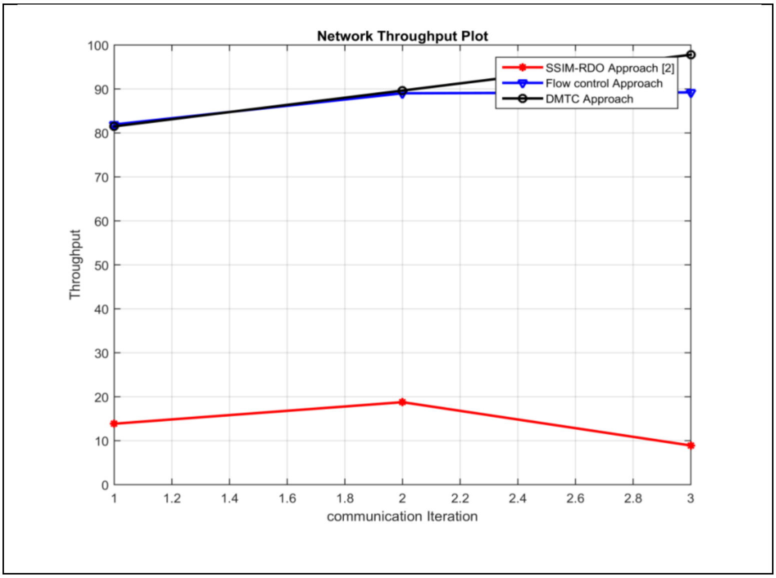


Figure 6-29: Network throughput plot

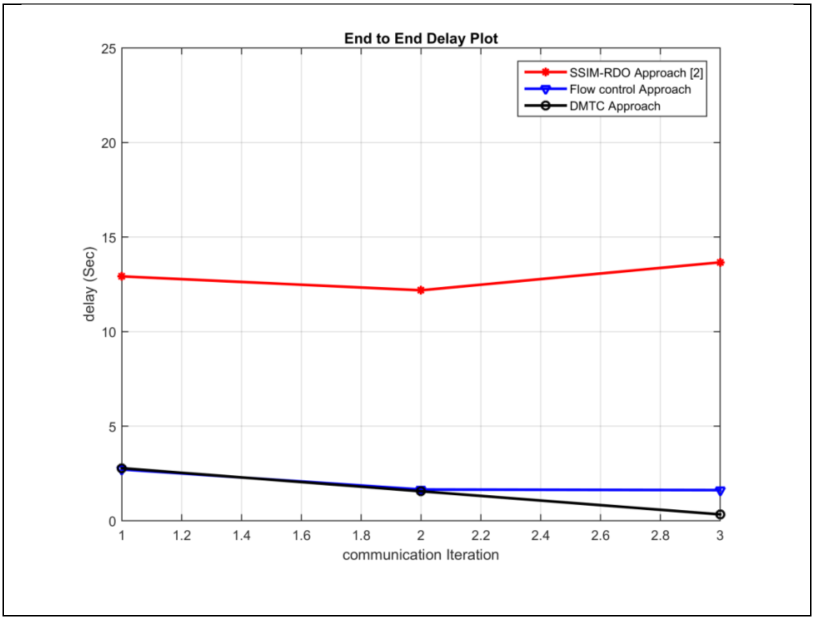


Figure 6-30: e2e delay plot

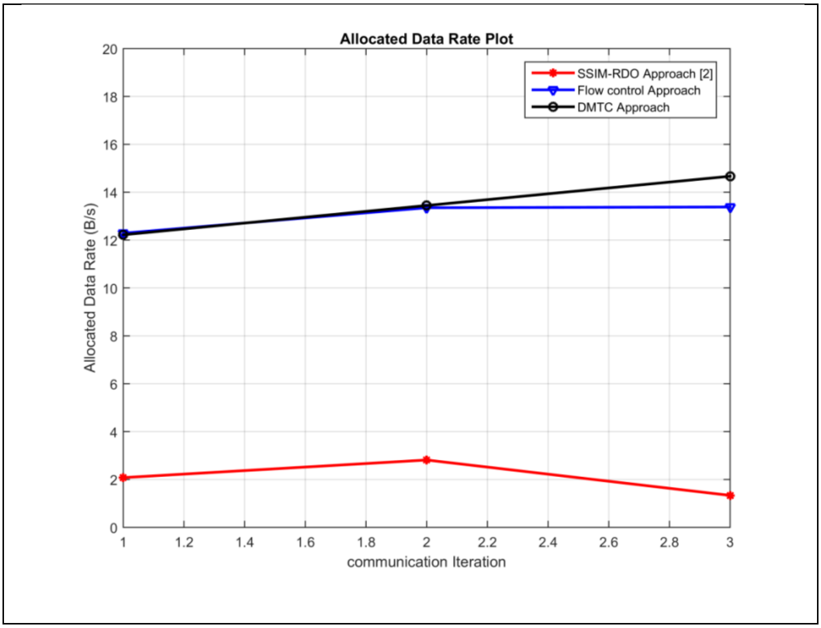


Figure 6-31: Allocated data rate plot

Case 3 Variance = 0.2 (Figs. 6.32, 6.33, 6.34, 6.35).

Fig. 6.32, Fig. 6.33, Fig. 6.34 and Fig. 6.35 show throughput, route overhead, allocated data rate and e2e delay of SSIM-RDO, FC with DMTC and without DMTC channel noise scheme at variance = 0.2.

