Analyzing & Designing a Framework of Video Editing

THESIS

SUBMITTED TO
BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
LUCKNOW

FOR THE DEGREE OF

Doctor of Philosophy

IN

COMPUTER SCIENCE

SUBMITTED BY

NEETISH KUMAR

ENROLLMENT NO.-1389/16

UNDER THE SUPERVISION OF

DR. DEEPA RAJ

ASSOCIATE PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE
SCHOOL FOR INFORMATION SCIENCE & TECHNOLOGY
BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
(A CENTRAL UNIVERSITY; NAAC- 'A' GRADE)
VIDYA VIHAR, RAEBARELI ROAD, LUCKNOW-226 025 (U.P.), INDIA

2021

Dedicated to my Parents…

CANDIDATE'S DECLARATION

I, **Neetish Kumar**, solemnly declare that the research work embodied in this thesis entitled **"Analyzing & Designing a Framework of Video Editing"** carried out by me under the guidance and supervision of **Dr. Deepa Raj, Associate Professor, Department of Computer Science, Babasaheb Bhimrao Ambedkar University (A Central University), Lucknow, India** is an original work and does not contain part of any work submitted for the award of any degree either in this University or any other University around the globe. It is further undertaken that the thesis is essentially free from all kinds of plagiarism.

Neetish Kumar

Date: 09/06/21

Place: Lucknow

(Neetish Kumar)
Research Scholar
Department of Computer Science
Babasaheb Bhimrao Ambedkar University, Lucknow

CERTIFICATE

This is to certify that the thesis titled **"Analyzing & Designing a Framework of Video Editing"** submitted by **Mr. Neetish Kumar** is an original research work and has not been previously submitted in part or full for the award of any other degree or diploma to this or any other university.

The thesis submitted to Babasaheb Bhimrao Ambedkar University Lucknow satisfies all the requirements as stipulated in the *Doctor of Philosophy (Ph.D.) regulations - 1999 as amended in 2013* and it is fit for submission and evaluation for the award of the degree of Doctor of Philosophy of the University.

Date: 09/06/21

Deepa Raj
Supervisor

9/6/21
Head of the Department

AKNOWLEDGEMENTS

ज्यों तिल माहि तेल है, ज्यों चकमक में आग।

तेरा साईं तुझमें है, जाग सके तो जाग ।।

Just as there is oil inside the sesame, and there is light inside the fire, just as our God is present inside us, if you can find it, find it – Sant Kabir.

Doctor of Philosophy is a rewarding journey, which would not be possible without the support of many people. As this journey comes to its end, I would like to take this opportunity to thank these amazing people who inspired me during the ups and downs of this pleasant experience.

First and foremost, I wish to express my sincere gratitude to my supervisor, ***Dr. Deepa Raj, Associate Professor, Department of Computer Science, Babasaheb Bhimrao Ambedkar University,*** *for her guidance, tremendous support and encouragement in completing this research and providing me extensive support for various research activities. She has been actively interested in my work and has always been available to advise me. Her invaluable insight, into technical and professional matters, has helped me immensely throughout my research work.*

My deep and sincere thanks to ***Prof. S. K. Dwivedi, Dean, School for Information Science and Technology and Head, Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, India.*** *His valuable suggestions made the contributions of the thesis more valuable.*

I will forever be thankful to ***Prof. Vipin Saxena, Department of Computer Science and Prof. R.A. Khan, Head, Department of Information technolog, Babasaheb Bhimrao Ambedkar University, Lucknow, India*** *for sparing their valuable time in giving insightful suggestions in publishing quality research papers.*

My deep regards and thanks to **Dr. Manoj Kumar, Dr. Narander Kumar and Dr. Shalini Chandra, Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, India** for extending all possible helps and support whenever required.

I am grateful to **Prof. Sanjay Singh, Vice Chancellor, Babasaheb Bhimrao Ambedkar University, Lucknow, India** for providing the excellent computation facilities in the University campus. Special thanks are extended to **University Grant Commission, India** for providing financial assistance to the Central University for doing research work.

My heartfelt thanks to my bosom friend **Jitendra Kumar Samariya**, **Snehlata** and my loving junior **Jayveer Singh** for their constant support and motivation in my research work.

I extend my sincere thanks to all the Research Scholars of **Department of Computer Science** for their generous help as and when needed during the research period.

I would also like to thank the non-teaching staff members of the **Department of Computer Science, Babasaheb Bhimrao Ambedkar University,** for keeping the lab resources up-to-date and doing their work in a systematic way.

Special thanks to my family for supporting, helping and providing me each and every facility during the period of research. My parents have been the source of my inner strength and self-confidence. They are the reason I could see the light of this day. I would like to express my deep hearted thanks to my brothers, **Santosh**, **Dr. Rajeev** and my sisters **Snehlata**, **Puja**, **Deepika** and **Priyanka** for supporting me in fulfilling my professional dreams.

Neetish Kumar

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

ABBREVIATIONS	Description
TV	Television
DIS	Digital image scanner
NIST	National Institute of Standards and Technology
NASA	National Aeronautics and Space Administration
DSP	Digital signal processing
GPUs	Graphic process devices
DVD	Digital versatile disc
CCD	Charge coupled device
CRT	Cathode ray tube
BMME	Block-based matching ME
SAD	Sum of absolute differences
ME	Motion estimation
MC	Motion Compensation
BM	Block matching
MCMC	Markov-chain Monte carol
VAE	Variational self-encoder
GAN	Generative adversarial network
FPS	Frames per second
DCT	Discerning cosine transform
EFP	Electronic field production
CCTV	Closed-circuit TV
RAM	Random access memory
DRAM	Dynamic RAM
SRAM	Static RAM
CPU	Central processing unit
LPC	Linear predictive coding
VBSME	ME variable block size
DCT	Discrete cosine transform
MV	Motion vectors
DS	Diamond seal

LDSP	Long search Diamond patterns
SDSP	Small Diamond System patterns
TSS	Three-step search
VGEG	Video coding expert group
MPEG	Moving picture Expert Group
NLE	Nonlinear editing systems

SYMBOLS

Cr	Critical ratio
Hz	Hertz
ms	Milli second
fps	Frames per second
Mb	Mega byte
kb	Kilobytes
fn	Pixel in frame n
tm	Time stamps of the last frame m
tn	Time stamps of the first frame
T_0	Time interval for first clip
T_1	Time interval for second clip
T_k	Time interval for kth clip

LISTS OF PUBLICATIONS

[1] Neetish Kumar, Dr. Deepa Raj (2017), “Video Processing and its Applications: A survey”, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 6, Issue 4, ISSN 2278-6856. - **(UGC Indexed)**

[2] Neetish Kumar, Dr. Deepa Raj (2018), “Video Generation using Still Images”, International Journal of Computer Sciences and Engineering (IJCSE), Volume 6, Issue 11, ISSN: 2347-2693. - **(UGC Indexed)**

[3] Neetish Kumar, Dr. Deepa Raj (2018), “A Study and Analysis of Images in Different Color Models”, International Journal of Advanced Studies In Computer Science And Engineering, (IJASCSE) Volume 7, Issue 1, - **(UGC Indexed)**

[4] Neetish Kumar, Dr. Deepa Raj (2019), “A New Pentagon Search Algorithm for Fast Block-Matching Motion Estimation”, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume 8, Issue 10, ISSN: 2278-3075 - **(Scopus Indexed)**

[5] Neetish Kumar, Dr. Deepa Raj (2020), “A Framework of Video Compression and Decompression”, International Journal of Grid and Distributed Computing Volume 13, Issue 2, pp. 235 – 245. - **(Web of Science Indexed)**

[6] Neetish Kumar, Dr. Deepa Raj (2020), “Algorithmic solutions for high quality Video editing software problems”, Journal of Critical Reviews, Volume 7, Issue 09, ISSN- 2394-5125 - **(Scopus Indexed)**

[7] Neetish Kumar, Dr. Deepa Raj (2021), “Shot Boundary Detection Framework For Video Editing Via Adaptive Thresholds And Gradual Curve Point”, Turkish Journal of Computer and Mathematics Education Volume12, Issue11, 3820- 3828 - **(Scopus Indexed)**

Summary

SUMMARY

In recent years, video development has experienced an enormous explosion, and we've seen people hovering on social networks very frequently. Almost in every field, digital image and video processing puts a live effect on things and is growing with time to time and with new technologies. Since the video process is multidimensional field we have touched the compression and editing part of it very minutely. Based on that algorithms are developed and detailed analysis is done. The entire research work is organized as follows:

CHAPTER I

INTRODUCTION

Video is termed as image sequence, represented by a time sequence of still images. In other words, it is said as the sequence of kinetic frames captured/ play backed into some sequences. Digital video is an illustration of moving visual images in the form of encoded digital data. The rapid development of video capturing systems changed the world from recording the still frames on tapes to the storing frame digitally through charge coupled devices. The Video recording system changes from past analog era to digital era which made editing and manipulation very easy. Any video recording systems do not actually record videos but the sequence of the still image at the desired frame rate. The human eye can’t distinguish delay between the processing of still frames if it can be processed with above 16 frames per second due to the persistence of vision. It gives an illusion of moving videos.

As multimedia and communication technology is on peak nowadays and keep growing day by day, the vast and voluminous content present on the web is video data. It needs to be compressed as it is of draconian in nature if not compressed. A lot of compression techniques are developed for handling the video data. If it is keenly observed then it is found that search techniques play a pivotal role in video compression. Video signals comprise of temporal, spatial and statistical correlations and the same needs to be exploited to achieve better compression. Spatial correlation is due to the similarity of neighbouring pixels within a frame, whereas temporal correlation in a video is due to the similarity of the adjacent frames as changes are

mainly due to camera or object motion. As adjacent video frames are highly correlated, independent coding of each frame generally generates a lot of redundant bits. Block based motion estimation serves as the backbone of video compression and the search techniques used in the block based method play a pivotal role in detecting the motion vector. Motion vector is detected by motion estimation and compensation. For the detection of motion vector some patterns are used. The impact of search pattern is a crucial part in the block-based motion estimation for finding the motion vector. An issue of distortion performance and search speed heavily depends upon the size and shape of search strategy applied. Performing the deep analysis for motion vector distribution on standard test videos, it is desirable to have such type of algorithm that meets the requirement of searching motion vector in less time. The encoding of motion vector and residual results in a compressed video form.

All the things mentioned above deal only with the pre production activities. A broader view of video processing is also there which is known as the post production, widely used in cinematography, film industry and news industry. Video editing, the part of post production activity, is not only used in professional movie making but also in various areas like education field, marketing fields etc (eg. Since lockdown due to covid-19 the education world started classes online). Today, we have more popular video cameras to manage the video data but algorithmic solution is needed to solve various problem related to the video editing. There are also more opportunities in quality degradation, video security, video copy detection etc.

Day to day huge volume of extended videos resulting from documentaries, movies, sports and surveillance cameras are evolving over video databases and Internet. Processing of these videos manually is hard, costly and time-consuming. For extracting these long-duration videos an automatic procedure is desperately needed. As a vital factor the Shot boundary detection (SBD) is considered for lot of video analysis tasks, for example video editing, indexing, summarization and action recognition.

Digital video standards are required for exchanging of digital video among different products, devices and applications. Fundamental consumer applications for digital video comprise digital TV broadcasts, video playback from DVD, digital cinema, as well as videoconferencing and video streaming over the Internet

CHAPTER II

REVIEW OF LITERATURE

This chapter covers the literary analysis of the evaluation and development of video editing systems. Literature analysis has been done from various reputed online journals, e-books and many more resources. By performing analytical studies from various resources we are able to find out some interested and vital research areas in video processing and video editing.

CHAPTER III

VIDEO GENERATION

This chapter proposed methods of video generation using still-image techniques. A video is the visual multimedia source that combines a sequence of images to form a moving picture. Any video recording systems do not actually record videos but the sequence of the still image at the desired frame rate. The human eye can't distinguish delay between the processing of still frames if it can be processed with above 16 frames per second due to the persistence of vision. It gives an illusion of moving videos. The video recording systems change from past analog era to digital era very rapidly. This chapter also briefs about the evolution of video camera. In this we have used Matlab platform to work with still image for generating the videos of different kinds. Matlab is a good platform for working on digital image processing application like compression, enhancement, and detection of an object in an image, segmentation and pattern recognition. In summary of experimental study, it is concluded that Matlab is also very suitable for working in video processing application. Easily images can be read from file or folder and the frame is also generated by calling simple function im2frame().Video file is generated and all the frames are stored in the video file and execution of the video takes less time and acquire a less space in the memory compare with other platforms.

The resources for this chapter were published in the International Journal of Computer Sciences and Engineering, Open Access Research Paper, Volume 6, Issue 11, (2018), E-ISSN: 2347-2693.

CHAPTER IV

MOTION VECTOR SEARCH AND MOTION ESTIMATION & COMPENSATION

This chapter describes the motion vector as the central element in the compression of video. In previous chapter we have seen that the video is the sequence of kinetic images. In those sequences of images, there are strong redundancies of different types e.g. spatial, temporal, perceptual and statistical. The compactness among the neighboring pixels is termed as spatial redundancy. There is a vast correlation among the neighboring pixels that needed to be removed for achieving the better compression ratio. On the other hand, there are also strong temporal redundancies between the frames in a video sequence. Temporal redundancy is resolved by block matching motion estimation techniques in successive video frames. In this chapter we have gone through the different search techniques for finding the motion vector. It is calculated through the method of motion estimation and compensation.

CHAPTER V

PENTAGON SEARCH TECHNIQUE

After various studies we proposed a new kind of search technique for finding the motion vector. This chapter deals with the justification behind the proposal of Pentagon search technique. The simulation results shows that it has better performance over the existing hexagon and diamond search (DS) algorithm in terms of lesser no of search points. The speed improvement rate also gets increased with respect to the DS algorithm. Motion vector finding the probability through pentagon search is better than the previously proposed search strategy.

This chapter's resources were published in the International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume 8 Issue 10, (2019), ISSN: 2278-3075.

CHAPTER VI

FRAMEWORK FOR VIDEO COMPRESSION AND DECOMPRESSION

This chapter explains the compression and decompression of videos with the help of pentagon search. Our proposed framework is a step forward in the field of compression and decompression. Basically at the back end i.e. search strategy, a new pentagon search method is used that searches the true motion vector efficiently that further encoding method gets benefitted by it. Before knowing the complex process of compression it must be investigated first that what a video frame consists of. Basically video is a sequence of kinetic frames. These frames are bundled together by its properties and these bundles are known as group of picture (GOP). In a GOP there are three types of frames I, P and B frames. All three types of frames are encoded in different ways. The first frame of the GOP is always an I frame that is encoded just like the still image. The P & B frames are then coded with the reference of other frames. The compression performance achieved for P and B frames through this method is larger than any other previously proposed framework. This framework is also applicable for compressing the real time video data.

This chapter's resources were published in the International Journal of Grid and Distributed ComputingVolume 13, Issue 2, (2020), pp. 235 – 245.

CHAPTER VII

VIDEO EDITING

This chapter deals with the video editing. It is very much required in film, news, online education industries. As we know that the process of video editing is also known as the post production phenomena. According to the need visual effects are added, corrupted frames are removed and some new frames are added. Some useful operations are carried out with the editing operations like clip insertion, bad frames removal, gray conversion, frame rate modifications etc.

In further study of video editing, we observed some issues in shot boundary detection. We have tried to sole out the problem related to the abrupt and gradual transition detection in shot boundary. Using the adaptive threshold the similarity features are obtained for detecting the abrupt transitions. For detecting the gradual transition the gradual curve was obtained. For this initially the average edge image was found. The minimum difference between sequences and local variance were considered. At last the gradual change point was identified.

The resources for this chapter were published in the journal of critical reviews Issn-2394-5125 vol 7, ISSUE 09, (2020) and Turkish Journal of Computer and Mathematics Education Volume 12, Issue 11, (2021), 3820- 3828.

CHAPTER VIII

CONCLUSIONS AND FUTURE SCOPE

This research work provides the overall idea of video production and video editing. Video compression is very much required in today's scenario. We have developed a framework for compression and decompression for that. There are maximum chance of a video to become error prone during the capturing side for this we have some video editing operation to overcome such issues. Video editing is not only done on distorted videos but if, one requires addition changes, special effects, etc. then editing play pivotal role. This chapter concludes all the research work discussed in the previous chapters and the potential scope of the book.

CHAPTER I

Introduction

CHAPTER I

INTRODUCTION

In this chapter, crucial feedback of research on video processing is presented to examine the video editing system. The outline of digital signals includes video processing and video editing. In today's era making video communications via mobile, video sharing, and video transmission are gradually increasing. Media contents have to be transmitted with less bandwidth in less time. The original video signal has to be compressed without affecting video quality or data loss by some encoding methods. Video editing helps in gathering all video content, including movies and TV for processing and arranging video files. This chapter focuses on the basic analysis of video processing, video compression and video editing terminologies. The analysis also results in the problem formulation along with the successful executions.

1. HISTORICAL PERSPECTIVE OF VIDEO PROCESSING

Advances in multidimensional communication, digital computers, graphics processing, analog, and digital transformers theories have helped in shaping image and video in real time. The history of video processing states the current scope of this area. A video signal is nothing but a time series of varying images.

1.1. HISTORY OF VIDEO PROCESSING

History of video processing is about centenary old event that started with Cinematography. Before that people were well aware about photography. The journey of video processing started with the invention of image capturing devices. i.e camera. A series of photographs gives illusion to human eyes of a motion pictures. The further investigation of motion picture leads to the evolution of video processing. we will discuss here some mile stone achievements in the video processing that shows a way to research community to explore in this field. In 1957, the first Digital Image Scanner (DIS) was established at the National Automated Computer Standards Office, working in partnership with the National Standards Office, the National Institute of Standards and Technology (NIST). This scanner was used to convert the analog image to a discreet pixel and store it in the Standards Electronic Automatic Computer (SEAC) memory. SEAC is used for image enhancement experiments with edge improvement filters. The development of modern computer technologies motivates these improvements and ultimately helps to develop the world of digital video processing.

In the 1960s, the National Aeronautics and Space Administration (NASA) Jet Propulsion Laboratory created electronic imagery using a monochrome charging system to enable silent electronic cameras. Digital cameras and digital video processing were the driving forces behind the scientists of NASA in using accurate video processing through space exploration. Today, the technology for industrial inspection and medical imaging is being developed for video treatment applications. Due to the comparison of low-and medium-level operations, the first computer for digital imaging had a thumb drive with massive and parallel mainframes. The video processor architecture should be built comparable with a large volume of data necessary for handling. In the past, multi-panel boards with many parallel processors have been used in real-time video device designs, especially for military and medical applications. In the 1980s, this concept was about to change with the emergence of programmable digital signal processing (DSP) technology.

Over the next decade, the first commercially available DSPs to boost signal calculations have been revealed. DSPs introduced the age of portable electronic computing. In the late 1980s, programmable logic devices, such as the programmable portal array, which combined the software with advanced hardware speeds, such as integrated application circuits, were developed using programmable logic.

In 1990s, the DSP’s and the Field Programmable Gate Arrays (FPGA’s) improved performance meets the requirements of multimedia devices. A chip-based instrument (SoC) was created to ensure that the signal processor had all the required computer processing power.

A recent trend exploits the enormous parallel computing abilities of graphic process devices (GPUs) used on most modern computers and laptops for computer-intensive video processing. Another invention was implemented in the late 1990s and early 2000s, the portable multimedia supercomputer which combines the high-performance parallel processing capacity using intermediate video operations with high energy efficient mobile integrated devices.

1.1.1. A History of Video Recording

Video recorders became available much later than Radio. In the 1950s, RCA produced the first 6-meter / sec tape rate unit. They used the longitudinal signal capture on the phone, making it possible to shut down the broadband T.V. signal quickly. Ampex Company created the first practical video recording system in 1953. He has used the scan process, which is still with us today, which simplifies the degree of recording without an awful speed of the tape.

Legal action by media corporations has forced Replay T.V. to eliminate features such as automatic commercial skip and uploading of recordings over the Internet. However, newer devices have continuously restored these features while introducing additional features, such as DVD recording and programming and remote control, using PDAs, networked P.C.s, and web browsers. For compressing digital video, Many DVRs use the MPEG format. Video recording capabilities have been an integral part of the new set-top box, as T.V. audiences decided to take care of their television experience. Since viewers have been able to pull together and the amount of video content on their set-tops delivered by conventional 'broadcast' cable, satellite, and terrestrial networks, as well as I.P. networks. The ability to catch and watch program whenever they want has become a must have for many consumers. Unlike VCRs, hard disk-based digital video recorders make "time-shifting" more convenient and also allow features such as paused live T.V., instant replay, chasing playback (seeing a recording before it's finished), and skipping advertisements during playback.

1.2. THE PROGRESSIVE AND INTERLACED SCAN PATTERN

In progressive scan the line by line scanning is done but in case of interlaced scan the scanning process is done based on odd and even lines. The even scanned lines are drawn first then the odd line are drawn. Figure 1.1 illustrates the gradual and interlaced patterns of scanning.

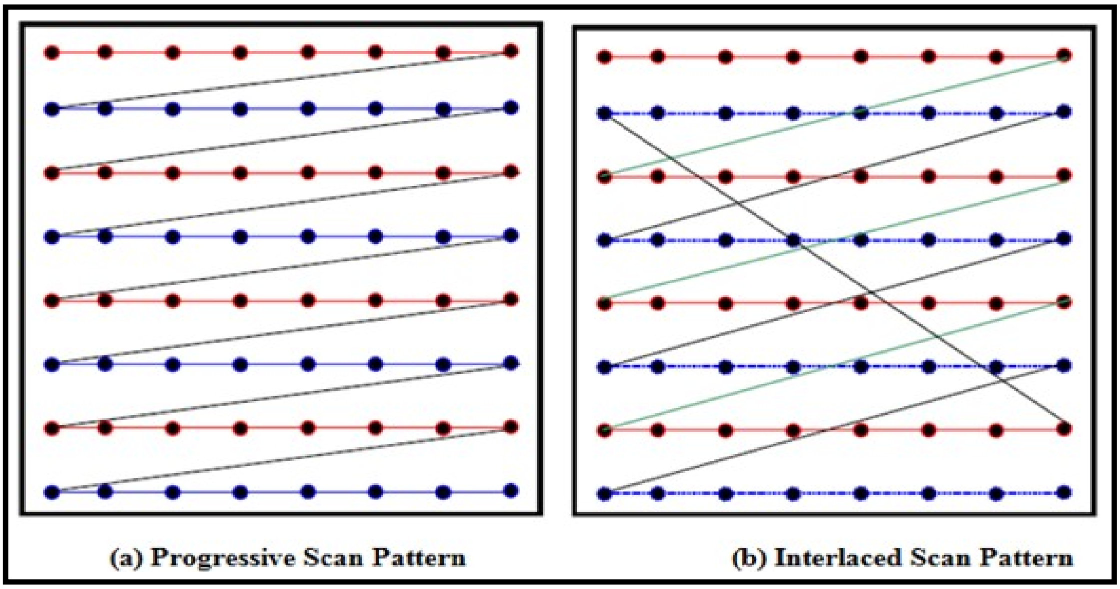


Figure: 1.1. Progressive and Interlaced Scan Pattern

1.3. FRAME TYPES

A video consists of various group of pictures commonly known as GOP. A GOP consists of three types of frames i.e. I frames, P frames and B frames. I frames are also known as intracoded frames because it not require any prediction for encoding. P frame is known as predictive frames because it requires I frame for prediction and lastly the B frames are known as Bidirectional frames which requires either P or I frame for its prediction.

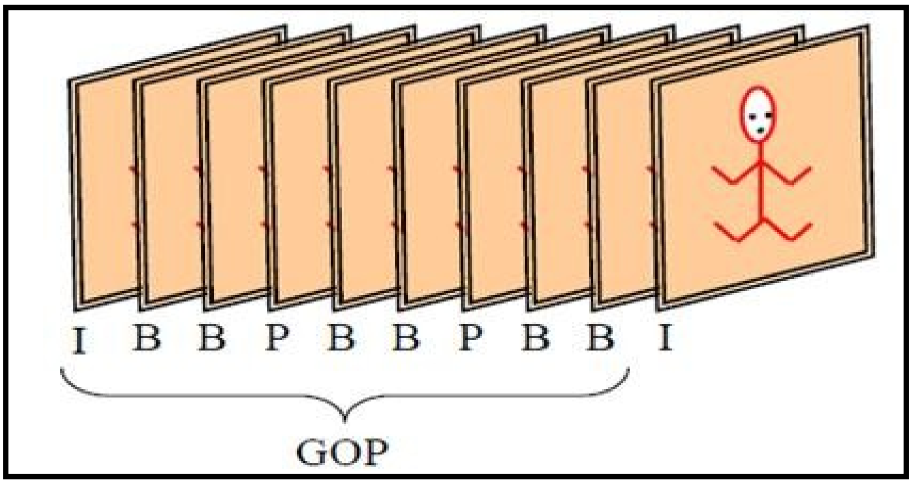


Figure: 1.2. Group of picture

1.4. VIDEO PROCESSING COMPONENTS

1.4.1. Noise

Noise is defined as any unwanted information that contaminates the property of a frame. While capturing the video by any type of devices, there are chances of frames to be get affected by noise may be due to the wrong focus of camera, atmospheric disturbance, camera motion e.t.c that results in quality degradation of a frame. The noise in pictures may come from numerous outside and internal sources (i.e., sensors), which is difficult to avoid in practical circumstances. Sometimes it is necessary to add noise for a faithful reconstruction of image during the digitization process. There are various types of noises present in a picture, namely thermal noise, periodic noise, Gaussian noise, salt and pepper noise etc and there are various types of denoising algorithms too for removing it.

Figure: 1.3. Image noise

1.4.2. Enhancing

Digital video today is a crucial part of daily life. Digital video is captured, handled, and used for various purposes, e.g., surveillance, general identity verification, trafficking, law compliance, civilian or military, etc. Frames can be enhanced by following methods:

1. The noisy frames are needed to denoised first in order to get an enhanced frame
2. The video signal information may degraded from travelling through various

channels. It may cause the blur as a result so before reaching at end user the frame boundary, edge, contrast, needed to be modified.

3. Frame enhancement can be done either in spatial domain or frequency domain.

4. Different types of filters are also used for smoothing the frames.

1.4.3. Object identification

Object recognition tools are used to identify objects in frames or videos. To recognize objects, deep learning and machine learning algorithms are used. The machine are trained in such a way so that it can quickly read characters, incidents, scenes, and visual details when viewing or watching a video clips. It is now a day's very commonly used in inspection of industrial products. It also allows in many applications to detect diseases in bio- images, industrial inspections, and robotic vision. We can also see this feature in smart cameras.

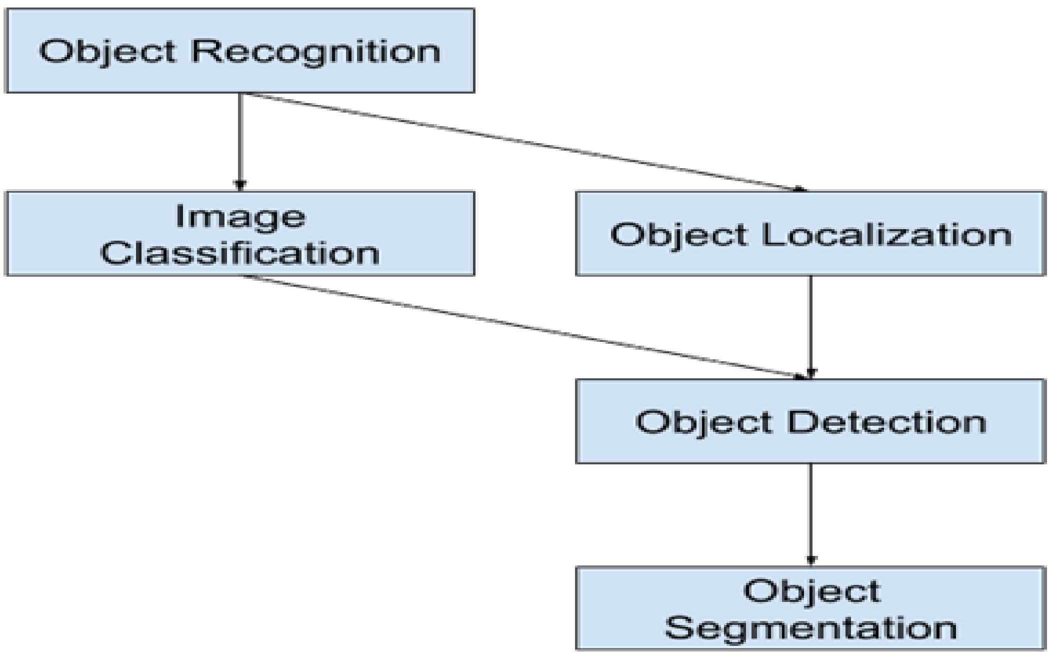


Figure: 1.4. Object Identification

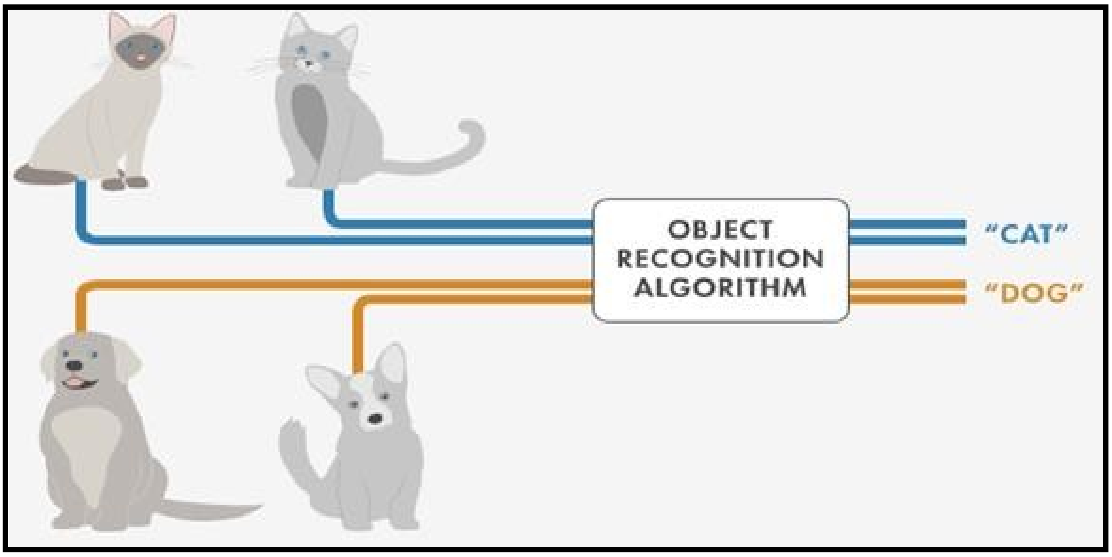


Figure: 1.5. Identifying various object groups using object recognition

1.4.4. Video segmentation

As in image segmentation the image is partitioned into non overlapping blocks, the same mechanism is applied to the video segmentation too. Firstly the video are segmented into shots, then these shots are further partitioned into a large no GOPs. In every GOP there are there are 3 types of frames namely I, P, and B frames. These frames are further divided into macro blocks for analysis purpose. We decompose the full video clips at such level wherever it is easy to analyze. As in case of motion estimation and compensation we segment the video clips at the macro block level and investigate the motion of a particular blocks by some references.

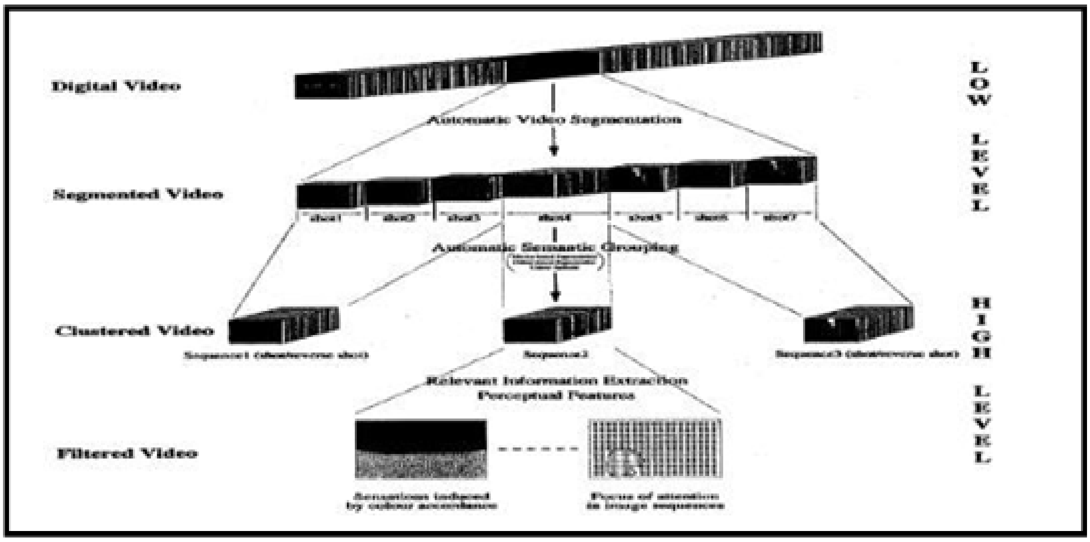


Figure: 1.6. Steps of video segmentation

1.4.5. Video compression

Videos compression plays a significant role in many digital media applications, such as live streaming, high-definition T.V., and video-on-demand. A still picture has only spatial redundancy but in case of video there are vast temporal redundancies present in every GOPs. In order to get a compressed video it needs to exploit all types of redundancies first. A video sequence comprises several frames. The video should be stored and transmitted in an efficient manner using temporal redundancy in consecutive frames. The correlation between adjacent frames are too strong whereas pixels correlation within a frame is also too high. The exploitation of both types of redundancies results in a compressed video. Video compression algorithms can be broadly divided into two forms.

1. Lossless video compression and
2. Lossy video compression.

The below figure depicts the basic correlation between candidate frame, previous frame and future frame. By various methods the redundancies among the frames are exploited and a compressed video is created.

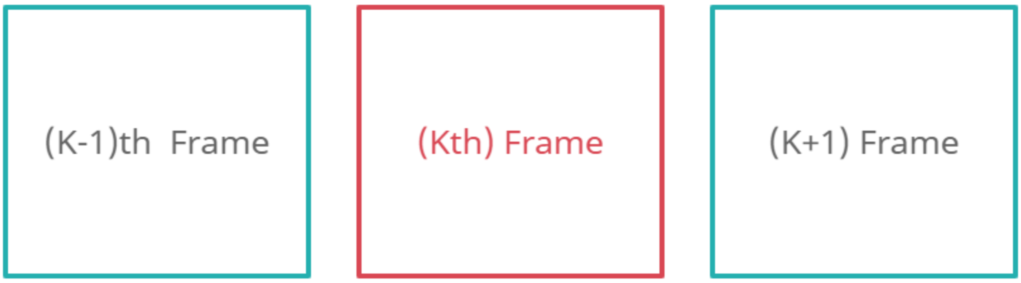


Figure: 1.7. Three consecutive frames of a GOP.

Intra frame coding

A frame is usually called intra frame coding to by eliminating spatial redundancy as like coding done for still images. Transform minimizes spatial redundancy within a frame. The most widely used transform is discrete cosine transform.

Inter frame coding

The inter frame coding eliminates the temporal redundancies between the frames. Inter frame coding takes advantage of video frame interdependencies. The inter frames coding requires some addition reference frame that are already coded. A frame is used as a reference to minimize the temporal variance. The reference frame may be a forward frame or backward frame

Figure 6 shows the general block diagram of a video encoder. Below are the explications of the various blocks.

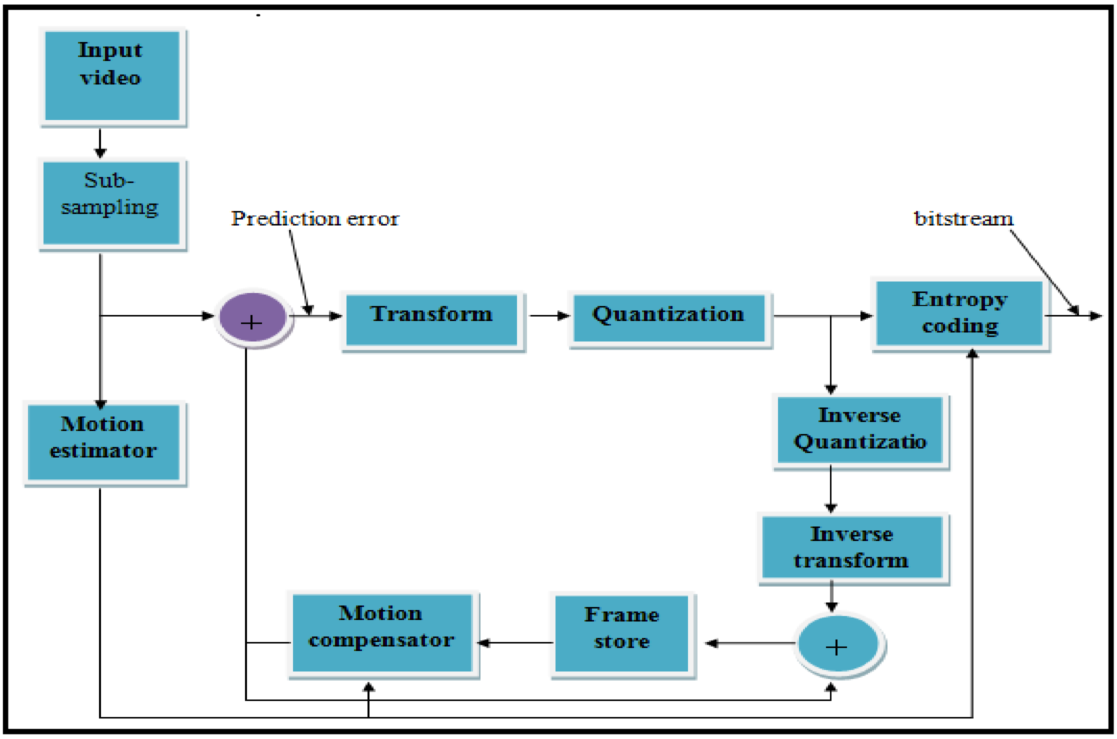


Figure: 1.8. Block diagram of general video coding scheme.

1.5. MOTION ESTIMATION AND COMPENSATION

Motion estimation and compensation plays a pivotal role in video compression. The motion estimation defines the method of measurement of movements in a picture sequence between two or more images. Motion compensation is a mechanism used to model and recreate a system using a specific frame of reference and a set of movement parameters. In general, the motion estimation and motion compensation are the most bulky part of video compression. It takes 80% time of total video compression and the computational complexity of performing these two operations is also too high. Motion estimation and compensation calculations were typically done with block-based approaches. The advantages are that they have a wide range of simple, easy-to-implement, and beneficial materials. The block-based estimation of motion is the most effective way to obtain true motion vectors. It divides frames into equal rectangular blocks and shows how a search window carries the closest intersection of the previous frame. The movement of the block which best matches is defined following block distortion or other criteria as the vector to the block within the actual framework. A cost function, like a Minimum of squared error (MSE), and minimum of absolute error MAE, or sum of absolute differences (SAD, for instance), decides the best match.

Figure: 1.9. Normal frame and motion compensated frame.

1.6. BLOCK-BASED MATCHING MOTION ESTIMATION (BMME)

Block-based matching ME (BMME) is the most widely used form for ME video encoding, as images can be rectangular in shape and block splitting. Earlier MPEG standard i.e. MPEG 1 and MPEG 2 uses a static format of either 16 x 16 or 8 x 8 size blocks for motion estimation. Later vast improvement was done and finally the dynamic block size motion estimation came into existence. MPEG 4 or other higher standard uses dynamic block size i.e. 4 x 4, 8 x 8 or 16 x 16.

The block size influences the efficacy and precision of the prediction.

Two basic assumptions are made in this technique.

1. The block size of the frame is smaller than the chosen window size.

2. The movement is assumed to be in real time (real time motion).

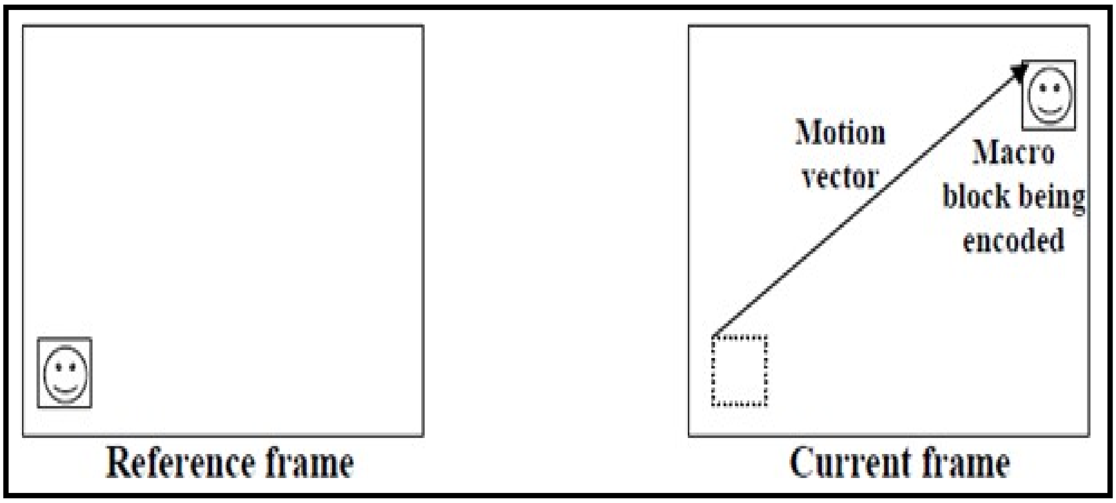


Figure: 1.10. Forward motion- compensated prediction

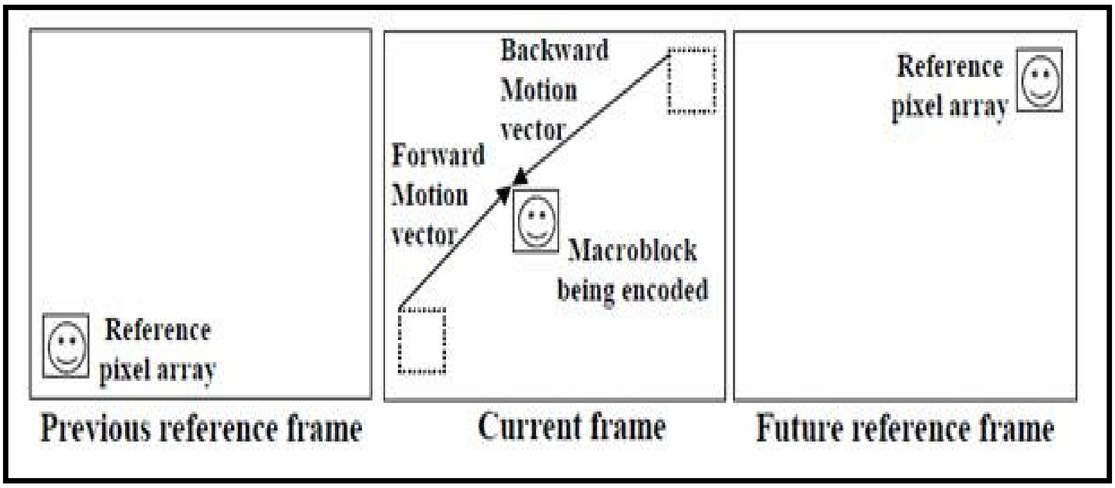


Figure: 1.11. Bidirectional motion-compensated prediction

1.6.1. Motion vector

A motion vector is the main element in the motion evaluation phase in video compression. It is used to represent a macroblock based upon the location of this (or similar) macroblock in a second image, called the reference image. The H.264 / MPEG-4 AVC standard defines a movement vector as a two-dimensional vector used for interdiction, which offsets the decoded image's coordinates into the reference image's coordinates. The motion vector prediction method leads to a modified residual signal that mainly consists of small values. Like the entropies we used in coding quantized transform coefficients with Huffman or arithmetic methods, we can also apply them to residue movement vectors. For Huffman coding and various probability distributions for arithmetic coding, it is evident that different VLC (Variable length coding) tables are required.

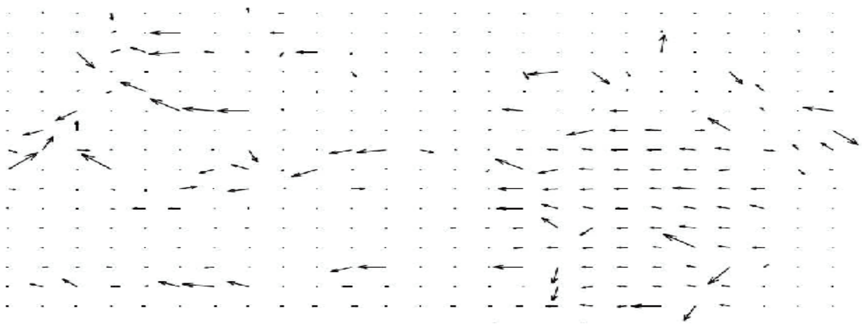


Figure: 1.12. Bidirectional motion-compensated prediction

1.6.2. Full motion search

In full search algorithm, a sufficiently large size of window is chosen with respect to the candidate block. The window is relocated to each candidate location in the search window when finding the best match. There are total positions (2w+1) * (2w+1) that must be checked, where w is the block search range. Where the MAD is least that is termed to be a best match. The entire search is heuristic where each and every search location within the window is searched, and it offers the highest match accuracy. But it has a drawback of having high computational complexity. We can’t use this search technique in real-time because in real time video compression we have only 30 ms time to disposal of motion information to the next frame. In those 30 ms the entire process of motion estimation and compensation can’t be done and there is a vast possibility for end viewer to see more corrupted frames in a single video. So we move towards fast motion search techniques. Fast search algorithms follow a particular pattern, which tests for fewer points instead of a full search. Full search Algorithms try all locations in the search window that why it guaranties the optimal solution for finding the motion vector.

1.6.3. Logarithmic Search Method

The logarithm search algorithm introduces pattern search. Definitely, it reduces the time complexity of finding the motion vector but it doen not guaranties the optimal solution. We have to compromise with some sub optimal solution. The five squares are on a pattern like the five crosses (+). The applicant blocks [0, 0], [0, + s], [0, -s], [-s, 0] and [+ s, 0] are thus tested when a search point located at coordinate [0, 0]. The steps are only halved if the minimum distortion point is found at the middle (x, y) in the last step, or if the actual minimum point reaches the boundary of the search window. Otherwise, the stage scale is the same. The minimum distortion point is the best fit for each step for the target block; if the step is reduced to one, then all eight blocks around the centre position are calculated.

1.7. VIDEO TRACKING

Video tracking is something about monitor the camera overtime for a moving target (or several objects) including interaction with human electronic systems, protection and monitoring, video and inspection, enhanced fact, traffic control, medical imaging, and visual editions.

Photo surveillance is designed to suit physical objects in consecutive camera frameworks. The relation can be especially tricky concerning the frame rate as objects move quickly. Another condition that raises the complexity of the problem is the change of the managed object's orientation over time. In these situations, video trackers typically use a motion model that explains how to change the target's picture to the potential movement for the object.

1.8. OBJECTIVE OF THE RESEARCH WORK

Our objectives are as follows:

1. Generating video from still images and analysing the video at different frame rates.
2. To develop a motion vector search algorithm in order to target the backbone of compression, i.e the development of motion vector search algorithm by block based motion estimation for finding the true motion vector.

3. Designing a framework for video compression and decompression.
4. Targeting the issues related to post-production activities i.e video editing, and making videos error-free, noise-free, and smooth at the playback side.
5. Detecting the shot boundary in abrupt as well as gradual transition condition which is useful in video editing operation.

1.9. THESIS ORGANIZATION

This thesis is organized as follows:

Chapter I i.e Introduction basically deals with the video definition, properties, types of frames and various aspects of video processing. Digital video is an illustration of moving visual images in the form of encoded digital data. The Video recording system changes from past analog era to digital era which made editing and manipulation very easy. Any video recording systems do not actually record videos but the sequence of the still image at the desired frame rate. Some measurement metrics are also discussed which are used to evaluate the performance of proposed methods

Chapter II i.e. Review of Literature, covers the literary analysis of the evaluation and development of video editing systems. Literature analysis has listed several online research archives and online journals. Video processing and frame editing are met with complicated problems.

Chapter III i.e. Video Generation deals with the generation of video using still-image techniques. A video is the visual multimedia source that combines a sequence of images to form a moving picture. Any video recording systems do not actually record videos but the sequence of the still image at the desired frame rate. The human eye can't distinguish delay between the processing of still frames if it can be processed with above 16 frames per second due to the persistence of vision. It gives an illusion of moving videos. In summary of experimental study, it is concluded that Matlab is also very suitable for working in video processing application. Easily images can be read from file or folder and the frame is also generated by calling simple function in frame. Video file is generated and all the frames are stored in the video file and executions of the video takes less time and acquire a less space in the memory compare with other platforms.

In **chapter IV** motion vector search and motion estimation and compensation are analyzed and discussed. It also describes the motion vector as the central element in the compression of video. The compactness among the neighboring pixels is termed as spatial redundancy. There is a vast correlation among the neighboring pixels that needed to be removed for achieving the better compression ratio. On the other hand, there are also strong temporal redundancies between the frames in a video sequence. Temporal redundancy is resolved by block matching motion estimation techniques in successive video frames. In this chapter we have gone through the different search techniques for finding the motion vector. It is calculated through the method of motion estimation and compensation.

In **chapter V** we proposed a new kind of search technique for finding the motion vector. This chapter deals with the justification behind the proposal of Pentagon search technique. The simulation results shows that it has better performance over the existing hexagon and diamond search (DS) algorithm in terms of lesser no of search points. The speed improvement rate also gets increased with respect to the DS algorithm. Motion vector finding the probability through pentagon search is better than the previously proposed search strategy.

In **chapter VI** framework for video compression and decompression is proposed. This chapter explains the compression and decompression of videos with the help of pentagon search. Our proposed framework is a step forward in the field of compression and decompression. Basically at the back end i.e. search strategy, a new pentagon search method is used that searches the true motion vector efficiently that further encoding method gets benefitted by it.

In **chapter VII** detailed discussion has been done on video editing. It is very much required in film, news, online education industries. As we know that the process of video editing is also known as the post production phenomena. According to the need visual effects are added, corrupted frames are removed and some new frames are added. Some useful operations are carried out with the editing operations like clip insertion, bad frames removal, gray conversion, frame rate modifications etc.

In further study of video editing, we observed some issues in shot boundary detection. We have tried to sole out the problem related to the abrupt and gradual transition detection in shot boundary. Using the adaptive threshold the similarity features are obtained for detecting the abrupt transitions. For detecting the gradual transition the gradual curve was obtained. For this initially the average edge image was found. The minimum difference between sequences and local variance were considered. At last the gradual change point was identified.

In **chapter VIII**, the thesis concludes the critical analysis of the works described in the previous chapters with the future scope in this field.

CHAPTER II

Review of Literature

CHAPTER II

REVIEW OF LITERATURE

To understand the video processing and video editing process, a systematic literature review was carried out. Work clarifies and simplifies the processing of visual signals by briefly describing video output. In this regard, the central theoretical gap was the understanding of how still images can process videos.

In the year 1993, Vassilis Seferidis et al. [119] experimented with general transformations of geometry pixels co-ordinates. A widespread method for block matching was developed from this perspective. It can not only be precisely calculated translational movement but also other deformations of moving objects. Its simplicity is the main advantage of the proposed solution and can incorporate the existing video encoding methods easily and in parallel. The only drawback is the considerable cost needed by the computer to measure the motion parameters. Computational complexity can be very high due to use of fast search algorithms. Bede Liu et al. [17] demonstrated the block motion vectors estimation and presented studies on modern fast algorithms. The MSE of motion compensated structure patterns reduces calculation complexity, a complete, exhaustive scan without a factor of 4. The models of the sub-samples are adjusted and searched at various places to estimate the motion vector. Jing Yang Chen et al. [125] conducted motion offset time filtering for pre-processing and noise reduction in digital video data. Using time compensated filters with motion vectors previously created and adaptive spatial filters in scene minimizes noise. Various types of noise can be inserted into the video until encoding and transmission. Added noise reduces system efficiency and eventually reduces system performance before image compression.

In the year 1994, Fillia Makedonet al. [44] addressed problems relating to multimedia and user interfaces in multimedia documentation, such as compression and digital video. A video distributed remote access model is defined. It defines the steps and requirements for multimedia production process environments, video function and the objective of digital video editing programming.

In the year 1995, S.M.Kulkarni et.al. [78] uttered about motion-based video compression on the entire move is the most computationally expensive and time-consuming process from their earlier study, they concluded that the three-step search techniques for the matching blocks are the most common and useful of numerous motion evaluation techniques. Jae HeonJeong et al. [66] aimed to reduce the estimates of movement in the HEVC with the improved test zone search (TZS) algorithm. In reducing ME complexity and maintaining comparable performance, TZS is of good quality. P.B. Penafielet al. [109] discussed the new video compression method that takes noise directly into account and seeks optimum video quality and compression. Compression is accomplished by minimizing spatial and temporal redundancies. A reference frame is selected and encoded before transmission from each block. In blocks of N-frames, the encoding takes place in a video buffer.

In 1996, Nuno Correia et al. [106] conducted a study on components for video processing applications. A tool kit for video segmenting, encoding and support are included in the new framework. It provides a video modeling and model reuse approach for evaluating experience-based visual knowledge. The Toolkit framework and additional tools (HTML generation and Java documents) emphasize WWW aspects. Lurng-Kuo Liu et al. [84] provided a block motion evaluation algorithm based on the video code (block based gradient decent search) BBGDS. The BBGDS analyses the values of the given objective function from a small centralized control block. The lowest segment in this test is downstream and is determined by the gradient in which the minimum level is to be found by the search head and control block location. The BBGDS cannot conduct complete full search (FS), three step search (TSS) search, and new three step search (NTSS). Experimental results indicate that the proposed solution offers reduced system complexity to competitive efficiency.

In the year 1997, S. Moni et al. [116] have only mentioned a summary form. Video conferencing, digital recording, telemedicine, HDTV are the most common applications of images and videos. The setting up of a family of algorithms for image compaction helped in meeting these various applications. They offered a compression system that compares the complexity, compression ratio and other family characteristics with the imaging family. This set was composed of the Zero Tree (EZW).

In the year 1998, T. Meier and K.N. Ngan[129] decomposed automatic moving object into segmentation video. Motivation was based on the new MPEG-4 video coding standard. A detailed review was carried out before introducing video object plane (VOP) techniques. Moving objects are taken from a current automatic segmentation video sequence algorithm. In the middle of this algorithm lies the object tracker matched to a binary (2-D) model for the following scheme. The most suitable match is the translation of the image and adjusts each frame's shape and rotation. A modern update strategy based on transferring components makes it possible to amend the original model in a relatively big way. The developed algorithm was improved with a filtering technique that eliminates the fixed context. Tien-Ying Kuo et al. [132] studied fast overlapped checkerboard block partitioning motion compensation. In this work, they concentrated on reducing the difficulty of overlapped block motion compensation (OBMC) scanning the end of the encoder in both motion vectors. Deepak Turaga et al. [37] compared multiple algorithms and provided some major observation. The basis of the assessment is two measures, CPU computational complexity time and average minimum absolute distance (AMAD) performance. As hierarchical algorithms have been believed to preserve the same complete search efficiency and reduce the computer's complexity. But the AMAD is not sure of us as a fair measure. Nicole S. Love et al. [104] presented an empirical study of several options in block matching techniques to transfer object detection. In videos with varying traffic and weather conditions, the MSD corresponding parameters overweigh the corresponding steps to detect moveable objects (special reminder) and the processing speed with a zero-motion biassing. Search methods are close to those of 2D- logarithmic search which works the best way in fog series. The simple block determination method does not look as robust as the hierarchical approach in the nebula and snow series. Still, it results in an average object motion error size and a 1-2 order magnitude increase in the processing speed.

In the year 1999, Abhijeet Golwelkaret al. [50] examined the motion-compensated time filtering and motion vector coding. A further coded increase in the rate of PSNR in long filters and a massive increase in the number of motion vectors (MVs) are responsible for a lower rate of PSNR. This is remedied by implementing the MCTF "sliding window" method.

In the year 2000, Gibson et al. [49] have proposed a handbook of image and video processing. Mathematically the variations of the model in the target pose, they have used deformation design theory components. The appropriate size, rotate, and translation of the prototype reflects all the target scene occurrences. Sets of geometric characteristics cover all dimensions, rotations, and translations possible. As described later, they have a group structure. In short, these transformations are implemented to make the models a suitable scenario. Shan Zhu et al. [121] outlines some of the latest trends for searches and strategies. The algorithm for estimating the fast block matching movement is based on this research and observation. The new DS (Diamond) search technique does not limit the searchable window compared with TSS, NTSS and 4SS. DS algorithms showed above the most common TSS algorithm and approached MSE when their estimation is approximately 22 percent lower than NTSS. Simulation experiments compared to other recent BMA's, our DS algorithm, like 4SS and BBGDS, is much better on the average. Hongjun Jia et al. [59] analysed the algorithm for the fast block motion estimate based on the scan pattern and suggested for a directional model: three new search patterns and three new BMA called DCDS. The suggested algorithm is much more robust and more accurate than other experimental and theoretical algorithms. Shan Zhu et al. [121] developed a new block motion algorithm that searches in Diamond shape. The results presents that DS algorithm performance goes beyond the standard 3-step search algorithm. The evidence also show that the DS algorithm is more robust than the 4-stage (4SS), (BBGDS).

H. Pan et al. [110] Studied in sports video with slow movement replay segments to classify highlights. In their slow-motion playback method, a Markov-hidden (HMM) model calculates the probability of a part for replay-slow movements and is also applicable to find the limits of this segment Inferenz algorithm. Frederic Dufaux et al. [41] investigated privacy video monitoring. They described a technique for protecting the privacy of those monitored by a video monitoring system. A research module identifies regions of interest. The sensitive detail that must be sensitive was scrambled. The scratch is pseudo-related to the signal of transforming coefficients while transforming domain encoding. It is also flexible to adjust the distortion between mere fluidity and full noise. Ut- VaKoc et al. [75] demonstrated that most

DCT video codec standards were introduced, inherent shortcomings to be addressed previously the traditional hybrid framework are proposed to be handled by a video codec framework based on alternative DCT systems. Where there is no need to address what requirements prior research would allow for these two dramatic interactive architectures following standards.

In the year 2002, Yao Nieet al. [145] illustrated evaluation of quick block matching movement predictions for an adaptive root pattern scan. The proposed algorithm consists of two sequential phases: (1) initial and (2) locally optimized search. Only once at the beginning is any macroblock (MB) used to find a solid base for sophisticated and continuous local searches. The first study focuses on usable MVs for adjacent MBs. It proposes adaptive log architecture and dynamically determines the dimensions of the ARP for each MB. A log size (URP) pattern is used several times before the last MV is detected during the advanced local search stage. Ce Zhu et al. [24] developed a new fast algorithm based on an experimental hexagon search pattern. The HEXBS proposed algorithm uses the hexagonal search model's two different sizes. HEXBS can search quicker, larger, or medium than DS. Eric Dubois et al. [40] addressed time-varying filtering compensated for motion. The work describes Spatio-temporal filtering with time-variable images and is paid for activity: the characteristics and the interactions between the motion trajectories and displacement fields.

In the year 2003, Ross Cutler [30] researched a system for passing, recording, and showing meetings from far away. The system includes a large number of 360'Ring Cam, Whiteboard Display, Capture Devices Cam summaries, PC graphics, and microphone arrays, which provide remote participants with a wide range of environments. Lap-PuiChau et al. [26] proposed Octagon-based search model (OCTBS), a new quick motion vector search algorithm. The proposed OCTBS algorithm contains a motion vector with fewer search points than DS. The results of the tests demonstrate the efficiency of the OCTBS algorithm for many other regular search algorithms. Iain E. G.Richardson [63] introduced a mile stone in video compression. With hundreds of scientists and developers have been working in recent years to standardize visual MPEG-4 and 10 AVC bits from H.264 / MPEG-4. These are also separate milestones. MPEG-4 is far-sighted, creative, with still a variety of visual compression features and tools. In conjure with current and modern multimedia software H.264 has addressed the problems more pragmatically and closely.

In the year (2004), GUI-Guang et al. [51] assessed the critical characteristics of the block motion fields. A new LSPS analytical and parallel processing algorithm, ME, is designed to suit fast blocks. The experimental results show that a similar prediction with the DS algorithm can be achieved in comparison to the LSPS algae. Consistency and machine complexity are reduced considerably. C.L. Madhwacharyulaet al. [21] modified semantic content-based video metadata. The signals are divided between audio board multimedia signal, video, and other multifunctional streams high-level symbols (metadata) are among the most crucial multimedia research fields. Their studies resolve the critical gap in combination with video editing operations using a new XML-based mechanism. Xian-Sheng Hua et al. [143] presented a home video editing system based on optimization techniques. The system automates the temporal structure and determines the segment of video of desired length of highlighted segments. In addition, with the help of the developed framework, best-matched music for a given video can be chosen for different editing styles.

In the year (2006), Yung-Lyul Lee et al. [80] examined the variable block fast motion vector search algorithm. The variable block and quarter-pixel movement in their codec are estimated. The proposed search difficulties with a minimal PSNR loss can be used by the JVT codec to minimize the computational activity. The suggested one may also add quality to many reference frames without significant video loss. The hierarchical sum norm and fast search motion estimation of three pixels reduce the machine complexity. Jianbin Song et al. [127] looked for big diamonds and tiny pentagon searches at fast motion estimation patterns. The suggested small Pentagon model would enhance the analysis rather than the short diamond and the small hexagonal pattern. The proposed algorithm provides a vector with fewer search points that maintain similar distortions irrespective of the movement, tiny, medium or essential. Compared to other conventional experimental results on fast algorithms, further developments in the LDSPS algorithm are primarily explained. Asad Islam et al. [12] proposed an optimized, compressed domain video editing algorithms for common editing operations for mobile devices. The algorithm developed by them is time and space efficient.

In the year 2007, Tsinghua Science and Technology [135] published analysis of the video structure in which structural analysis is an essential requirement for most material in video editing and processing systems. Their article provides a method to examine the frame segmentation video structure quickly. The video is divided into different objects, including color, position and architecture, through structural analysis. Through technology, users can also increase productivity. This approach has shown that interactive video analysis that is fast and stable.

In the year 2008, Daniel Schonberg et al. [35] worked toward encoding and encrypting images and video sequences. They proposed a scheme of encrypted image and video file encoding. In previous studies, encrypted data have shown that it is as compressible as non-encrypted data as a distributed source code. However, these theoretical results pose two significant problems. So, they build their own methodological models and captured the basic statistical structure.

In the year 2009, Brendan Calandra et al. [19] investigated video editing to cultivate novice teaching practice. Video editing has been studied for the advancement of new technology so they found less evident teaching concept digitally. Martin Haller et al. [54] assessed global motion estimate (GME) for camera motion characterization and experimented on pixel and motion vector-based motion measurement. They assessed GME methods using a camera motion system with PSNR measures, based on pixel and motion vectors characterization. Experimental results show that GME methods based on vectors examined in motion globally are not very precise. The GME simplified pixel gives the best direction for vectors. This is also less computer-based. Imbaby I. Mahmoud et al. [64], proposed an object tracking algorithm (OPPEN) that supersedes OS, HEXBS, LOSP and EDS because the search points have smallest. Therefore, the algorithm has a small number of points, the monitoring of artifacts is less complicated and fast than any other. Sherief M. Hashimaa et al. [124] proposed DS hardware algorithm in which diamond search, motion assessment and object tracking were applied through hardware. A new architecture has been developed. The hardware implementation improved the performance of diamond search. The results of the test show that DS can be used in real time for low energy consumption. Li Xiaoliet al. [81] used another hardware support for fast searching of motion vector in order to get compressed video. According to them the compression algorithms need to be improved and a suitable control mechanism should be placed in place to improve the video compression method in real time with efficient use of hardware resources. Their method takes the advantages of tubes pit taping to maximize a double-core DSP polling surveillance. Koohyar Minoo et al. [76] used semitropical code for fast and efficient compaction of video data and applications using a model of parametric source. Their method provides an approach which can be expressed in a parametric distribution model with effective entropy. The entropy code achieves coding effects with the use of max posteriori or max likelihood (ML) methods by checking the codified data parameters on a statistical model using the proposed system.

In the year 2010, Siddhartha ahluwalia et al. [7] addressed with the logarithmic circular 2-d Motion estimation search algorithm. The circular 2-D logarithmic resolution estimation algorithm is an overall successful technique over previous methods, such as diamond search and the search algorithm for other effective block movements. This algorithm is suitable for both large and short video sequences and does not vary in size. However, the larger the motion vector, the more effective will be our proposed process, as experimental results verify it. Zhen ‘gang Wei et al. [149] used the DES-based video mixing system. Effective ways to edit and fuse images should be tested for large-scale video clips. However, non-linea rvideo editing such as MPEG-2 is very complex. It provides a multi format editing framework for MPEG, AVI, WMV and VOB files. A video cutting point can be cancelled and recovered by an algorithm. The app modifies the schedule and previews in real time the video clips added to the schedule.

In the year 2011, Sherief M. Hashimaaet al. [123] done experimental comparisons between fast block matching motion evaluation and object tracking algorithms. Cixun Zhang et al. [28] introduced a new SVT algorithm to improve the coding efficiency of video coders in their article. SVT provides video codes to change the block location rather than state-of-the-art images. In addition to adjusting the position of this transforming block, the transformative factor can also be modified to use the SVT correlations in order to recognize predictive errors.

In the year 2012, Dr. Anil Kokaram [39] conducted a study about introduction to digital video processing. The basic concepts of a video signal have been defined here.

Replication of motions is particularly essential, and therefore video processing algorithms should be careful not to harm the motion series. The measurement act of moving objects between the frames is referred to as the motion measurement, for extracting pixel data on the moving path from several structures. Motion compensated data are typically processed for video enhancement or compression. Ankita P. Chauhan et al. [9] explained comparative analysis in the diamond quest motion. For all movement statistics the diamond search algorithm is the best algorithm. Methods for diamond quest, DLTS, diamond- based FME algae and diamond-based diamond quest models (DHS) as well as MDSs (MDSS) for diamond search models. The DHS and line / triangle search models of the diamond search study are calculated by RFM. James Nightingale et al. [68] introduced the new video compression format developed by the Joint ITU-T / ISO Video Coding Team (JCT-VC) in the generation ahead is high performance video coding. The capacity for growth to grow will require standardization and the adoption of HEVC for many consumer products. The new HEVC norm covers areas like radio television, multimedia streaming, mobile and video content storage. Rajeshwar Dass et al. [115] developed new technique of video compression. They concluded with the key video compression techniques (H.264 / AVC) in their document and the latest video compression technology. They shown that H.264 / AVC created ISO /IEC (MPEG) and ITU-T (VCEG) are of up to quarte pixel accuracy. It incorporates numerous advancements in coding efficiency such as flexible, stable and application areas. Carlos Cuevas et al. [23] offered temporary segmentation tools for high-quality real-time video editing. Quick and effective editing tools were required with the increased use of video editing software. For applications in their study, they provided a high-quality, lightweight video indexing tool.

In the year 2013, Jamie Zigelbaum et al. [69] has determined the tangible video editor: designing for collaboration, exploration, and engagement. The tangible video editor (TVE) is a tangible tool for editing multimedia clip sequences. The TVE can be used on any surface, under usual illumination conditions, with the inclusion of maneuverable elements in interfaces without the need for an improved area surface, a creative computer or the video projection device. Comparison findings were discussed between TV and Microsoft Movie Maker. This result shows that their TVE is a collaboration, learning, user engagement, and cognitions platform delivery. ZHENG YUAN [150] investigated on advanced video processing techniques in video transmission systems and proposed advanced techniques of video processing. Their study analyzes four innovative technologies in the area of video transmission. The first two methods, video retargeting and video synthesis improve video material adaptability on different platforms. The other two techniques, the evaluation of the perceptual quality and the encoder optimization, are handling human visual features to reconfigure the video system's components. Razali Yaakob et al. [144] studied a comparison of various motion matching algorithms. They compared four blocks with the search algorithms (ES, ARPS, NTSS and SES) with respect to time and PSNR . It is the longest time for ES and PSNR estimation. Compared to other algorithms, the ARPS has a logical calculation time and PSNR. They also investigated into effects in step size p and macroblock size, and PSNR. Mohammed Ebrahim [95] studied on Multi-reference motion estimatation techniques. Their study concentrates on one of the most used video encoding techniques, the hybrid blocks motion compensation (HBMC). There are two prominent video compression standards in both the H.26X family and the Moving Picture Experts Group (MPEG). It solves a variety of problems, such as bit rate, complexity, picture quality and resilience. Erik Cuevas et al. [29] performed with an ABC (Artificial Bee Colony) block matching algorithm for motion estimation. A new algorithm explained, that reduces the number of search sites in the BM system, Artificial Bees Colony (ABC). The algorithm relies on a simple fitness calculation i.e nearest neighbor's interpolation (NNI) algorithm. By identifying which fitness value can only be measured or measured, it saves machine time. The method can subsequently be substantially reduced in numbers and retaining good search power. M. Jakubowski et al. [67] compared different block matching algorithms and calculated efficiency computational complexity and suggested the application of them. Yuan Gao et al. [147] experimented an algorithm based on a motion vector distribution prediction which is a fast motion estimate. Vector distribution research was performed based on the hexagon search-algorithm and the motion statistics. They also planned on the distribution of vectors and the motion forecasting method for a more comprehensive search. Tung Nguyen et al. [137] described HEVC and coding of H.265 / HEVC technology. The H.265 / HEVC offers a flexible and sensitive representation by the image quarter method of the block structures in a motion-compensated forecast and transformation coding. Improved coefficients for different block sizes in H.265 / HEVC have also been given in entropy codes. Gangadhar Tiwari et al. [48] performed comparative investigation of image and video compression techniques. Their study explores, for example, various compression techniques and results. JPEG, H.264, MPEG, and many more. Goal and subjective methods are employed to compare image codification strategies. While PSNR offers detailed image measurements, SSIM offers emotional consistency with human visual perception considerations. Miller et al. [94] suggested video editing for students are much more accessible when anyone speaks about the latest video editing applications on desktops and laptops, such as the iMovie, GoProStudio and Windows Movie Maker. For marketing students, the differences between applications and their use for professional projects and individual projects are significant. They also suggested advanced new technology is required to equip the students with a toolkit during the marketing education research.

In the year 2014, Salonir. Mistry et al. [118] performed motion estimation with logarithmic search. There are studied variants of block matching algorithms. For complete study and quick search, they examined BMA and suggested that the comprehensive search algorithm gives the best cost of reproduction more machine dynamics. The design complexity of the device needs to be reduced to boost the spectrum of hardware growth. Muhammad Aakif Shaikh etal. [120] addressed using motion correction algorithm technique. Based on the SATD video process results are achieved in efficient and accurate manner. It is possible with SATD method i.e the most redundant information that can be performed in temporary domain frames. This means the reduction in size and finally in video compression. T. Toivonen et al. [133] performed video code-based statistical distribution motion vector refinement. In this method after the first rough motion estimation, a pre-calculated static table is generated, and after that the associated parameter distribution over several motion vectors obtained. The number of improvement points openly chosen to enhance the compatibility of the machine's quality with real-time encoders. Péter Tamás Kovács et al. [77] advised the applicability of light field applications in real time. In their study, at least three issues that needed reflection are taken and further analyzed to develop LF compression methods based on usage and treatment considerations. Mihir Mody

[93] designed new VCF (firmware video) architecture for growing power requirements for video codecs, including MPEG2, H.264, and HEVC, that lead to local RISC control solution. The local control firmware uses a basic sequence type video processor model as there is no real time operating system (RTOS). Author works leads resolution in video conferencing. P. Razvan et al. [108] finally found other areas where movement measures are theoretically useful, such as pattern recognition and medical imaging. In their work, motion vectors between consecutive images can be used for food analysis and functional analysis.

In 2015, Vineet Gandhi et al. [139] described a computational framework for vertical video editing. They also implemented procedures for the operation of a virtual pan, tilt, zoom during post production using key frame interpolation to track and identify actors based on their visual appearance. A vertical video editing of a computer system has been investigated by them for contribution is an interactive system that can help the film publisher in choosing and randomly generating complex films, taking into account of specific basic film. Li Yao et al. [82] realized a summary recordings of a temporary structure. They demonstrated the time structure use of frames. In the standard Youtube2Text and DVS data sets, the metrics were empirically checked on each form. Their experiments show that each model employs an enhanced approach to the base model and provides the better results for all systems. Zekun Hao et al. [57]. experimented with controllable video generation sparse trajectories. They presented pixels figure and hallucinate for generating a static picture with local gradients of 4 pixels. B. Girod [14] described motion compensated analysis of multi hypothesis. As suggested by him he pronounced that B-frames are examples of the multi-hypothesis movement correction in which movement-compensated signals overlap the video codec bites. It analyses the effect of motion compensation on the effectiveness of multi-hypothesis compensation, the effect of the residual noise level and the value of the perfect combination of N hypotheses. Suk Ju Kang et al. [128] addressed motion compensated frame rate using increased bilateral motion conversion. They introduced an updated algorithm of the motion compensated frame rate using an expanded movement route to improve accuracy. The algorithm proposed by them also expresses an adaptable motion vector to optimize the fake motion vector correction. Kwon Moon Nam et al. [98] developed a quick hierarchical vector assessment mean pyramid algorithm. The algorithm was based on a hierarchy vector-search strategy. The proposed hierarchal motion estimation approach reduces machine complexity as it uses a pyramid search pattern too. It minimizes the potential of a minimum local error measures to boost performance by applying the form of a motion for several candidate frames. Dr. Anil Kokaram [38] reviewed motion prediction, image and video processing algorithms. The problems of aperture orientation and movement of camera are the two main problems with capturing corrupt images/ video frames. The motion measurement in the corners is, therefore, the most vital issue. The drawbacks of gradient dependent motion are that it does not usually estimate movement. However, gradient dependent estimates are traditionally lower than BM and more accurate than computational cost. Dhara Patel et al. [112] carried out handheld camera equipment on motion compensation. They suggested that feature based techniques compensate the motion estimation. The technique is designed to stabilize extraction functions using fast matching techniques. This makes it simpler than other methods for obtaining more features. H. Fang et al. [53] detected and created a piece subtitle in video. A CNN is added to the pictures, and MIL information is used for word recognition. The sentence structure is then extracted directly from subtitles to remove previous sentence structure assumptions. Finally, the question of the development of picture subtitles is a question of optimization.

In 2016, Pei-Yu Chi et al. [27] designed video based interactive instructions. Their work demonstrated computer driven methods to promote the creation and learning of tutorials in software. To include little guidance from the author's presentations, a range of interactive technologies was introduced. In design a new learning variables are taken into account. Anudeep Gandam et al. [10] analyzed a review paper of video processing & its applications in which motion estimation and compensation, video processing techniques are discussed. Motion compensation is the spot for reducing calculation costs in future research to maximize estimation efficiency and reliability. Carl Vondrick et al. [22] concentrated on recording output based on scene dynamics in which they feel, computer vision is essential. Their work explores how complex dynamics from large unlisted quantities can be known for video capitalization on adverse learning phases. Since annotations are costly, they believe the promising way to learn from unlabeled information. Dhanashree Bhujbal et al. [18] tested video stabilization techniques using block-based motion vectors and explored video stabilization strategies and methods. The system has shown the main steps in the video stabilization pipeline, including movement assessment. Various motion estimates are discussed, such as the pixel and functional movement estimating process. They addressed several blocks covering algorithms in their work, such as Full Search, TSS, NTSS, 4SS, FSS, DS, and HBMA one-by-one.

In the year 2017, Neetish Kumar and Dr. Deepa Raj [102] investigated video processing and its applications. Their study focuses on possible attempts to tackle how protection should be appropriately represented and how multimedia data should be adequately secured. Leake Mackenzie [88] has explored computational video editing for dialogue driven scenes. Editing dialog driven live scenes are challenging to pick clips and organize them in the story-transferring sequence employing frame based video editing instruments.Yitong Lietal [146] experimented on text video creation. They present the text as a VAE-GAN hybrid framework for the generation of images. The emphasis of the method contributes to the static value of text input videos. Saman Naderipariz et al. [97] researched on camera architectures. Their research is continuing on long-term computer-based video streams via mobile cameras. Given the high requirements of computation, general visual treatment is currently required. They developed a conventional mobile video processing pipeline called "Glance frame" to support inflexibility, success and precision. Luming Zhang et al. [83] demonstrated how to intelligently transform coherent video clip to a personalized album. Their proposed structure comprises scenic recognition model. Olivia Wiles et al. [141], examined frames that can also be manipulated as video editing tool. Their model achieved all of the editing strategies without any load/facial/depth information annotations. Sonam T. Khawase et al. [73] checked for a description of the motion vector estimation block matching algorithms. Their literature shows the usage of vector specific algorithms for motion estimation of various applications, including medical, security, scientific and psychological field studies, etc. A.V. Paramkusam [6] suggested the layered motion vectors that are evaluated by partial distortion steps for the computational motion rate. The successful movement estimation algorithm is designed to speed up the movement estimate in the layer without impacting its accuracy. D. Selvanayagi et al. [33] demonstrated compression-decompression of image processing through wavelet polynomial transformation. Quantized noise and poly aromatic wavelet transformation are used for decompression of compression/image content. The technology of Vector QWT (VQ-WTD) ensures higher rates of imaging compression. It increases image quality.

In 2018, Neetish Kumar and Deepa Raj [99] generated video using still images through Matlab. They discussed it in the experimental study which indicates that Matlab is also beneficial for working with video programs. It is an excellent platform for digital image processing techniques such as image compression, enhances, and objects detection in image segments and samples. Neetish Kumar and Deepa Raj [102] reviewed study and analysis of images in different color models. The study provides a brief overview of the various colors model, i.e RGB, HSC, etc. There were also discussed many types of compression techniques widely used for image and video transmission. Per Erik Eriksson and Akademisk avhandling [113] described video processing key questions on the feasibility and efficacy of vibrant action as an architecture source. In their study, the suggestions for a videographer are also given how to record perfect frames in video. Alexandra Trujillo J. et al. [8] determined about summarization of video from feature extraction method using image processing and artificial intelligence. Neural networks have spread to different locations because of their capacities to inform them about ratings and pattern recognition problems. In this type of system, labeling and handling images, identifying traffic signals, segmentation of biological models, have been particularly useful. Katsunori Ohnishi et al. [107] provides an FTGAN with two GAN, Texture GAN and Flow GAN, and discussed information, movement and participation. It only takes a few minutes for the video editing. Besides, video representation can be dramatically enhanced, and all other discrimination data are checked. They also suggested that it is important not only to keep videos in mind but to make videos. . Yatin Dandi et al. [34] managed residual vector picture and video generation jointly. To construct similar vectors, they implemented a new approach in codifying a description vector and model for each variation. In video production, this contributes to the generated image models and latent interpolation of the image space. In both video and imaging generation, experimental studies display changes in the baselines. Sergey Tulyakov et al. [136] introduced the MoCoGAN motion system with the decomposed video content. It automatically disengages the unattended movement of materials with video training information. MoCoGAN can isolate the human person's facial expressions from a speech to synthesize words performed by a new or new video and recognize names. Stefan Felsner et al. [43] found new evidence through pentagon maps. The coordinates for the pentagon corners can be calculated precisely since they depend only on the linear equation method solution's values. If the test had a polynomial limit on the number of iterations before the end, the algorithm would run strongly during polynomial times. Chao-Yuan Wu et al. [142] introduced photo interpolation reborn image compression. Their study includes the top-to-bottom video codec. It continually concentrates on deep interpolation of the picture. To decrypt interpolation, they encrypt compressible data parts not generated by nearby key rames. It leads to a faithful reconstruction instead of pure hallucination.

In 2019, Neetish Kumar and Deepa Raj [103] examined a new pentagon search algorithm for fast block-matching motion estimation. Their study explores the search pattern and methods of several block matching algorithms, and provided a new pentagon block matching algorithm that shows higher efficiency than the current hexagon and DS algorithm in terms of lower search points. B. Gnana Priya et al. [15]. studied the development of a 3D image with monocular 2d image depth indicators they provided a way for monocular sharpness and contrast of converting 2D to 3D images. They used the water shift algorithm to separate the image and build a depth chart with the profile indicators. Smoothing is used for the bilateral cross filters. No solution can be found to convert 2D to 3D images. The findings are only better for the KB dataset. With natural scenes and more realistic 2D images, this model should be an improvement. Yuki Nakahiraet al. [148] developed a model for the 3D information generation. To make the model more detailed, they included a video architecture for the generation of GAN photos. In the first half of thier architecture, the video depth is created, while one color video in the following half is produced by resolving the domain's deep color translation. By modeling scene dynamics and concentrating on depth data, they are able to create videos with better quality than a conventional approach. Junting Pan et al. [111], introduced a single semantic label chart that is rendered by video generation. The incorporation of semantic knowledge significantly improved the static content of the motion forecast. Comparison of impressive results shows the importance of other baselines in the proposed video creation phase. Alexis M et al. [134] carried out improving block dependent motion estimation-predictive motion vector adaptive field search technique. They suggested a new, highly successful block calculation algorithm and called it the predictive movement. The algorithm involves two techniques of generalized movement, MVFAST algorithm with adaptive thresholds and vector predictors and PMVFAST for rapid motion estimation algorithms. Siwei Ma et al. [126] experimented more compact visual signals and high-quality preservation on image and video compression. It is increasingly important for the excellent visible data era. In this neural network research, image and video compression methods were examined, particularly regarding the latest methods of correct analysis and video compression. Their study demonstrates the pioneering end – the neural network's full images in its early days

In the year 2020, Neetish Kumar and Deepa Raj [101] reviewed on algorithmic solutions for high-quality video editing software problems. Thier work provides a brief account and solution for problem editing. Video editing is becoming more critical in every human life, including journalism, movies, and online classes. There may be a chance of frame corruption to the recording side due to any type of camera movement, atmospheric situation etc. If an error occurs, then is corrected by the editing process. The animation and film industries rely entirely on the video editing process, and other sectors require moderate editing. For example, during the COVID-19 pandemic, the education sector was not in a standstill state. Video editing is also a blessing for catching mistakes.