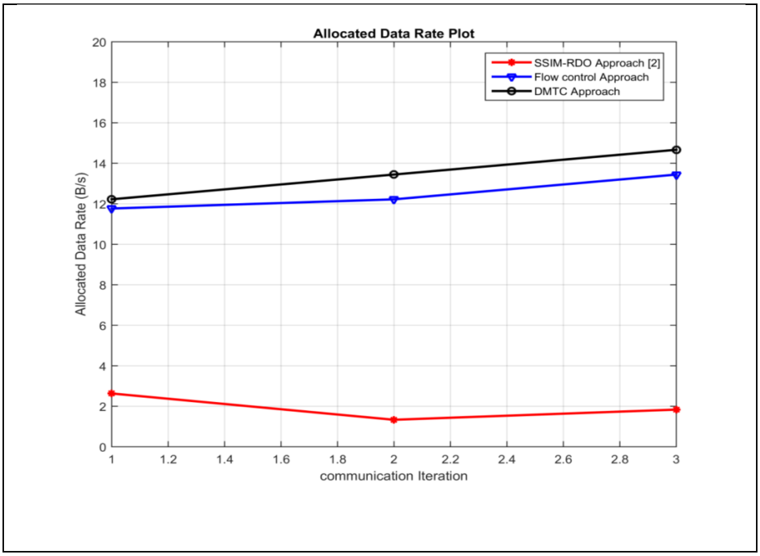


Figure 6-32: Allocated data rate plot

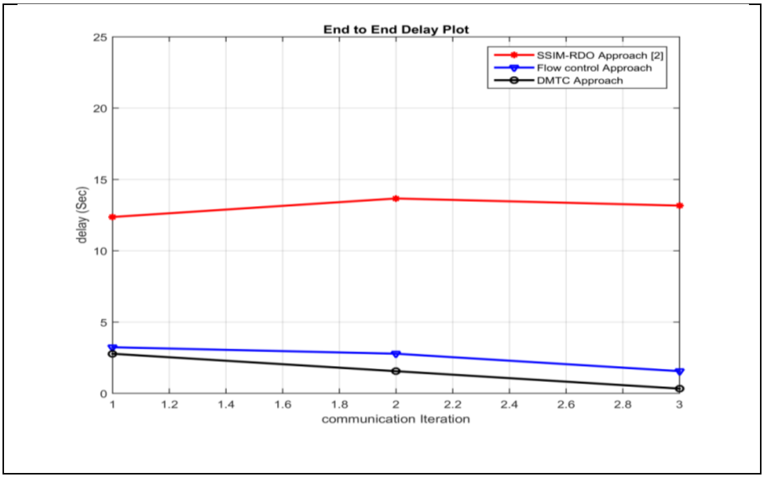


Figure 6-33: End-to-end delay plot

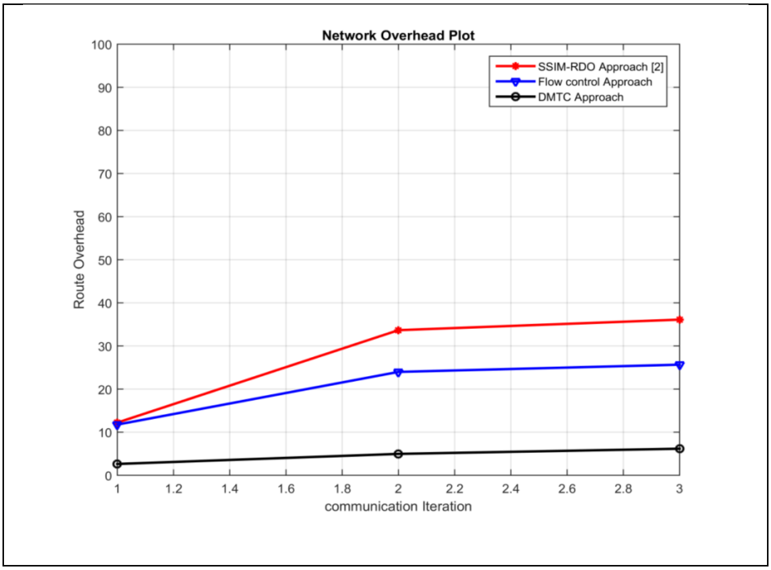


Figure 6-34: Route overhead plot

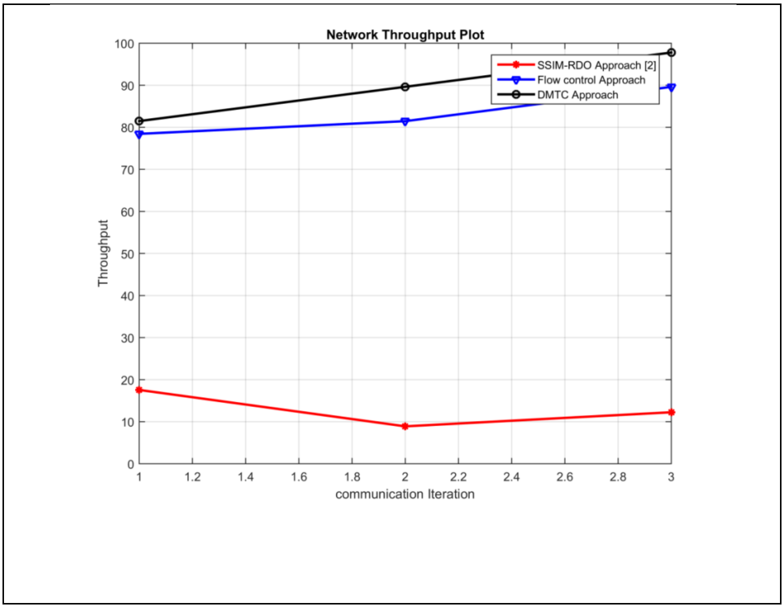


Figure 6-35: Network throughput plot

6.5 SUMMARY

The innovation of the suggested DMTC approach is a combination of the SSIM-RDO with data traffic congestion metric, and the modeling of ERC with a increased traffic flow was offered. A dynamic DFC with probabilistic route density was introduced to manage the flow of captivated video data over a multi-hop IEEE 802.11e network representation. Accordingly, in this scheme, the video quality enhancement was accomplished with ERC by means of the SSIM constraint. The ERC was then enhanced for increased throughput by means of DFC via ROA design. Moreover, from the investigational outcomes, it was obvious that an enhancement in system throughput together with video quality has been attained. Regardless of the efficiency of the implemented algorithm, there exists a scope for further development in this work. For example, with enhanced motion vector prediction or enhanced variable block size segmentation, H.265 codec can be deployed, thereby minimizing the communication overheads and improving the data compression.

CHAPTER 7. CONCLUSIONS AND FUTURE SCOPE

7.1 INTRODUCTION

The rapid growth of urban population throughout the globe necessitates the urgency to find the smarter ways to address the associated challenges. This concept has been accepted under the banner of ‘Smart City’. Out of many avenues of smart city, we tried to address some of the issues of ‘Smart Transportation’. Important and foremost technology in intelligent transport system is vehicle detection. This information can help central server to supervise the flow of traffic and road congestions in real time.

In several real time applications, traffic surveillance has expanded its significance depending on the speedy development in vehicular traffic density and the seriousness in scrutinizing it. In automatic analysis of traffic, the captured information regarding the traffic video need to be investigated based on the remote monitoring unit. In legacy realizations, numerous stationary video cameras were implemented at the traffic junctions or on highways in order to capture the traffic information of moving vehicles. From each camera, the video streams were provided to the centralized scrutinizing network. Thus, it was found to be desirable to direct the information to the scrutinizing station at a better rate and with high visual precision. Therefore, in our work we employed the rotating camera in place of stationary one, further reducing the cost of ownership.

In the transmission process, the video frames were forwarded in a multi-hop behavior in which, each position of link situated at a particular distance directs to the information to the subsequent target. In this routing process, the possibility of traffic occurs depending on the uninterrupted and volumetric streaming of data. Hence, several nodes were required to be made data traffic controlled in order to attain maximum throughput.