CHAPTER III

Video Generation

The content of this chapter is published in-

- Neetish Kumar, Dr. Deepa Raj (2018), "Video Generation using Still Images", International Journal of Computer Sciences and Engineering (IJCSE), Volume 6, Issue11, E-ISSN: 2347-2693. - **(UGC Indexed)**

CHAPTER III

VIDEO GENERATION

3.1. INTRODUCTION OF VIDEO GENERATION

A video is a multimedia visual source, which incorporates pictures with movements. Any video capture system does not capture video, but the sequence of pictures at some desired frame rate. The human eye can not differentiate between similar frames if the frames are passing at high speed, because there are too much similarity present in consecutive frames. It gives an illusion of moving frames. Due to visual persistence, if series of similar frames played at more than 20 fps (frames per sec), human eyes cannot distinguish the processing delay of frames and it feels the frames as video. There is a long journey of video recording technologies from past analog to current digital era. Analog technologies, e.g., kinescopic recording, supported the first video recording system but due to low quality recordings it cannot be reached to stations with cable or microwave interconnects. Due to the popularity of audio tape recorder, the consumer market was entirely occupied with it, in the fifties. It provides a sense of the video recording logically related to the audio but much more challenging than the audio frequency that deals with video recording issues. The digital era has useful video recording resources for broadcasters. In the analog type, electromechanical irregularities were omnipresent. Noise is most vulnerable to phase modulation recordings, jitter transport, noise amplifier and inappropriate equipment layout during operation with analog signals. The above issue has been digitally fixed at some extent.

Text pictures are cultured, but video clips based on text have not been thoroughly explored. This section is based on video generation. One significant factor in video generation is that both the broad picture and object movement are to be decided by the text input. An empirical consequence of adapting the text to the picture system is a video that is not influenced by motion. The video forecast aims to learn a nonlinear transformation function between particular frames to predict the following structures. It is also essential for video production. However, for future stands to be provided, an entire video clip cannot be sufficient.

The journey from still images to video clips can be categorized in three steps. These are preproduction, production (also referred to as critical photographs) and post-production. Before shooting starts, preproduction includes any element of video preparation like scripting, logistics and other administrative responsibilities. Production is when video material (moving photos/videos) is captured, and the video object(s) are subjected to camera. Lastly, when the capturing process is over the raw video needed to be checked thoroughly. It is most important phase of a video generation activity in which the corrupted frames are removed, additional features may be invoked, and special effects can be added. The post production activity really plays a crucial role of a video clip for reaching among wider audience.

3.2. FRAME RATE

The frame rate is the number of video frames a camera captures per second. These running frames actually give illusion to the human eyes of video clip. There are lots of empirical studies done in this regards, if similar kind of frames are moving continuously then human eye feels these static frames as a motion pictures. A psycho visual experiment suggests that having a frame rate of 20 fps is sufficient to fool our eyes. The frame rates of 24 fps, 25fps, 30 fps, 60fps and 120fps, can be supplied in different standard forms. The frame rate is the basis of the movie and the first silent movies. The frame rate can be seen as a book of images with each paper is drawn as a frame it gives illusion of constant movable picture when it flipped around.

3.2.1. Human vision

Human eyes are very good detector of color and intensity up to a certain level. Beyond that it cannot distinguish the color pattern or intensity pattern. The whole image processing tasks are performed within this limit only because our eyes are the ultimate detector. The sensitivity of pixel resolution depends on the visual stimuli. If 10 to 12 images can be processed and perceived by the visual system simultaneously, it understands as motion pictures. The modulated light (eg a computer display) is seen as stable to above 50 Hz due to the flicker fusion. If frames are running at 30fps then there is only 33 milli sec times is devoted to a single frame only. In these 33 milli sec eye cannot distinguish much more.

3.2.2. Silent video

Silent films are displayed initially at the frame rate of 16 to 24 frames/ second, because their cameras were hand-wrapped, the quality was always altered to the scene's mood. Additionally at earlier times blade shutters were used to fasten the frame rate. Later, the frame rate shifted but it the result was not up to the mark. To minimize the eye's pressure, second and third blades were displayed two or three times, thus it reduced the perceived flicker projection. For silent films, the image rate rose between the middle and the late 1920s to between 20 and 26 FPS.

3.2.3. Audio-video

Video with audio changed the scenario of film making world. The acceptance of audio in video are widely appreciated, this leads to the multimedia development. There were no improvements in film speed when a sound film was released in 1926, as the human ear was vulnerable to change sometimes in place of the eye. Many theatres showed silent movies in 22-26 FPS, so 24 FPS were chosen as a solution for sound films. The 24 FPS rate was standard for sound films from 1927 to 1930, with many studios upgrading the different devices. The Film measures 456 mm (18.0 pounds) a second at 24 FPS through the projecting method.

3.2.4. Animation

With the development of fast computers the images are generated synthetically. This further leads to the animated video. Animation world changed the cartoon, gaming, video word entirely. In animation, sketches moves at different frames rates as like in case of real video. Animation industries not only changed the cartoon world but it also incorporates the VFX technologies that are widely used in modern motion pictures making.

3.3. VIDEO CAMERA

The Nipkow mechanical disc made early video cameras for radio transmission research from the 1910s to the 1930s. In the '30s, Vladimir Zworykin and Philo Farnsworth Image Dissector replaced the Nipkow method using a fully electronic camera tube system. This has been so recurrent until the 1980's when room technology issues such as burn-in or streaming were typically eliminated by reliable image sensors like CHC and the later active pixel CMOS (CMOS sensor). Solid-state image sensors are based on the metal oxide seminal technique, which comes from the 1959 MOSFET formation by Bell Labs. This resulted in the development of photo sensors for semiconductor systems, such as CCD and CMOS pixel sensors. A load of Bell Laboratories' coupling machine in 1969 based on MOS capacitor technology was the first image sensor for semiconductors. Later on, the 1985 Olympus designed the active NMOS Pixel sensor, and in 1993 NASA Jet Propulsion Laboratory created the CMOS pixel active sensor.

Available digital cameras were also permitted by practical progress in video compression due to the almost high requirements for uncompressed video storage and bandwidth. The discrete cosine transform (DCT) is the essential compression algorithm proposed for the first time in 1972. Since 1988, DCT-based video compression standards like H.26x and MPEG have been implemented to make realistic digital video cameras possible. Digital video cameras have been strengthened with the transition to digital TV. In the early 21st century, there were several digital cameras. The sporadic process between professional video cameras and film cameras has vanished with the advent of digital capture. Currently, only professional TV and other (except films) video cameras are used for medium-range cameras.

A video camera is a film camera for the TV industry that was initially created. However, it is popular in other applications, unlike a video camera collecting images from the video.

Video cameras are primarily used in two modes. Early broadcast is live TV, where the camera sends pictures in real-time directly to the monitor. Some cameras still serve live TV production, but most live streaming involves remote viewing that are required for defense, military/tactics, and industrial operations. For several years, video cassettes were used, but optical drives, hard discs, and flash memories were substituted the previous one. Images were saved in a storage unit for archiving and processing in the second mode. The video is captured in television production and is routinely tracked for future studies in unattended situations.

3.3.1. Types and uses

Modern video cameras have been developed and used in several different ways:

- Video cameras used in TV production and electronic field production (EFP) can be used with Professional or Handheld television Studio Cameras. These cameras typically provide the camera operator with excellent manual control, generally without auto-use.

Three sensors are usually used for separately recording red, green, and blue. Cams for TV, home video, electronic newsgathering (ENG), citizen journalism and similar are commonly used.

- The combination of camera and VCR and other recording devices with camcorders can be achieved in one unit. Since they transitioned to digital video cameras, most cameras have integrated media recording. Cameras with action also have 360 ° capability for filming. For defense, tracking and surveillance purposes, Closed-circuit TV (CCTV) typically is used with a PTZ camera. These cameras are designed for small, quickly overlapping cameras and can function carelessly. Sometimes, they must be used in conditions that are not generally available or inconvenient for people, which harden these hostile environments (e.g., radiation, high heat or harmful chemical exposure). For scientific or industrial purposes, they are used.

- Live streaming of video cameras from phone.

- Many mobile devices can capture 4 K of video even at high performance and have integrated video cameras.

- Unique camera systems are used for scientific studies, such as satellite and spacecraft analysis on artificial intelligence, robotics and medical applications. The purpose is frequently to provide infrared (night-vision and sensing) or non-visible X-rays (Medical and Video Astronomy) radiation.

3.4. ARCHITECTURE OF VIDEO CAMERA

Figure 3.2 illustrates how a video camera uses images and generates a video. It displays the various hardware sections of a typical video camera. The outside view is captured on the charged coupled devices (CCD). Some screens and drivers link the playback side screens to store the frame sequence. Some functions allow the user to set the desired video frame rate. Audio records and limits the different audio codecs through a separate cable line. Some software elements, including JPEG and JPG decoding that are used to produce images for quantifying transformations and entropy. There are a wired LCD and USB port for the whole machine, drivers, light flash and display screen drivers.

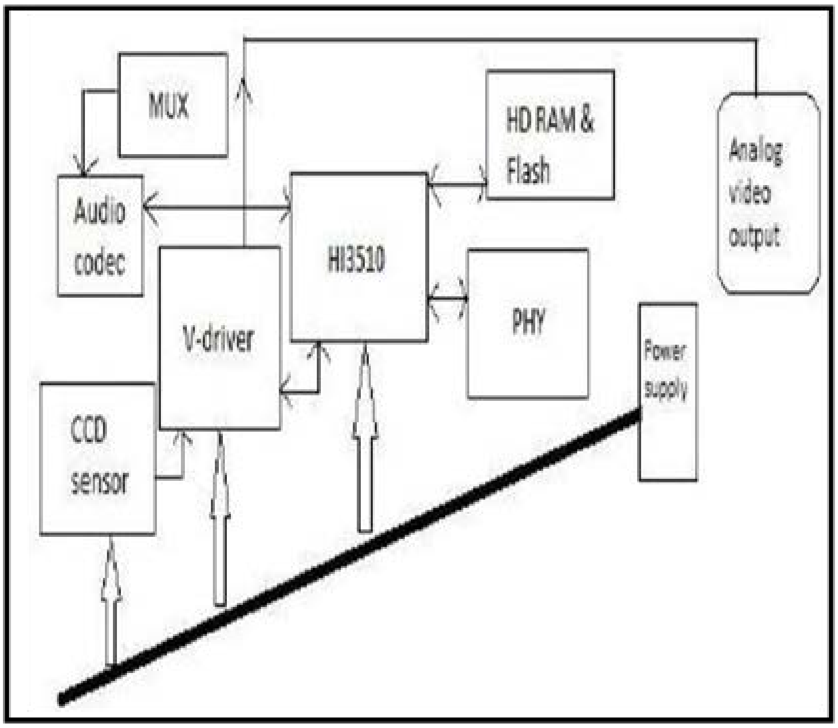


Figure: 3.1. Architecture of video camera

3.4.1. CCD Sensor

Willard Boyle and George E invented a 'CCD' in the Bell Laboratories in 1969 as an electronic module capable of holding electronic charges and being a 'charge-pair unit.' Nevertheless, the data storage method was adopted for storage. Their invention was also used to take several small steps to the central A/D converter into a semi-conductive exposure.

3.4.2. V-Driver

The driver is a computer program that acts or controls a computer-adjusted system of a certain kind. A driver offers the interface software to access the hardware and other software of operating systems without being familiar with the hardware.

The driver is connected to the device via a computer bus or a communication hardware system. When a calling program calls a driver's routine, the driver issues computer commands. After the data is returned to the driver, the driver invokes routines in the original software. The drivers are an operating system and hardware-based.

3.4.3. V-Driver

The driver is a computer program that acts or controls a computer-adjusted system of a certain kind. A driver offers the interface software to access the hardware and other software of operating systems without being familiar with the hardware.

The driver is connected to the device via a computer bus or a communication hardware system. When a calling program calls a driver's routine, the driver issues computer commands. After the data is returned to the driver, the driver invokes routines in the original software. The drivers are an operating system and hardware-based.

3.4.4. HD RAM & Flash

Memory Flash is mainly used for storage, whereas storage data are computed by RAM (Random Access Memory). Flash memory and RAM are quicker than alternative storage devices, such as hard disc and tape. RAM is quicker, but it's also cheaper.

3.4.5. Audio codec

Audio codec refers to audio encoder or decoder. A compression and decompression algorithm is used in a specific audio file or media audio format. Audio codec enables digital data streams to be encoded or decoded. The algorithm desired a minimum number, for maintaining the consistency of the highly stable audio signal. The storage and bandwidth needed for the transmission of audio files can be reduced effectively. The most sophisticated audio compression algorithms are linear predictive coding (LPC).

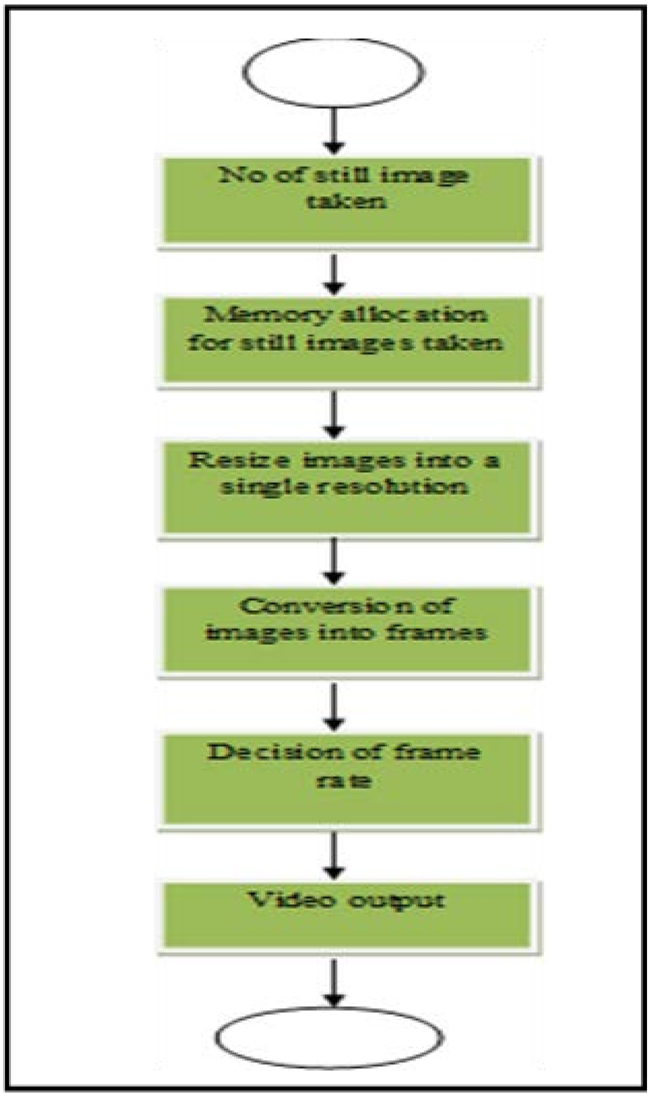


Figure: 3.2. Flow diagram of generation of video from still image

3.5. EXPERIMENTAL RESULTS AND ANALYSIS

To generate a video from still images, following steps are followed in MATLAB environment. At first, the n numbers of still images are taken and desired memory space are allocated by using cell () function. All still images are read by imread function from the hard disk or from folders. The images taken are may be of different sizes so it is needed to resize all in similar size as 256*256. The reason behind it is to make the process computationally efficient. Then all images are converted into frames by im2frame function. After that, the decision of frame rate is decided according to one's need. With the function writeVideo (), frames are stored into a video file and implay() function is used to execute the video that gives human eye an illusion of motion pictures. Pictorial representation is depicted in Figure 3.2.

Two video generation tools Movie maker and Animoto have been taken for performance measurement in terms of time and space in comparison with Matlab generated software. four test video is taken as Baby_video11, Baby_video22, Baby_video_grey11 and Baby_video_grey22. Still, images are depicted for Baby_video11 in Figure-3.3.

Figure: 3.3. Still images of Baby_Video11

For time consideration time elapsed for creating a video Baby_video11 having 11 frames, it takes 2.649 seconds without any additional effects. The size of the video generated is 591 KB. At this instant, it can be said that MATLAB provides more flexibility in terms of desired frame rate, so the video sequence at the playback side met guaranties of viewers need. For this video, Movie maker takes 11 seconds to execute it. The output video file is shown at the bottom rightmost corner of the above Figure 3.4.

Figure: 3.4. Still images of Baby_Video22

Another video Baby_video22, it contains 22 frames, still, images are depicted in Figure-3.5. MATLAB takes 3.584 seconds to execute it. The size taken of the video is 1.03 Mb while movie maker software takes 22 seconds to execute it. The output video file is shown at the bottom rightmost corner of the above Figure 3.4.

Figure: 3.5. Still images of Baby_video_grey11

The grey video Baby_video_grey11 that have 11 frames takes slightly less time with respect to the color frames processed. 11 frames of the grey image video take 1.124 sec and size acquired is 494 KB. The grey image of another video Baby_video_grey22, having 22 frames takes 2.982 sec and occupies 923 KB of space. The size of the video generated by the grey image sequence is also less than the video generated by the color images. Figure 3.5 and Figure 3.6 show the still images of baby_video_grey11 and Baby_video_grey22. The output video file with 11 frames and 22 frames is shown at the bottom rightmost corner of the above Figure 3.5 and Figure 3.6 respectively.

Figure: 3.6. Still images of Baby_video_grey11

The moviemaker tool takes the total no of time equal to the no of frames processing i.e 22 seconds. Another software of video generation tool Animoto, for generating the video baby_movie11 takes a 31 second for the completion with the inclusion of special audio and video effects. Without Effect of video and audio, it is not working. After this study, we can say that Matlab is a good plate form for generating any video or still images of any sizes and any quantity very fast and space complexity is also less.Table-1 one shows the performance analysis of the Matlab generated tools with Animoto and movie maker and Figure. 3.7 shows the pictorial representation of their performance. Figure 3.8 shows the comparative time of different video with movie maker and MATLAB.

Table: 3.1. Comparison analysis

Test videos	**MATLAB software**	**moviemaker**	**No of frames**	**size**
Baby_ video 11	2.649 sec	11 sec	11	591 KB
Baby_ video 22	3.584 sec	22 sec	22	1.03 KB
Baby_ video 11	1.124 sec	11 sec	11	494 KB
Baby_ video 22	2.982 sec	22 sec	22	889 KB

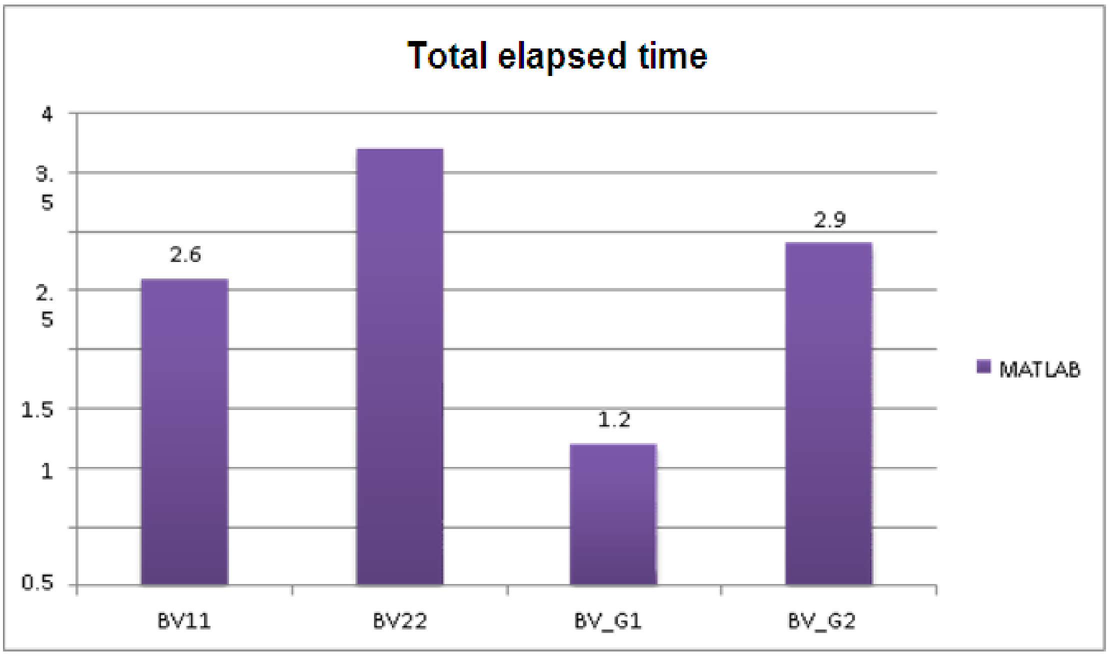


Figure: 3.7. Total elapsed time for each video

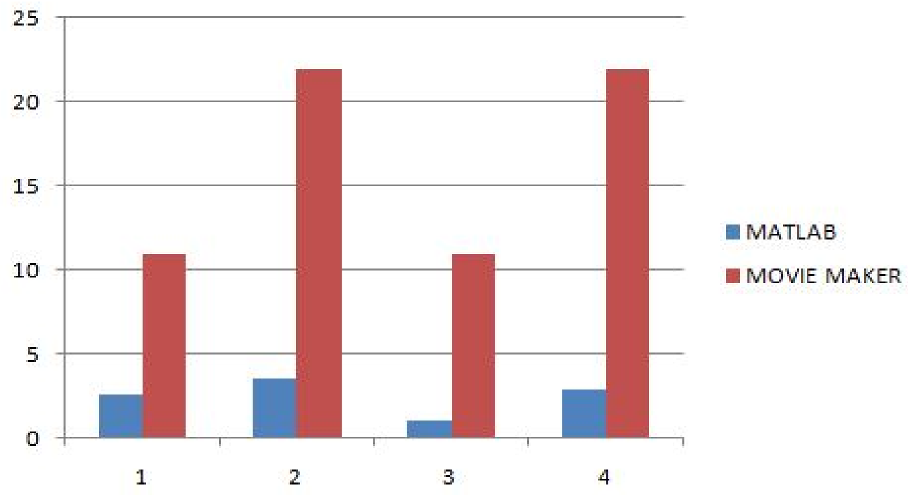


Figure: 3.8. Total elapsed time for each video through movie maker and MATLAB

3.6. MAJOR FINDINGS

For digital image processing Matlab is an optimal platform for compression, enhancement, and object detection application. The research shows that Matlab can also be used for video applications. The above experiments demonstrate that the video file is created using all its frames in a video file. Video file is generated and all the frames are stored in the video file and execution of the video takes less time and acquires a less space in the memory compare with Animoto and movie maker.

CHAPTER IV

Motion Vector Search and Motion Estimation & Compensation

CHAPTER IV

MOTION VECTOR SEARCH AND MOTION ESTIMATION & COMPENSATION

4.1. MOTION VECTOR

A motion vector is the key element of the video compression. Basically, motion vector depicts about displacement of blocks. The changes in consecutive frames in a video clip are not measured by frame to frame basis. As it is well known that in a frame there are two much similarity in pixel intensity among neighboring pixels, it is also extremely time complex job to analyze a frame on pixel basis, so the concept of block measurement evolved. During the process of motion estimation and compensation, the motion of each blocks are measured. That displacement of block from previous to current block is known as motion vector. Actually the process of motion estimation (ME) and motion compensation (MC) are done for finding the displacement vector that is known as motion vector.

In ME and MC the motion vector are searched in consecutive frames. There is a long history of evolution of complete search or full search to modern fast search techniques for finding the motion vector. There are different approaches performed to reduce the machine complexity. Quick searching algorithms adopt the specific patterns that tests for less point number, instead of a full search. The locations in the search window are evaluated using a complete search algorithm. It is desirable to have the best match for a specified search range.The cost function is used to find motion vectors in order to estimate movements in search windows for each possible position.

4.2. MOTION VECTOR SEARCH

The motion vectors may be measured either in cases of forward motion estimation or backward motion estimation. The below figure shows the measurement of movement of blocks and finding the motion vector.

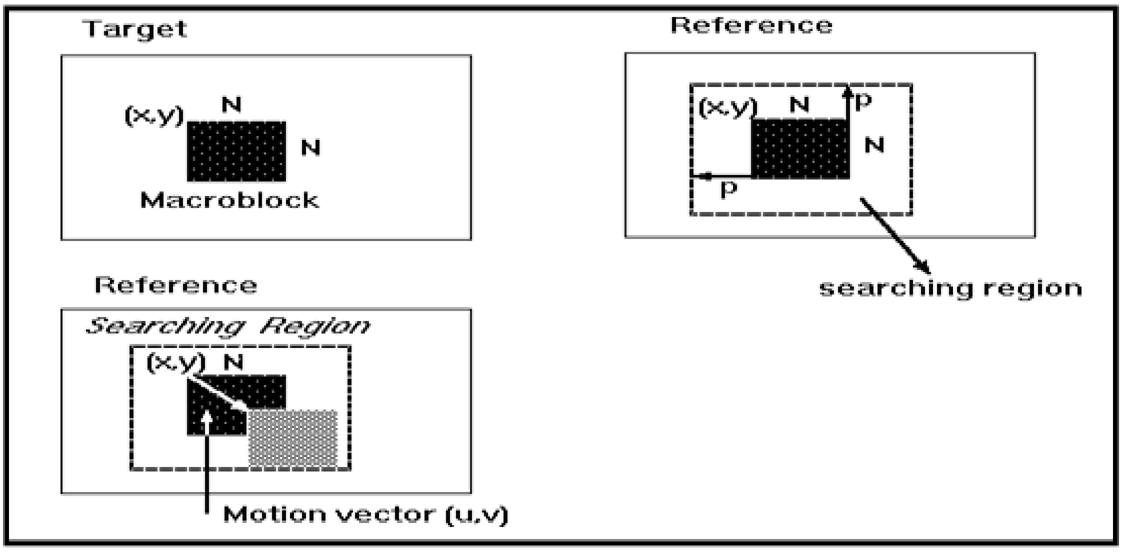


Figure: 4.1. Motion Vector Search

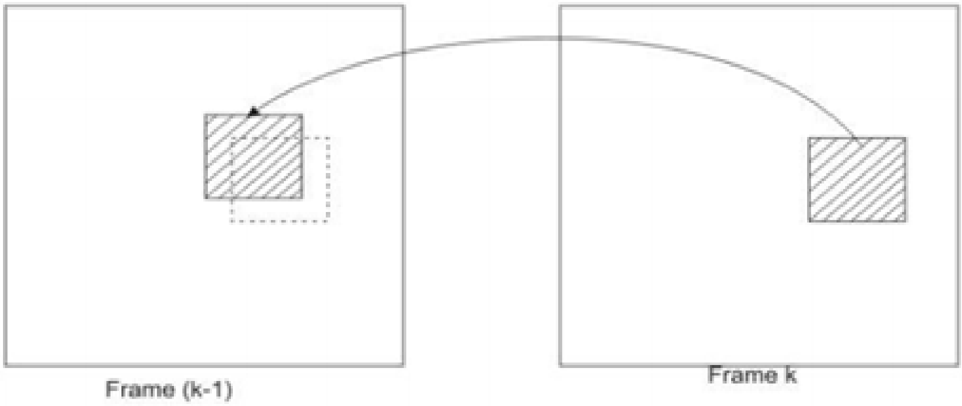


Figure: 4.2. Backward Motion Vector Search

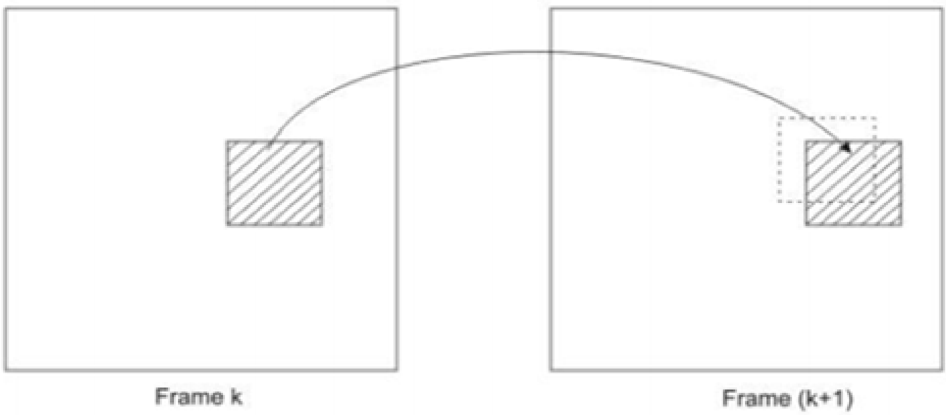


Figure: 4.3. Forward Motion Vector Search

4.2.1. Full Search Method

The complete search algorithm is a simple estimation tool for the measurement of movement. In this search technique every possible position are searched in a window. The window is relocated inside the search window, searching for the best fitness for the position of the candidate block. Absolute positions (2p+1) * (2p+1) need to be checked, where p is the search range of the block. There are different approaches for machine complexity reduction. A quick searching algorithm adopts a particular pattern for less search point, instead of a full search. All locations in the search window are evaluated using a complete search algorithm. It is desirable to have abest match for a specified search range. It uses the cost function to find motion vectors to estimate movements in search windows for each possible position. The full search algorithm needs more computations and for finding optimal results.

There are many issues with complete search methods. Some of these are as follows:

1. Finding the whole search points from [-p,p].
2. Operationally expensive,
3. Computationally challenging.

Of course there is a need of true motion vector in order to compress the video but not at a cost of too much delay. Because if a video clips are play backed at 30 fps then we have only 33ms time for compression of a frame. In this 33 ms we have to perform motion estimation and compensation for finding the motion vector and the idea of complete search algorithm is to search each and every possible position. So it is not feasible in real time scenario. The complexity of complete search is $O(n^2)$.

4.2.2. Logarithmic Search Method

Through the complete search approach the probability for finding the true motion vector is almost high but the biggest drawback of this method is that it is too much time consuming. After the 90’s various approaches are developed to deal with the finding of motion vector with less time, so fast searching algorithm came into existence. Logarithmic search is one of the fast search technique for finding the motion vector without searching the whole places. The algorithm was based on principle of optimality, in which we have to satisfy with only some sub optimal solution. The evolution of logarithmic search plays a pivotal role in fast search algorithms. This search concentrates on a particular area, while the center is placed as a diamond by four others. The search center and four adjacent points take the cost function. In the first iteration it searches 5 points including the centre. After finding the lowest value among those points, the same process is performed assuming the lowest as centre. Brief description of this algorithm is as follows. The search field will be evaluated in place [0, 0], in [0, + s], [0, -s], and [-s, 0], where [+ s, 0] is scanned. Step size is halved only if the midpoint of the preceding stage is the minimum deformation point (x, y) or if the real minimum point is located inside the scan boundary. Otherwise, the stage scale is the same.

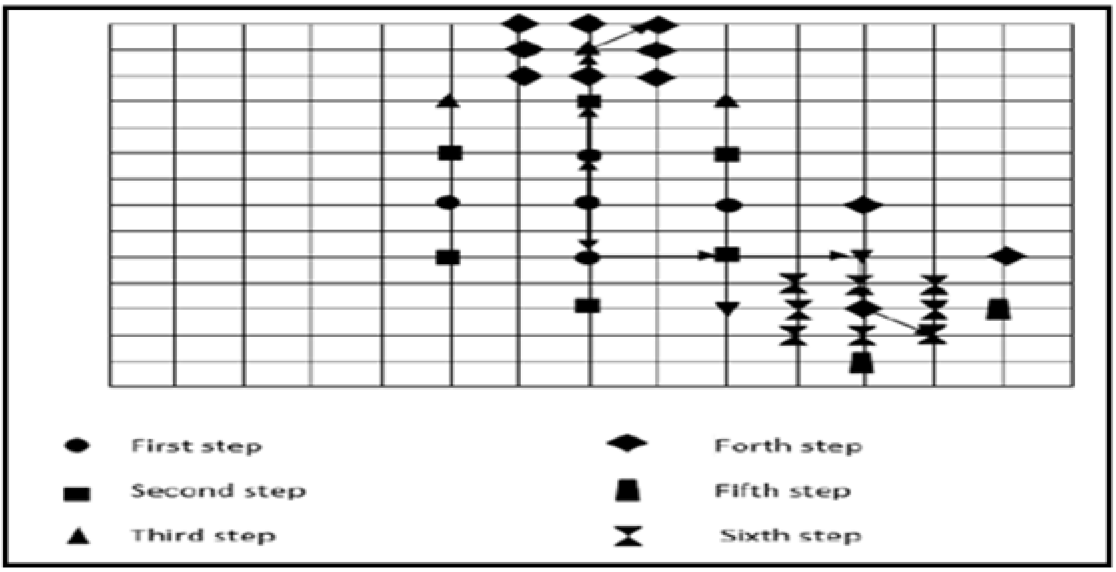


Figure: 4.4. Logarithmic search

These are tested and calculated to be the best fit for the objective block to the minimum the distortion measurement point and then stop the algorithm. If not (step larger than one), the candidate blocks positions [x, y], [x - s, y], [x - s, y], [x, y - s], and [x, y - s] the distortion is measured. The search takes place in different stages by successively growing the search area to a low level in each step. 2D logarithmic search has many levels.

Stage 1

The maximum allowable displacement is d is searched from centre i.e d displacement from left, right, up and down. The search block and the four candidate blocks of the x and y-axis are compared to the target block to determine the best match. The five places are like the five cross points (+). Therefore when the search centre is located at [0, 0] at the [0, 0], [0, + s], [0,-s], [-s, 0] and [+ s, 0] the candidate blocks are evaluated.

Stage 2

If the middle suits well [x, y], the stage 's' will be split. If the best match is in one of 4 external places, the next stage will be centre. This is [x, y] = [a, b] where [a, b] is the most suitable match.

Stage 3

When the steps 's' are equal to one, the best match is decided for the goal block in all nine blocks around the centre position. That is [x-1,y-1], [x-1,y], [x-1,y+1], [x,y-1, [x, y], [x,y+1], [x+1,y-1], [x+1,y], and [x+1,y+1] are examined the algorithm is placed. If not (step larger than one), the candidate blocks positions [x, y], [x+s, y], [x-s, y], [x, y+s], and [x, y-s] are compared.

4.2.3. Hierarchical Estimation of the motion vector field

The hierarchical evaluation of the motion vectors (also known as pyramid search) requires a popular approach for motion estimation. The computing complexity is low when it is used. In a single frame, the pixel block is checked to find the best match and the respective motion vector. In heuristic algorithms, there is an absolute presumption that only selected block numbers are comparable to the original block form of the vector field.

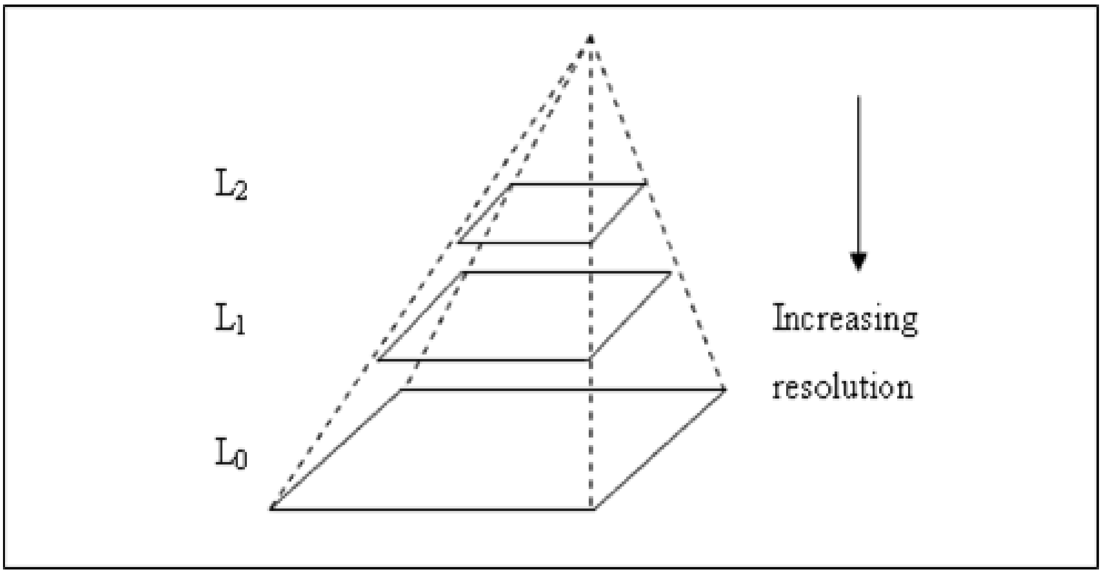


Figure: 4.5. The hierarchical (pyramid) structure of an image.

To generate a less resolution image from the initial image, two approaches, mean strength or sub- sample can be used. A block of 4 pixels with a medium intensity approach is replaced by one of the images with a grey level.

Sub-sample sampling is another popular method which uses pixel block. This method results in less consistency but it is significantly faster than the medium intensity. Every 0-level block on the first frame is placed once the both pictures which have been created for higher-level block.

To compare the blocks, measurements of the block difference must be made. The most widely used block distance measure is the average absolute difference. Both components of the block distance equation shall be the mean and pixel difference classification. The entire search algorithm is specified intensively. The decline in the resolution therefore helps one to find the best match in a large area without having unnecessary comparisons. The highest level search algorithm makes a rough approximation of the motion vector. The initial approximation must be improved for each successive step in achieving the final vector. There are various methods for this function. A general approach is to spread the results to the next level and heuristically search for a specific region to find the best match.

The vector is repeatedly propagated and optimized until the completion of a level 0 test. Another approach is to consider the four blocks in level 1 of level L-1, which is used to simplify the vector of movement. This approach is faster but less efficient as well. One common problem in the hierarchy is that the faulty match at a higher level is spread always downward and leads to a defective motion vector. This is because the low resolution of the target is still uncertain. That means that multiple candidate blocks are evaluated identically and those currently matched are not slightly better at level 0 compared to the higher level. This can be prevented by spreading several vectors to level 1. When a lot of BDM blocks appear at the highest level, the best-selected levels can be optimized.

4.3. MOTION ESTIMATION

Motion estimation (ME) is the most computationally complex component of video compression and video enhancement systems. In an image sequence, ME checks the movement of objects to get the estimated motion vector (MV). Methods based on pixels are computationally very complex and are not suitable for real-time applications. Object-based techniques significantly minimize the complexity of the computer but don’t achieve high efficiency. Block-based ME uses matching blocks that are ideal for producing high-quality hardware performance. The most popular technology is now blocking matching.

4.4. BLOCK-BASED MATCHING MOTION ESTIMATION

The most frequently used video coding method for ME is Block-oriented ME (BMME).In the MPEG 1/2, 8 x 8 and 16 x 16 block size dimensions are used. For MPEG-4 and H.264 / AVC, the block sizes are 4 x 4, 8 x 8 and 16 x 16 are used. ME aims to predict the next complete frame by connecting the motion vector to predict the macro blocks accurately.

Prediction efficiency and accuracy are calculated by the block size. The most popular and robust technology is the Block Matching (BM).

Two conditions are made in this technique.

1. Constant change in movement through small blocks (say 8 × 8 or 16 × 16). The smallest object often reaches the selected block dimension.
2. The scope is available for vertical and horizontal sections. The same is applied to the artefact series. This lowers the vector scope and reduces the algorithms efficiency.

The frame is divided into N x N blocks, usually of the same size. Each block is assigned to a motion vector. The motion vector is selected in the previous frame with a set of blocks of the same size, employing a corresponding blocking structure. In the previous frame, we can define the DFD between a pixel and the motion of the compensated pixel motions if we possess a possible vector (v) = [dx;dy].

$$DFD(x, v) = In(x) - In\text{-}1(x+v) \quad (4.1)$$

$$MAE(x, v) = \frac{1}{N^2} \sum v \in Block\, |DFD(x, v)| \ldots \ldots \ldots (4.2)$$

In n-1, the vector v gives the sequence of motions. The location of each block fits the graphical grid for effective motion estimation. If any interpolation occurs that needs to be accomplished. In most cases, there is substantial bilinear interpolation.

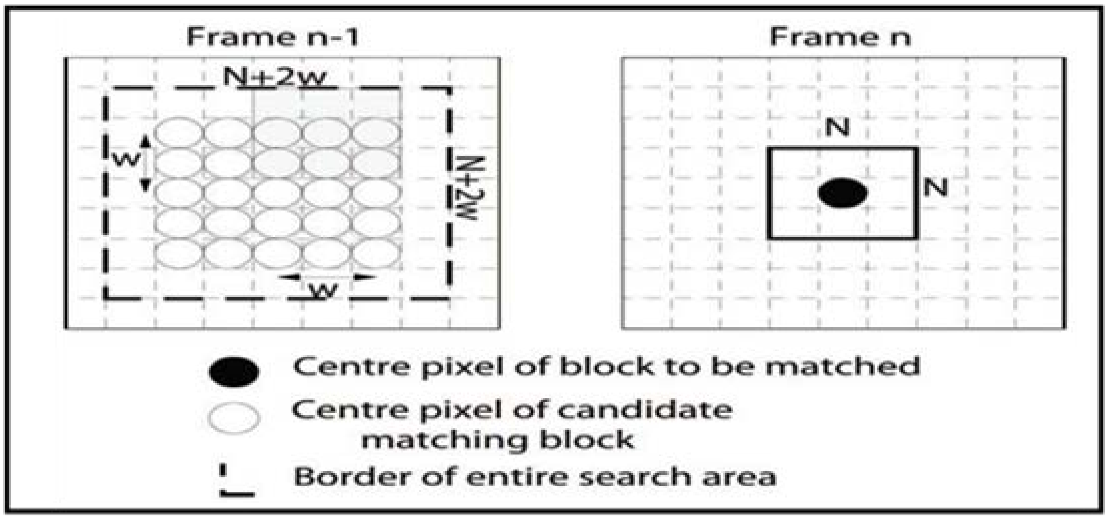


Figure: 4.6. Block matching motion estimate.

The positions shown by O in frame n – 1 are sought for a match with the N block in the frame n.

A search field is used for a full movement search. Compared to all size (2w+N) (2w+N) blocks the magnitude is estimated with maximum movement ± w for horizontal and vertical components.

4.4.1. Motion Estimation Procedure

For each current frame (CF) block (16 x 16, 8 x 8 or 4 x 4), the following procedure is performed.

1. The reference frame (RF) has a search area for each CF block. The macro blocks size in the search area which is usually taken 2-3 times larger (16 x 16) because in real time there is some kind of restriction in translational movement of any object in 33ms. The best match is found in the field after searching within the window. The best match is the motion vector (MV) i.e minimal motion.
2. The motion vectors and residue between the present and the reference blocks are chosen when the best match is obtained.
3. The MVs are encoded and transmitted for decoding.
4. The process of restoring the original image is reversed on the decoder.

The RF can be a previous frame, a future frame or a combination of two or more frames which has been previously coded.

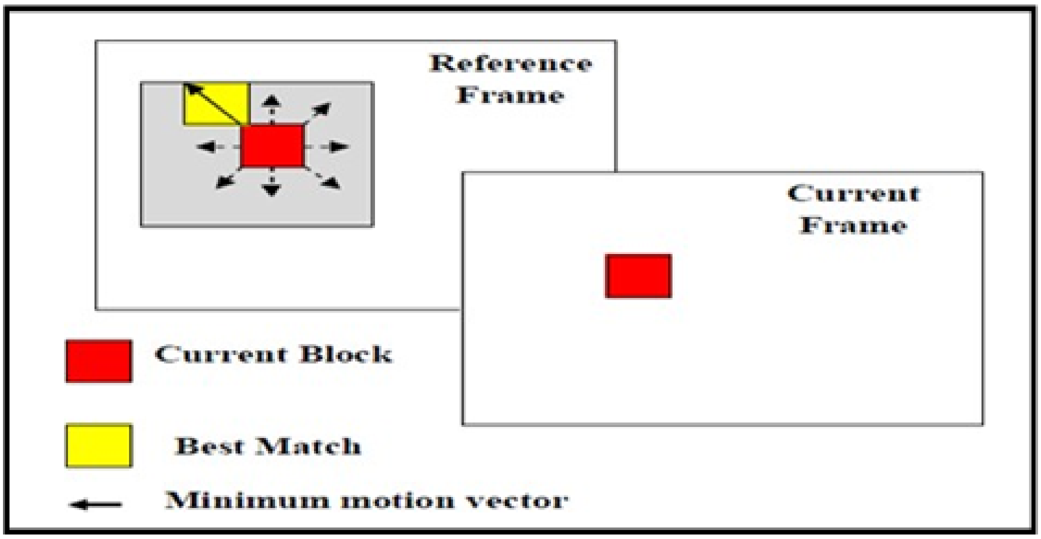


Figure: 4.7. Block – Based matching motion estimation

4.5. CRITERIA FOR MATCHING

The whole motion estimation and compensation process are very computationally expensive. To measure the effectiveness of performance, some matching criteria are used. The performance of matching criterion is not same. It further depends upon the internal computational power of its own criterion. A study is carried out concerning the mean absolute error, the mean of average square error, minimum block distortion etc.

4.5.1. (MAD) - Mean Absolute Difference

The MAD cost function is given in Equation.

$$\text{MAD (m,n)} = \frac{1}{N \times N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}|CB(i, j) - RB(i+m, j+n)| \qquad (4.3)$$

The scale of the block is N x N. CB (i, j) is the current block locations pixel (i, j), the pixel value for the new block is the block pixel (i, j) value. The motion vector of the MB is the (m, n) at the search position. The advantage of the MAD feature is its fast deployment of hardware. Unfortunately, MAD tends to prioritize small improvements and offers a more preliminary result than MSE.

4.5.2. Mean Square Error (MSE)

MSE is a tool for measuring energy and costs that are still included in the discrepancy. The MSE cost function is defined by equation below.

$$\text{MSE (m,n)} = \frac{1}{N \times N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}[CB(i, j) - RB(i+m, j+n)]^2 \qquad (4.4)$$

The scale of the block is N x N. The pixel intensity of the current block is CB (i, j), and the pixel intensity is RB (i, j). MSE (m, n) for MB at the search location is the MSE displacement vector (m, n) and (i+m,j+n). The advantage of the MSE is its accuracy, but its complexity is high for software and hardware implementation.

4.5.3 (SATD) - Sum of Absolute Transformed Difference

The SATD cost function is defined in equation below.

$$\text{SATD (m,n)} = \frac{1}{N \times N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}|CBT(i,j) - RBT(i+m, j+n)| \tag{4.5}$$

Block size is N x N, Location of the current block is CBT (i, j), location (i, j) and pixel strength (i, j) is converted into position (i, j). SATD (m, n) is the moving vector of an MB in the search direction (m, n), SATD significantly increases the efficiency, but the hardware needs to be incorporated into the ME loop which increases the hardware complexity.

The ME latency and efficiency have declined due to the addition of transforming hardware. This technology is used in a four-pixel precision ME H.264 / AVC system. The ultimate step of ME requires high accuracy (Richardson2003).

4.6. VARIABLE BLOCK SIZE MOTION ESTIMATION (VBSME)

In variable block size the blocks are made at run time according to the density of motion vector. For example, the ME variable block size (VBSME) is applied to select an appropriate block size in compliance with the most recent coding standards. The image (frame) in H.264 (Wieg et al. 2003) is divided into macro blocks. The 7 different block sizes are sub blocks (4 x 4, 4 x 8, 8 x 4, 8 x 8, 8 x 16, 16 x 8, and 16 x 16) which are separated further by each macro-block as shown in Figure 4.22. Each macro block consists of 41 sub blocks. For every sub block, an MV is generated in the VBSME. A total of 41 MVs forms the macro block. As the number of MVs in the VBSME has increased from 1 to 41, the number of operations needed for measuring SAD has also increased. This rise is insignificant for software implementation, but the hardware implementation is greatly increased and the number of benchmarks is up from 1 to 41.

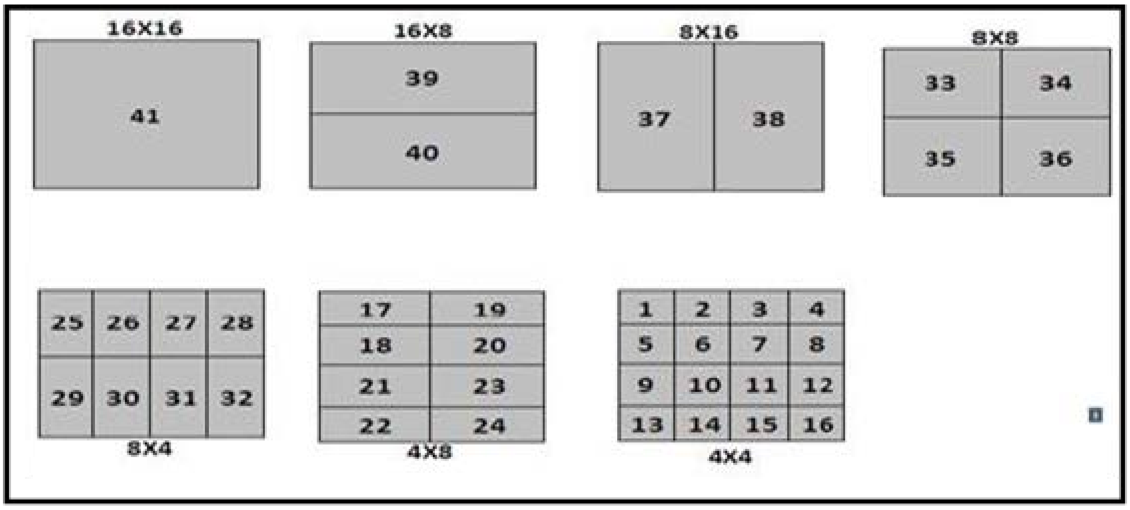


Figure: 4.8. Variable block size motion estimation

4.7. METHOD OF BLOCK MATCHING

We estimate a displacement vector in the positions of n1, n2, and n in the scope of the target. The prototype of the block centered is employed in the matching blocks described here. The search area (M2 and M1) will be defined in the reference frame, based on the pixel location (n1, n2) and then the multiple searches (2M1 C1 and 2M2 C1) are to be carried out. The block match in this method is not regularly done for each pixel.

MAD is often used in motion vector search applications because MSE is susceptible to external effects. An additional advantage of MAD is its more simple calculation, also it is less vulnerable to statistical outliers. Two examples of the previous technique include 2-D log search and three-stage search which uses fast search is discussed separately. Initially, there are four quadrants of the search window with three grids.The top right corner of the first phase, the third and last turn in the top right corner is the best matches in the lower right-hand angle. It’s noticed that all the three strategies do much better than differentiating the frame, which is zero for all moving vectors.

4.8. MOTION COMPENSATION

The process of constructing blocks by motion vector and residual frames is known as motion compensation. For the best equivalent block optimal motion vectors are measured by displacement in the new window through block-based algorithms. During image or video sequence creation, the filter/encoder is helpful in video processing which is based on the moving paths. Motion corrections are often used in all video encoding algorithms since it is easy to encode the least predictive error. The points of reference of the vector v or vector d are measured for this purpose. To understand the motion compensation in easy way firstly, the $Image_t(x,y)$ is split in block $Blocks_t[n](x',y')$. The function of motion compensation is to generate $Blocks_t[n](x',y')$ from any part of $Image_{t-1}(x,y)$. So, we have to find out the best possible matching block $Blocks_{t-1}[n](x'+mx,y'+my)$ not strictly nearby of 16x16 boundary. Here, mx,my is called motion vectors. Basically, the encoder transmits the error and the motion vector for each block. The error can be calculated between the target and reference as $Err_t[n](x,y)=Blocks_t[n](x',y')-Blocks_{t-1}[n](x'+mx,y'+my)$. The processor estimating $mx,my[n]$ for every n such that $Err_t[n](x,y)$ is minimized is called motion estimation. Whereas, the process of constructing Blockst $[n](x',y')$ from It image pixels and $(mx,my)[n]$ is called motion compensation.

4.9. BLOCK-BASED MOTION-COMPENSATED PREDICTION (MCP)

A video frame can be compensated by projecting the current frame from the reference background. This frame is called a guide and the context is used to calculate the motion compensation. The current system is predicted and the distance between the design and the existing structure is compressed which called hybrid is encoding. The principal video codec of hybrid nature are MPEG-4, H.26x, SMPTE VC-1 and China AVS.

4.9.1. The motion is two types

- **Global motion:** In this scenario the prediction is more complicated which shifts the camera and value shifts from frame to frame around each pixel in the image.
- **Local movements:** The camera concentrates on a moving human, camera installation and object direction but the rest of the scene is stationary as the videoconference.

The next frame is divided into rectangular pixel blocks to suitable boundaries in the local motion. This is referred as the block motion-controlled estimation. In MCP, the contrast of the existing structure is blocked by an obstruction with the next picture. The next frame is not transmitted by a series of motion vectors (MV) when the terminal compresses or sends video frames. The relationship can be defined by explaining the sum of the absolute differences (SAD) and the sum of the quadratic discrepancies (SSD), i.e. SAD and SSD.

4.10. MOTION-COMPENSATED FILTERING

In a video system the noise is minimized by applying time-compensated motion filtering and by adaptive space filtering. Different forms of noise enter into the video before compression and transmission. Objects emerge during encoding, recording and signaling, terrestrial or orbital contact. The noise introduced before image compression interferes with system and impairs the systems performance. Filtration usually reduces noise in an image; the definition of the edge leading to a loss of focus can also be reduced. Filtering system is always increases the computational complexity. As defined by the pixel group it is more difficult to reduce the noise by the moving objects in frames.

The recent breakthrough also improves the filtering methods and enhances FIR space filtering to reduce noise while maintaining edge parameters. It also uses motion vectors which were previously determined to reduce P-frame and B-frame and relates pre-processing computational complexity as part of the first-move images encoding.

4.11. COMPARATIVE ANALYSIS OF TSS (THREE STEP SERACH) AND DS (DIAMOND SEARCH)

The above mentioned two search techniques drastically reduce the time in searching the true motion vectors. Earlier the heuristic method was used in which the probability of finding the motion vector is too high but the time taken for entire search was a big issue in it. The development of fast search methods results in the finding the true motion vector in vary less time.

Table 4.1. Performance comparison of TSS and DS search

TSS (THREE STEP SERACH)	DS (DIAMOND SEARCH)
It was developed by KOGA.	It was developed by Shan Zhu
An initial step size is picked. Eight blocks at a distance of step size from the centre (around the centre block) are picked for comparison.	The DS algorithm employs two search patterns. The first pattern, called large diamond search pattern (LDSP) in which total 9 search points are checked
In the next step, the step size is reduced and again 8 points are searched	If the minimum is found at the edge points then 3 new points needed to be searched and if the minimum is found at the corner point then then 5 new points needed to be searched
If the minimum is found at the center then the size of the step is reduced but the new search points remains to be the same that is 8.	If the minimum is found at the centre then the LDSP is reduced to SDSP (Small diamond search pattern) in which only 4 new search points are checked.
The efficiency of this algorithm is quite good in searching and finding the true motion vector	This algorithm performance is 20% higher than the TSS algorithm

4.12 MAJOR FINDINGS

In this chapter different types of searching techniques and different metrics have been analyzed for finding the motion vector. Motion estimation and compensation are the basic need for video compression in real time. After performing analytical observations from various literatures it is found that there is still a lag in finding the true motion vector. The developed search techniques for getting a better compressed frame by reducing number of search points are still a goal for research. All the search techniques developed yet are not much optimally feasible. These studies and observations further motivated us to look for an optimal solution for finding the motion vector.

CHAPTER V

Pentagon Search Algorithm

The content of this chapter is published in-

- Neetish Kumar, Dr. Deepa Raj (2019), “A New Pentagon Search Algorithm for Fast Block-Matching Motion Estimation”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) Volume 8, Issue 10, ISSN: 2278-3075. - **(Scopus Indexed)**

CHAPTER V

PENTAGON SEARCH ALGORITHM

5.1. INTRODUCTION OF THE PENTAGON SEARCH ALGORITHM

Digital video shows significant redundancies in various image forms, such as space, time, perception and statistics. The compactness between the adjacent pixels is called spatial redundancy. To achieve the best compression ratio, the neighboring pixels redundancies have to be exploited. On the other hand, there are also strong temporal redundancies between the frames in a video sequence. The block matching motion estimation techniques manage the temporary overlap in successive video systems. Similarly, perceptual redundancies cannot be interpreted by the human eye so, without losing picture quality, it can be abandoned. The techniques for the MPEG and H.26x series block matching motion calculations form a critical part of motion related video encoding processes. The concept behind this method is to transform a frame into non overlapping block video sequence. The candidate block must be situated in the particular range behind placing the motion vector in the current context. In general the range is (2W+1)*(2W+1) window size.

Time-consumption is the crucial drawback of heuristic methods. The whole picture blocks are searched in the heuristic search that's why finding the motion vector takes so much time. It is not used in real time due to its time consumption so for finding the motion vector with suboptimal solution different fast search methods are developed. Various computer-efficient variants have been developed as the most realistic implementation, usually 3-step (TSS), New 3-step search (NTSS etc.). With the help of fast search, not only the coarse, but even the fine vector was also measured. Although the TSS performance was good but after the critical analysis of it, researchers moved towards the another variant of TSS i.e. NTSS. The center-specific motion vector results in the development of NTSS. Motion estimation (ME) represents a critical video coding technique, which eliminates redundancies in temporary compression. The ME calculations escape the implementation of the software in video-coding systems in actual time. ME is implemented in real-time applications with VLSI video encoding systems.

The BMA is exceptionally suitable for VLSI implementation due to its regularity and simplicity (2). Three variables affect the BMA's achievement: 1) the search process, 2) the search scope, and 3). multiple instances. Search methods, including Three Step Search, Bloc Based Decent Gradient Search and DS, can be collected individually with the FBMA motion vector. Many of them are divided into algorithms of FBMA.

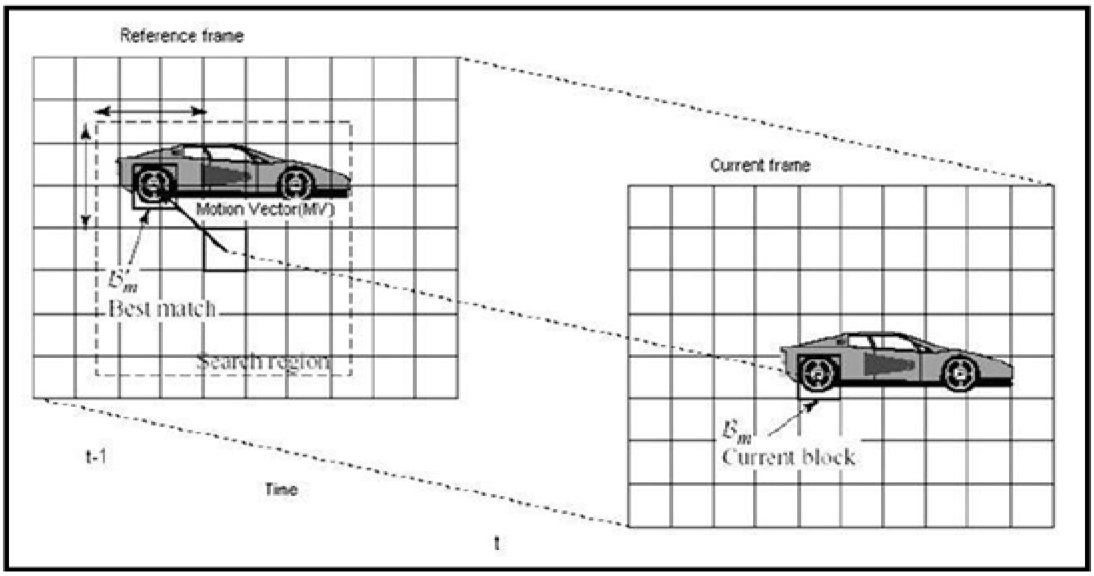


Figure: 5.1. Motion vector between the following two frames

Different algorithms were developed keeping in mind about the complexity of each is less than $O(n^2)$, for example the Diamond Search (DS) and the BBGDS. The DS Algorithm is preserved in diamond-shaped research, and in hexagonal search, the six surrounding points are searched. Complete search method is renowned for its high performance and precision and it is widely used in those contexts where time is not a big issue. FSBM (Full Search Block Matching) typically employs a systematic array architecture, which includes many parallel components for real-time processing. Diamond search (DS) offers an appropriate image consistency and it is the best option for real-time applications. The result is a promising search performance with the aid of fast algorithms such as DS. Although these algorithms are sufficiently short to be deployed for real-time machinery applications, the battery- powered mobile devices still require low energy hardware. For the new class of algorithms in stock costs and memory access patterns, existing VLSI architectures like the systolic array or tree architecture are not realistic. Due to its variable data stream, the device array architecture cannot be implemented.

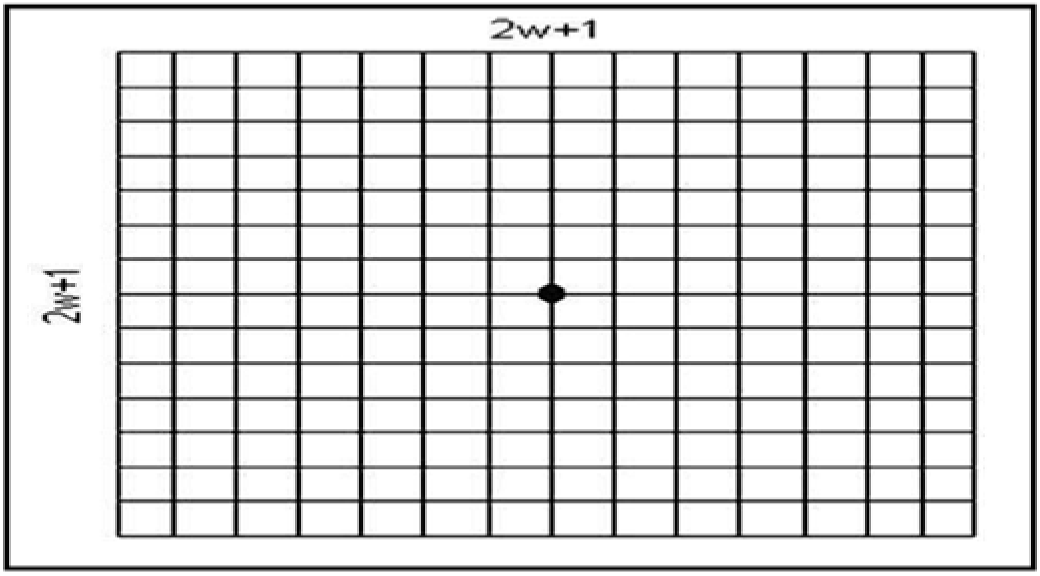


Figure: 5.2. Scale of (2W+1) search window

5.2. TECHNIQUE FOR THE SEARCH OF THE MOTIONVECTOR

It takes decades for a search pattern to measure motion vectors. In 1994, Tsai et al. proposed a video-sequence Angiogram based on a full DWT medicine system. This is achieved through the frame differentiation process to attain a high compression ratio. Gibson and others have suggested eliminating objects from the last image, which will reduce digital angiograms leading to loss of wavelets. Following quick-finding algorithms for real-time testing of video movement. For all different types of I, P or B frames, block search techniques are created. Below are some standard and effective techniques.

5.2.1. Hexagon algorithm

This approach was proposed in 2002. The number of search points after the algorithm's implementation is smaller than the previously used, and the expense of the calculation is also decreased. The center consists of six hexagonal terminal points and seven control points for the scan including the centre. The all six endpoints of the hexagon are not equidistant from the center. The figure 5.3 displayed that there are also two fewer test points in the hexagonal search pattern than the DS 9-point pattern. The hexagonal search pattern moves to one of the six endpoints with the center of the search. The small, shrunk hexagonal pattern contains four checkpoints (left, right, right, up and down at a distance 1 around the center) in the moving sector. The shrunk pattern of the hexagon is the same as the shrunk pattern of the diamond.

5.2.2. HEXBS technique: description steps

Step 1) A block of minimal distortion is evaluated for the broad seven points (0.0) hexagon and all seven points. There are two choices for this:

If Minimum Block Difference (MBD) point is found to be at the centre of the hexagon, proceed to (3) (end) else Go to (2) (Search) otherwise.

Step 2) A new MBD is chosen as center and a new hexagon is formed according to it. This newly constructed MBD hexagon is linked once again to a check at all seven points. If it remains the same then terminate (3), otherwise repeat it continuously.

Step 3) the large hexagon is switched into a small hexagon. The four new points associated with it is compare to the present MBD point. Newly initialized MBD point leads to end stage of motion vector.

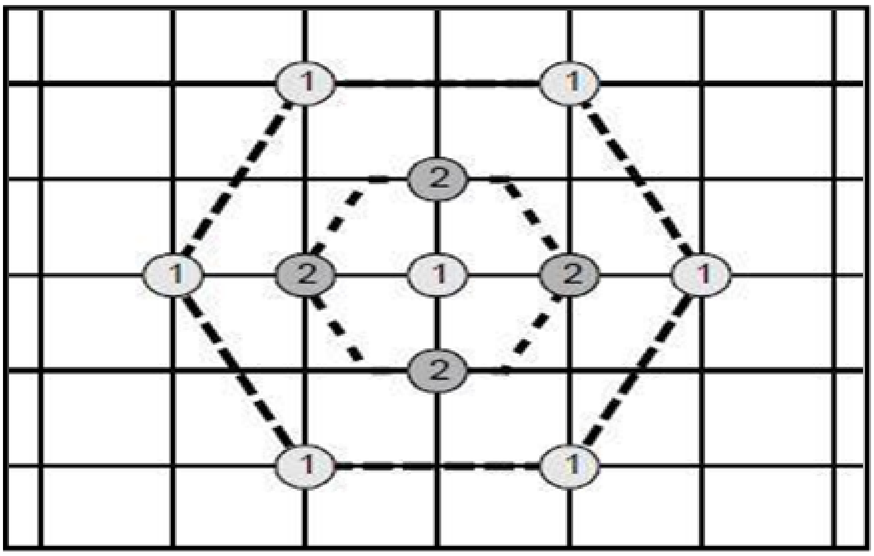


Figure: 5.3. Search for hexagon (HEXBS): hexagonal patterns big (1) and small (2)

5.3. DIAMOND SEARCH ALGORITHM

It was proposed in 2000. Diamond Search is the same as Four step search (4SS), except the diamond is shifted between the triangle and the square.

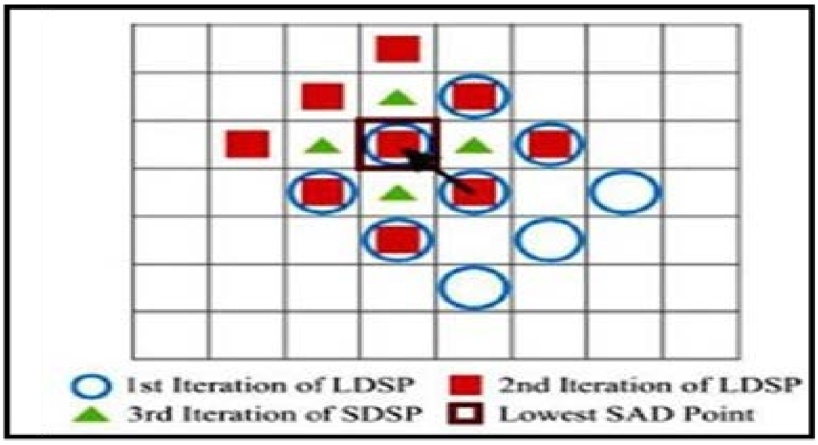


Figure: 5.4. DS method

The searching method uses two various forms of fixed patterns.

- LDSP - Large Diamond Search patterns.
- SDSP - Small Diamond Search patterns. The following is the algorithm:

LDSP- Large Diamond search patterns.

1. Begin with the search center;
2. Enable stage S = 2size;
3. Evaluate 9 places (X, Y) or Scan 8 positions with centre checkpoints (0, 0);
4. From the 9 locations searched, choose one with a minimum cost function;
5. If the minimum distortion is found at the center of the search screen, go to the SDSP level;
6. If the minimum value is one of eight locations outside the middle, place the new source in this position;
7. Reduction of LDSP.

SDSP - Small Diamond System patterns

1. Set the current search root;
2. Specify a new phase size S = S/2 (that is S=1);
3. Repeat the search for the lowest value;
4. Choose a position with the least value in motion vector.

This algorithm is very accurate, as the search architecture is limited. The diamond search algorithm is equivalent to a comprehensive analysis signal/ratio with considerably lower costs. LDSP is the first step. The fourth step is when the lowest point of the quest is key.

5.3.1. The steps of diamond search algorithm

The Diamond Search algorithm illustrated as:

*Step 1) l*et centre of LDSP assumed as origin, where we've 9 points to be checked on origin of search window. if the MBD points found at centre, then moved to (3) else proceed for (2);

Step 2) the new LDSP is created with the previous MBD point found, assuming the MBD as centre. According to previously described step s newly MBD point should be on found on centre, if it's not then moved to (3), else recursively repeat this step;

Step 3) LDSP is converted into SDSP (Small Diamond Search Pattern). The four-point of SDSP is tested. The new MBD point is the final solution of the motion vector i.e. best matching block.

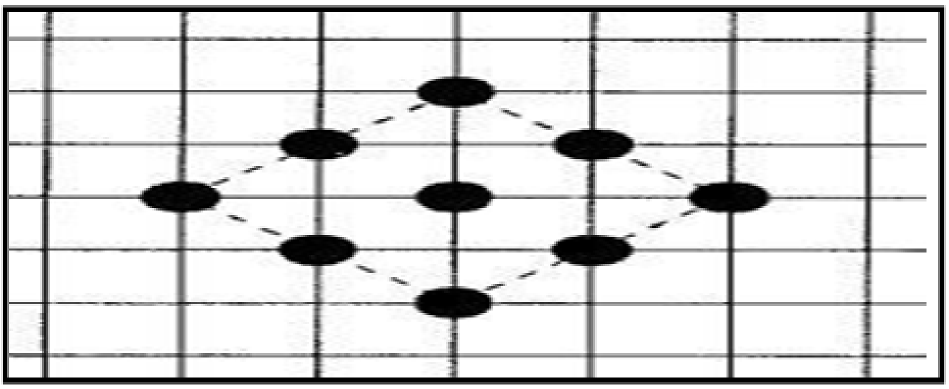


Figure: 5.5. Large pattern of diamond search

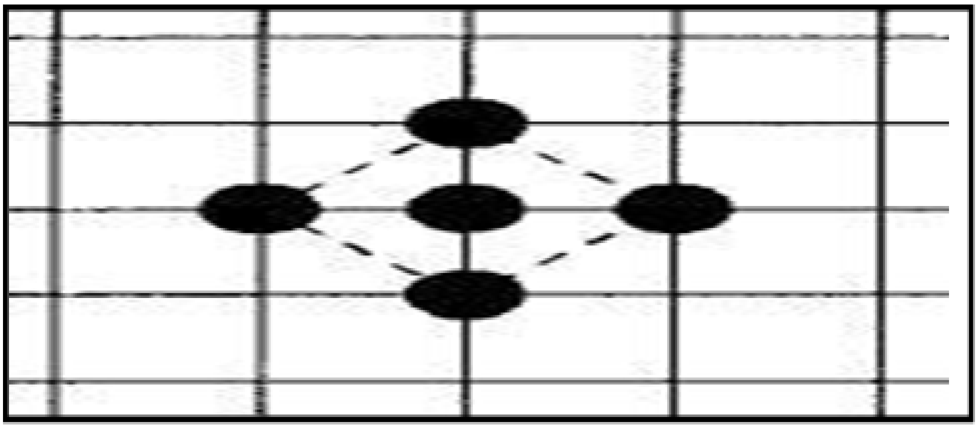


Figure: 5.6. Small Search Diamond Pattern

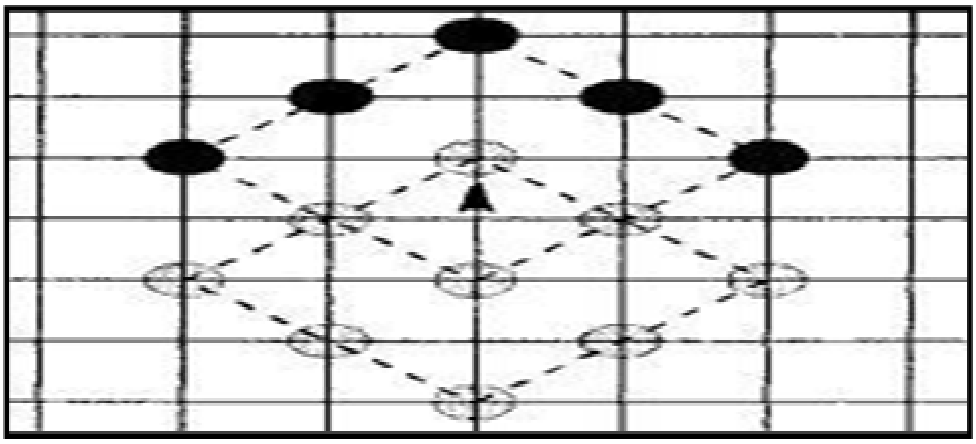


Figure: 5.7. The corner point of diamond search

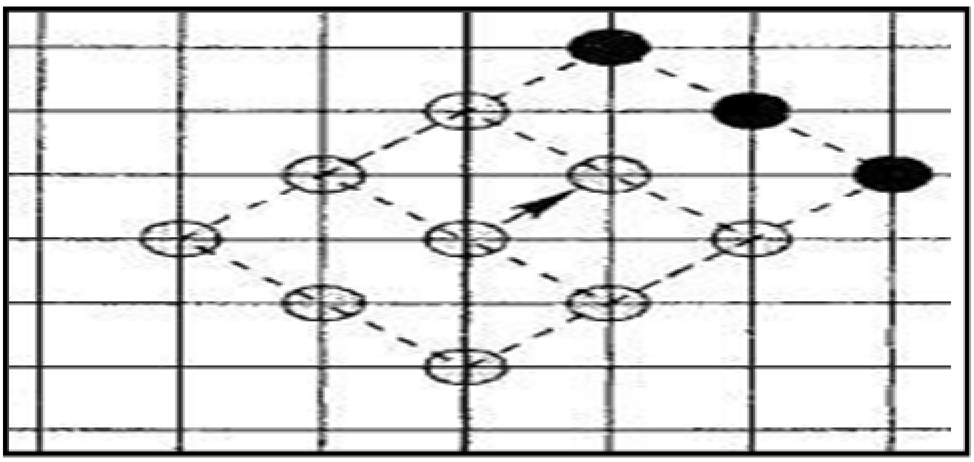


Figure: 5.8. The Edge point of diamond search

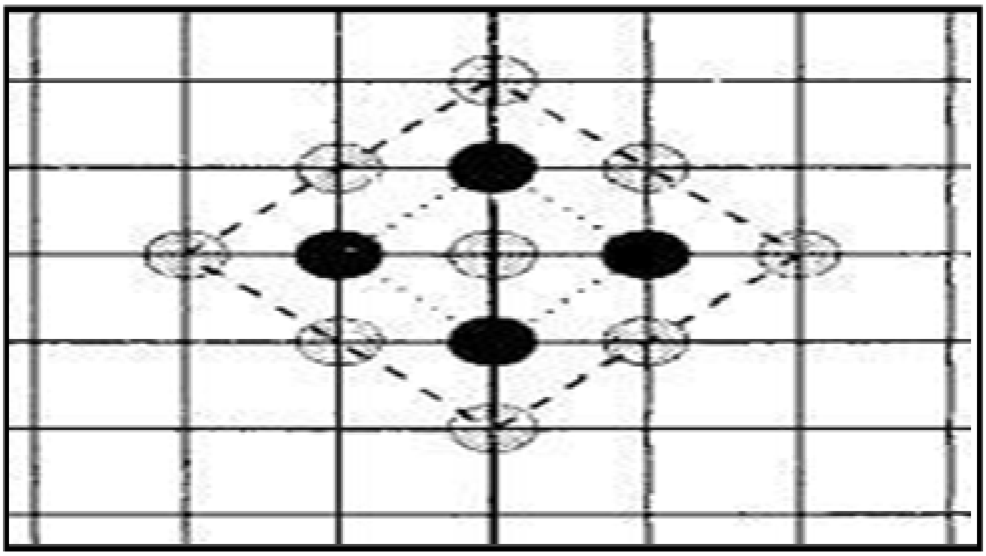


Figure: 5.9. The LSDP center point

The above algorithm sets the minimum number of points in the LDSP angle position on 5 additional points. 3 different points must verify the minimum number if the minimum is found on edge. The LDSP is converted to SDSP, when minimum is found at the central place. In SDSP, 4 new points are needed to search.

5.4. THREE-STEP METHOD SEARCH

The algorithm was proposed by Al and Koga in year 1981. Its search strategy is one of the best in the rough research pattern; this algorithm's efficiency is almost perfect. TSS uses a rigid mechanism and could miss small steps. NTSS offers a new search system and provides halfway stops to reduce calculation costs. It gives an enhanced TSS version. It has been widely adopted and used to meet earlier requirements, including MPEG1 and H.261, for the first fast algorithms. The steps in the algorithm are:

Step 1) Pick the first level scale. In a stage away from the core (the central block around), there is a contrast between 8 blocks;

Step 2) The process size is halved. The center is forecasted at the lowest distortion level. The two steps described above are repeated to less than 1 °. This algorithm, therefore, converges directly.

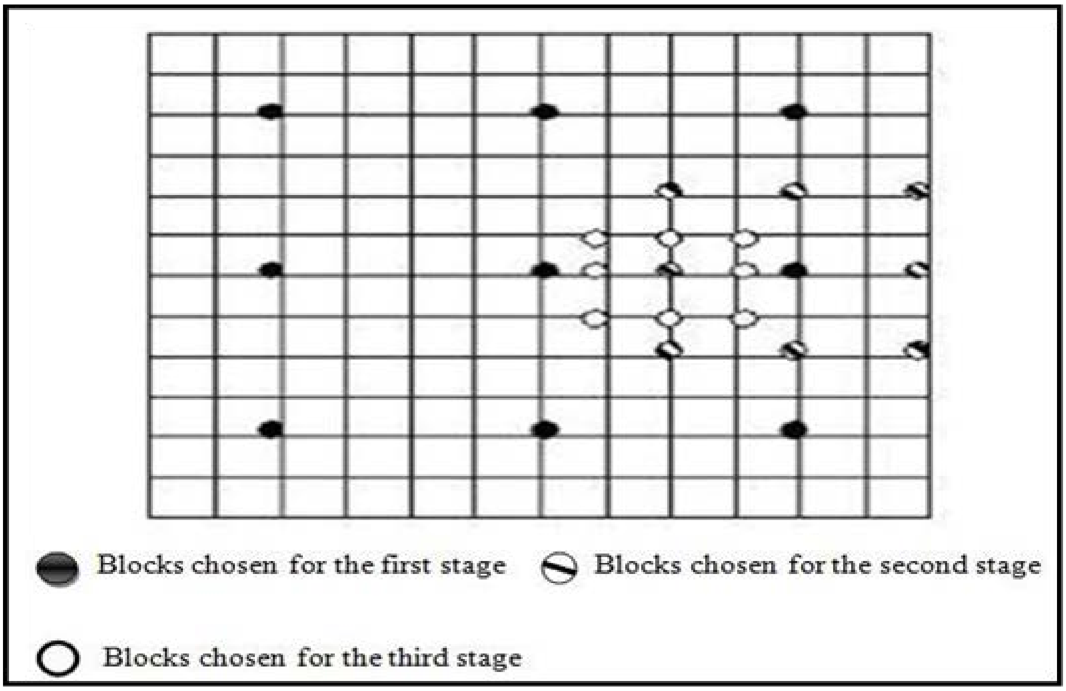


Figure: 5.10. Different stages of TSS Algorithm

5.4.1. The following is the algorithm

1. Choose one with a minimum cost feature from the 9 locations searched;
2. Enter the new source in the position above;
3. Place S = S/2 for the latest moves;
4. Repeat S = 1 scan.

The concept behind TSS is that a macro block which has a unimodal surface. A unimodal surface is a bowl surface which monotonously elevates cost weights from the word minimum. However, there cannot be two minimum quantities of a unimodal surface so that calculations can be combined and processed compared to an 8-point, fixed model quest for TSS. SES is the TSS extension that considers this sentence.

5.5. PROPOSED ALGORITHM

A new fast pentagon algorithm is proposed based on the previous observation. The algorithm is represented as figure 5.12.