In many computer vision applications, strong and authentic instant foreground identification approach was considered to be as a critical concern. There has been substantial effort on identifying the objects and examining the motion. On the other hand, such efforts presupposes non-rotating camera. On the contrary, the difficulties related to the identification of objects over an extensive area depending on a rotating camera were addressed. Thus, a background model was deployed in order to discriminate between the background and the foreground objects. Background subtraction was defined to be as a distinctive technique for identifying the foreground by comparing each new frame based on an implemented scheme of the background scene in image series, which is obtained from a camera. Typically, motion compensation was made essential with respect to the relevant background subtraction which was provided to the motionless background. In fact, it was proved to be complicated to apprehend it to satisfactory pixel accurateness. Motion identification with a motion based viewing sensor has established substantial consideration of the researchers. Here, real-time foreground partitioning was also regarded to be as a challenging difficulty. This application includes computerized visual surveillance, vehicle-borne VS, identification of objects and tracking with camera and other capturing devices. In such circumstances, background subtraction approach cannot be applied in a direct manner. Motion compensation is essentially required to balance for the motion that occurs due to the moving sensors. Generally, a motion based approach related to the background was assumed and hence, the parameters related to the motion were predictable. Then, the background was recorded perfectly and the foreground was identified based on the pixel level.

Several essential postulations were done so that, the motion based approach was made adequately precise and thus, the constraints of the motion based approach was also projected precisely. Similarly, the sensing lenses were considered to be free from deformations. Thus, these postulations were regarded to be more complicated for recognition. These approaches were considered to be time-uncontrollable and thus, it was inappropriate for real-time applications. Depending on the estimation of the motion based approaches, the existing and the background image cannot deform and record in an efficient manner. This difficulty was also met with respect to the sequential differentiation approach. The exploitation of background modeling for the identification of the motion based object is very familiar in various applications. In the scene such as VS, the background modeling approach was implemented by attaining a background image without motionless objects and such circumstances were probably infrequent. Similarly in some circumstances, the background was considered to be inaccessible and there arises a variation in the clarification circumstances. Moreover, object was eradicated or established from the scenes. Thus, several background modeling approaches were implemented based on these challenges and thus, it was required to make them more speedy and adaptive.

This chapter concludes the research work related to the detection of moving objects in the VS system using several motion estimation approaches. This approach was formulated to improve the performance in terms of detecting the moving objects in the VS system. Here, it was seen that motion estimationapproaches pooledwith coding schemeswere the most promising image processing method for increasing robustness against the streaming of video information with high quality. Thus the assortment of this motion estimation based coding approach in VS system has emanated as an indispensable research concept.This research work utilizes several techniques for the effective detection of moving objects as well as the efficient streaming of video information in VS system using several motion estimation approaches. Thus, this research work involves the effective streaming of video information in VS system by processing several advanced techniques.

- The first work introduces the statistical background subtraction model for identifying the moving objects in case for a rotating camera. Here, the background model was evaluated in both spatial and temporal domain with respect to the distribution of each pixel in the background. With respect to this statistical model, the current frames with each pixel were classified in to foreground and background classes. From the simulation results it was evident that, this approach attains better accuracy in terms of template localization and thus leads to the improvement of accurateness in coding and coding speed when compared with the other standard approaches.

- The second work investigates about the improved identification of objects with respect to the prediction of region in online VS system using improved identification of objects depending on R-FSBMA approach. Here, a least mean estimator approach was deployed and thus, the intermittent full search motion estimator logic was defined for attaining the foreground moving substances from the video sequences, which was captured from a rotating sensor. From the simulation results it was obvious that, this approach achieves accurate detection rate in case of authentic moving object, thus minimizes the redundant information by eradicating the unnecessary background information. Thereby increasing the compression level and making suitable for real time application.
- The third work inspects about the consequence of using energy interpolated template coding for the identification of objects in case of compressing the video in traffic surveillance applications. Here, the dynamic assortment of templates was attained depending on the energy interpolation of consecutive frame information over some instant of time, rather than only two consecutive frames information. Thus, a coherent histogram approach was also implemented in order to create a precise template to attain enhancement in compression technique. From the experimental results it was clear that, this proposed approach attains enhanced reduction in processing time when compared with the other traditional approaches without affecting the quality of received data.
- The fourth work scrutinizes about the diffusion of information through the wireless media and leads to the progressive streaming of video information for the traffic surveillance. Here, during the streaming of video, high quality was maintained in spite of the compressed transmission of information. This leads to the implementation of a novel dual metric traffic control method, in which both the metrics such as traffic in data and deformation were considered. Thus, it mainly depends on the enhanced SSIM approach in which it was integrated with the rate of allocation approach which then, serves as a function. The experimental results show that, the performance of the developed method was analyzedin terms of video quality and data throughput in comparison with the existingtechniques. This

ensures the employability of the approach for edge computing in smart city infrastructure.

7.2 MAIN FINDINGS

- In moving object detection with statistical method, the performance measures of the proposed methodology was estimated in terms of positive measures such as accuracy, sensitivity, specificity and precision. The performance of the implemented methodology in terms of negative measures denotes that, the False Positive Rate (FPR) of the suggested scheme was 89.47% better than conventional approach and on considering False Discovery Rate (FDR), the adopted scheme was 37.14% superior to the traditional scheme. Thus the enhanced computation of the proposed approach has been validated successfully.
- In the enhanced object detection based on R-FSBMA, the redundant coefficients of the proposed scheme for mean filter, median filter and LMS filter was 42.1%, 47.37% and 47.37% better than FSBMA schemes. In addition, the motion element detection for mean filter, median filter and LMS filter was 47.37%, 48.65% and 39.47% superior to FSBMA techniques.
- In template coding based object detection, the computation time, data overhead, motion element detected, MSE, PSNR, redundant coefficients and SSIM were computed in terms of various observations. The computational time of the proposed EI-HIST at first observation was 82.85% better than TMP and 70% better than HIST approaches. Similarly, the data overhead at first observation was 67.27% superior to TMP and 60% superior to HIST scheme.
- In data diffusion through wireless media, it was found our proposed DMTC have better throughput, route overhead, allocated data rate and e2e delay than SSIM-RDO, FC (DMTC without channel noise) scheme.

7.3 FUTURE SCOPE

The future scope of this research work for the effective and efficient detection of moving objects and streaming of video information in VS system using motion estimation approaches is enlisted as follows,

- In case of robotic or automated vehicle support system, there arises a continuous variation in the background with respect to the motion of camera, which leads to the functioning of various enhanced and adaptive segmentation approaches.
- Computations based on nature inspired algorithms, such as genetic one, can be developed for MV estimation for better performance.
- In case of non-static camera, a template dependent segmentation approach can be deployed in combination with the other tracking approaches.
- It would be considered idyllic, if the system can automatically update its model with less manual supervision. Thus, several self-updating and learning algorithms can be used to benefit the system.
- An enhancement in ‘speed of computation’ for ‘block search’ scheme could be achieved by executing the controller with respect to the benefits of the imminent tendencies in multiprocessor system, thereby making more suitable for edge computing.
- The background modelling and foreground object identification component can be investigated through fuzzy models, which offers a good potential.
- An enhancement in our proposed steaming of video sequence approach could be achieved for H.265 codec system.
- Since last few decades an important network research area has been quality of service (QoS) support, specifically throughput and end to end delay. We believe, still this is an open problem for time variant wireless channel.

REFERENCES

[1] H. Kim and H. Lee, “A Low-Power Surveillance Video Coding System with Early Background Subtraction and Adaptive Frame Memory Compression,” IEEE Transactions on Consumer Electronics 63, pp. 359–367, 2017.

[2] T. Nishi and H. Fujiyoshi, “Object-based video coding using pixel state analysis,” Proc. - Int. Conf. Pattern Recognit., Proceedings of the 17th International Conference on, vol. 3, pp. 306-309. IEEE, 2004.

[3] T. Liu, Z. Wu, M. Zeng, Q. Jiang, and L. Hu, “More successful recognition: Seeking the relation of video object detection performance with video coding parameters,” 2015 12th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process. ICCWAMTIP 2015, pp. 184–187, 2016.

[4] Lingchao Kong and Rui Dai, “Multimedia Capturing, Mining, and Streaming,” Object-Detection Based Video Compression for Wireless Surveillance Systems, IEEE, pp. 76–85, 2017.

[5] S. H. Shaikh, K. Saeed, and N. Chaki, “Moving Object Detection Using Background Subtraction,” pp. 15-23. Springer, Cham, 2014.

[6] Karasulu, Bahadir, and Serdar Korukoglu. "Moving object detection and tracking in videos." Performance Evaluation Software. Springer New York,. 7-30, 2013.

[7] Lokeswari, P. N., K. ChandraSekhar, and Mr Sathiyaraj. “Adaptive Video Data Streaming And Sharing in Cloud,” International Journal of Computer Science and Mobile Computing, vol. 3, no. 7, pp. 133–139, 2014

[8] M.Sona , D.Daniel , S.Vanitha, “A Survey on Efficient Video Sharing and Streaming in Cloud Environment Using Vc,” International Journal of Innovative Research in Computer and Communication Engineering, pp. 1775–1780, 2013.

[9] Konda, Krishna Reddy, et al. "Real-time moving object detection and segmentation in H. 264 video streams." Broadband Multimedia Systems and Broadcasting (BMSB), International Symposium on. IEEE, pp. 1-6, 2017.

[10] R. V. Babu and A. Makur, “Object-based Surveillance Video Compression using Foreground Motion Compensation In Control, Automation, Robotics and Vision, , ICARCV'06. 9th International Conference on, pp. 1-6. IEEE, 2006.

[11] E. Pereira et al., "Imported from Como brasileiros profissionalizaram a criação de imagens de humor. Veja mais no UOL. Acesse: http://uol.com/bbj8TH," Int. Encycl. Commun. Theory Philos., vol. 11, no. 1, pp. 39–59, 2016.

[12] S. Pudlewski, N. Cen, Z. Guan, and T. Melodia, "Video transmission over lossy wireless networks: A cross-layer perspective," Journal of Selected Topics in Signal Processing, vol. 9, no. 1, pp. 6–21, 2015 Feb.

[13] X. Zhu and B. Girod, "Video Streaming over Wireless Networks," . In Signal Processing Conference, 15th European , IEEE, pp. 1462–1466, 2007.

[14] K. Lin, J. Song, J. Luo, W. Ji, M. Shamim Hossain, and A. Ghoneim, "Green Video Transmission in the Mobile Cloud Networks," IEEE Trans. Circuits Syst. Video Technol., vol. 27, no. 1, pp. 159–169, 2017.

[15] F. Fitzek and P. Seeling, Martin Reisslein "Video streaming in wireless internet,"Electrical Engineering and Applied Signa l Processing Series pp. 1–102, 2004

[16] V. Tsakanikas and T. Dagiuklas, "Video surveillance systems-current status and future trends," Elsevier, Comput. Electr. Eng., vol. 0, pp. 1–18, 2017.

[17] K. Kardas and N. K. Cicekli, "SVAS: Surveillance Video Analysis System," Expert Syst. Appl., Elsevier, vol. 89, pp. 343–361, 2017.

[18] W. Huang, H. Ding, and G. Chen, "A novel deep multi-channel residual networks-based metric learning method for moving human localization in video surveillance," Signal Processing, Elsevier, vol. 142, pp. 104–113, 2018.

[19] K. Zhang, Z. Huang, and S. Zhang, "Using an optimization algorithm to establish a network of video surveillance for the protection of Golden Camellia," Ecol. Inform., Elsevier, vol. 42, pp. 32–37, 2017.

[20] Z. Sun, Q. Zhang, Y. Li, and Y. Tan, "DPPDL: a Dynamic Partial-Parallel Data Layout for Green Video Surveillance Storage," IEEE Trans. Circuits Syst. Video Technol., vol. 8215, no. c, pp. 1–1, 2016.

[21] S. Javanbakhti, S. Zinger, and P. H. N. de With, "Fast semantic region analysis for surveillance video databases," 2017 IEEE Int. Conf. Consum. Electron., pp. 25–26, 2017.

[22] H. Seibel, S. Goldenstein, and A. Rocha, “Eyes on the Target: Super-Resolution and License-Plate Recognition in Low-Quality Surveillance Videos,” IEEE Access, vol. 5, pp. 20020–20035, 2017.

[23] M. Bilal, A. Khan, M. U. K. Khan, and C.-M. Kyung, “A Low-Complexity Pedestrian Detection Framework for Smart Video Surveillance Systems,” IEEE Trans. Circuits Syst. Video Technol., vol. 27, no. 10, pp. 2260–2273, 2017.

[24] X. Chen, J. N. Hwang, D. Meng, K. H. Lee, R. L. De Queiroz, and F. M. Yeh, “A Quality-of-Content-Based Joint Source and Channel Coding for Human Detections in a Mobile Surveillance Cloud,” IEEE Trans. Circuits Syst. Video Technol., vol. 27, no. 1, pp. 19–31, 2017.

[25] J. Sérot, L. Maggiani, F. Berry, and C. Bourrasset, “Dataflow object detection system for FPGA-based smart camera,” IET Circuits, Devices Syst., vol. 10, no. 4, pp. 280–291, 2016.

[26] J. Chen, Y. Wang, and H. Wu, “Coded aperture compressive imaging array applied for surveillance systems,” Journal of Systems Engineering and Electronics ., vol. 24, no. 6, pp. 1019–1028, 2013.

[27] S. B. Lee and Y. S. Ho, “Temporally consistent depth map estimation for 3D video generation and coding,” China Communication. IEEE, vol. 10, no. 5, pp. 39–49, 2013.

[28] F. L. Lian, Y. C. Lin, C. T. Kuo, and J. H. Jean, “Voting-based motion estimation for real-time video transmission in networked mobile camera systems,” IEEE Trans. Ind. Informatics, vol. 9, no. 1, pp. 172–180, 2013.

[29] L. Liu, Z. Li, and E. J. Delp, “Efficient and low-complexity surveillance video compression using backward-channel aware Wyner-Ziv video coding,” IEEE Trans. Circuits Syst. Video Technol., vol. 19, no. 4, pp. 453–465, 2009.

[30] Z. Liu, L. Li, Y. Song, S. Li, S. Goto, and T. Ikenaga, “Motion feature and hadamard coefficient-based fast multiple reference frame motion estimation for H.264,” IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 5, pp. 620–632, 2008.