From the center point of the Pentagon all the distance is equal of 2 units except the lower corner ends that are $\sqrt{5}$.

The Pentagon algorithm steps are as follows:

Step 1: The large pentagon with six checking points centered at (0,0), is tested for minimum block distortion with all six points. The MBD is tested and if, identified at the center of the Pentagon, proceed 3. (Ending); otherwise, go to step 2) (searching)

Step 2: Now a new large pentagon is created by assuming the MBD point of the previous search as the center. All new candidate points are checked and the MBD point is again identified. If the MBD point is still at the center point of a newly formed pentagon, then proceed to step 3) (ending); otherwise perform this continuous step repeatedly.

Step 3: Reduce the size of a search pattern from the large pentagon to four nearest neighbour points. The new four-point in the small pentagon tested for comparison with the present MBD point. The present MBD point termed as final solution for the motion vector.

The search technology is usually as follows.

Npent (mx, my) = 6 + 3*n, + 4

N = no iteration in the new statement.

mx, my = coordinates of vector motion

Npent = the proposed tool has no influence.

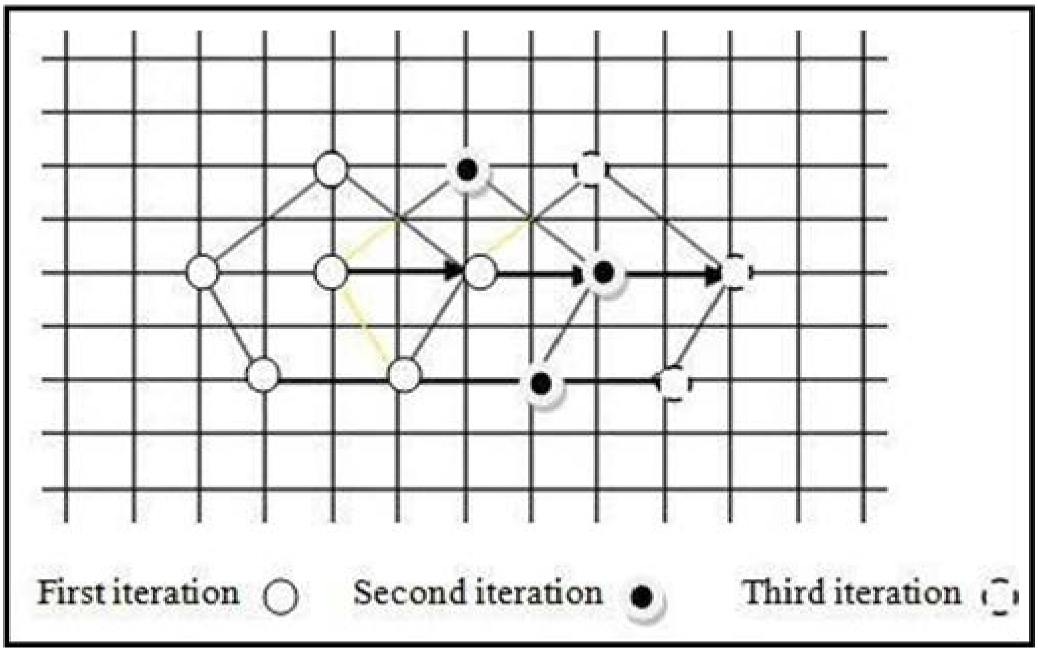


Figure: 5.11. Different stages of pentagon search

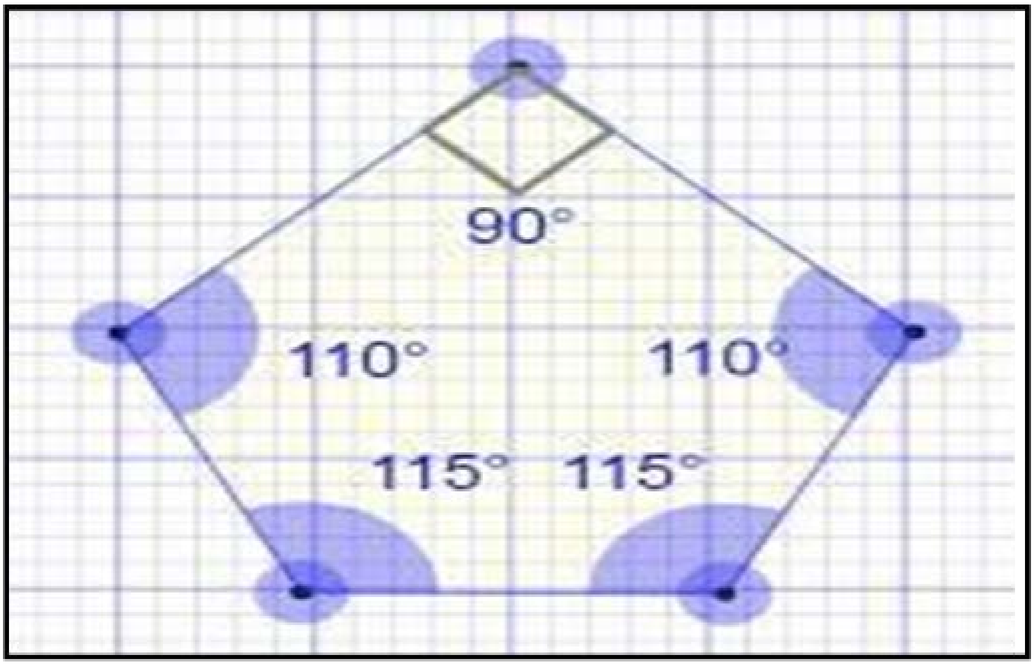


Figure: 5.12. Proposed pentagon shape

This pentagon's angles are 90, 110, 110, 115, and 115.

A. phase 1

If the minimum is found at the left or right hand side, only 3 new search points are available in the newly formed pentagon. The minimum should be checked in all three external levels.

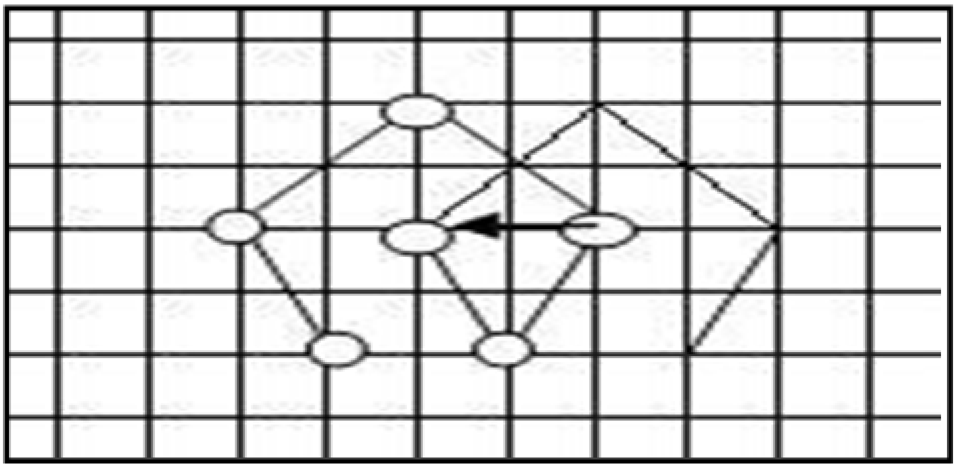


Figure: 5.13. Minimum MAD located on the left or right corner

B. phase 2

If the minimum block distortion found in the bottom left or right corner then it is tested for 3 new points. One point inside the previous Pentagon need not be checked because there is a correlation between neighbouring blocks.

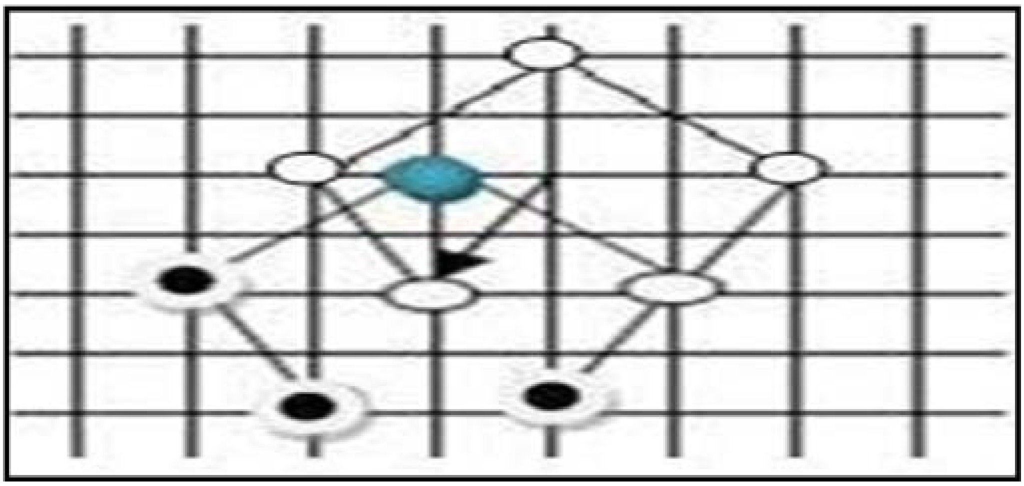


Figure: 5.14. Left place Minimum MAD found.

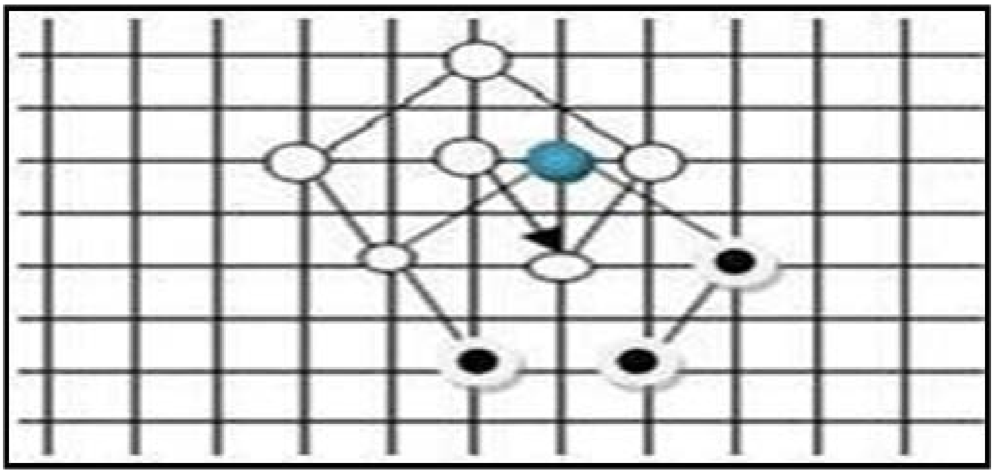


Figure: 5.15. Minimum MAD at the right position

C. phase 3

Another case may arise if the minimum found at the top most corner of pentagon then two points of the newly formed pentagon are inside the previous one so it is not checked.

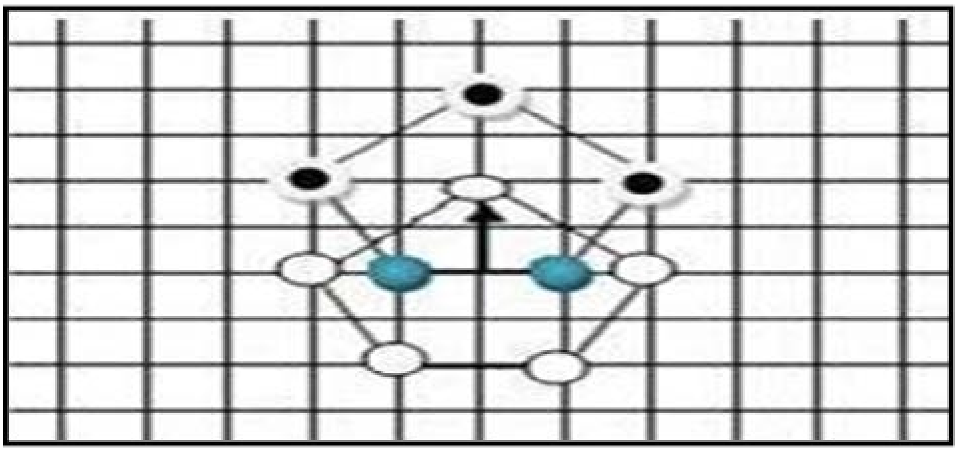


Figure: 5.16. Minimum in the Pentagon's top position

D. phase 4

If the minimum level is at the central position of the pentagon, the pentagon size is decreased. The new search position will be the closest position from the centre i.e. left right top and bottom.

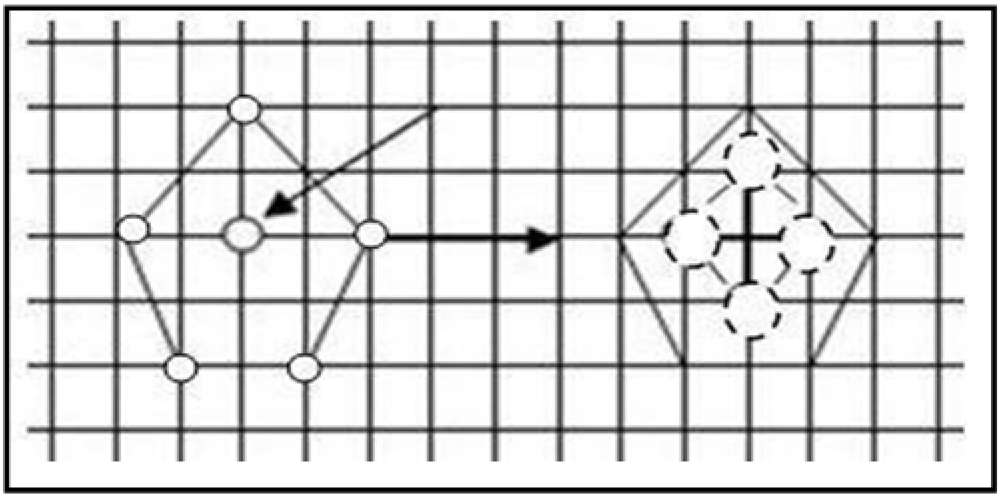


Figure: 5.17. Minimum located in the middle

5.6. RESULTS AND DISCUSSIONS

Two video sequences are taken for analysis purpose “missa” and “caltrain”. In caltrain video sequence there are 32 frames of size 512*400. In missa video sequence there are 148 frames of 360*288 size.

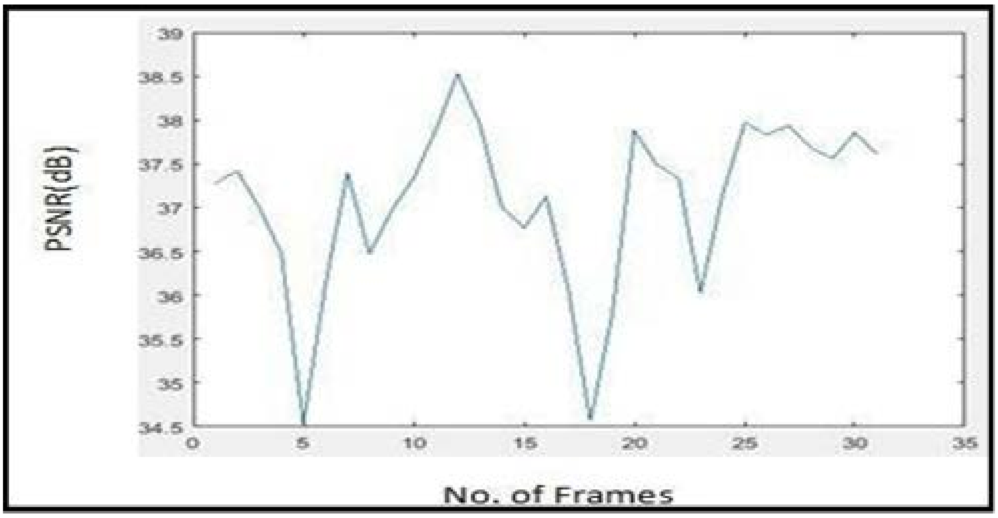


Figure: 5.18. PSNR value of 32 frames of “missa” sequence having total 149 frames

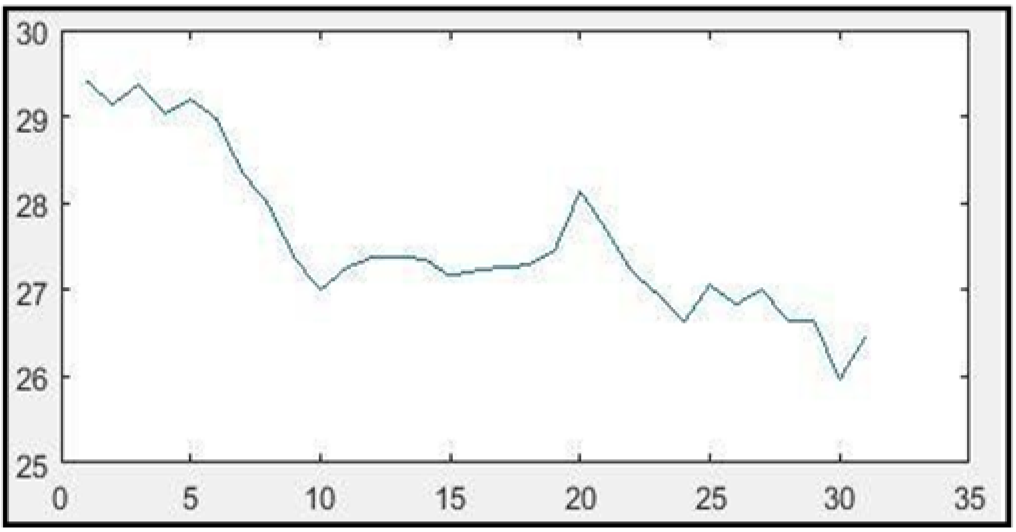


Figure: 5.19. PSNR value of "caltrain" sequence having 32 frames

$$SIR = \frac{(9 + M \times n + 4) - (6 + M \times n + 4)}{6 + 3 \times n + 4} \times 100\% \qquad (5.1)$$

$$SIR = \frac{MSE(DS) - MSE(PENT)}{MSE(PENT)} \times 100\% \qquad (5.2)$$

Phase - 1... (a) 6+3*1+4= 13(PENT) (b) 9+3*1+4= 16(DS) SIR= ((16-13)/3)*100 = 23%

Phase - 2... (a) 6+3*2+4 = 16(PENT) (b) 9+3*2+4= 19(DS) SIR= ((19-16)/16)*100 = 18.75%

Phase -3 ....SIR = ((9-6)/6)*100 = 50 % (in both DS andPENT)

Avg SIR = (23+18+50)/ 3 = 31%

By looking at minimum search points available in DS and Pentagon, three items are saved on average. SIR has thus accomplished a 31 percent boost over a diamond search. A comparative performance for the video series is given below.

Table: 5.1. Motion vector and total time elapsed for caltrain video sequence

	Motion vector	Frame size	Total time
DS	143	512*400	17.684
Hexagon	165	512*400	16.73
Pentagon	179	512*400	15.98

Table: 5.2. Motion vector and total time elapsed for missa video sequence

	Motion vector	Frame size	Total time
DS	82	360*288	13.9419
Hexagon	91	360*288	13.8512
Pentagon	97	360*288	12.6734

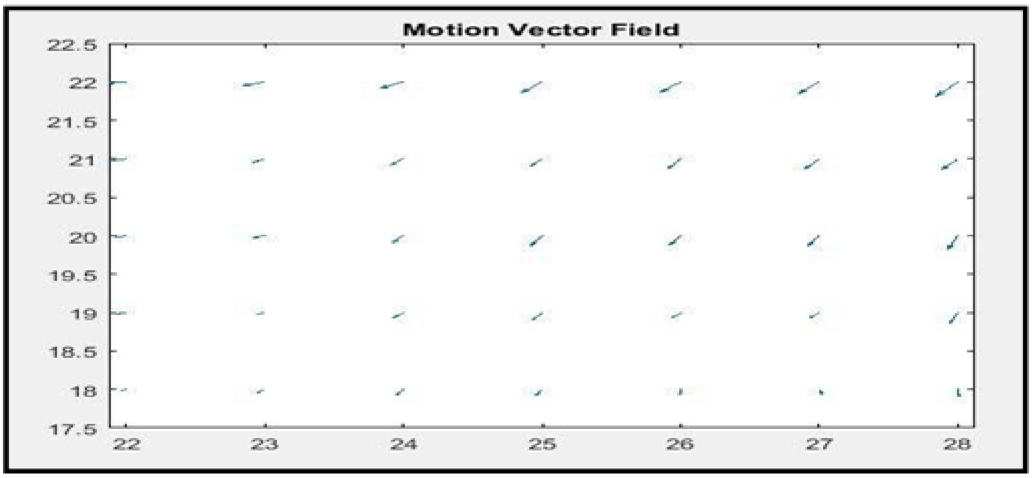


Figure: 5.20. Caltrain video series motion vector

Table: 5.3. Average MAD per pixel for a particular frame rate for missa and caltrain

Mad per pixel		
	Frame diff1	Frame diff3
missa	4.809	4.984
caltrain	2.561	3.882

Table: 5.4. Average check per frame for the diff. frame rates of missa and caltrain video

Average search per frame		
	Frame diff1	Frame diff3
missa	0.604	0.843
caltrain	5.774	9.997

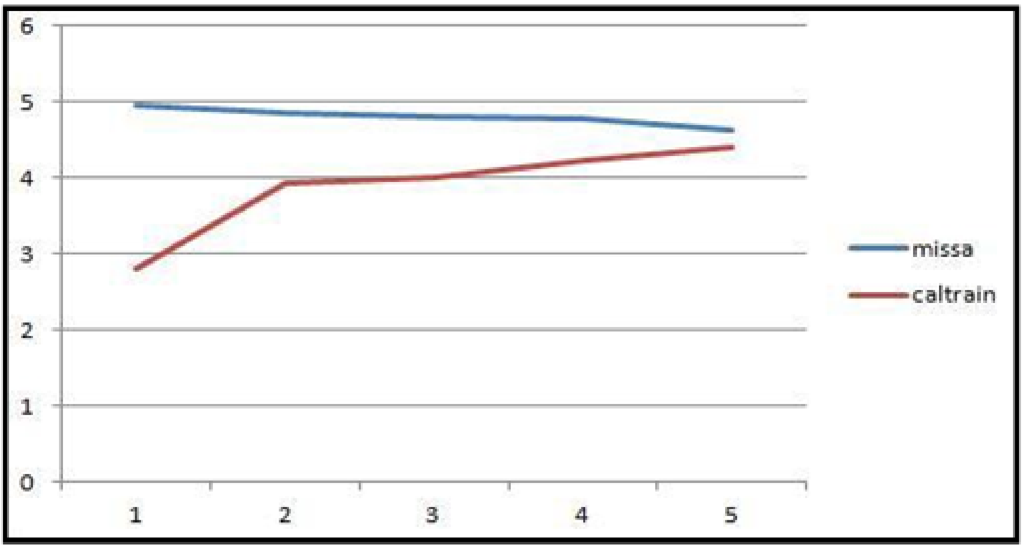


Figure: 5.21. Graph showing MAD for missa and caltrain sequence

5.7. MAJOR FINDINGS

This chapter analyses how various motion vector search algorithms works. This chapter focuses on the implementation of a new Pentagon search algorithm and the results show that it works better in terms of lower search points than the current hexagon and DS algorithm. All the simulations are performed in MATLAB environment. The speed improvement rate is about 31% with respect to the DS algorithm. Like the DS and other algorithms it may be applied for the latest video compression codec.

CHAPTER VI

Framework for Video Compression and Decompression

The content of this chapter is published in-

- Neetish Kumar, Dr. Deepa Raj (2020), "A Framework of Video Compression and Decompression", International Journal of Grid and Distributed Computing, Volume 13, Issue 2, pp. 235 – 245. - **(Web of Science Indexed)**

CHAPTER VI

FRAMEWORK FOR VIDEO COMPRESSION AND DECOMPRESSION

6.1. INTRODUCTION

The multimedia and networking technologies are currently developing every day, in which one of the vast and voluminous content is video data. It must be in compacted, since it is draconian. Many techniques have been developed for compressing video data. When analyzed carefully, it becomes clear that search techniques characterize video compression. In addition to search techniques, the redundancies exploitation can also be used in video compression. There are several types of redundancies in the actual content of the uncompressed video. Spatial, temporal, and statistical redundancies occur based on their characteristics and are exploited without losing real video effect. Spatial and time-specific redundancy in the frame series is used to eliminate or code an algorithm for video compression. Similarities can be detected only by storing the frame differences or by applying the discerning features of human vision. Block-based motion estimation is the foundation of video compression, and plays a key role in evaluating motion vector. The motion vector and residual encoding results in a compressed video form. This chapter explores a new way of using a video compression technology in a very different nature. The most popular method used in this chapter is to look for pentagon block matching search algorithm and apply on collection of images.

6.1.1. Introduction about the video

A video is also known as the sequence of kinetic images. A standard video includes different kinds of shots. A shot mainly contains GOPs (Group of picture), which have different frame types, such as I, P and B frames. The macro block, accompanied by blocks and pixels, is an essential element of a video to study its data structure. A semantic diagram of the video is shown in Figure 6.1.

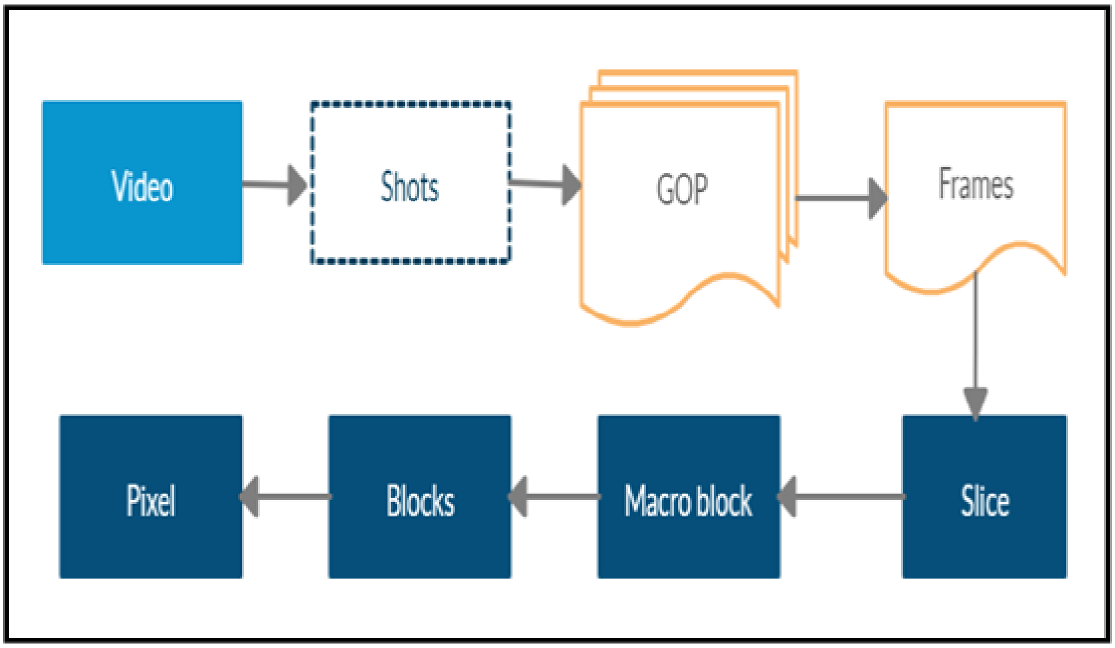


Figure: 6.1. Organization of video data

Motion and motion compensation: As we have seen the detailed discussion about the motion estimation and motion compensation, so we have used it with pentagon search technique in the proposed framework. The pentagon search is used for the encoding and for the analysis of the movement vector in his chapter. The pentagon search pattern is shown in Figure 6.2.

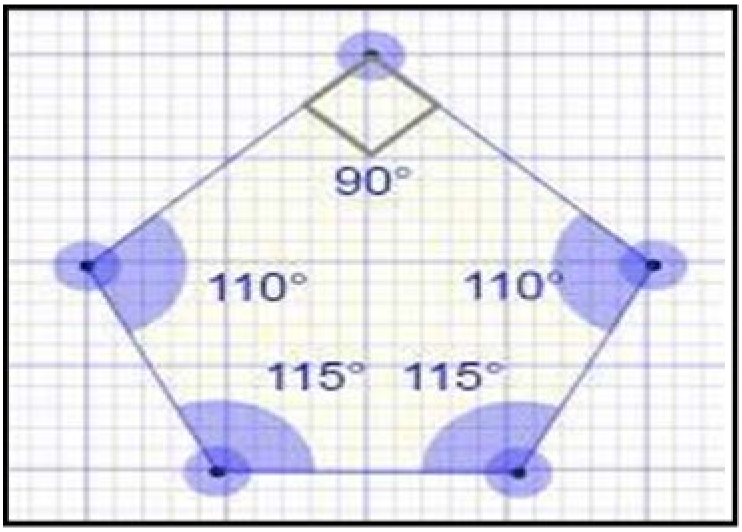


Figure: 6.2. Shape of the pentagon

6.2. VIDEO COMPRESSION AND DECOMPRESSION

Compression of video is the removal of unnecessary data. It is needed for storing the large video files in lesser bits. With the help of compression the volume of data used for transmission is decreased. Compression is valuable because it prevents the use of expensive facilities such as bandwidth and hard disc space. On the other hand, compressed data needed to decompressed the decoder end. Decompression is the process used to restore the compressed data to its original state. Data decompression, including lossless and lossless compression, it is necessary for almost all compressed data. Data decompression is also based on different data compression algorithms. Decompression is used extensively in communication with data, multimedia, audio, video and file. Numerous techniques and algorithms are required for data decompression. For each compression technique, there is a corresponding decompression technique. The majority of decoding software uses a decoder to decompress data. In the case of lossless compression, the original data is obtained without loss of reduction.

6.3. BLOCK DIAGRAM OF ENCODER/DECODER

The encoding and decoding phenomena are performed by the Encoder and Decoder. The encoder converts the source data to a compressed form (reduces the number of bits) before copying or saving, and the decoder makes the compressed data to the source data. The encoder/decoder pair is also known as CODEC which is shown in Figure 6.3.

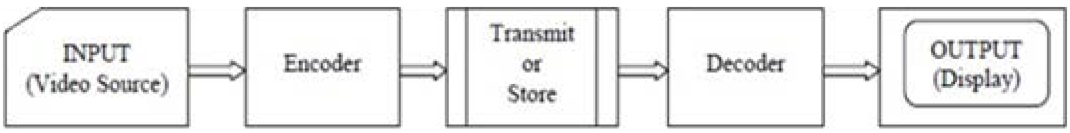


Figure: 6.3. Block Diagram of CODEC (Encoder/Decoder)

6.4. COMPRESSION RATIO

The compression ratio is a statistical term that uses a data compression algorithm to calculate the reduction size of the data. The compression ratio is identical to the suppression ratio of the compression measure. The compression ratio is specified between the undisclosed size and the compressed size:

$$\textbf{Compression ratio} = \frac{\textbf{Uncompressed size}}{\textbf{Compressed size}} \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (6.\ 1)$$

If a 10 MB video file is compressed up to 2 MB file then the a compression ratio is 5 (10/2 = 5, 5:1, or 5/1).

$$\textbf{Space saving} = 1 - \frac{\textbf{Uncompressed size}}{\textbf{Compressed size}} \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (6.\ 2)$$

This would save 1-2/10 = 0.8 space, which is reported as 80 percent by a representation which comprises a 10 MB file to a 2 MB file.

6.5. TYPES OF COMPRESSION

There are two types of compression involved in image and video processing.

1. Lossless Compression
2. Lossy Compression

6.5.1. Lossless Compression

Lossless compression is a class of reduction that reconstructs the compressed data from the original data. No loss of information is estimated. The lossless compression is applied in the original, and the restored data is the same, for example, text, letter, zip-file, PNG or GIF.

6.5.2. Lossy Compression

A lossy compression may differ from the original in which compression and decompression can be restored, but it is relatively similar in many respects. Lossy compression ensures that any loss of information happens then the data cannot be recovered.

6.6. METHODS OF VIDEO COMPRESSION

Video compression is used to reduce the duplication of video data using modern coding techniques. The majority of video compression and codec algorithms combine to reduce the spatial images with compensation from time to time. Video compression is a practical implementation of source video code. Many video codecs also use audio compression techniques combined with a single module of video compression. Lossless video compression codecs are compressed with a factor of 5-12, and a compression factor between 20 and 200 for MPEG-4 loss video compression. There is a trade-off between video quality, time and cost of encoding and processing of decompression. The video that is highly compressed may contain distracting or visible artifacts. Some video compression schemes typically run on neighboring pixels classes, also known as macroblocks. These pixel groups or pixel blocks are opposed to the next frame. High-frequency information usually leads to a reduction in the quality or an improvement in the variable bit rate during explosions, shooting, animal flocks or in individual panning shots.

6.7. VIDEO CODEC

A video codec is a digital electronic circuit or hardware compressor. It converts raw (uncompressed) digital video into a compressed format. Codec is a video encoding concatenation of encoders and decoders. An encoder is commonly called a machine that only compresses the video data. The diagram of the video encoder block is shown in figure 6.4. It consists of three functional core units: a temporary model, an entropy encoder and a space model. The time model entry is a non-compressed video sequence. The tempo model avoids temporal duplication, using the similarity between the neighboring video frames, usually by approximating the current video frame. The output of a time model comprised of a residual frame and a collection of model parameters that naturally reflect motion compensation due to the removal of the forecast from the real structures. The restful framework forms the spatial model's input with the similarities between the samples in the residual frame. The transformation converts the parts into another region where they are expressed in the transformation coefficients. A compressed sequence contains codified parameters of the motion vector, encrypted rest information and header information.

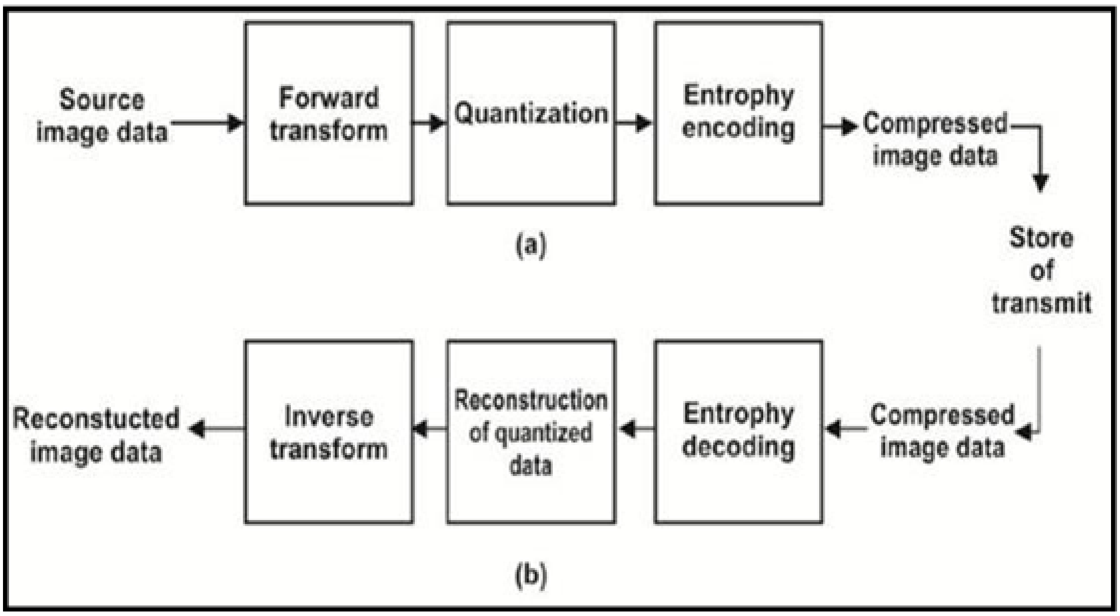


Figure: 6.4. Block Diagram of video codec: (a) Encoder and (b) Decoder

The coefficients and motion vectors are decoded by an entropy decoder. The decoder uses motion vector parameters to approximate the current structures with one or more previously decoded frames and reconstructs the frame with the rest of the frames.

6.8. TYPES OF VIDEO COMPRESSION

Video compression is the combination of spatial reduction and time-motion compensation involving to reduce the data length. It is based on the fact that the data before compression is of excellent perceptual quality. For example, DVDs use standard MPEG-2 video coding that can compress about two hours or more of video data between 15 and 30 times, without affecting the image quality, which for legal video definitions which will be of high quality. Video compression normally operates in square groups of neighboring pixels, often known as macroblocks. Pixel groups or blocks of pixels are compared between frames, and the video Codec (Encode / Decode Scheme) displays only the difference in these blocks. In a video with more movement, more pixels deviates from one frame to another frame.

6.8.1. Intra-frame Video Compression

The intraframe compression is a compression applied to still images such as diagrams which uses a redundancy called spatial image redundancy. Intra-frame compression techniques can be applied to single frames of the video series mostly of I-frames.

6.8.2. Inter-frame Video Compression

Interframe compression is applied on P and B type frames. Interframe compression uses temporal redundancy between consecutive frames to reduce the amount of information required for compression. Several interframe compression techniques of various complications are available.

6.9. VIDEO COMPRESSION STANDARDS

Two international bodies define the standards for video coding. The first is ISO and the second is ITU (International telecommunication union). A range of video compression standards has been introduced that affect the visual communication industry's growth. The following two groups had significant contributions:

- Motion Pictures Expert Group (MPEG) established by ISO / IEC
- H.26x series developed by ITU-T

6.9.1. MPEG FAMILY

Steps in video compression

- Motion estimation & Motion compensation: These two crucial parts are discussed in above section.
- Transform and quantization: The basic transform coding process is almost similar to that of every codec mechanism. The process includes a forward transform and quantization followed by zigzag ordering and entropy coding.
- Inverse transform: It is used in reconstruction process. The reconstruction process includes an inverse quantization and inverse transform followed by motion compensation.

There are some limitation on certain simulator to analyse the result of a video frames. Typically, these frames are of big size and take too long to analyze. As suggested in various literatures, all the simulation for video compression is performed on QCIF video format due to prevent the heavy computational complexity. Time taken for the simulation of normal/ HD video takes through MATLAB usually takes much more time. Hence for analysis purpose CIF or QCIF are used. Spatial resolution and other properties for different format are given in Table 1.

Table: 6.1. Different formats of videos and its properties

Format	Spatial resolution	Frames/second	Uncompressed data Rate
QCIF	176×144	15	4561920
CIF	352×240	30	30412800
ITU-R601	720×480	30	124416000
HDTV	1280×720	30	331776000
HDTV	1920×1080	30	746496000

MPEG 1

The Motion Picture Experts Group (MPEG) adopts MPEG-1 as the first public standard for video compression. ISO created the contraction of the 26:1 and 6:1 video without an excessive Quality loss. It was used for the reduction of audio and video.

MPEG 2

In 1995, an upgrade of MPEG-1 was implemented in the MPEG-2 codec. It includes a video and the range of bandwidth usage. It is suitable for applications for optical broadcast TV. The TV Bit Rate can be used from 4-15 Mbps (up to 100 Mbps) for television broadcasting, such as HDTV (high definition digital television), cable TV and ISM (cable TV). Profiles and levels were added in MPEG-2 for better efficiency.

MPEG 4

MPEG-4 is the method of encoding audio and visual digital data. The MPEG-4 continues to be an evolving standard and is divided into different parts for its versatility and development. Its additional areas are MPEG-4 Part 2, MPEG-4 Part 10 and so on.

It was developed in October 1998 and enabled sound and video in low-bit networks to communicate with the devices. The encoding of video artifacts is one of the critical features of the MPEG-4 that permits for the arbitrary shape, despite the rectangular frame.

Table: 6.2. Different compression techniques of MPEG family

S.NO	Year	Standard	Publisher	Popular implementation	Bit Rate
1	1992	MPEG - 1	ISO	Video - CD	1.5 Mb/s
2	1994	MPEG - 2	ISO, ITU - T	DVD video, digital video broadcasting	>2Mb/s
3	1998	MPEG - 3	ISO	Video in internet, DivX	variable

6.9.2. H.26x Family

H.261

The International Telecommunications Union (ITU) adopted this standard in 1990. The handling of temporal prediction is easy in H.261. A limited amount of visual information available supports two resolution frames, the CIF (Popular Interface Format) with a frame size of 352x288 frames and the QCIF frames (QCIF). The coding and decryption process are done at p x 64 kbit /s which ranges from 1 to 30. It uses the discrete cosine transformation of (DCT) and codes both INTRA and INTERS frames. The H.261 provides prominent features for supply, block transformation, quantization and entropy coding without degrading efficiency. It is mainly used in applications for video conferencing. It uses video conferencing, computer training, studio video conferencing, monitoring and telemedicine.

H.263

The ITU-T Video Coding Experts Group (VCEG) is designed and developed H.263 with some improvement on H.261's application to conduct video conferences. A step forward H.263 standard not only support video conferencing but video telecommunication too. It was approved in early 1996. Variable block size compensation and overlapping block motion compensation are essential aspects of the H.263 standard. While the H.261 provides the regular or wireless network's ability to achieve a moderate quality of 18 to 24 kbps, the H.263 has a high rate of 28 to 33 kbps. It is used for various web applications such as Flash video content. There are five supporting resolutions: QCIF, CIF, and SQCIF, 4CIF, and 16 CIF.

H.264

Advanced Video Coding (MPEG-4 AVC), Part 10 H.264 or MPEG-4 has been established with the ITU-T Expert Group (VGC) along with the ISO / IEC. The creators of MPEG and H.26X series have brought the best codec for this project together. The H.264 encoder boosts compression more than ever. The video compression standard is focused, based on motion compensation. As AVC expands the existing MPEG-4 format, it can also support MPEG-4 infrastructure resources (e.g., audio and system layer). For most of the applications demanding the highest degree of compression and quality over the current MPEG-4 video compression standard is the Advanced Simple Profile (ASP), MPEG-4 / AVC. The format for encoding Visual and H.264 are MPEG-4 Visual (including advanced video encoding). H.264 is mainly used in video conferencing, TV encoding of the blue-ray disc content.

Table: 6.3. H.26x family series

S.NO.	Year	standard	publisher	Popular implementation	Bit Rate
1	1990	H.261	ITU - T	Video conferencing, Video telephony	P*64 kb/s
2	1995	H.262	ITU - T	Video conferencing, Video telephony, Video on mobile phones	<33.6 kb/s
3	2003	H.263	ISO, ITU - T	Blu-ray, DVD, Digital video broadcasting, HDTV, Apple TV	10’s to 100’s kb

6.10. A PROPOSED FRAMEWORK FOR VIDEO COMPRESSION & DECOMPRESSION

Now there are a lot of video codecs developed, our proposed framework is a step forward in this regard. Basically at the back end i.e search strategy, a new pentagon search method is used that searches the true motion vector so much efficiently that further encoding method gets benefitted by it. Before knowing the complex process of compression it must be investigated first that what a video frame consist of. Basically video is a sequence of kinetic frames. These frames are bundled together by its properties, these bundels are known as GOP. In a GOP there are three types of frames I, P and B frames. All the three types of frames are encoded in different ways. First frame of the GOP is always an I frame that is encoded just like the still image.

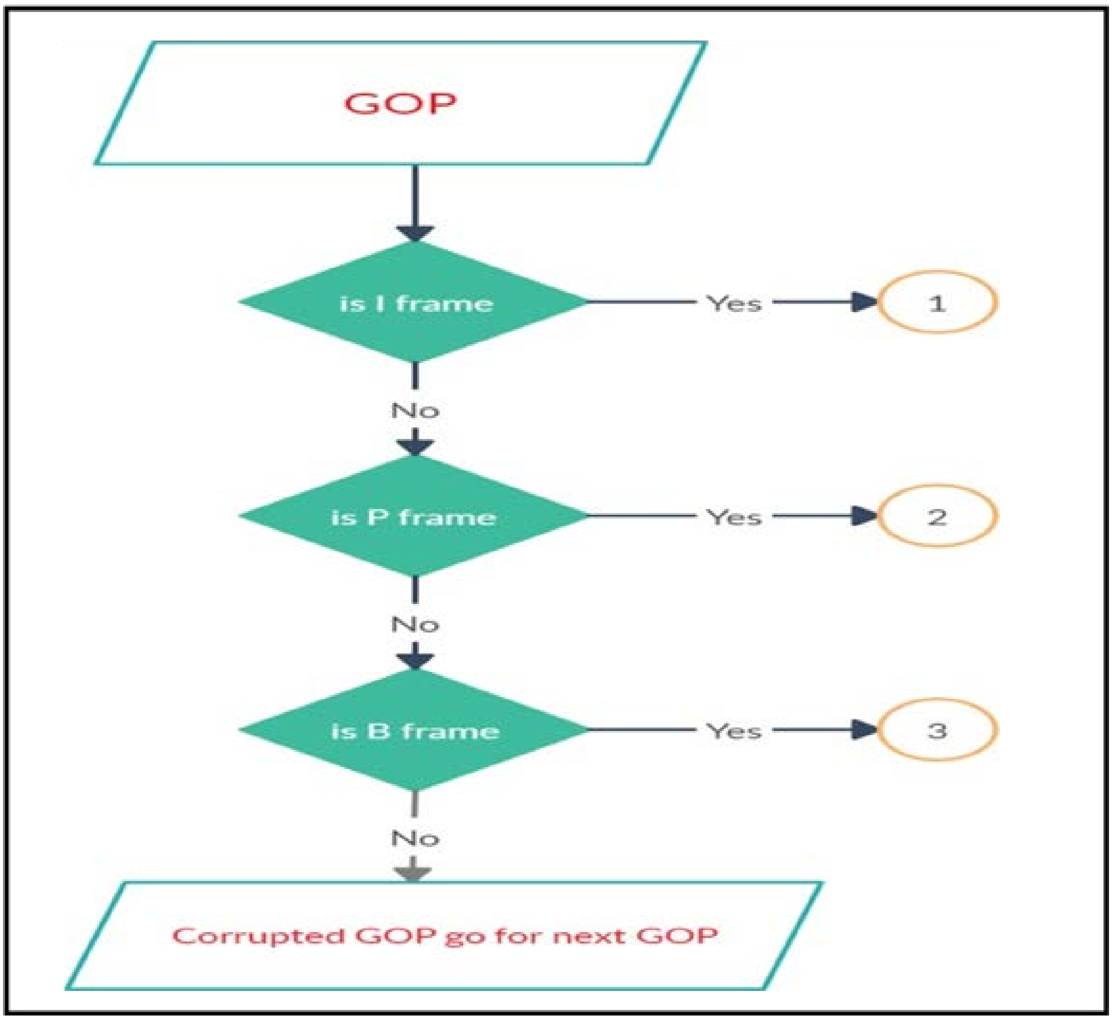


Figure: 6. 5. Encoding of different frame through GOP

6.11. Encoding of different Frames

a) Encoding of I frame

I frame encoded like the JPEG file. I frame is the every GOP's first frame. The INTRA system is also known for its spatial redundancy. It is the only framework recognized as an error-free system that is coded because no forecast is required in this stage to encode. For the next P & B frames, it acts as a frame of reference. The whole frame of DCT is 8 * 8 blocks.

The AC coefficients are coded by DPCM. A model diagram in which the I system is encoded is shown in Figure 6.6.

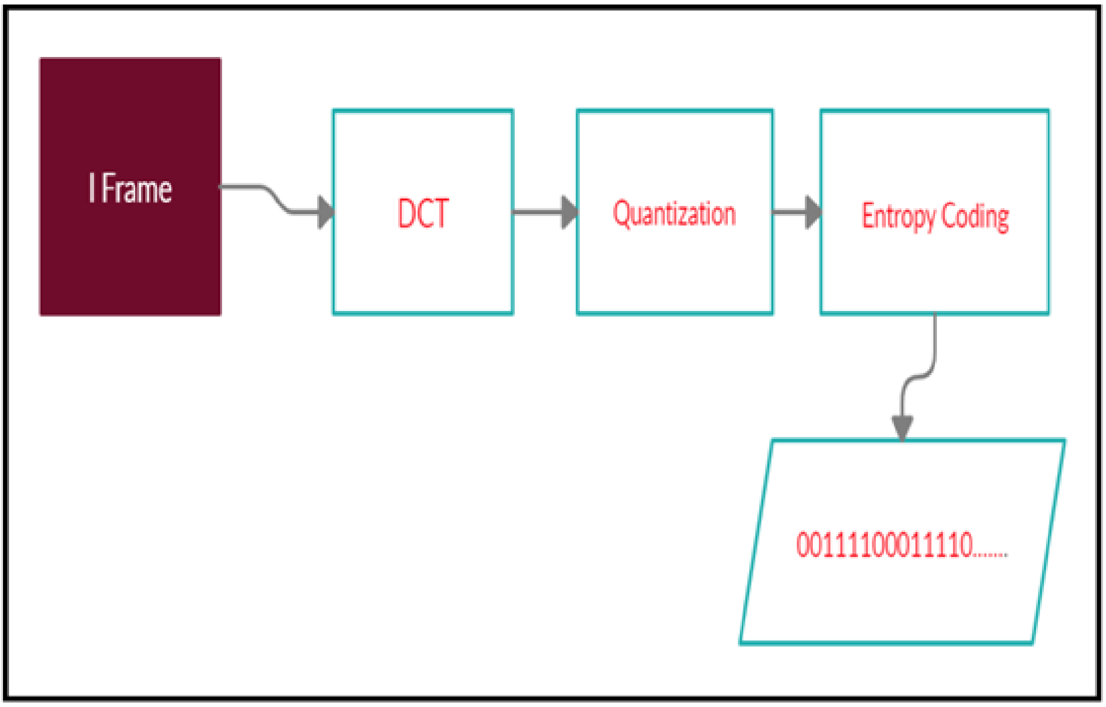


Figure: 6. 6. Encoding of I frame

b) Encoding of P Frame

P frame is also known as the forward-looking frame. Firstly, the I frame is coded then taking it as a reference the P frame is predicted. The technique to locate the motion vector is Pentagon search, which provides a high output for finding a motion vector. A large window predicts the movement of a specific block. The corresponding block will be tested on the window. Different metrics decide the exact match between the actual block and the frame of reference. The resulting change in the block position is called the motion vector. Various techniques have been developed for detecting the motion vector. This encodes new data and saves information about the remaining motion vectors in the buffer. For such blocks, DCT, Quantitation and entropy coding are done in the same way. At the end of the encoding, the visual data is converted to the bit stream. The compression ratio of the P-frame is between 20 and 30:1.

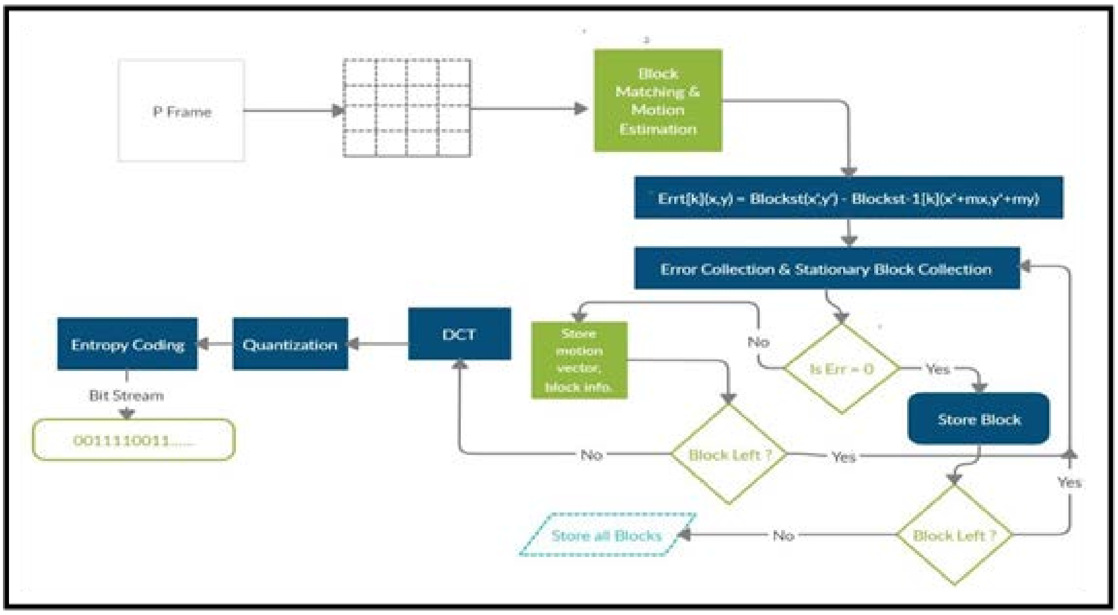


Figure: 6.7. Encoding of P frame

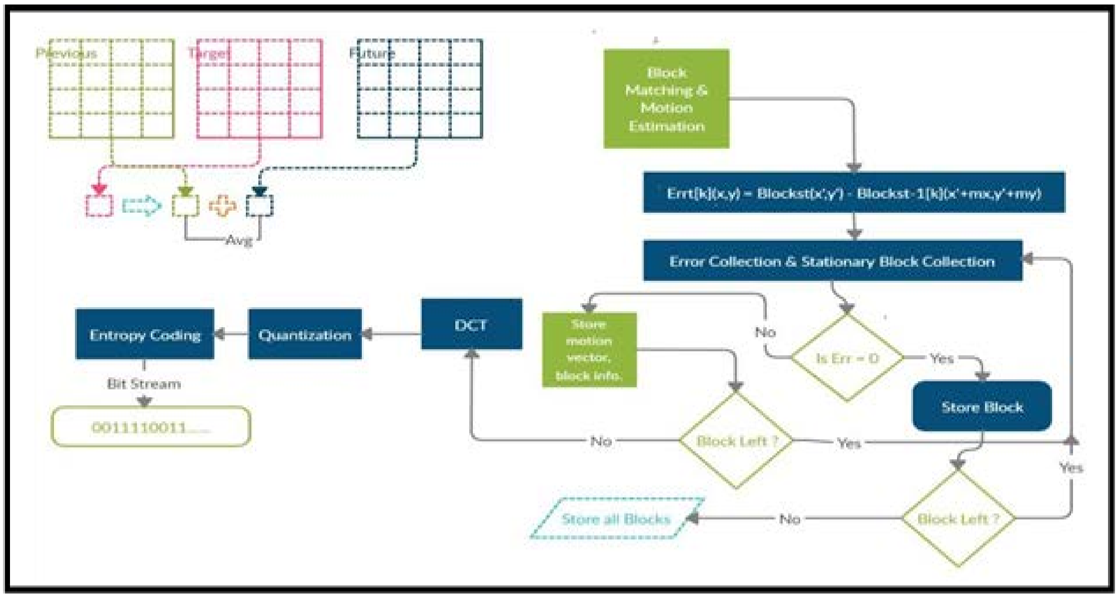


Figure: 6.8. Encoding of B frame

Encoding of B frame

The bi-directional frame is commonly known as the B frame. The entire block matching process, motion evaluation and motion compensation is carried out same as P frames. First of all the window is sufficiently chosen larger to describe the motion vector. The first step is to choose six blocks and calculate the error. The pentagon search for the motion vector is used for finding the motion vector that is the most efficient search technique to test the motion vector. There is a slight change in the block matching process in B frame. As in P frames only past frame is considered as the reference but here past and future both frames are taken into consideration. So, in block matching phase average of these two is taken. Thus compression of B frame takes more space in buffer. It typically requires very fewer bits than that of I and P frames. The compression ratio of B frames is very high and nearly about 30 to 50:1.

6.12. DECOMPRESSION STRATEGY OF FRAMES

The decompression strategy of I frame is same as decompression of a still image but this approach is very different for P and B frames. There is a subtraction in the encoder of p and b frames between the input and the projected image, it is then applied at the end of the reconstructed picture. Exact reconstruction of the structure cannot be done because the quantization is an irreversible process at the end of the encoder. Figure 6.9 displays the video decoding map. The video error will be decoded, dequantized, and inverse to DCT when the bit stream is encoded. After interpreting the error, the video is applied to the restored frame. The decoding of I, P and B frames have discussed separately.

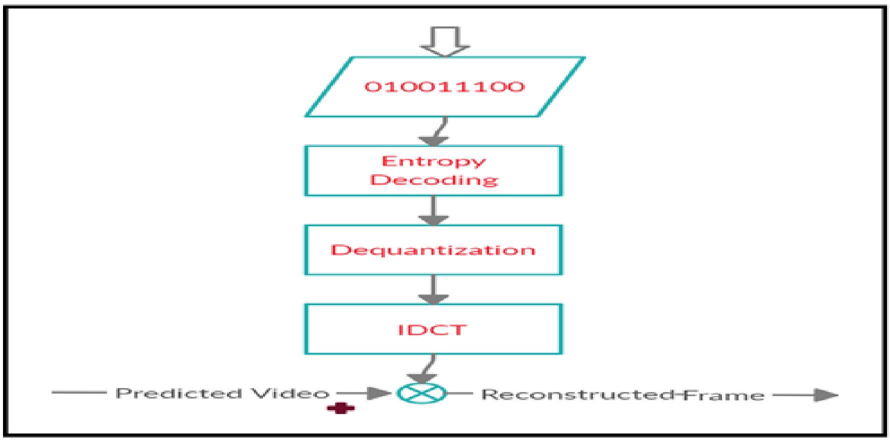


Figure: 6.9. Decompression strategy of Frames

a) Decoding of I frame:

Like the still image technique (JPEG), I decode the image. The encoded bitstream is interpreted by the entropy decoder, dequantizers and IDCT. The I Frame is also known as the error free frame because it is fully intra-coded, which means there is no need for past or future frames. The decoding of I frame is shown in Figure 6.10.

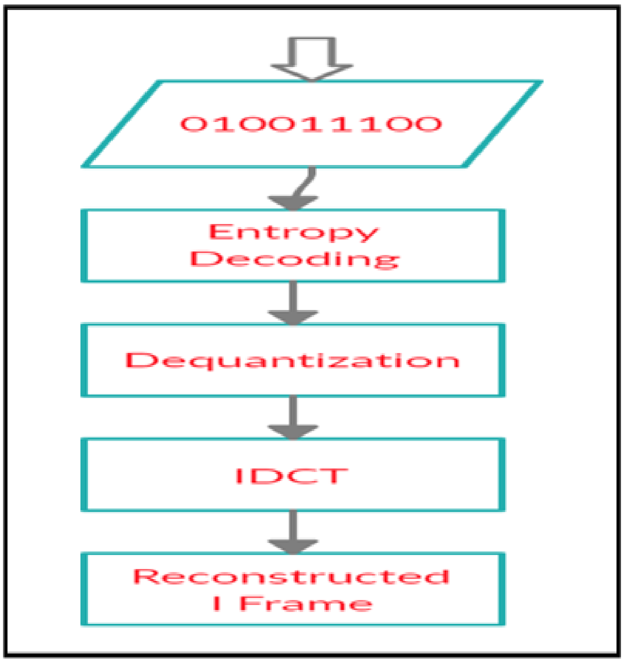


Figure: 6.10. Decompression of I frame

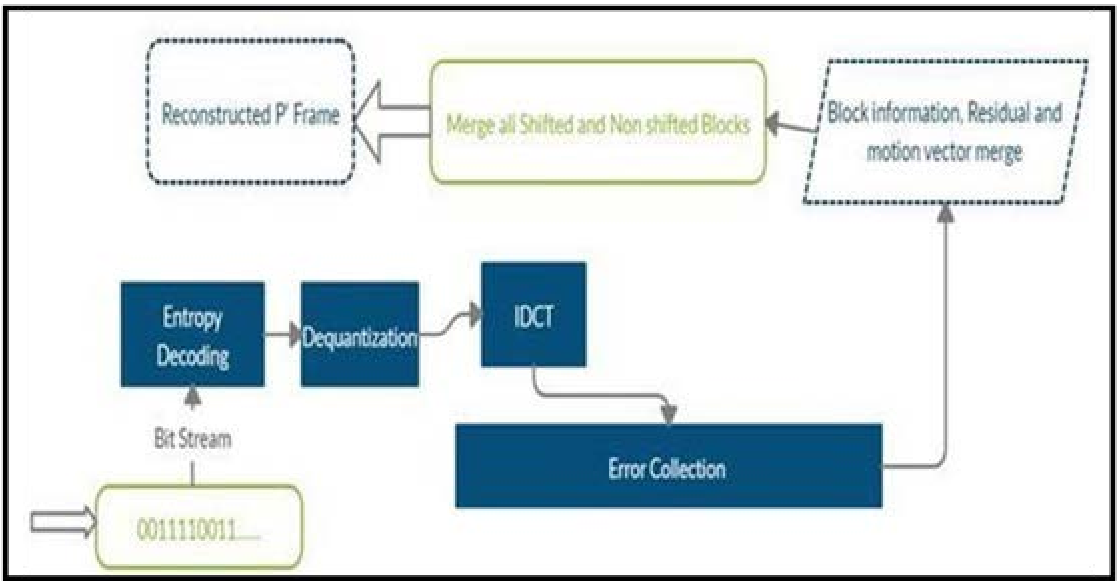


Figure: 6.11.Decompression of P frame

b) Decoding of P frame:

Initially, the bit stream at the end of the decoder is entropy decoded, followed by dequantization and reverse DCT. The Error signal is exactly what is obtained after all these steps. The corresponding motion vector and the entire block must be applied to the error signal for a faithful reconstruction. After combining the error signal, motion vector and unchanged block, a p-frame at the decoder end is reconstructed. The P frame decoding method is shown in Figure 6.11.

c) Decoding of B frame

For the B-frame and for the P frame, the same decoding approach is used. A little bit difference occurs after the error accumulation, as there is only one anchor used as reference while in case of B frame two frames are used as reference frame. The method for decoding the B frame is shown in Figure 6.12

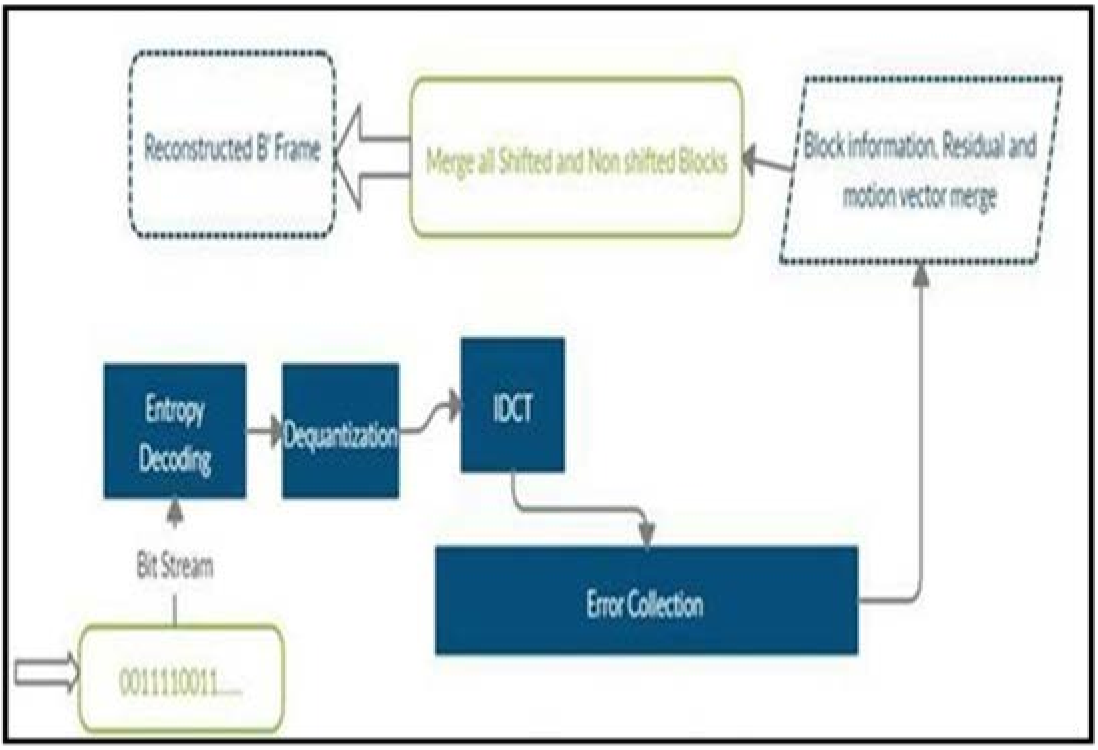


Figure: 6.12.Decompression of B frame

The B frame coding has certain benefits that reflect the coding quality. In most cases, B frames will result in less bits being coded overall. Quality can also be improved in the case of moving objects that reveal hidden areas within a video sequence. Backward prediction in this case allows the encoder to make more intelligent decisions on how to encode the video within these areas. Also, since B frames are not used to predict future frames, errors generated will not be propagated further within the sequence.

The B frames management has a variety of inconveniences. One is directly related to memory. Because two reference frames are predicted, additional memory space will be required. Usually, the memory requirement is multiplied when handling the B frames, so it is not feasible. There is a disadvantage in the case of video conferences.

6.13. RESULT AND DISCUSSION

The proposed framework operates efficiently concerning P and B frames. We ‘he already analysed that search pattern of pentagon method for block matching motion estimation gives a satisfactory performance over different video frames. Search strategy used in this method is pentagon search that searches the motion vector in efficient manner. Motion vector finding probability through pentagon search is better than previously proposed search strategy. Accumulating the error and process through this method is a different idea in both encoding and decoding. It also provides the better simulation result compared to others. The compression performance achieved for P and B frames through this method is larger than any other previously proposed framework. This framework is also applicable for compressing the real time video data.

6.14. MAJOR FINDINGS

This chapter introduces an optimized video compression frame for estimated block motion with the Pentagon search. As the motion vector's search time is decreased, it can be used directly for the video compression method. Pentagon search is a better option for the back end of newly proposed framework of video compression and decompression. For different domain areas, e.g., wavelets, the same compression technique can be used in the future.

CHAPTER VII

VIDEO EDITING

The content of this chapter is published in-

- Neetish Kumar, Dr. Deepa Raj (2020), “Algorithmic solutions for high quality Video editing software problems”, Journal of Critical Reviews, Volume 7, Issue 09, ISSN- 2394-5125 - **(Scopus Indexed)**

- Neetish Kumar, Dr. Deepa Raj (2021), “Shot Boundary Detection Framework For Video Editing Via Adaptive Thresholds And Gradual Curve Point”, Turkish Journal of Computer and Mathematics Education Volume 12, Issue 11, 3820- 3828 - **(Scopus Indexed)**

CHAPTER VII

VIDEO EDITING

7.1. INTRODUCTION OF VIDEO EDITING

Video editing is the post production phenomena in which corrupted frames are removed or reinserted, additional effects can be added, clips are rearranged to get an improved video. The video editing process gives extra chance to the user to modify the frames that get distorted during the capturing process by the internal or external agents. It also includes adding title, animation, sound and various activities. A video profile includes film, TV, video publications and video themes, as well as all video content. In recent years, there has been major democratization of the editing of personal computer video editing applications. Video editing and designing new software can be difficult and tedious to assist people in this job. Pen-based video editing applications allow users to edit the video more intuitively and easily. Improved quality, performance and ingenuity have made it possible to capture new video and audio material on existing magnetic tapes.

7.1.1. Video editing backgrounds

Video editing is a process for post-production editing of film segments with special effects and sound recordings. Image editing is a pioneer to video editing. The editors use video editing to express fictional events and non-fiction events. The purpose of the editing is to monitor these actions and to optimize the conflicts with the original intent. It's a work of art. In the beginning of the 1950s, there was only option to cut the film tapes that gets corrupted during the capturing. Later on the logical sequence of clips are added in the form of connecting the film strips or tape to make a short film. After that the magnetic tapes are used to capture the quality and durability which supports to remove the video and audio content. A linear editing protocol has been used. When the length of the scene is altered during the commencement of the video card, all the locations on the card must be kept in the display.

The view mixer (video switch) plays the source so that more complex transitions can be made between the scenes. U-Matic (called the U-shaped tape path) was the standard system of the 1970s-80s for such transitions. The machine used two tape players and a tape record with an automated backup to prevent the update from rolling or going wrong. More complex controls and electronic synchronization of smaller beta devices (b-shaped) were used in the late 1980s and 1990s.

Video editing software, along with motion graphics software, is becoming more familiar with the evolution of high definition videos. Video clips are scheduled, tracked and the titles are added, digital graphics are displayed, unique effects are built, and video is made ready. Photos can be distributed in a number of ways such as DVDs, Internet broadcasts, short films, iPods and CD-ROMs.

7.1.2. Types of editing

Video Editing Software was once seen as the province for expensive computers, but for other devices and workstations, video editing requires cutting down, re-sequencing, and transitional sections and other special effects.

- LVD - Linear video editing
- (NLE) - Non-linear editing
- Offline editing
- Online editing
- Cloud-based editing
- Vision mixing

7.1.2.1. LVD types of video editing

Linear video editing is a video editing technique for post-production used to pick, arrange, and modify images and sound in a predefined series. The shot using a video camera or a tapeless video camcorder are captured and the content can be accessed. Most video editing applications have replaced the linear editing. Linear editing has been done in the past. Movie and cartoons were cut into long stripes and eventually stuck or glued together to create a logical series of films. It takes more time and it is too repetitive for linear video editing. But today, it is significant for these reasons mentioned below:

- Fast and economical method.
- It is the fastest and most straightforward way to merge two sections of the video clips sequentially.
- Improve the knowledge and versatility in linear editing as video editor. Many professional publishers with the first linear editing master aim to become all-around expert publishers.

By the early 1990s, linear video editing was just an ordinary video editing. Computer-based random access (NLE) in USA, a two-inch, four-pin video cable was commonly recognized for the first time at 15 cm / s. There are four video recorders and playback heads of a tape of two-inch width to achieve an excellent head-to-tap pace.

Figure: 7.1. Linear editing configuration

Initially, the video was edited by watching the track, such as a film version which is captured by Ferro fluid, and then cut by a bolt cutter or guillotine cutter. This method was too much complicated. Each segment has been painted in a thin iron filing solution with a carcinogenic toxic and carcinogenic carbon tetrachloride solution. The magnetic tracks were planned to be aligned with a splicer which was visible by a microscope.

In the vertical retraction, the canals were cut without disturbing the odd area. The cut must also have the same angle. As the reader heads of the video and the audio have extra inches, physical video and audio cannot be edited. The cut was done for the video and sound repeated by the same method to render the 16 mm film with a magnetic soundtrack in the right direction.

7.1.2.2. Nonlinear editing systems (NLE)

The first correct, nonlinear editor, the CMX 600, was introduced in 1971 by CMX Systems in a joint CBS-Memorex project. A video was captured and played on the modified disc size in the "ski area" with analog drives in black and white. They are also used to store data on mainframe computers for about half an hour. Two screens were mounted on the 600. The editor was used to minimize and edit the video using a Light Pen on the right screen.

The left screen was used to display the edited video. The whole machine DEC PDP-11 has been used as a computer. The original material will not be changed during offline editing. Modifications are explained by advanced nonlinear software editing. The versions are displayed with a pointer image playlist in the Edit Decision List (EDL).

Audio, video or photography changed from the source and the requested editing phases are recovered.

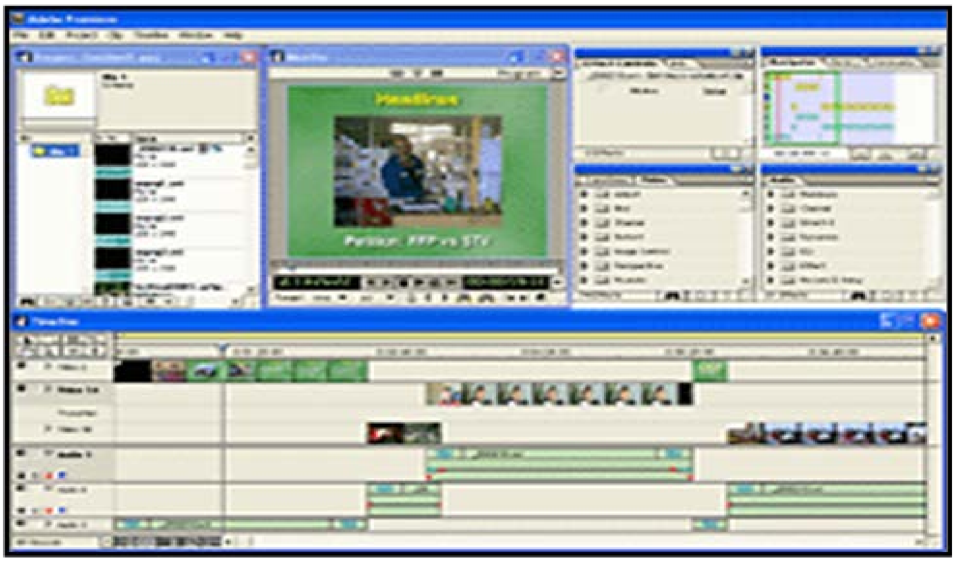


Figure: 7.2. Non-linear editing configuration

7.2. VIDEO HIERARCHICAL STRUCTURAL

In the new digital age, multimedia has an impact on every aspect of human life. Due to its global acceptability, it draws the scientists to investigate different areas of video analysis. For the transmission of video over a wired or wireless system, compression aspects are taken into account. Some fields require modification and special effects to show the videos. Frames are changed for the video editing process, external noise is extracted from a specific structure, and undesired objects are removed. Such practices are crucial in education, film, online classes, news, YouTube, etc. To understand the basic operation of video editing, firstly it is needed to know about video structure.

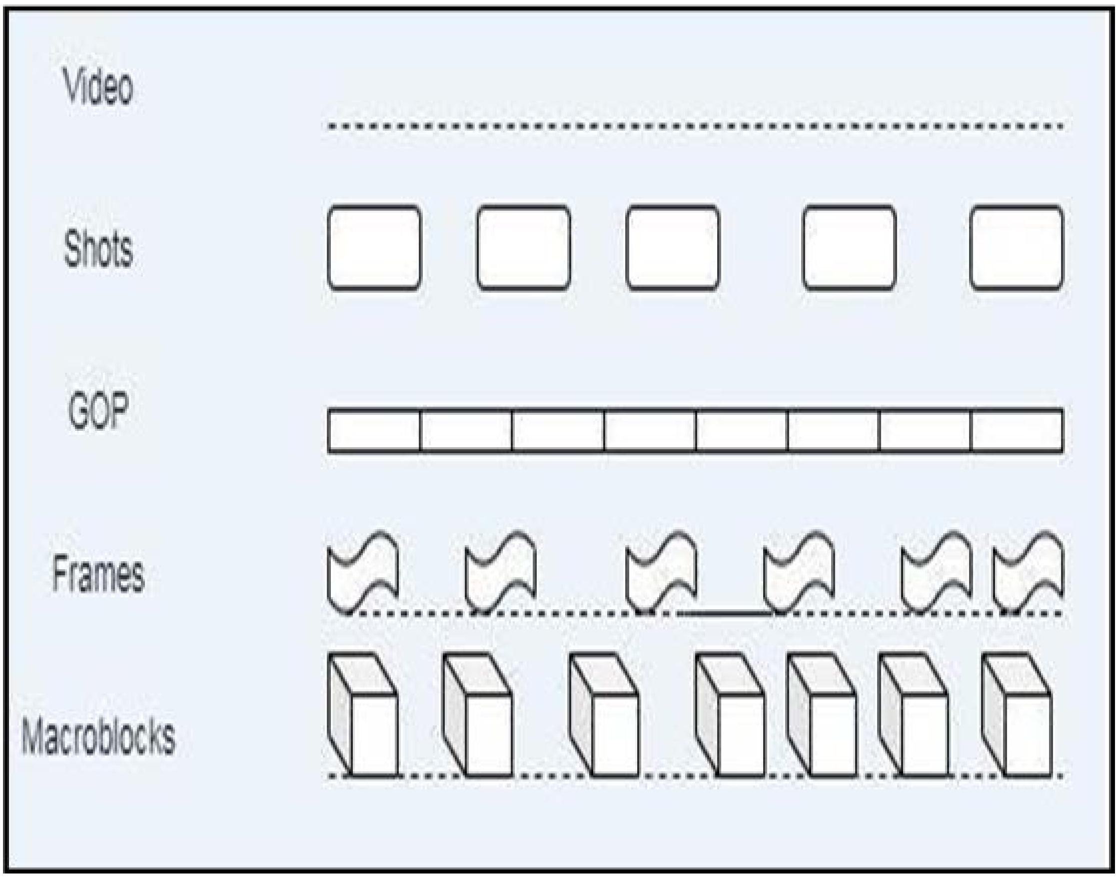


Figure: 7.3.Video hierarchical structural

The above figure shows the layout of a video. Video is nothing but the collection of shots. Shots are the bundle of pictures commonly known as group of pictures (GOP). In GOP, there are 3 types of frames i.e. I, P and B frames. The basic principle of editing was to be taken further at the macro blocks level. Matching, addition and deletion of frames are heavily depends upon the variation of blocks. The pixel-level is the fundamental unit for the editing process. Pixels are clubbed as the concept of neighborhood similarity into a uniform block.

7.3. CONDITIONS FOR CLEAR VISUALIZATION OF VIDEO FRAMES

Initially, all sorts of viewers want to watch the video without disruptions. The structures strictly govern the spatial variance of the pixels.

i) Assume Cmin = minimum pixel intensities in a frame and

cmax= maximum pixel intensities in a frame,

t' = threshold for spatial variation. Then,

$|$ Cmax -Cmin $|$ $\geq$ t' , for $0 < t' \leq 255$ (1)

ii) Again assume η= average pixel intensity of the frame

ω> 0. Then,

(Cmin+ ω) ≤ η≤ (Cmax- ω) (2)

Both of the above condition must be satisfied for clear visualization of the frame. The first condition checks about the meaning information about the intensity distribution where as the second condition depicts the smooth distribution of pixel intensity in order to ensure the maximum visibility. Literatures have suggested that the value of t' and ω nearly 20 for good results.

7.3.1. Frame and video clip Insertion

If there is fault in the captured frame, it is a standard editing process. It can be done in various ways. There is vast similarity between the successive frames in a video sequence. To insert a new frame these steps are followed.

I) firstly the past and next frame is recorded.
II) Difference between both are calculated
III) On the basis of difference the average motion vector is added to the previous frame

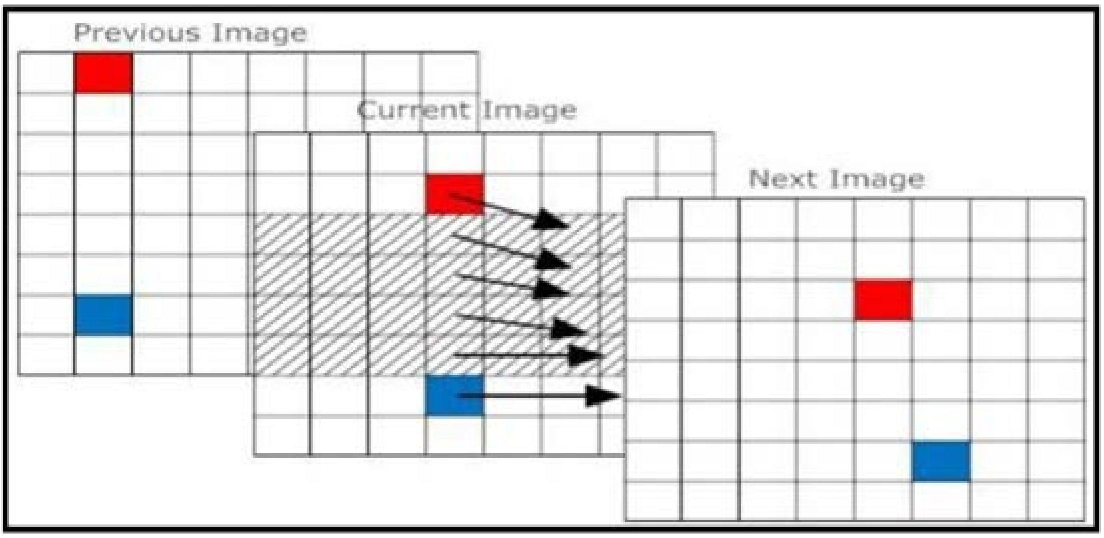


Figure: 7.4. Frame insertion

The same procedure follows the positions for adding a group of frame at a time in the past and next frames. For the observers illusion, the intersection must have the frame I. I frame comprise of a great deal of rich knowledge which is not present in the P and B frames. So it needed to add the video clips that must have I frames.