[31] R. Zhang, S. Zhang, and S. Yu, “Moving objects detection method based on brightness distortion and chromaticity distortion,” IEEE Trans. Consum. Electron., vol. 53, no. 3, pp. 1177–1185, 2007.

[32] Koskinen L, Paasio A, Halonen KA. Motion estimation computational complexity reduction with CNN shape segmentation. IEEE transactions on circuits and systems for video technology. (6):771-7.2005

[33] J. H. Ko, B. A. Mudassar, and S. Mukhopadhyay, “An Energy-Efficient Wireless Video Sensor Node for Moving Object Surveillance,” IEEE Trans. Multi-Scale Comput. Syst., vol. 1, no. 1, pp. 7–18, 2015.

[34] S. Goel, Y. Ismail, and M. Bayoumi, “High-speed Motion Estimation Architecture for Real-time Video Transmission,” The Computer. Journal., vol. 55, no. 1, pp. 35–46, 2012.

[35] L. Zhou, B. Zheng, a. Wei, B. Geller, and J. Cui, “Joint Routing and Rate Control Scheme for Multi-Stream High-Definition Video Transmission over Wireless Home Networks,” The Computer. Journal., vol. 52, no. 8, pp. 950–959, 2008.

[36] J. Wu, B. Cheng, and M. Wang, “Improving Multipath Video Transmission With Raptor Codes in Heterogeneous Wireless Networks,” IEEE Trans. Multimed., vol. 9210, no. c, pp. 1–16, 2017.

[37] Z. Zhang, D. Liu, X. Ma, and X. Wang, “ECast: An enhanced video transmission design for wireless multicast systems over fading channels,” IEEE Syst. J., vol. PP, no. 99, pp. 1–12, 2015.

[38] M. Azimi, R. Boitard, M. T. Pourazad, and P. Nasiopoulos, “Performance evaluation of single layer HDR video transmission pipelines,” IEEE Trans. Consum. Electron., vol. 63, no. 3, pp. 267–276, 2017.

[39] M. A. Kourtis, H. Koumaras, G. Xilouris and F. Liberal, "An NFV-Based Video Quality Assessment Method over 5G Small Cell Networks," IEEE Multi Media, vol. 24, no. 4, pp. 68-78, 2017.

[40] Jianhua Lu and M. L. Liou, "A simple and efficient search algorithm for block-matching motion estimation," IEEE Transactions on Circuits and Systems for Video Technology, vol. 7, no. 2, pp. 429-433, Apr 1997.

[41] M. Wang, X. S. Hua, J. Tang and R. Hong, "Beyond Distance Measurement: Constructing Neighborhood Similarity for Video Annotation," IEEE Transactions on Multimedia, vol. 11, no. 3, pp. 465-476, April 2009.

[42] Pinghua Zhao, Yanwei Liu, Jinxia Liu, Song Ci and Ruixiao Yao, "SSIM-based error-resilient rate-distortion optimization of H.264/AVC video coding

for wireless streaming," Signal Processing: Image Communication, Elsevier, vol. 29, no. 3, pp. 303-315, March 2014.

[43] Arun Sankisa, Katerina Pandremmenou, Peshala V. Pahalawatta, Lisimachos P. Kondi and Aggelos K. Katsaggelos, "SSIM-Based Distortion Estimation for Optimized Video Transmission over Inherently Noisy Channels," International Journal of Multimedia Data Engineering and Management, vol. 7, no. 3, 2016.

[44] L. Zhou, B. Zheng, A. Wei, B. Geller and J. Cui, "A Robust Resolution-Enhancement Scheme for Video Transmission Over Mobile Ad-Hoc Networks," IEEE Transactions on Broadcasting, vol. 54, no. 2, pp. 312-321, June 2008.

[45] J. Wang, S. Wang and Z. Wang, "Asymmetrically Compressed Stereoscopic 3D Videos: Quality Assessment and Rate-Distortion Performance Evaluation," IEEE Transactions on Image Processing, vol. 26, no. 3, pp. 1330-1343, March 2017.

[46] Q. Xu, Z. Wu, L. Su, L. Qin, S. Jiang and Q. Huang, "Bridging the gap between objective score and subjective preference in video quality assessment," 2010 IEEE International Conference on Multimedia and Expo, Suntec City, pp. 908-913, 2010.

[47] Changick Kim and Jenq-Neng Hwang, "Fast and automatic video object segmentation and tracking for content-based applications," IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 2, pp. 122-129, Feb 2002.

[48] Zhijun Lei and Nicolas D.Georganas, "Adaptive video transcoding and streaming over wireless channels," Journal of Systems and Software, vol. 75, no. 3, pp. 253-270, March 2005.

[49] W. Xiang, P. Gao and Q. Peng, "Robust Multiview Three-Dimensional Video Communications Based on Distributed Video Coding," IEEE Systems Journal, vol. 11, no. 4, pp. 2456-2466, Dec. 2017.

[50] V. Mezaris, I. Kompatsiaris and M. G. Strintzis, "Video object segmentation using Bayes-based temporal tracking and trajectory-based region merging," IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 6, pp. 782-795, June 2004.

[51] O. Perkasa and D. H. Widyantoro, "Video-based system development for automatic traffic monitoring," 2014 International Conference on Electrical Engineering and Computer Science (ICEECS), Kuta, pp. 240-244, 2014.

[52] Sunhun Lee and Kwangsue Chung, "Combining the rate adaptation and quality adaptation schemes for wireless videostreaming," Journal of Visual Communication and Image Representation, vol. 19, no. 8, pp. 508-519, 2008.

[53] Ling Shao, Simon Jones and Xuelong Li, "Efficient Search and Localization of Human Actions in Video Databases," IEEE transcactions on circuits and systems for video technology, vol. 24, no. 3, 2014.

[54] C. E. Erdem, B. Sankur and A. M. Tekalp, "Performance measures for video object segmentation and tracking," IEEE Transactions on Image Processing, vol. 13, no. 7, pp. 937-951, July 2004.

[55] Tianmi Chen, Xiaoyan Sun and Feng Wu, "Predictive Patch Matching for Inter Frame Coding ," Visual Communication and image processing, vol. 7744, p 774412, 2010.

[56] Ivan Laptev, "On Space-Time Interest Points," International Journal of Computer Vision, vol. 64, n0. 2-3, pp. 107-123, 2005.

[57] James W. Davis and Aaron F. Bobick, "The Representation and Recognition of Action Using Temporal Templates," IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97), no. 42, 1997.

[58] Chalidabhongse TH, Kim K, Harwood D, Davis L. A perturbation method for evaluating background subtraction algorithms. InJoint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (pp. 11-12),2003.

[59] L. Maddalena and A. Petrosino, "A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications," IEEE Transactions on Image Processing, vol. 17, no. 7, pp. 1168-1177, July 2008.

[60] R. H. Evangelio, M. Patzold, I. Keller and T. Sikora, "Adaptively Splitted GMM With Feedback Improvement for the Task of Background Subtraction," IEEE Transactions on Information Forensics and Security, vol. 9, no. 5, pp. 863-874, May 2014.