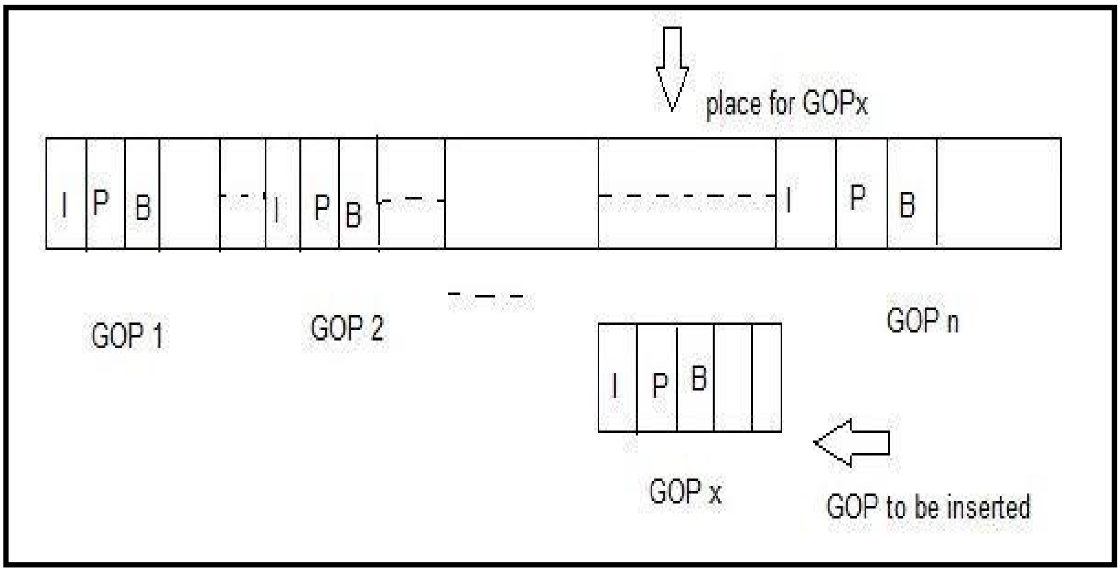


Figure: 7.5. GOP insertion

7.3.2. Splicing

A mixture of 2 or more video clips is required for the process. The specifications for the service have the same format and resolution in the video clips. The figure below shows two videos V0 and V1, are stitched at the stitching point. m is assumed to be the last frame of the V0 and n is the first frame of the V1 video clip. The stiching operation will be successful if the following conditions are met.

Let t_m =time stamps of the last frame

m in V0 t_n = time stamps of the first frame n in V1

t_{v0} = the duration of V0

t_m = the time duration of frame m

Then for perfect stitching point this condition must satisfy $t_n = t_m + t_m = t_{v0}$ (3)

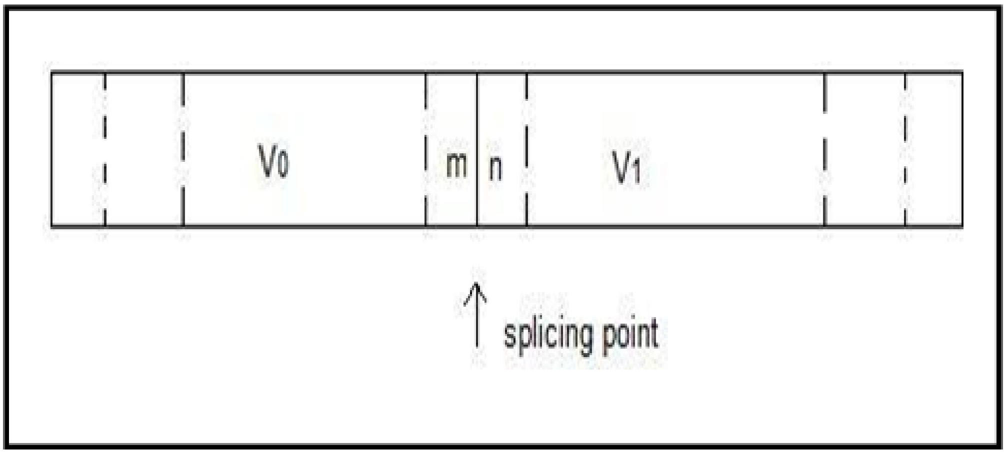


Figure: 7.6. Splicing point & merging of two video clips

7.3.3. Cutting

Opposite to the splicing operation the result of the cut video clips are of same format and resolutions. Firstly the cut section is to be identified i.e. the cut in point and cut out point is detected. Secondly at the cut in point the previous P frame of the last

GOP is converted into I frame. Then at the cut out point the clip length is needed to be removed after a certain time stamp. For a perfect cut the condition are as follows

Borrowing the above equation from the slicing part the equation is

$$t_m = \sum_{k=0}^{m-1} T_k \quad \ldots\ldots (7.1)$$

Where,

t_m =time stamps of the last frame

m in V0 tn = time stamps of the first frame

n in V1 T0= time interval for first clip

T1= time interval for second clip

Tk= time interval for kth clip

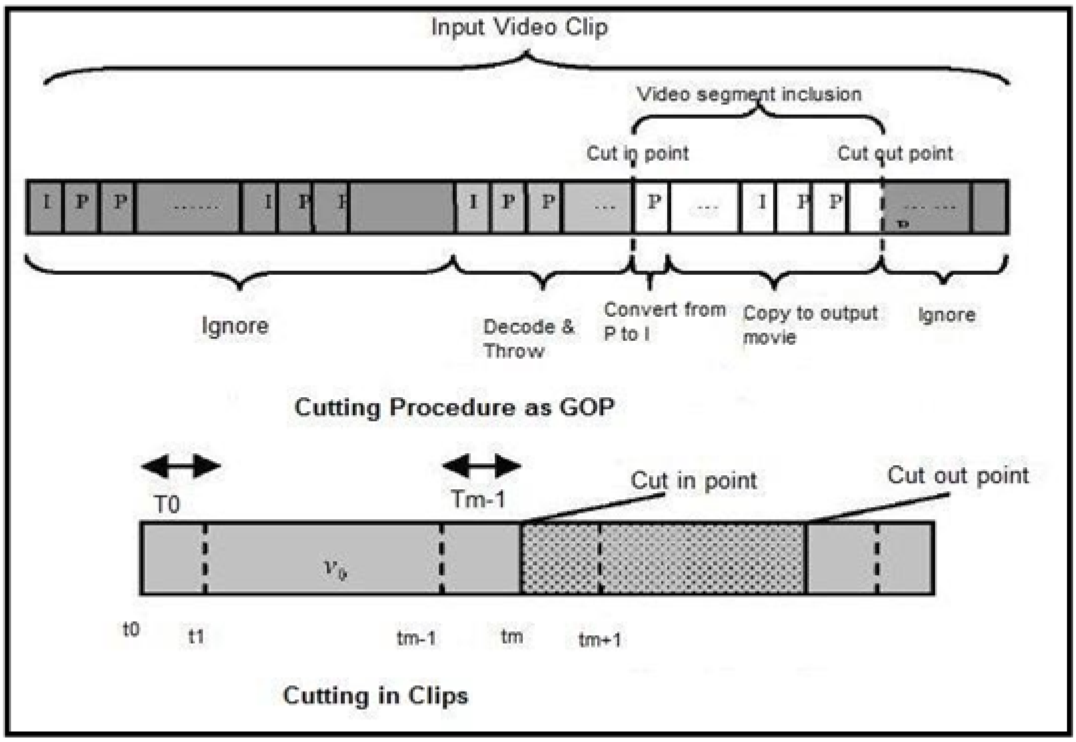


Figure: 7.7: Cutting procedures

7.3.4. Gray scale conversion:

It is the spatial effects of the video editing. The primary color that make the video colorful are the red, green and blue. The intensity combination of the three colors is used to gain certain color. It requires 24 bit to represent color image 8 for all three primary color. As we know that human eye is more sensitive towards intensity rather than color. The color scale can easily be converted into gray scale by simple averaging method.

$$\text{Grayscale} = (R + G + B) / 3 \qquad (5)$$

7.3.5. Slow Motion & Fast motion:

It is the temporal effect of the video clip that is done by modifying the frame rate. Frame rate is the rate at which the frames are played at the playback side. Frame rates are the essential thing used in both the sides of video processing i.e. at capturing side and playback side. If there is no change in the frame rate at the playback side the video clips runs normal but if there is a change in capturing and playback side then it results into the fast or slow motion videos.

7.4. ISSUES IN VIDEO EDITING AND ITS SOLUTIONS

7.4.1. Shot Boundary detection (SBD)

SBD of Temporal video is task of video segmentation into significant shots by identifying the transition among sequential frames, and the boundary between two consecutive shots are marked by the transition. There are two types of transition, i.e. abrupt and gradual. An abrupt transition is said as a rapid change in the video contents in which there is no intersection of frames among the margins of two shots. There is a slow change in the content of the frames in the gradual transition, some of the frames intersect in the consecutive shots and these frames duration is named as a gradual transition period. It is not easy to detect abrupt transition as it includes lot of complexities such as OCM and fast illumination causing extreme false positive results. The video data in huge volume has triggered a considerable desire for effectual tools which handle, deploy and store the overall data. The structural analysis of a video is a difficult task because of certain video attributes. The Video structure analysis splits the video into its simple elements. A hard transition (HT) is formed when two shots are attached directly. In the past two decades SBD has attracted the attention of many researchers. Which has classes: compressed and uncompressed domains. When compared to compressed domain the latter has gained lot of interest for valuable and fabulous visual information. Though, additional processing time is required by the uncompressed domain-based algorithms owing to the video frames decoding practice. Some of the methods employed for feature extraction are motion-based, edge-based, Pixel-based and histogram-based.

Currently in the transform domain the analysis of SBD is done rather than in the time domain. Since the signals are viewed in different domain which is allowed in the transform domain more over this provides a great shift concerning its potentability for evaluating the components of the signal. An image/signal is converted using the orthogonal polynomials from time/spatial domain into transform (moment) domain. To signify the visual information the scalar quantities such as transform coefficients, orthogonal moments are used. The projection of signal on orthogonal polynomials (OPs) are represented by these moments. The OPs capability is considered by the energy compaction and their localization properties. In this work we proposed a SBD in which abrupt transition can be detected using gradient and color information. The luminance distortion and the gradient similarity is calculated to measure each frames contrast and structural changes. In the abrupt detection phase the adaptive threshold across the videos are used to extract the transition. The Gradual transition is found mainly using the gradual curves using the average edge details. Using the optimal edge detector the average edge frame is obtained.

7.4.2. Prerequisite for finding SBD

a) Gradient similarity

Information from images can be extracted using Image gradients [21]. The change in intensity is measured by every pixel of a gradient image of that same point in a given direction of the original image. To attain the absolute range of direction the gradient images in the x and y directions are found.The gradient similarity presented is well-defined in Eqn 6.

$$G(x,y)=\frac{2(1-R)+T}{1+(1-R)^2+T} \qquad (6)$$

Where

$$T=\frac{T'}{MAX\left(G_x,G_y\right)}$$

$$R=\frac{\left|G_x-G_y\right|}{MAX\left(G_x,G_y\right)}$$

The gradient values for the central pixel of image blocks *x* and *y*, are signified as G_x and G_y. *G(x, y)*is the gradient similarity between *x* and *y* and its values lies in between [0, 1].

The Gradient value G_x for each block in an image (similar for G_y) is found as the supreme weighted average of difference.

b) Luminance similarity

Visible distortion is produced due to the luminance changes. When they are compared with the contrast or structure changes they are not so frustrating [21]. Using the Eqn 7 the Luminance similarity is described.

$$E(x_i,y_i)=1-\left(\frac{x_i-y_i}{L_u}\right)^2 \qquad (7)$$

In image blocks x, y the pixels at position i are represented as x_i, y_i, the pixel values dynamic range is denoted as L_u. Among the image pixels x_i and y_i within the range of [0, 1] the luminance similarity is represented as $E(x_i, y_i)$.

The overall quality indicator $Q(x_i, y_i)$ be defined as:

$$Q(x_i,y_i)=W(G,E).E+\left(1-W(G,E)\right).G \qquad (8)$$

The shortened forms of $Q(x_i, y_i)$, $G(x_i, y_i)$ and $E(x_i, y_i)$ are denoted as Q, G, and E respectively. The relative importance of the two components are adjusted using the weighting function W(G, E).Using Eqn 4 W(G, E)is calculated.

$$W(G,E) = P.G \qquad (9)$$

The positive weighting parameter is denoted as ‘P’. Since P also is considered to be lesser than 0.5 and G is in the range of [0, 1]. In this paper the value of P is taken as 0.1.

c) CIEDE 2000 color difference

This colour-difference formula is based on the colour space of CIELAB [22]. Eqn 10 defines the CIEDE2000 colour difference.

$$\Delta F = \Delta F(l_1^*, a_1^*, b_1^*, l_2^*, a_2^*, b_2^*)$$

$$= \sqrt{\left(\frac{\Delta l'}{k_L s_L}\right)^2 + \left(\frac{\Delta c'}{k_C s_C}\right)^2 + \left(\frac{\Delta h'}{k_H s_H}\right)^2 + R_T\left(\frac{\Delta c'}{k_C s_C}\right)^2 + \left(\frac{\Delta h'}{k_H s_H}\right)^2} \quad (10)$$

In CIEDE2000 the samples differences in chroma, lightness and hue are represented as l, c and h.

The rotation function is denoted as R_T and the blue region is responsible for the interface among the chroma and hue differences. In CIEDE2000 five corrections on CIELAB have been presented and they are chroma (s_c), lightness (s_L) and the hue (s_H) weighting functions. The CIEDE 2000 Colour difference in a Video Sequence is used as a rate to find the relationship among the frames

7.4.3. Proposed methodology

In this section the approach presented is explained in detail. Fig. 1illustrates the block representation of the presented scheme. This scheme involves of the feature extraction, thresholding, abrupt and gradual transition detection.

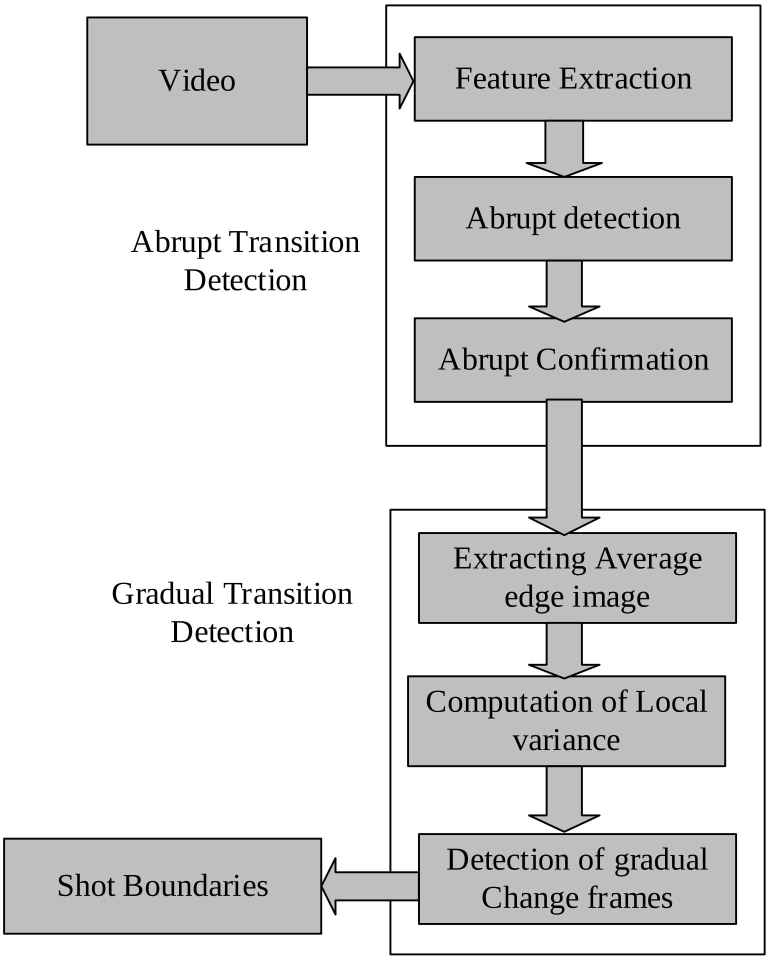


Fig. 7.8 The proposed block Diagram

a) Feature extraction

Feature extraction is the initial process of the proposed system in which the features are extracted from the equivalent frames such as Luminance Similarity Comparison (E) and Gradient Similarity Measurement (G) using Eqn (1) and (4). Combining these two features the equivalent frames quality similarity Measure (q) is found using (3). The features Q and E are used to find the probable detection of abrupt transition. The abrupt sections Lab colour difference is used in the confirmation stage.

b) Thresholding

The abrupt or gradual changes are clarified using a threshold. Beyond a certain threshold when there is a distance between the consecutive frames then a transition is declared. In a video, for detecting transitions we have employed two adaptive thresholds i.e. probable abrupt threshold (α) and final threshold (λ), represented in Eqn(11) and (12).

$$\alpha = \mu_q + (C_1 \times \sigma_q) \quad (11)$$

$$\lambda = \mu_{\Delta E} - (C_2 \times \sigma_{\Delta E}) \quad (12)$$

The possible abrupt transition is detected using the threshold α and in the abrupt confirmation stage λ is used. C_1 used in Eqn 6 is a constant and the value lying in between the range of [-3.2, -2.8] is selected likewise in Eqn 7 the constant C_2 having a range of [1.6, 2].

7.4.4. Abrupt section

Generally into two stages this abrupt section is categorized i.e., Probable Abrupt Detection Stage and Abrupt Affirmation Stage.

7.4.4.1 Probable abrupt detection stage

All the video frames in this stage are categorized into two i.e., probable transition frames and non-transition frames. In this stage the transition like behaving frames are held and remaining is rejected. The quality indicator (Q) is used for this classification using Eqn 3. As given in Eqn 10 the probable abrupt frames (PA) are found using the probable abrupt threshold (α).

The irregular and the former frames of fPA_i are given by $fPA_{i\pm\eta}$. In Lab colour space fPA_i is the probable transition frame. Experimentally η is set as 2 which is a user defined constant**.**

$$PA = \begin{cases} \text{Pr}\,obable\ \text{tran., if } \text{q}_\text{i} \geq \alpha \\ non - tran., otherwise \end{cases} \quad (12)$$

7.4.4.2 Abrupt confirmation stage

The probable abrupt transition frames (fPA_i) in this stage are stated as actual abrupt transition (A) based on the Eqn 13.

$$A = \begin{cases} True, \ if \ \Delta \mathrm{E}\left(\mathrm{fPA}_{\mathrm{i}-\eta}, fPA_{i+\eta}\right) \geq \lambda \\ False, \qquad otherwise \end{cases} \qquad (13)$$

E(fPA(i-η) is represented as the CIEDE2000 colour difference among the frames fPA(i-η) and fPA(i+η). The $\pm\eta^{th}$ frames from the i^{th} frame is indicated by μ *which is a* constant. For investigation the value of μ is considered as 2.

7.4.5. Gradual transition detection stage

This stage for gradual detection uses the gradual curves which are considered by the distribution of average edge information.

7.4.5.1 Typical edge image extraction

For each frame based on the pixels an average edge image was built. It is found that than the average intensity the pixels have more intensities. Than the original image these images are smoother and distinct. With certain changes the typical edge image found was stated from the effect average gradient (EAG). The average edge of image is obtained using the following steps.

1. Conversion of a color image into a gray image.

2. Employing the optimal edge detector the edge image is attained with threshold 100.

3. Using equation (10) calculate the average gradient (AG)

$$A_G = \sum_{x,y} F(G)(X,Y) / \sum_{x,y} R(X,Y) \qquad (14)$$

Where $R(X,Y) = 1, if \ \mathrm{F(G)(X,Y)} > 0$

$R(X,Y) = 0, if \ \mathrm{F(G)(X,Y)} = 0$

4. Based on the average gray (A_G) value Obtain an average edge image; using equation (11) consider a threshold value which is new.

$$F(X,Y) = \begin{cases} F\,\mathrm{G(X,Y)}\,if\ \mathrm{F\,G(X,Y) < A_G} \\ \quad 0, \qquad \mathrm{if\ F\,G(X,Y) \le A_G} \end{cases} \tag{15}$$

7.4.5.2 Gradual point Analysis

Considering the average gradient image the Gradual point detection is done. The variance was computed in every twenty frames. The variance is almost the same for the gradual sequence. Based on equations (12) and (13) the variance is calculated.

$$VAR(x) = \frac{1}{T-1}\sum_{k=x}^{x+T-1}\left(AGI(k) - mean(k)\right)^2 \tag{16}$$

$$Mean(x) = \frac{1}{T}\sum_{k=i}^{x+T-1} AGI(k) \tag{17}$$

Where x = 1, 2 ...n – T+1

In a window the total number of frames is denoted as T

In the kth window AGI (k) is the variance of the AGI.

7.4.5.3 Detection of the gradual change point

One of the frames in a gradual change of the local parabolic sequence has a least value. Based on the depth and width of the sequence the gradual change point was detected. A gradual change point is confirmed once the equations (18) and (19) are fulfilled.

$$D_{VAR} = \left|\omega LocMax[i \pm 1] - \omega LocMin[i]\right| > 0 \tag{18}$$

$$D_{FRAME} = \left|FrmLocMax[i \pm 1] - FrmLocMax[i]\right| < 20 \tag{19}$$

Where, i = 1, 2, 3...n

The variances of ith frame are denoted as $\omega LocalMin[i]$ and $\omega LocalMax[i]$ having the local minimum and maximum. The has local maximum frame number id denoted as $FrmLocMax[i]$

7.4.6. Results and discussion

This section presents a detailed explanation about the dataset used, simulation and execution of the approach proposed.

Dataset

The dataset used for experimentation is the TRECvid 2007 video databases which a benchmark dataset which is provided by Netherlands *Institute for Sound and Vision* on request. Compressed MPEG videos are used. We employ the Recall (R_{ec}), Precision (P_{re}) and F1 Score (F_1) factors for the performance evaluation of the proposed approach using Equations (20), (21) and (22).

$$R_{ec} = \frac{NC}{NC + NM} \times 100 \tag{20}$$

$$P_{re} = \frac{NC}{NC + NF} \times 100 \tag{21}$$

$$F_1 = \frac{2 * R_{ec} * P_{re}}{R_{ec} + P_{re}} \tag{22}$$

In a video the transitions which are correctly, missed and falsely detected are represented as *NC*, *NM* and *NF*. *Using the* chosen videos of TRECVid 2007 experimentation is conducted.

Table 1 provides the performance of some of the videos that are collected. Each videos computation time is provided in the Table 1. Figure 2 presents the transitions which are detected correctly from the *NAD*58.*mpg* video. Using the videos the proposed systems average F_1 scores are 95.6% and 80.9% for both the transitions with an average overall performance of 91.6%.

Table 7.1. Results of the Proposed System for TRECVid 2007 Database

Video	Computation time(sec)	Abrupt			Gradual			Overall		
		Rec	**Pre**	**F1**	**Rec**	**Pre**	**F1**	**Rec**	**Pre**	**F1**
BG 3027	1630	84.3	100.0	92.2	100.0	51.0	65.7	95.9	92.2	95.0
BG 3097	1535	86.9	100.0	92.7	-	-	-	88.9	100.0	92.6
BG 3314	1155	84	100.0	89.9				84	100.0	90.0
BG 16336	84	92.0	100.0	93.7				96.0	100.0	96.4
BG 37309	305	100.0	100.0	100.0	85.5	62.6	74.5	95.5	82	88
BG 37770	506	100.0	100.0	100.0	91	78.4	87.3	94	84	89
BG 22677	507	97.2	100.0	99	84.4	87.3	88.8	92.3	95	92
BG 36658	871	94.2	97.2	96	86	89	87	92.7	95.4	94.0
BG 8947	545	94.6	97.3	95.9	72.2	100.0	83.9	90.3	97.7	92.8
BG 4455	772	95.2	98.6	96.9	85.3	90	88	93	96	94
BG 35153	872	92.0	100.0	94.2	87	82.0	81.6	87.2	91	91
Clip	198	89.5	100.0	94.2				87.5	100.0	94
Average	**664.0**	**93**	**97**	**96**	**89.2**	**81.3**	**83.2**	**91.5**	**92.6**	**92.8**

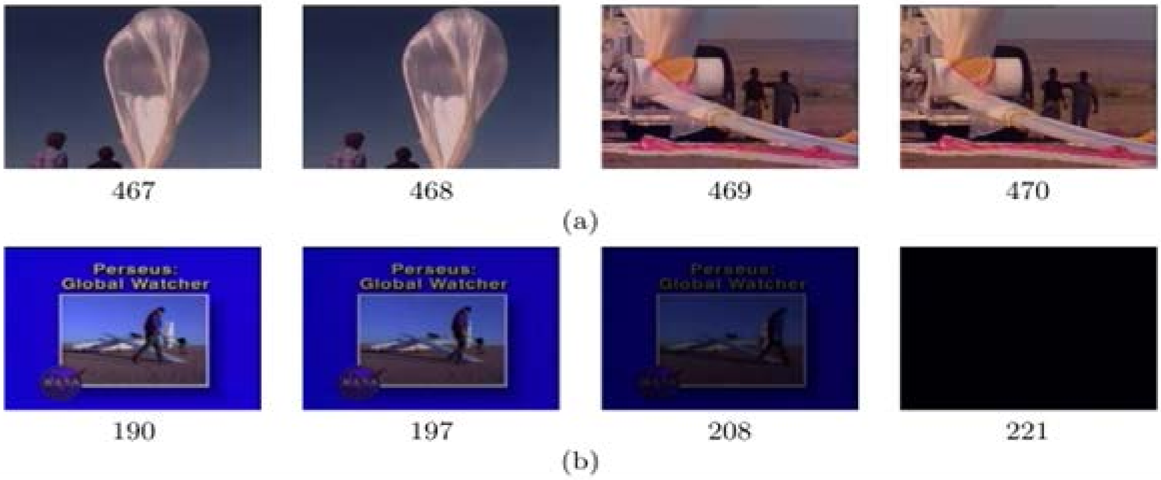


Fig. 7.9 Detected transitions a abrupt and b gradual transition

To validate proposed system, three SBD techniques proposed by the authors are considered.

Table 7.2. Performance of the proposed scheme compared with the other approaches

Methods	Parameter for evaluation	videos				Average
		Anni006	**Anni009**	**Anni010**	**NAD58**	
Proposed	R_{ec}	83.9	85.5	87.3	90.2	86.5
	P_{re}	77.3	90.2	93.6	94.1	91.7
	F_1	83.6	87.5	92.4	92.5	89.6
Eigen value decomposition and Gaussian transition detection method	R_{ec}	93.8	84.5	91.2	93	92.3
	P_{re}	85.2	82.1	78.4	92	84.5
	F_1	89	81.5	83.5	91.4	87
Temporal segmentation method	R_{ec}	84.3	87.9	88.5	92.8	89.1
	P_{re}	92.2	86	84.2	91.5	86.7
	F_1	91.1	84.2	85.0	93.7	89
3D convolutional networks method	R_{ec}	92.8	94.3	84.5	90.2	91.7
	P_{re}	95.5	80.6	85.6	90.6	87.3
	F_1	90.2	88.2	88.9	92.1	89.7

In Table 1 the F_1 score is 90.3% andfor gradual detection the obtained percentage is 85.7%. Table 2 displays the comparison of the proposed with the existing methods.

7.5. CONCLUSION AND FUTURE WORK

A new SBD technique is presented in this chapter for identifying the abrupt and gradual transitions. Using the adaptive threshold ( ) the similarity features are obtained for detecting the abrupt transitions. For detecting the gradual transition the gradual curve was obtained. For this initially the average edge image was found. The minimum difference between sequences and local variance were considered. At last the gradual change point was identified. The high-quality real-time results proves the effectiveness of the proposed that the existing strategies for video editing based SBD which needs quality and with speed

7.6. MAJOR FINDINGS

This chapter discusses the problem editing and its possible solutions. Video editing is much more required in all humans' life, such as journalism, film, online lessons, etc. It not only gives the viewers a descending image but also allows the publisher to add more effects to the needs of their visitors. There are some disadvantages to the captured side. If there is a mistake, it will be corrected in the editing process. Brief overview and a new frame work is also given for shot boundary detection. As a vital factor the Shot boundary detection (SBD) is considered for lot of video analysis tasks, for example video editing, indexing, summarization and action recognition. Based on this, an effective SBD approach is proposed. The gradient and color information are used for abrupt transition detection. For Gradual transition detection the average edge information of the gradual curves in the sequence of frames are obtained. From the optimal edge detector an average edge frame is gained. The computational complexity is reduced by this approach by processing only the transition regions. The proposed approach when compared to the exiting work done achieves improved results in terms of precision, recall and F1.

CHAPTER VIII

Conclusions and Future Scope

CHAPTER VIII

CONCLUSIONS AND FUTURE SCOPE

The aim of the thesis is to provide a research framework for the design and testing of video editing software. Different aspects of video editing mechanism has been analyzed through design and presentation of video processing, generation of videos (video generation through still images), research on motion vector, motion estimation and compensation. Different problems have been solved in this thesis pertaining to the video generation, motion estimation and compensation and finally an enhanced framework for video editing is developed. In video editing, the shot boundary detection method are critically analyzed and a better shot boundary detection algorithm has been proposed.

The following significant observations are inferred based on the presented results:

1. The thesis mainly focused on approaches of video processing. The backbone of video compression i.e. the motion estimation and compensation are deeply studied and a new pentagon search technique is proposed for the motion vector search
2. The algorithm proposed for the result-oriented methods based on video compression in the video system resolved the finding of true motion vectors.
3. Before discussing the any kind of problems, comprehensive literature was conducted in the thesis.
4. The entire work is based on approaches which are used to solve the problems mentioned in this thesis in order to gain satisfactory results.
5. The study results are optimized and compared with the current literature findings.
6. The tables and graphs presented are based on the optimized data.
7. Further studies in the same direction would also find very effective conclusions.

This work revealed that it can be generalized in various ways and some of the prominent directions are given below:

1. This research shows that Matlab can also be used in video applications. The experiments mentioned in the work indicates that a video file is generated using all of its frames. Some other platforms may also be incorporated for the evaluation of the video generation using still images.

2. In this dissertation a new Pentagon block algorithm is proposed and the results proved that it works better when compared with the existing hexagon and DS algorithm. In further studies the direction of pentagon may be rotated for analysis purpose.

3. In this work an optimized video compression and decompression framework is presented for estimated block movements with the Pentagon search. Our study is done on DCT mechanism it may be analyzed for different domain areas, e.g., wavelets, etc.

4. The study also figures out the issues of video editing and its potential solutions. The one major issue deeply analyzed is shot boundary detection and the possible solution is also provided for abrupt and gradual transitions. Feature extraction method is used for abrupt confirmation and average edge extraction is used for graduation confirmation. Some other methods may also be used for viewers need as the video editing limitations increasing day by day.

References

REFERENCES

[1] A. Graps (1995), "An introduction to wavelets," Computational Science and Engineering, IEEE [see also Computing in Science & Engineering], vol. 2, pp. 50-61.

[2] A. K. Jain (1988), Fundamentals of Digital Image Processing: Prentice Hall.

[3] A. Said and W. A. Pearlman (1996), "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 6, pp. 243-250, 1996.

[4] A. Said and W. A. Pearlman (1996), "An image multiresolution representation for lossless and lossy compression," in IEEE Transactions on Image Processing, vol. 5, no. 9, pp. 1303-1310, Sept. 1996, doi: 10.1109/83.535842.

[5] A. Skodras, C. Christopoulos, and T. Ebrahimi (2001), "The JPEG2000 still image compression standard," Signal Processing Magazine, IEEE, vol. 18, pp. 36-58.

[6] A.V. Paramkusam (2017), “Efficient motion estimation algorithm on the layers”,Electronics Letters (Volume: 53, Issue: 7 , 3 30 2017), Page(s): 467 – 469, Date of Publication: 30 March 2017, Print ISSN: 0013-5194, INSPEC Number: 16759448,DOI:10.1049/el.2016.3869.

[7] Ahluwalia, Siddhartha, Sourabh Rungta, and A. N. U. P. A. M. Shukla. "Circular 2-D Logarithmic Search Algorithm For Motion Estimation." Journal of Theoretical and Applied Information Technology 11, no. 1 (2010).

[8] Alexandra Trujillo J. et al. (2018), "Summarization of video from Feature Extraction Method using Image Processing and Artificial Intelligence" https://www.researchgate.net/publication/322578191, Alexandra Trujillo J. on 18 January 2018.

[9] Ankita P. Chauhan1, Rohit R. Parmar2, Shahida G. Chauhan3 (2012), "Comparative Study on Diamond Search Algorithm for Motion Estimation", International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 10, December- 2012, ISSN: 2278-0181.

[10] Anudeep Gandam1 and Dr. Jagroop Singh Sidhu2 (2016), "Review paper: Video processing & its applications" International Research Journal of Engineering andTechnology (IRJET), e-ISSN: 2395 -0056 Volume: 03 Issue: 08 | Aug-2016 www.irjet.net p-ISSN: 2395-0072.

[11] Arun H. Virage1, RameshJain2, and Terry E Weymouth3 (1995) "Feature Based Digital Video Indexing" S. Spaccapietra et al. (eds.), Visual Database Systems 3© Springer Science + Business Media Dordrecht 1995.

[12] AsadIslam1, Fehmi Chebil2 and Ari Hourunranta3 (2006), "Efficient Algorithms for Editing H.263 and MPEG-4 Videos on Mobile Terminals", 2006 International Conference onImage Processing, Date of Conference: 8-11 Oct. 2006, Date Added to IEEE Xplore: 20 February 2007, INSPEC Accession Number: 9462026, DOI:10.1109/ICIP.2006.313045.

[13] B. Furht, J. Greenberg and R. Westwat (1996), "Motion Estimation Algorithms for Video Compression", 1 ed.: Springer.

[14] B. Girod(2015), "Efficiency analysis of multi-hypothesis motion-compensated prediction for video coding", IEEE Transactions on Image Processing (Volume: 9, Issue: 2 ,Feb 2000), DOI: 10.1109/83.821595.

[15] B. GnanaPriya, M. Arulselvi (2019), "3d Image Generation from Single 2d Image using Monocular Depth Cues", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-2, December, 2019.

[16] B. Haskell, A. Puri, and A. Netravali (1996), "Digital Video: An Introduction to MPEG-2. Chapman & Hall", New York.

[17] Bede Liu and Andre Zaccarin (1993), "New Fast Algorithms for the Estimation of Block Motion Vectors", IEEE transactions on circuits and systems for video technology, vol. 3, no. 2, april 1993. 1051-8215/93$03.00.

[18] Bhujbal, Dhanashree, and B. V. Pawar. "Review of video stabilization techniques using block based motion vectors." Int. J. Adv. Res. Sci. Eng. Technol 3, no. 3 (2016): 1741-1747.

[19] Brendan Calandra1, Laurie Brantley-Dias2, John K. Lee3 and Dana L. Fox4 (2009), "Using Video Editing to Cultivate Novice Teachers' Practice" Journal of Research on Technology in Education, Copyright © 2009, ISTE (International Society for Technology in Education), 800.336.5191. (U.S. & Canada) or 541.302.3777 (Int'l),

[20] C. Andrew Segall and Aggelos K. Katsaggelos (2000), "Pre- and Post-Processing Algorithms for Compressed Video Enhancement", Department of Electrical and Computer Engineering, 2145 N. Sheridan Road Evanston, Illinois 60208-3 1 18, USA, 0-7803-615~3/00/$10.0002000 JEEE.

[21] C. L. Madhwacharyula1; M.S. Kankanhalli2and P. Millhem3 (2004), "Content basedediting of semantic video metadata", 2004 IEEE International Conference on Multimedia andExpo (ICME) (IEEE Cat. No.04TH8763), Date of Conference: 27-30 June 2004, Date Added to IEEE Xplore: 22 February 2005, Print ISBN: 0-7803-8603-5, INSPEC Accession Number: 8281968, DOI:10.1109/ICME.2004.1394118.

[22] Carl Vondrick1, Hamed Pirsiavash2 and Antonio Torralba3 (2016), "Generating Videos with Scene Dynamics", 1See http://mit.edu/vondrick/tinyvideo for the animated videos, 29th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

[23] Carlos Cuevas and NarcisoGarcía(2012), "High-quality real-time temporal segmentation tool for video editing software", 2012 IEEE International Conference on Consumer Electronics(ICCE), Date of Conference: 13-16 Jan. 2012, Date Added to IEEE Explore: 01 March 2012, INSPEC Number: 12589163, DOI: 10.1109/ICCE.2012.6161777.

[24] Ce Zhu1, Xiao Lin2, and Lap-Pui Chau3 (2002), "Hexagon-Based Search Patternfor Fast Block Motion Estimation", IEEE transactions on circuits and systems for video technology, vol. 12, no. 5, May 2002. 1051-8215/02$17.00

© 2002 IEEE.

[25] Cen, Hao, Kenneth Koedinger, and Brian Junker. "Learning factors analysis– a general method for cognitive model evaluation and improvement." In International Conference on Intelligent Tutoring Systems, pp. 164-175. Springer, Berlin, Heidelberg, 2006.

[26] Chau, Lap-Pui, and Ce Zhu. "A fast octagon-based search algorithm for motion estimation." Signal Processing 83, no. 3 (2003): 671-675.

[27] Chi, P. Y. (2016). "Designing Video-Based Interactive Instructions "(Doctoral dissertation, UC Berkeley).

[28] Cixun Zhang1; Kemal Ugur2; Jani Lainema3; Antti Hallapuro4 and Moncef Gabbouj2 (2011), "Video Coding Using Spatially Varying Transform", IEEE Transactions on Circuits andSystems for Video Technology (Volume: 21 , Issue: 2 , Feb. 2011),Page(s): 127 – 140, Date of Publication 2011, INSPEC Number: 11834665, DOI: 10.1109/TCSVT.2011.2105595.

[29] Cuevas, Erik, Daniel Zaldívar, Marco Pérez-Cisneros, Humberto Sossa, and Valentín Osuna. "Block matching algorithm for motion estimation based on Artificial Bee Colony (ABC)." Applied soft computing 13, no. 6 (2013): 3047-3059.

[30] Cutler, Ross. "The distributed meetings system." In 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03)., vol. 4, pp. IV-756. IEEE, 2003.

[31] D. Marr (1983), "Vision: A Computational Investigation into the Human Representation and Processing of Visual Information". San Francisco: W. H. Freeman.

[32] D. Santa-Cruz, R. Grosbois, and T. Ebrahimi (2002), "JPEG2000 performance evaluation and assessment" Signal Processing: Image Communication, Vol. 17, No.1, pp. 113-130.

[33] D. Selvanayagi, and Dr. S. Pannirselvam (2017), "Image Compression –

Decompression using Polynomial Based Wavelet Transformation in PDF Document", International Journal of Engineering Development and Research (www.ijedr.org), IJEDR1702136, Volume 5, Issue 2 | ISSN: 2321-9939.

[34] Dandi, Yatin, et al. "Jointly trained image and video generation using residual vectors." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020.

[35] Daniel Schonberg1; Stark C. Draper2; Chuohao Yeo3and Kannan Ramchandran4 (2008), "Toward Compression of Encrypted Images and Video Sequences", IEEE Transactions onInformation Forensics and Security (Volume: 3 , Issue: 4 , Dec. 2008), Page(s): 749 – 762, Date of Publication: 18 November 2008, INSPEC Accession Number: 10313211, DOI:10.1109/TIFS.2008.2007244.

[36] David Horowitz (2017), "Teaching video editing and motion graphics with Photoshop", Innovative Marketing, Volume 13, Issue 3, 2017, and DOI: http://dx.doi.org/10.21511/im.13(3).2017.02, ISSN PRINT 1814-2427, ISSN ONLINE 1816- 6326.

[37] Deepak Turaga, and Mohamed Alkanhal (1998), "Search Algorithms for Block- Matching in Motion Estimation", Mid-Term project, 18-899, spring,1998.

[38] Dr. Anil Kokaram (2015), "Motion Estimation, Image and Video Processing", anil.kokaram@tcd.ie, 2.1 Problems with Pixel Difference.

[39] Dr. Anil Kokaram et al. (2012), "Introduction to Digital Video Processing" www.mee.tcd.ie/. ISD. Recording pictures and creating the video.

[40] Dubois, Eric. "Motion-compensated filtering of time-varying images." Multidimensional Systems and Signal Processing 3, no. 2 (1992): 211-239.

[41] Dufaux, Frédéric, and Touradj Ebrahimi. "Scrambling for video surveillance with privacy." 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06). IEEE, 2006.

[42] Eric Laurier1, Ignaz Strebel2& Barry Brown3 (2008), “Video Analysis: Lessons from

[43] Felsner, Stefan, Hendrik Schrezenmaier, and Raphael Steiner. "Pentagon contact representations." Electronic Notes in Discrete Mathematics 61 (2017): 421-427.

[44] Fillia Makedon1, James W. Matthews2, Charles B. Owen3and Samuel A.Rebelsky4 (1994), “Multimedia authoring, development environments, and digital video editing”, Proceedings Volume 10278, Defining the Global Information Infrastructure: Infrastructure, Systems, and Services: A Critical Review; 102780H (1994) https://doi.org/10.1117/12.192187 Event: Photonics for Industrial Applications, 1994, Boston, MA, UnitedStates.

[45] G. K. Wallace (1992), "The JPEG still picture compression standard," Consumer Electronics, IEEE Transactions on, vol. 38, pp. xviii-xxxiv.

[46] G. Karlsson and M. Vetterli (1998), "Three dimensional sub-band coding of video," in Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on, 1988, pp. 1100-1103 vol.2.