[61] Z. Huang, R. Hu and Z. Wang, "Background Subtraction With Video Coding," IEEE Signal Processing Letters, vol. 20, no. 11, pp. 1058-1061, Nov. 2013.

[62] F. Chen, H. Li, L. Li, D. Liu and F. Wu, "Block-Composed Background Reference for High Efficiency Video Coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 27, no. 12, pp. 2639-2651, Dec. 2017.

[63] K. V. Sriharsha and N. V. Rao, "Dynamic scene analysis using Kalman filter and mean shift tracking algorithms," 2015 6th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Denton, TX, pp. 1-8, 2015.

[64] Dar-Shyang Lee, "Effective Gaussian mixture learning for video background subtraction," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pp. 827-832, May 2005.

[65] W. Xiang, C. Zhu, C. K. Siew, Y. Xu and M. Liu, "Forward Error Correction-Based 2-D Layered Multiple Description Coding for Error-Resilient H.264 SVC Video Transmission," IEEE Transactions on Circuits and Systems for Video Technology, vol. 19, no. 12, pp. 1730-1738, Dec. 2009.

[66] D. Mukherjee, Q. M. J. Wu and T. M. Nguyen, "Gaussian Mixture Model With Advanced Distance Measure Based on Support Weights and Histogram of Gradients for Background Suppression," IEEE Transactions on Industrial Informatics, vol. 10, no. 2, pp. 1086-1096, May 2014.

[67] Z. Chen, P. V. Pahalawatta, A. M. Tourapis and D. Wu, "Improved Estimation of Transmission Distortion for Error-Resilient Video Coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 4, pp. 636-647, April 2012.

[68] Y. Shen et al., "Real-Time and Robust Compressive Background Subtraction for Embedded Camera Networks," IEEE Transactions on Mobile Computing, vol. 15, no. 2, pp. 406-418, Feb. 1 2016.

[69] L. Maddalena and A. Petrosino, "Stopped Object Detection by Learning Foreground Model in Videos," IEEE Transactions on Neural Networks and Learning Systems, vol. 24, no. 5, pp. 723-735, May 2013.

[70] H. Bhaskar, L. Mihaylova and A. Achim, "Video Foreground Detection Based on Symmetric Alpha-Stable Mixture Models," IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, no. 8, pp. 1133-1138, Aug. 2010.

[71] Yi Liu and Yuan F. Zheng, "Video Object Segmentation and Tracking Using Si-Learning Classification," IEEE Transactions on circuits and system for video technology, vol. 15, no. 7, 2005.

[72] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149), Fort Collins, CO, vol. 2, pp. 252, 1999.

[73] Thierry Bouwmans, Fida El Baf and Bertrand Vachon, "Background Modeling using Mixture of Gaussians for Foreground Detection - A Survey," A Survey. Recent Patents on Computer Science, Bentham Science Publishers, vol. 1, no. 3, pp. 219-237, 2008.

[74] Q. Zhu, Z. Song, Y. Xie and L. Wang, "A Novel Recursive Bayesian Learning-Based Method for the Efficient and Accurate Segmentation of Video With Dynamic Background," IEEE Transactions on Image Processing, vol. 21, no. 9, pp. 3865-3876, Sept. 2012.

[75] W. Wang, J. Luo and H. Qi, "Action recognition across cameras via reconstructable paths," 2013 Seventh International Conference on Distributed Smart Cameras (ICDSC), Palm Springs, CA, pp. 1-6, 2013.

[76] R. Zhang, W. Gong, V. Grzeda, A. Yaworski and M. Greenspan, "An Adaptive Learning Rate Method for Improving Adaptability of Background Models," IEEE Signal Processing Letters, vol. 20, no. 12, pp. 1266-1269, Dec. 2013.

[77] T. H. Thi, J. Zhang, L. Cheng, L. Wang and S. Satoh, "Human Action Recognition and Localization in Video Using Structured Learning of Local Space-Time Features," 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, Boston, MA, pp. 204-211, 2010.

[78] H.B.Kazemian and K.Ouazzane, "Neuro-Fuzzy approach to video transmission over ZigBee," Neurocomputing, Elsevier, vol. 104, pp. 127-137, 2013.

[79] H. A. Abdelali, F. Essannouni, L. Essannouni and D. Aboutajdine, "Algorithm for moving object detection and tracking in video sequence using color feature," 2014 Second World Conference on Complex Systems (WCCS), Agadir, pp. 690-693, 2014.

[80] J. C. Schmidt and K. Rose, "Jointly optimized mode decisions in redundant video streaming," 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, pp. 797-800, 2009.

[81] W. Wang, J. Yang and W. Gao, "Modeling Background and Segmenting Moving Objects from Compressed Video," IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, no. 5, pp. 670-681, May 2008.

[82] J. A. Robinson and Y. Shu, "Zerotree pattern coding of motion picture residues for error-resilient transmission of video sequences," IEEE Journal on Selected Areas in Communications, vol. 18, no. 6, pp. 1099-1110, June 2000.

[83] S. Chen, J. Zhang, Y. Li and J. Zhang, "A Hierarchical Model Incorporating Segmented Regions and Pixel Descriptors for Video Background Subtraction," IEEE Transactions on Industrial Informatics, vol. 8, no. 1, pp. 118-127, Feb. 2012.

[84] A. Ghahremani and A. Mousavinia, "An efficient adaptive energy model based predictive Motion Estimation algorithm for video coding," 2014 IEEE International Conference on Image Processing (ICIP), Paris, pp. 3185-3189, 2014.

[85] H. Li, J. Tang, S. Wu, Y. Zhang and S. Lin, "Automatic Detection and Analysis of Player Action in Moving Background Sports Video Sequences," IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, no. 3, pp. 351-364, March 2010.

[86] B. Kamolrat, W. A. C. Fernando, M. Mrak and A. Kondoz, "Flexible motion model with variable size blocks for depth frames coding in colour-depth based 3D video coding," 2008 IEEE International Conference on Multimedia and Expo, Hannover, pp. 573-576, 2008.

[87] J. M. McHugh, J. Konrad, V. Saligrama and P. M. Jodoin, "Foreground-Adaptive Background Subtraction," IEEE Signal Processing Letters, vol. 16, no. 5, pp. 390-393, May 2009.

[88] E. A. Bernal, Q. Li, O. Bulan, W. Wu and S. Schweid, "Model-less and model-based computationally efficient motion estimation for video compression in transportation applications," 2016 IEEE Winter Applications of Computer Vision Workshops (WACVW), Lake Placid, NY, pp. 1-8, 2016.

[89] K. Muthuswamy and D. Rajan, "Particle filter framework for salient object detection in videos," IET Computer Vision, vol. 9, no. 3, pp. 428-438, 2015.

[90] Q. Zhang and K. N. Ngan, "Segmentation and Tracking Multiple Objects Under Occlusion From Multiview Video," IEEE Transactions on Image Processing, vol. 20, no. 11, pp. 3308-3313, Nov. 2011.

[91] G. Zhao, G. Ming, S. Wang and T. Wang, "Unequal Error Protection Schema for Wireless H.264 Video Transmission Based on Perceived Motion Energy Model," 2008 Second International Conference on Future Generation Communication and Networking Symposia, Sanya, pp. 158-161, 2008.

[92] Yubing Han, Zhihui Xu and Xiaoli Wang, "Video dynamic interpolation based on weighted shift interframe motion model," 2010 The 2nd International Conference on Industrial Mechatronics and Automation, Wuhan, Chinapp. 117-122, 2010.

[93] C. Lijun and H. Kaiqi, "Video-based crowd density estimation and prediction system for wide-area surveillance," in China Communications, vol. 10, no. 5, pp. 79-88, May 2013.

[94] Krystian Mikolajczyk and Cordelia Schmid, "An Affine Invariant Interest Point Detector," European Conference on Computer Vision ECCV: Computer Vision — ECCV, pp. 128-142, 2002.

[95] P. Dollar, V. Rabaud, G. Cottrell and S. Belongie, "Behavior recognition via sparse spatio-temporal features," 2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, pp. 65-72, 2005.

[96] Kalpana Seshadrinathan and Alan C. Bovik, "Motion-based Perceptual Quality Assessment of Video," Laboratory for image and video engineering, 2000.

[97] D. Koller et al., "Towards robust automatic traffic scene analysis in real-time," Proceedings of 1994 33rd IEEE Conference on Decision and Control, Lake Buena Vista, FL, vol.4, pp. 3776-3781, 1994.

[98] P.KaewTrakulpong and R. Bowden, “An improved adaptive background mixture model for real-time tracking with shadow detection”, In Second European Workshop on Advanced Video Based Surveillance Systems (AVBS2001), Sept 2001.

[99] C. H. Cheung and L. M. Po, “A novel cross-diamond search algorithm for fast block motionestimation,” IEEE Trans. Circuits Syst. Video Technol., vol. 12, no. 12, pp. 1168–1177, 2002.

[100] Jonathan Fabrizio, Séverine Dubuisson and Dominique Bereziat, "Motion compensation based on tangent distance prediction for video compression", Signal Processing: Image Communication,Elsevier, vol.27, no.2, pp.153-171, February 2012.

[101] Mengyao Ma, Oscar C. Au, Liwei Guo, S.-H. Gary Chan and Ling Hou, "Alternate motion-compensated prediction for error resilient video coding", Journal of Visual Communication and Image Representation, Elsevier, vol.19, no.I7, pp.437-449, October 2008.

[102] Ling Shao, Simon Jones, and Xuelong Li,“ Efficient Search and Localization of Human Motions in Video Databases”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 3, pp. 504-512, March 2014.

[103] Joaquin Zepeda, Mehmet Turkan, Dominique Thoreau, “Block Prediction using Approximate Template Matching”, 23rd European Signal Processing Conference (EUSIPCO)IEEE, 2015.

[104] Tianmi Chen, Xiaoyan Sun, and Feng Wu, “Predictive Patch Matching for Inter Frame Coding”, Visual Communications and Image Processing, Proc. of SPIE vol. 7744, 2010.

[105] Patil, S., Sanyal, R., & Prasad, R, “Efficient video coding in region prediction in online video surveillance”, In The 2015 International Conference on Image Processing, Computer Vision, & Pattern Recognition (IPCV), pp. 210–216, 2015.

[106] Chen X, Canagarajah N, Nunez-Yanez JL, Vitulli R. Lossless video compression based on backward adaptive pixel-based fast motion estimation. Signal Processing: Image Communication, Elsevier, 1;27(9):961-72, 2012.

[107] Zhao, M. C., Gong, X. Y., Que, X. R., Wang, W. D., & Cheng, S. D. (2012). Context-aware adaptive active queue management mechanism for improving video transmission over IEEE 802.11E WLAN. The Journal of China Universities of Posts and Telecommunications, 19(Suppl. 2), 65–72.

[108] Mohammad Ali Alavianmehr, "Video Foreground Detection Based on Adaptive Mixture Gaussian Model for Video Surveillance Systems", Journal of Traffic and Logistics Engineering, March 29, 2015.

CO-AUTHOR STATEMENTS

Co-author statements for the below mentioned scientific contributions are attached in the following pages.

Serial #	Contribution Details	Page
A	**Journal Publications**	
A.1	Patil, S., Sanyal, R., & Prasad, R, **“Progressive Streaming of Video Data for Traffic Surveillance”**, Springer’s Journal of Wireless Personal Communications, Vol. 100, Issue 2, pp 283-309, May 2018.	150
A.2	Patil, S., Sanyal, R., & Prasad, R,“**Energy Interpolated Template Coding For Video Compression In Traffic Surveillance Application”,** Journal of Mobile Multimedia – (Accepted)	151
A.3	Patil, S., “**Moving Object Detection Using Statistical Background Subtraction For A Rotating Camera**”, International Monthly Refereed Journal of Research In Management & Technology, pp. 67-71, Vol. 2, (ISSN – 2320-0073), Sept. 13.	152
A.4	Tonde, V., Patil, S., “**Real Time Background Subtraction On GPU Using CUDA**”, International Journal of Next Generation Computer Applications, Volume 1, Issue 5, 2013 (ISSN 2319-524X), Jan. 13.	153

B	**Conference Publications**	
B.1	Patil, S., Sanyal, R., & Prasad, R, "**Efficient video coding in region prediction in online video surveillance**", In The 2015 International Conference on Image Processing, Computer Vision, & Pattern Recognition (IPCV), pp. 210–216, World Congress (http://www.world-academy-of science.org/), USA, July 2015.	154
B.2	Bhate, S., Kulkarni, V., Lagad, S., Shinde, M., Patil, S. " **IoT Based Intelligent Traffic Signal System for Emergency Vehicles**", In Proceedings of 2nd International Conference of Inventive Communication and Computational Technologies (ICICCT 2018)IEEE Xplore Compliant, pp. 786-791, (ISBN: 978-1-5386-1974-2), April 2018.	155

AARHUS BSS | SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Declaration of co-authorship*

Full name of the PhD student: Shivprasad P. Patil.

This declaration concerns the following article/manuscript:

Title:	PROGRESSIVE STREAMING OF VIDEO DATA FOR TRAFFIC SURVEILLANCE
Authors:	Shivprasad P. Patil, Rajarshi Sanyal, and Prof. Ramjee Prasad

The article/manuscript is: Published ☒ Accepted ☐ Submitted ☐ In preparation ☐

If published, state full reference: Springer's Journal of Wireless Personal Communications, May 2018, Vol. 100, Issue 2, pp 283-309, USA.

If accepted or submitted, state journal:

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

A. Has essentially done all the work
B. Major contribution
C. Equal contribution
D. Minor contribution
E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	B
2. Planning of the experiments/methodology design and development	A
3. Involvement in the experimental work/clinical studies/data collection	A
4. Interpretation of the results	B
5. Writing of the first draft of the manuscript	A
6. Finalization of the manuscript and submission	C

Signatures of the co-authors

Date	Name	Signature
17/5/18	Prof. Ramjee Prasad	Ramjee Prasad
15/5/18	Dr. Rajarshi Sanyal	Rajarshi Sanyal

Date: 17/5/18

In case of further co-authors please attach appendix

AARHUS BSS

SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Declaration of co-authorship*

Full name of the PhD student: Shivprasad P. Patil.

This declaration concerns the following article/manuscript:

Title:	Energy Interpolated Template Coding For Video Compression In Traffic Surveillance Application
Authors:	Shivprasad P. Patil, Rajarshi Sanyal, and Prof. Ramjee Prasad

The article/manuscript is: Published ☐ Accepted ☒ Submitted ☐ In preparation ☐

If published, state full reference:

If accepted or submitted, state journal: Journal of Mobile Multimedia

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

A. Has essentially done all the work
B. Major contribution
C. Equal contribution
D. Minor contribution
E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	B
2. Planning of the experiments/methodology design and development	A
3. Involvement in the experimental work/clinical studies/data collection	A
4. Interpretation of the results	B
5. Writing of the first draft of the manuscript	A
6. Finalization of the manuscript and submission	C

Signatures of the co-authors

Date	Name	Signature
17/5/18	Prof. Ramjee Prasad	Rajee Prasad
15/5/18	Dr. Rajarshi Sanyal	Rajarshi Sanyal

Date: 13/5/18

In case of further co-authors please attach appendix

Signature of the PhD student

*As per policy the co-author statement will be published with the dissertation.

AARHUS BSS | SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Declaration of co-authorship*

Full name of the PhD student: Shivprasad P. Patil.

This declaration concerns the following article/manuscript:

Title:	Moving Object Detection Using Statistical Background Subtraction For A Rotating Camera
Authors:	Shivprasad P. Patil

The article/manuscript is: Published ☒ Accepted ☐ Submitted ☐ In preparation ☐

If published, state full reference: International Monthly Refereed Journal of Research In Management & Technology, pp. 67-71, Vol. 2, Sept. 13 (ISSN – 2320-0073)

If accepted or submitted, state journal:

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

- A. Has essentially done all the work
- B. Major contribution
- C. Equal contribution
- D. Minor contribution
- E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	A
2. Planning of the experiments/methodology design and development	A
3. Involvement in the experimental work/clinical studies/data collection	A
4. Interpretation of the results	A
5. Writing of the first draft of the manuscript	A
6. Finalization of the manuscript and submission	A

Signatures of the co-authors

Date	Name	Signature

In case of further co-authors please attach appendix

Date: 17/5/18

Signature of the PhD student

*As per policy the co-author statement will be published with the dissertation.

AARHUS BSS SCHOOL OF BUSINESS AND SOCIAL SCIENCES AARHUS UNIVERSITY

Declaration of co-authorship*

Full name of the PhD student: Shivprasad P. Patil.

This declaration concerns the following article/manuscript:

Title:	Real Time Background Subtraction On GPU Using CUDA
Authors:	Vandana Tonde, Shivprasad P. Patil

The article/manuscript is: Published ☒ Accepted ☐ Submitted ☐ In preparation ☐

If published, state full reference: International Journal of Next Generation Computer Applications, Volume 1, Issue 5, 2013 (ISSN 2319-524X)

If accepted or submitted, state journal:

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

A. Has essentially done all the work
B. Major contribution
C. Equal contribution
D. Minor contribution
E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	C
2. Planning of the experiments/methodology design and development	D
3. Involvement in the experimental work/clinical studies/data collection	D
4. Interpretation of the results	C
5. Writing of the first draft of the manuscript	D
6. Finalization of the manuscript and submission	C

Signatures of the co-authors

Date	Name	Signature
17/5/18	Vandana Tonde	VTonde

In case of further co-authors please attach appendix

Date: 17/5/18

Signature of the PhD student

*As per policy the co-author statement will be published with the dissertation.

AARHUS BSS | SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Declaration of co-authorship*

Full name of the PhD student: Shivprasad P. Patil.

This declaration concerns the following article/manuscript:

Title:	EFFICIENT VIDEO CODING IN REGION PREDICTION IN ONLINE VIDEO SURVEILLANCE
Authors:	Shivprasad P. Patil, Rajarshi Sanyal, and Prof. Ramjee Prasad

The article/manuscript is: Published ☒ Accepted ☐ Submitted ☐ In preparation ☐

If published, state full reference: In The 2015 International Conference on Image Processing, Computer Vision, & Pattern Recognition (IPCV), pp. 210–216, July 27-30, 2015. Las Vegas, USA, Congress (http://www.world-academy-of science.org/).

If accepted or submitted, state journal:

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

- A. Has essentially done all the work
- B. Major contribution
- C. Equal contribution
- D. Minor contribution
- E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	B
2. Planning of the experiments/methodology design and development	A
3. Involvement in the experimental work/clinical studies/data collection	A
4. Interpretation of the results	C
5. Writing of the first draft of the manuscript	A
6. Finalization of the manuscript and submission	C

Signatures of the co-authors

Date	Name	Signature
17/5/18	Prof. Ramjee Prasad	Ramjee Prasad
15/5/18	Dr. Rajarshi Sanyal	Rajarshi Sanyal

In case of further co-authors please attach appendix

Date: 17/5/18

Signature of the PhD student

*As per policy the co-author statement will be published with the dissertation.

AARHUS BSS **SCHOOL OF BUSINESS AND SOCIAL SCIENCES**
AARHUS UNIVERSITY

Declaration of co-authorship*

Full name of the PhD student: Shivprasad P. Patil.

This declaration concerns the following article/manuscript:

Title:	IoT Based Intelligent Traffic Signal System for Emergency Vehicles
Authors:	Shubhankar Bhate, Prasad Kulkarni, Shubham Lagad, Mahesh Shinde, Shivprasad P. Patil

The article/manuscript is: Published ☒ Accepted ☐ Submitted ☐ In preparation ☐

If published, state full reference: In Proceedings of 2nd International Conference of Inventive Communication and Computational Technologies (ICICCT 2018) IEEE Xplore Compliant, pp. 786-791, (ISBN: 978-1-5386-1974-2), April 2018.

If accepted or submitted, state journal:

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

A. Has essentially done all the work
B. Major contribution
C. Equal contribution
D. Minor contribution
E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	C
2. Planning of the experiments/methodology design and development	D
3. Involvement in the experimental work/clinical studies/data collection	D
4. Interpretation of the results	C
5. Writing of the first draft of the manuscript	D
6. Finalization of the manuscript and submission	C

Signatures of the co-authors

Date	Name	Signature
15/5/18	Mahesh Shinde	
15/5/18	Prasad Kulkarni	
15/5/18	Shubhankar Bhate	
15/5/18	Shubham Lagad	

In case of further co-authors please attach appendix

Date: 17/5/18

Signature of the PhD student

*As per policy the co-author statement will be published with the dissertation.