[47] G. L. Chen, J.-S. Pan and J.-L. Wang (1996), "Video encoder architecture for MPEG2 real time encoding," IEEE Transactions on Consumer Electronics, vol. 42, pp. 290-299, 1996.

[48] Gangadhar Tiwari1, Debashis Nandi2 and Madhusudhan Mishra3 (2013), “A Comparative study on Image and Video Compression Techniques”, IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 3, Issue 3 (Sep. – Oct. 2013), PP 69-73 e- ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197 [www.iosrjournals.org.](http://www.iosrjournals.org)

[49] Gibson, Jerry D., and Al Bovik, eds. "Handbook of Image and Video Processing." (2000).

[50] Golwelkar, Abhijeet, and John W. Woods. "Motion-compensated temporal filtering and motion vector coding using biorthogonal filters." IEEE Transactions on Circuits and Systems for Video Technology 17, no. 4 (2007):

417-428.

[51] Gui-guang, Ding, and Guo Bao-long. "Motion vector estimation using line-square search block matching algorithm for video sequences." EURASIP Journal on Advances in Signal Processing 2004, no. 11 (2004): 1-7.

[52] H. Chung-Lin and H. Chao-Yuen (1994), "A new motion compensation method for image sequence coding using hierarchical grid interpolation," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 4, pp. 42-52.

[53] H. Fang, S. Gupta, F. Iandola (2015), “From captions to visual concepts and back,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, June 2015.

[54] Haller, Martin, Andreas Krutz, and Thomas Sikora. "Evaluation of pixel-and motion vector-based global motion estimation for camera motion characterization." In 2009 10th Workshop on Image Analysis for Multimedia Interactive Services, pp. 49-52. IEEE, 2009.

[55] Haller, Martin, Andreas Krutz, and Thomas Sikora. "Evaluation of pixel-and motion vector-based global motion estimation for camera motion characterization." In 2009 10th Workshop on Image Analysis for Multimedia Interactive Services, pp. 49-52. IEEE, 2009.

[56] Haney, Brian Seamus. "Blockchain: Post-Quantum Security & Legal Economics." NC Banking Inst. 24 (2020): 117.

[57] Hao, Zekun, Xun Huang, and Serge Belongie. "Controllable video generation with sparse trajectories." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[58] Haoran Wang1, Yue Zhang1, and Xiaosheng Yu2 (2020), “An Overview of Image Caption Generation Methods Haoran”, Hindawi Computational Intelligence and Neuroscience, Volume 2020, Article ID 3062706, 13 pages, https://doi.org/10.1155/2020/3062706.

[59] HongjunJia, and Li Zhang (2000), “Directional Cross Diamond Search Algorithm for Fast Block Motion Estimation”, Tsinghua University, Beijing, China.

[60] Hosam, Osama. "Motion compensation for video codec based on disparity estimation." In 2012 IEEE Symposium on Computers and Communications (ISCC), pp. 000559-000565. IEEE, 2012.

[61] I. Daubechies (1992), “Ten Lectures on Wavelets”, SIAM: Society for Industrial and Applied Mathematics.

[62] I. E. G. Richardson (2003), “H.264 and MPEG-4 Video Compression”, VideoCoding for Next Generation Multimedia, 1 ed.: Wiley.

[63] Iain E. G. Richardson (2003), “H.264 and MPEG-4 Video Compression”, The Robert Gordon University, Aberdeen, UK, Copyright C _ 2003 John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, Telephone (+44)1243 779777. [107].

[64] Imbaby I. Mahmoud1, SheriefM. Hashimaa2, and Atef A. Elazm3 (2009), “Proposed One Point Pentagon inner search fast block matching algorithm”, 26th national radioscience conference, NRSC'2009 Future University, Sfu Compound, NewCairo, Egypt, March17- 19, Faculty of Engineering, Future Univ., Egypt.

[65] J. S. Lim (1990), “Two-Dimensional Signal and Image Processing”, Prentice Hall, Inc.

[66] Jae HeonJeong; NidhiParmar; MyungHoonSunwoo(2015), “Enhanced test zone search algorithm with rotating pentagon search”, 2015 International SoC Design Conference (ISOCC), Date of Conference: 2-5 Nov. 2015, Date Added to IEEE Explore: 11 February 2016, INSPEC Accession Number: 15799097, DOI:10.1109/ISOCC.2015.7401754, Publisher: IEEE.

[67] Jakubowski, M., and G. Pastuszak. "Block-based motion estimation algorithms—a survey." Opto-Electronics Review 21, no. 1 (2013): 86-102.

[68] James Nightingale and Qi Wang ; Christos Grecos(2012), "HEVStream: a framework for streaming and evaluation of high efficiency video coding (HEVC) content in loss-prone networks", IEEE Transactions on Consumer Electronics (Volume: 58 , Issue: 2 ,May 2012), Page(s): 404 – 412, Date of Publication: 05 July 2012, INSPEC Accession Number: 12834574, DOI: 10.1109/TCE.2012.6227440.

[69] Jamie Zigelbaum1, Michael Horn2, Orit Shaer3, and Robert J. K. Jacob4 (2013), "Tangible Video Editor: Designing for Collaboration, Exploration, and Engagement." Jamie Zigelbaum, Michael Horn, Orit Shaer, Robert J. K. Jacob, Tufts University, Medford, Mass.USA,

[70] Jason Li and Helen Qiu (2020), "Comparing Attention-based Neural Architectures for Video Captioning", jasonkli@stanford.edu, Jason Li and Helen Qiu (2020), "Comparing Attention-based Neural Architectures for Video Captioning", jasonkli@stanford.edu.

[71] Jianbin Song, Bo Li, Dong Jiang and Caixia Wang (2006), "Large diamond and small pentagon search patterns for fast motion estimation", ICNC'06: Proceedings of the Second international conference on Advances in Natural Computation - Volume Part IISeptember 2006 Pages608–616https://doi.org/10.1007/11881223_75.

[72] K. Asada1, H. Ohtsubo2, T. Fujihira3, and T. Imaide4 (1997), "Development of low power MPEG1/JPEG encode/decode IC," Consumer Electronics, IEEE Transactions on, vol. 43, pp. 639-645. G. K. Wallace (1992), "The JPEG still picture compression standard," Consumer Electronics, IEEE Transactions on, vol. 38, pp. xviii-xxxiv.

[73] Khawase, Sonam T., Shailesh D. Kamble, Nileshsingh V. Thakur, and Akshay S. Patharkar. "An Overview of Block Matching Algorithms for Motion Vector Estimation."In RICE, pp. 217-222. 2017.

[74] Kim, Doyeon, Donggyu Joo, and Junmo Kim. "TiVGAN: Text to Image to Video Generation With Step-by-Step Evolutionary Generator." IEEE Access 8 (2020): 153113- 153122.

[75] Koc, Ut-Va, and KJ Ray Liu. "Motion compensation on DCT domain." EURASIP Journal on Advances in Signal Processing 2001, no. 3 (2001): 1-16.

[76] KoohyarMinoo andTruong Nguyen (2009), “Entropy Coding via Parametric Source Model with Applications in Fast and Efficient Compression of Image and Video Data Entropy Coding via Parametric Source Model with Applications in Fast and Efficient Compression of Image and Video Data”, 2009 Data Compression Conference, Date of Conference: 16-18 March 2009, Date Added to IEEE Xplore: 26 May 2009, INSPEC Accession Number: 10666374, DOI: 10.1109/DCC.2009.80

[77] Kovács, Péter Tamás, Zsolt Nagy, Attila Barsi, Vamsi Kiran Adhikarla, and Robert Bregović. "Overview of the applicability of H. 264/MVC for real-time light-field applications." In 2014 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), pp. 1-4. IEEE, 2014.

[78] Kulkarni, S. M., D. S. Bormane, and S. L. Nalbalwar. "Coding of video sequences using three step search algorithm." Procedia computer science 49 (2015): 42-49.

[79] Lee, Ho Min, Donghwi Jung, Ali Sadollah, Do Guen Yoo, and Joong Hoon Kim. "Generation of benchmark problems for optimal design of water distribution systems." Water 11, no. 8 (2019): 1637.

[80] Lee, Yung-Lyul, Yung-Ki Lee, and HyunWook Park. "A fast motion vector search algorithm for variable blocks." In International Conference on Advanced Concepts for Intelligent Vision Systems, pp. 311-322. Springer, Berlin, Heidelberg, 2006.

[81] Li Xiaoli1; Liu Chao2; Wang Qiang3; and Ding Wenrui4 (2009), “Design and Implementation of Real-time Video Compression System Control Framework”, 2009 WRIInternational Conference on Communications and Mobile Computing, Date Added to IEEE Explore: 04 March 2009, Print ISBN: 978-0-7695-3501-2, INSPEC Number: 10501488, DOI:10.1109/CMC.2009.244.

[82] Li Yao1, Atousa Torabi2 and Kyunghyun Cho3 (2015), "Describing Videos by Exploiting Temporal Structure", arxiv: 1502.08029v5 [stat.ML] 1 Oct 2015, 1https://www.youtube.com/yt/press/statistics. Html accessed on 2015-02-06.

[83] LumingZhang ; Peiguang Jing ; Yuting Su ; Chao Zhang ; Ling Shaoz (2017), "SnapVideo: Personalized Video Generation for a Sightseeing Trip", IEEE Transactions onCybernetics (Volume: 47 , Issue: 11 , Nov. 2017), Page(s): 3866 – 3878, Date of Publication:19 July 2016, INSPEC Accession Number: 17243814,DOI:10.1109/TCYB.2016.2585764.

[84] Lurng-KuoLiu and Ephraim Feig (1996), "A Block-Based Gradient Descent Search Algorithm for Block Motion Estimation in Video Coding", IEEE transactions on circuits and systems for video technology, vol. 6, no. 4, august 1996, 1051-8215/96$05.00 0 1996 IEEE.

[85] M. J. Narasimha and A. M. Peterson (1978), "On the computation of the discrete cosine transform," IEEE Trans on Communications, vol. 26, pp. 934-926.

[86] M. K. Mandal (2003), Multimedia Signals and Systems: Kluwer Academic Publishers.

[87] M. Ravinder1, T. Venu gopal² and T. VenkatNarayana rao³ (2010), "Video indexing and retrieval - Applications and challenges" Oriental Journal of Computer Science &Technology, Vol. 3(1), 125-137 (2010).

[88] Mackenzie leake1, abe davis2, and truong3, Maneesh agrawala4 (2017), "Computational Video Editing for Dialogue-Driven Scenes" ACM Transactions on Graphics, Vol. 36, No. 4, Article130. Publication date: July 2017, DOI: h.p://dx.doi.org/10.1145/3072959.3073653.

[89] Mary Beisiegel1, Rebecca Mitchell2, and Heather C. Hill3 (2018) "The Design of Video- Based Professional Development: An Exploratory Experiment Intended to Identify Effective Features" Journal of Teacher

Education 2018, Vol. 69(1) 69–89, © 2017

[90] Mazumdar, Amrita, Brandon Haynes, Magda Balazinska, Luis Ceze, Alvin Cheung, and Mark Oskin. "Perceptual compression for video storage and processing systems." In Proceedings of the ACM Symposium on Cloud Computing, pp. 179-192. 2019.

[91] McArthur, E., Kubacki, K., Pang, B. &Alcaraz, C. (2017). The Employers' View of 'Work-Ready' Graduates: A Study of Advertisements for Marketing Jobs in Australia. Journal of Marketing Education, 39(2), 82-93. https://doi.org/10.1177/0273475317712766.

[92] Michael s1. Lew, Nicu Sebe2, and Paul C. Gardner3 (2001) "Video Indexing and Understanding" M. S. Lew (ed.), Principles of Visual Information Retrieval © Springer- Verlag London 2001.

[93] MihirMody (2014), "Video codec framework (VCF): Novel firmware architecture for video hardware", 2014 Twentieth National Conference on Communications (NCC),Date of Conference: 28 Feb.-2 March 2014, Date Added to IEEE Xplore: 08 May 2014, Electronic ISBN: 978-1-4799-2361 - 8, INSPEC Number: 14283079, DOI:10.1109/NCC.2014. 6811242.

[94] Miller, F. L., Mangold, W. G., Roach, J., & Holmes, T. (2013). Building the Technology Toolkit of Marketing Students: The Emerging Technologies in MarketingInitiative. Marketing Education Review, 23(2), 121-136. http://www.tandfonline.com/doi/abs/10.2753/MER1052-8008230202.

[95] Mohammed Ebrahim (2013), "Motion Compensated Prediction", volume 3, issue 3.

[96] MohdAveshand Rajeev Srivastava (2016), "Parametric study on the performance of active suspension system for variable passenger size and repeated road bumps", 2016 10thInternational Conference on Intelligent Systems and Control (ISCO), Date of Conference: 7-8 Jan. 2016, Date Added to IEEE Explore: 03 November 2016, INSPEC Accession Number: 16444060, DOI:10.1109/ISCO.2016.7726894.

[97] Naderiparizi, Saman, et al. "Glimpse: A programmable early-discard camera architecture for continuous mobile vision." Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services. 2017.

[98] Nam, Kwon Moon, Joon-Seek Kim, Rae-Hong Park, and Young Serk Shim. "A fast hierarchical motion vector estimation algorithm using mean pyramid." IEEE Transactionson Circuits and Systems for Video technology 5, no. 4 (1995): 344-351.

[99] Neetish Kumar and Deepa Raj (2018), "Video Generation using Still Images" International Journal of Computer Sciences and Engineering Vol.6 (11), Nov 2018, E- ISSN: 2347-2693.

[100] Neetish Kumar and Deepa Raj (2018), "A Study and Analysis of Images inDifferent Color Models" WWW.IJASCSE.ORG, Volume 7, Issue 1, 2018.

[101] Neetish Kumar and Deepa Raj (2020), "algorithmic solutions for high quality video editing software problems" journal of critical reviews. ISSN- 2394-5125 VOL 7, ISSUE 09, 2020.

[102] Neetish Kumar and Dr. Deepa Raj (2017), "Video Processing and its Applications: A survey" International Journal of Emerging Trends & Technology inComputer Science (IJETTCS), Volume 6, Issue 4, July- August 2017 ISSN 2278-6856.

[103] Neetish Kumar1, and Deepa Raj2 (2019), "A New Pentagon Search Algorithmfor Fast Block-Matching Motion Estimation" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-10, August 2019.

[104] Nicole S. Love and Chandrika Kamath (1998), "An Empirical Study of BlockMatching Techniques for the Detection of Moving Objects", Center for Applied Scientific Computing Lawrence Livermore National Laboratory, 7000 East Ave, Livermore, CA94550.

[105] NirasCheeckottu Vayalil1 and Yinan Kong2 (2017), “VLSI Architecture of Full- Search Variable-Block-Size Motion Estimation for HEVC Video Encoding”, IET Circuits, Devices &Systems (Volume: 11 , Issue: 6 , 11 2017), Page(s): 543 – 548, Date of Publication: 20 November 2017 , INSPEC Number: 17320524, DOI: 10.1049/iet- cds.2016.0267.

[106] Nuno Correia1 and Nuno Guimadies2 (1996), "Components for video processing applications" INESC, R.AlvesRedol, 9,60.,1000 © Springer-Verlag/Wien 1996.

[107] Ohnishi, Katsunori, et al. "Hierarchical video generation from orthogonal information: Optical flow and texture." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 32. No. 1. 2018.

[108] P. Razvan, G. Caglar, K. Cho, and B. Yoshua (2014), “How to construct deep recurrent neural networks,” Computer Science, 2014, http://arxiv.org/abs/1312.6026.

[109] P.B. Penafieland N.M. Namazi (1995), “A new framework for noise-resistant video compression using motion-compensated prediction”, 1995 International Conference onAcoustics, Speech, and Signal Processing, Date of Conference: 9-12 May 1995, Date Added to IEEE Xplore: 06 August 2002, Print ISBN: 0-7803-2431-5, Print ISSN: 1520- 6149, INSPEC Accession Number: 5137660, DOI: 10.1109/ICASSP.1995.479908.

[110] Pan, Hao, P. Van Beek, and M. Ibrahim Sezan. "Detection of slow-motion replay segments in sports video for highlights generation." 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221). Vol. 3. IEEE, 2001.

[111] Pan, Junting, Chengyu Wang, Xu Jia, Jing Shao, Lu Sheng, Junjie Yan, and Xiaogang Wang. "Video generation from single semantic label map." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3733-3742. 2019.

[112] Patel, Dhara, Dixesh Patel, Devang Bhatt, and Kalpesh R. Jadav. "Motion Compensation for Hand Held Camera Device." Int. J. of Research in

Engineering andTechnology 4 (2015): 771-775.

[113] Per Erik Eriksson1, Akademisk avhandling2 (2018), "video graphy as design nexus critical inquires into the affordances and efficacies of live-action video instructions" Copyright © Per Erik Eriksson, 2018, ISBN 978-91-7485-391-9, ISSN 1651-4238, Printed by E- Print AB, Stockholm, Sweden.

[114] Professional Video Editing Practice", © 2008 FQS http://www.qualitative-research.net/fqs/ Forum Qualitative Sozialforschung / Forum: Qualitative Social Research (ISSN 1438-5627)

[115] Rajeshwar Dass1, Lalit Singh2 and Sandeep Kaushik3 (2012), "Video Compression Technique", international journal of scientific & technology research volume 1, issue 10, november 2012 ISSN 2277-8616, www.ijstr.org.

[116] S. Moni; S. Sista (1997), "A framework for application specific image compression", Proceedings DCC '97. Data Compression Conference, Date of Conference: 25-27 March 1997, Print ISBN: 0-8186-7761-9, Print ISSN: 1068-0314, INSPEC Accession Number: 5658325, DOI:10.1109/DCC.1997.582119.

[117] Sadiq H. Abdulhussain1, Abd Rahman Ramli2, M. Iqbal Saripan3, Basheera M. Mahmmod4 Syed Abdul Rahman Al-Haddad5 and Wissam A. Jassim6 (2013), "Methods and Challenges in Shot Boundary Detection: A Review", Received: 25 January 2018; Accepted: 27 February 2018; Published: 23 March 2018, Entropy 2018, 20, 214; doi:10.3390/e20040214 www.mdpi.com/journal/entropy.

[118] Saloni r. Mistry1, heni s. Modi2, rahul n. And gonawala3 (2014), "logarithmic search for motion estimation", International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982 Volume-2, Issue-6, and June-2014.

[119] Seferidis, Vassilis E., and Mohammad Ghanbari. "General approach to block- matching motion estimation." Optical Engineering 32, no. 7 (1993): 1464-1474.

[120] Shaikh, Muhammad Aakif, and Sagar S. Badnerkar. "Video Compression Algorithm Using Motion Compensation Technique." International Journal of Advanced Research in Electronics and Communication Engineering 3, no. 6 (2014): 625-628.

[121] Shan Zhu ; Kai-Kuang Ma (2000), “A new diamond search algorithm for fast block-matching motion estimation”, IEEE Transactions on Image Processing (Volume: 9 , Issue: 2), Page(s): 287 – 290, Date of Publication: Feb 2000, INSPEC Number: 6527070, DOI:10.1109/83.821744.

[122] Shan Zhu, X. Sun, X Gao, J Wang, N Zhao, J Sha (2019)"Equivalent circuit model recognition of electrochemical impedance spectroscopy via machine learning", Journal of Electroanalytical Chemistry. Publication date 2019/12/15, Elsevier.

[123] Sherief M. Hashimaa; Imbaby I. Mahmoud ; Atef A. Elazm (2011), “Experimental comparison among Fast Block Matching Algorithms (FBMAs) for motion estimation and object tracking”, 2011 28th National Radio Science Conference (NRSC), Date of Conference: 26-28 April 2011, INSPEC Number: 12061704, DOI: 10.1109/NRSC.2011.5873609.

[124] Sherief M. Hashimaa1, Imbaby I. Mahmoud2 and Atef A. Elazm3 (2009), “hard ware implementation of diamond search Algorithm for motion estimation and object Tracking” Proceedings of the 7th Conference on Nuclear and Particle Physics, 11-15 Nov. 2009, Sharm El- Sheikh, Egypt.

[125] Siu-Wai Wu1 and Jing Yang Chen2 (2001), “Motion compensated temporal filtering for noise reduction pre-processing of digital video data”, Priority to US09/929,553

[126] Siwei Ma1, Xinfeng Zhang2, Chuanmin Jia3, Zhenghui Zhao4, Shiqi Wang4 and Shanshe Wang5 (2019), “Image and Video Compression with Neural Networks: A Review”, ieee transactions on circuits and systems for video technology.

[127] Song, Jianbin, Bo Li, Dong Jiang, and Caixia Wang. "Large diamond and small pentagon search patterns for fast motion estimation." In International Conference on Natural Computation, pp. 608-616. Springer, Berlin, Heidelberg, 2006.

[128] Suk-ju Kang1 Kyoungrok Cho2and Young Hwan Kim3 (2015), “Motion CompensatedFrame Rate Up-Conversion Using Extended Bilateral Motion Estimation”, IEEE Transactions on Consumer Electronics (Volume: 53 , Issue: 4 , Nov. 2007), DOI: 10.1109/TCE.2007.4429281.

[129] T. Meier; and K.N. Ngan (1998), “Automatic segmentation of moving objects for video object plane generation” IEEE Transactions on Circuits and Systems for VideoTechnology (Volume: 8 , Issue: 5 , Sep 1998), INSPEC Accession Number: 6045575, DOI:10.1109/76.718500.

[130] T. Wiegand1, G. J. Sullivan2, G. Bjntegaard3, and A. Luthra4 (2003), "Overview of the H.264/AVC video coding standard," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 13, pp. 560-576.

[131] Thomas Meier and King N. Ngan (1998), “Automatic Segmentation of Moving Objects for Video Object Plane Generation”, IEEE transactions on circuits and systems for video technology, vol. 8, NO. 5, september 1998, 1051–8215/98$10.00 ã 1998IEEE.

[132] Tien-ying Kuo1 and C.-C. Jay Kuo2 (1998), “Fast Overlapped Block Motion Compensation with Checkerboard Block Partitioning”, IEEE transactions on circuits and systems for video technology, vol. 8, NO. 6, october 1998, 1051–8215/98$10.00 ã 1998 IEEE.

[133] Toivonen, T., and J. Heikkilä. "Motion Vector Refinement for Video Coding Based on Statistical Distribution." In Proc. 13th International Conference on Systems. in: Proc. 13th International Conference on Systems, Signals & Image Processing (IWSSIP 2006), Budapest, Hungary, 39-42.., 2006.

[134] Tourapis, Alexis Michael, Oscar Chi Lim Au, and Ming Lei Liou. "Predictive motion vector field adaptive search technique (PMVFAST): enhancing block-based motion estimation." In Visual Communications and

Image Processing 2001, vol. 4310, pp. 883- 892. International Society for Optics and Photonics, 2000.

[135] Tsinghua Science and Technology, (2007), “Video Structure Analysis”, Vol. 12, Issue (6): 714-718, doi: 10.1016/S1007-0214(07)70180-3, http://tst.tsinghuajournals.com.

[136] Tulyakov, Sergey, Ming-Yu Liu, Xiaodong Yang, and Jan Kautz. "Mocogan: Decomposing motion and content for video generation." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1526-1535. 2018.

[137] Tung Nguyen1, Philipp Helle2, Martin Winken3, Benjamin Bross3, Detlev Marpe4, Heiko Schwarz5, and Thomas Wiegand6 (2013), “Transform Coding Techniques in hevc”, ieee journal of selected topics in signal processing, vol. 7, no. 6, december 2013, digital object identifier 10.1109/jstsp.2013.2278071.

[138] V. Bhaskaran and K. Konstantinides (1995), “Image and Video Compression Standards: Kluwer”, Academic Publishers.

[139] Vineet Gandhi1 and Remi Ronfard2 (2015) "A Computational Framework for Vertical Video Editing" https://www.researchgate.net/publication /277775714, DOI: 10.2312/wiced.20151075, The Euro graphics Association 2015.

[140] W. Pennebaker and J. Mitchell (1993), “JPEG Still Image Data Compression Standard”, Van Nostrand Reinhold, New York.

[141] Wiles, Olivia, A. Koepke, and Andrew Zisserman. "X2face: A network for controlling face generation using images, audio, and pose codes." In Proceedings of the European conference on computer vision (ECCV), pp. 670-686. 2018.

[142] Wu, Chao-Yuan, Nayan Singhal, and Philipp Krahenbuhl. "Video compression through image interpolation." In Proceedings of the European Conference on Computer Vision (ECCV), pp. 416-431. 2018.

[143] Xian-Sheng Hua1,Lie Lu2and Hong-Jiang Zhang3 (2004), “Optimization-based automated home video editing system”, IEEE Transactions on Circuits and Systems for VideoTechnology (Volume: 14 , Issue: 5 , May 2004) Page(s): 572 – 583, Date of Publication: 04 May 2004, INSPEC Number: 7968861, DOI:10.1109/TCSVT.2004.826750.

[144] Yaakob, Razali, Alihossein Aryanfar, Alfian Abdul Halin, and Nasir Sulaiman. "A comparison of different block matching algorithms for motion estimation." Procedia Technology 11 (2013): 199-205.

[145] Yao Nie; Kai-Kuang Ma (2002), “Adaptive rood pattern search for fast block-matching motion estimation”, IEEE Transactions on Image Processing (Volume: 11 , Issue: 12 , Dec 2002), Page(s):1442– 1449, Date of Publication:Dec 2002, INSPEC Accession Number: 7539403, DOI:10.1109/TIP.2002.806251.

[146] Yitong Li, Martin Renqiang Min, DinghanShen, David Carlson and Lawrence Carin (2017), “Video Generation from Text”, Duke University, Durham, NC, United States,27708

[147] Yuan Gao1, Pent-up Liu2and Ke-bin Jia3 (2013), “A Fast Motion Estimation Algorithm Based on Motion Vector Distribution Prediction”, JOURNAL OF SOFTWARE, VOL. 8, NO. 11, november 2013, © 2013 academy publisher,doi:10.4304/jsw.8.11.2863- 2870.

[148] Yuki Nakahira; Kazuhiko Kawamoto (2019), “DCVGAN: Depth Conditional VideoGeneration”, 2019 IEEE International Conference on Image Processing (ICIP), INSPEC Accession Number: 19212354, DOI: 10.1109/ICIP. 2019.8803764.

[149] Zhen'gang Wei1, Gang Wang2 and Gang Wang3 (2010), “The realization of video mixing edit method based on DES”, July 2010, DOI: 10.1109/ICSESS. 2010.5552347.

[150] Zhengyuan (2013) "advanced video processing techniques in video transmission systems," Doctoral dissertation, University of Florida 2013.

[151] https://www.mathworks.com/................(Accessed on 12/06/2017)

[152] https://creately.com/...............................(Accessed on 03/01/2018)

[153] https://www.coursera.org/learn/digital...(Accessed on 10/04/2018)

[154] https://www.youtube.com/.....................(Accessed on 05/02/2017)

[155] https://www.pyimagesearch.com/...........(Accessed on 22/07/2018)

Appendix

A Study and Analysis of Images in Different Color Models

Neetish Kumar, Deepa Raj
Department of computer science
BBAU, Lucknow, India

***Abstract*— Apart from having lots of methods for compression of color images, the importance of lossless color image is irreplaceable in many important situations. Today is the world of electronic gadgets which generates voluminous data that includes the various forms of multimedia data. In these data collections, images need more space to store the content and need more bandwidth to transmit through a network. By using Image compression the size of the image can be reduced which helps in less utilization of memory and transmitted across a network with less bandwidth. Due to the efficient compression techniques the execution time for transmission drastically decreases and throughput of machine attains at a satisfactory level. This leads to the wider acceptance of the compression performance in much wider domain.**

Keywords- Compression, Enhancement, Transmission

I. INTRODUCTION

In the current scenario, in our daily life not a single field id untouched with the digital technology. Advancement in the digital technology has also changed the way of communication. Due to this change visual appearance are more preferred over the text. It leads to the motivation of reduction of the size of multimedia data. Compression techniques are widely explored in two categories e.g. lossy and lossless. According to their name, in lossy compression same extent of inappropriate information is discarded while in case of lossless compression the original image can be perfectly reformed without loss of data. The motivation behind the color image processing deals with various factors one thing may be the more sensitivity of human visual systems towards the color images. A nother factor assures that color is a powerful deceptor that often simplifies object identification and extraction from a scene. Second, humans can discern thousands of color shades and intensities, compared to about only two dozen shades of gray.

The resolution if an image is highly related to the picture element, commonly known as pixel. The more the no of pixels of a particular image have the more and more clarity it gets. Each pixel stores the color information of an image in terms of bits.

Color image processing comprises of two major areas namely full color and pseudo color processing. In full-color processing, the images in question typically are acquired with a full-color sensor, such as color TV camera or color scanner. In the second category, the problem is one of assigning a color to a particular monochrome intensity or range of intensities.

II. Literature survey

Seyun kim et. al [4] presented a lossless color image compression called the hierarchical prediction and context-adaptive arithmetic coding. In this approach, YCuCv color space of the image is generated from an RGB image. Later this method is compared with VSBHP (variable size block hierarchical prediction and the performance somehow was increased. Zhang, Zhou et al [1,2,3] developed algorithms for lossy as well as lossless compression of images, among those the most widely used are lossless CALIC, JPEG, JPEG-LS, LOCO-I, JPEG2000 (lossless). Among these the development of LOCO-I and CALIC are in the process of state of art standardization JPEG, most of the useful ideas in LOCO-I are accepted for JPEG-LS, however, at the cost of more computations, CALIC can achieve better compression performance. WDR based approach is also used for compression by C.Wang [10]. In this scheme, an original RGB image is transformed into YCbCr color space, due to this transformation, the correlation among RGB planes is reduced. By empirical evaluation, it can be observed most of the image information lies in luminance i.e. Y plane that helps for achieved high compression performance in chrominance color components. While Cb and Cr components had much less information that of Y. In some other wavelet-based approach, a wavelet-based simultaneous fusion, encryption and compression technique is used by C. Wang et al [9, 13]. The sparse representation of source images using DWT has helped in keeping properties of source images in the reconstructed fused image. The low and high frequency are fused using different rules and both frequency coefficients are simultaneously encrypted and compressed with the help of pseudo random number and pseudorandom permutation. S.Sathappan and P.suresh babu [14] used Image compression by Reversible Color Transform (RCT) method. The size of the image can be reduced which helps in less utilization of memory and transmitted across a network with less bandwidth. The

Modified Hierarchical Prediction (MHP) encodes Cu and Cv of chrominance channel, then they are decomposed into sub-images to predict the vertical, horizontal and diagonal pixels for effective image compression. After that luminance image Y is encoded using a technique called Variable Sized Blocks Hierarchical Prediction (VSBHP) coding scheme. This gives much better result than the previous existing approach.

III. DIFFERENT COLOR MODEL

The goal of the color model is to expedite the blueprint in some standard and mostly in a generalized way.

A. *RGB Color Model*

In RGB color model each color appears in its primary components of red, green, and blue. This model is based on a Cartesian coordinate system in which at the origin of the cube the color is black, the farthest point from the origin the color is white and the line joining between the two represent grayscale. The different corners depict different colors of the cube. The effect of using RGB color model is that it expresses very close to the form we grasp colors using alpha and gamma receptors in our retinal. It is the basic and crucial color model used in any other medium that projects the color or television. And widely used for Web graphics, but it cannot be used for print production.

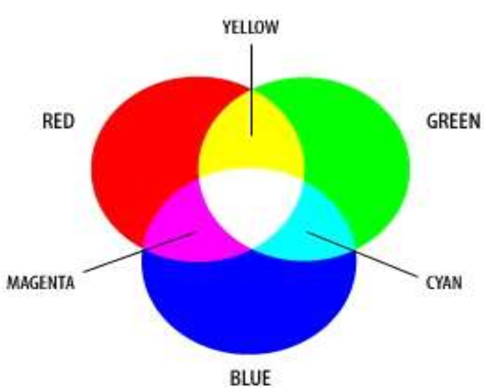


Figure 1. RGB color model

B. *CMYK Color Model*

It is the model representing the image in secondary color namely Cyan, Magenta, and yellow. When a surface coated with cyan pigment is illuminated with white light, no red is reflected from the surface. The RGB to CMY conversion is performed with a simple operation.

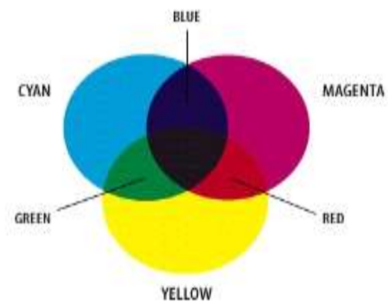


Figure 2. CMYK color model

C. *HSI Color Model*

The RGB and CMY color model systems are best suited for the hardware implementation. Unfortunately both the color model not suits for describing colors for human interpretation. When human view a color object, it can be described by its Hue, Saturation, and intensity. The HSI model also has the advantage that it decouples the color and gray scale information in an image, making it suitable for many of the gray scale techniques developed yet.

IV. COMPRESSION STRATEGY

The data compression refers to the process of reducing the amount of data required to represent a given quantity of information. The two types of compression strategy are as follows.

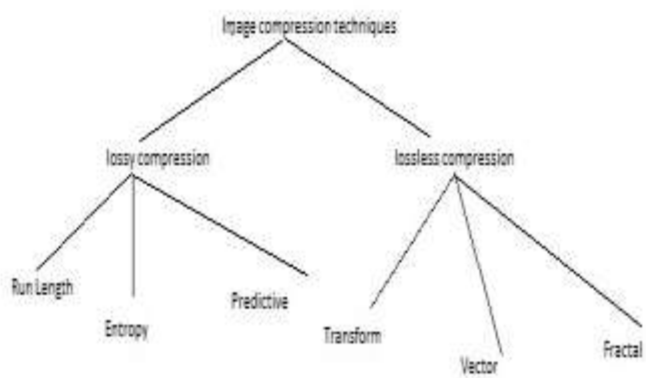


Figure 3. Compression techniques

A. *Lossy vs Lossless*

The lossless technique is meant for restoring the data files similar to the original one. In certain cases it becomes very necessary viz. processing for word file, executable processing codes, tabular formats records, etc. [12]. In such cases, loss of a single bit of information can utter the intent of the original contents. The same doesn't apply to the file could be stored and transmitted without having entire detail intact. Mostly, real-world data contains noise and removing them hardly impact the data and information contained. The technique which brings alteration to the original content during compression is termed as lossy. Lossy compression is distinct from lossless compression in terms of effectiveness, the former brings more efficient. High compression ratio is a clear indication of added noise to the data.

Lossless	Lossy
run-length	CS&Q
Huffman	JPEG
delta	MPEG
LZW	

a. Lossless or Lossy

Method	Group size: input	output
CS&Q	fixed	fixed
Huffman	fixed	variable
Aritlmetic	variable	variable
run-length, LZW	variable	fixed

b. Fixed or variable group size

Figure 4. Compression techniques[15]

V. Redundancy in the Data

There are mainly three types of redundancies that can be identified while dealing with the two-dimensional intensity arrays namely Coding redundancy, Spatial and temporal redundancy and irrelevant information. Through the different compression techniques, the above redundancies are either eliminated or reduced. There are two algorithm which deals with coding redundancy

A. Huffman Coding

It is the quite good and most prominent technique to eliminate coding redundancy [5]. The smallest possible number of code symbols is produced by this method per source symbol. There is a no of steps followed for the reduction symbols to achieve the smallest code. In the first step of this approach, a series of source reductions are created by ordering the probabilities of the symbols. Further combining the lowest probability symbol into a single symbol takes place that replaces them in the next source reduction. In the second step to code each reduced source, starting with the smallest source and working back to the original source. This procedure generates the optimal code for a set of symbols and probabilities subject to the constraint the symbols be coded one at a time.

B. LZW Coding

It is a dictionary-based approach that follows these steps.

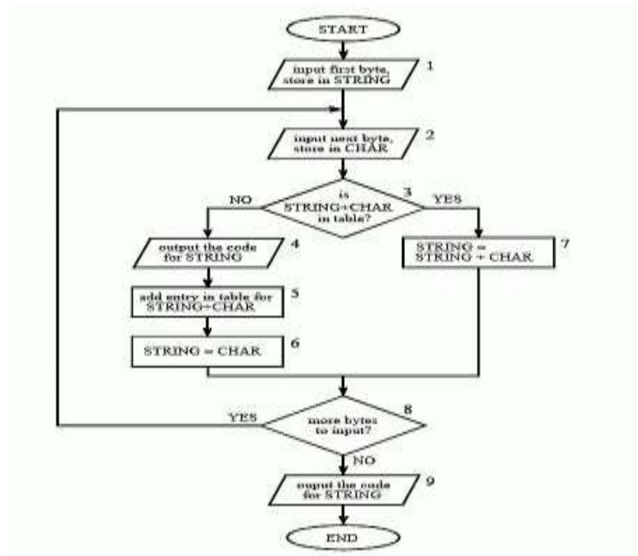


Figure 5. LZW Compression technique[15]

VI. Image Enhancement

Image Enhancement is the process of improving the image quality for a specific application. The process of enhancement is divided into two broad categories, namely spatial domain method and frequency domain method. Spatial domain method works directly on pixels whereas in frequency domain the image is first converted into Fourier transform. There and techniques of enhancement through which we get a more enhanced image respective of previous one like histogram processing, histogram equalization, histogram matching, bit plane slicing, contrast stretching etc [6][7]. Another most important tool for image enhancement is filtering.

A. Filtering

A filter is a device or a material that is used for suppressing or minimizing waves or oscillations of certain frequencies. It is used to remove different types of noises [8]. Through Convolution and Correlation, we perform filtering operation. For filtering, we have to generate a mask and use convolution or correlation to find the sum of product at each point. There are various types of filters e.g. low pass filter or high pass filter and they are further subdivided as the Gaussian filter, Laplacian filter, Butterworth filter etc.

VII. Issues and Challenges

Challenges related to compression: Now a day, we are dealing with a voluminous amount of storage that needed to occupy less bandwidth during efficient transmission. Data warehousing cost are drastically reduced by compressing the files because more files can be stored in available storage space when they are compressed. Transmission time also reduces as compared to previous. There is a lot of algorithms for image compression; still, there is a need for such compression techniques that cause less degradation of an image during compression. In terms of transmission time, error rate and many other concerns there is a vast scope of compression and needed to look into it.

VIII. CONCLUSION

This paper gives a brief review of the different color models. In terms of efficient data transmission, it is highly desirable to occupy less bandwidth during transmission so some types of compression strategies also discussed that are generally used for effective transmission of data. Usage of lossy and lossless methods is briefly analyzed. The paper tells about the enhancement of images at some instant

REFERENCES

[1] Y. Zhang, L.Y. Zhang, J. Zhou, L. Liu, F. Chen, X. He, "A review of compressive sensing in information security field", IEEE Access. 4 (2016) 2507–2519.

[2] N. Zhou, A. Zhang, F. Zheng, L. Gong, "Novel image compression–encryption hybrid algorithm based on key-controlled measurement matrix in compressive sensing", Opt. Laser Technol. 62 (2014) 152–160.

[3] N. Zhou, A. Zhang, J. Wu, D. Pei, Y. Yang, "Novel hybrid image compression— encryption algorithm based on compressive sensing", Optik 125 (2014) 5075–5080.

[4] Kim, Seyun, and Nam Ik Cho. "Hierarchical prediction and context adaptive coding for lossless color image compression", *IEEE Transactions on image processing* 23.1 (2014): 445-449.

[5] R. Huang, K.H. Rhee, S. Uchida, "A parallel image encryption method based on compressive sensing", Multimedia. Tools Appl. 72 (1) (2012) 71–93.

[6] N. Zhou, H. Li, D. Wang, S. Pan, Z. Zhou, "Image compression and encryption scheme based on 2D compressive sensing and fractional", Mellin transform, Opt. Commun. 343 (2015) 10–21.

[7] Alfalou, C. Brosseau, "Optical image compression and encryption methods", Adv. Opt. Photonics 1 (3) (2009) 589.

[8] Alfalou, A. Mansour, M. Elbouz, C. Brosseau, "Optical Compression scheme to multiplex and simultaneously encode images", in: Optical and Digital Image Processing Fundamentals and Applications, Wiley, New York, 2011, p. 463.

[9] N. Zhou, S. Pan, S. Cheng, Z. Zhou, "Image compression–encryption scheme based on hyper-chaotic system and 2D compressive sensing", Opt. Laser Tech- nol. 82 (2016) 121–133.

[10] C. Wang, J. Ni, Q. Huang, "A new encryption-then-compression algorithm using the rate-distortion optimization", Signal Process., Image Commun. 39 (2014) 141–150

[11] Pearlman, William A., et al. "Efficient, low-complexity image coding with a set-partitioning embedded block coder", *IEEE transactions on circuits and systems for video technology* 14.11 (2004): 1219-1235..

[12] Ahumada Jr, Albert J and Peterson, Heidi A. "Luminance-model-based DCT quantization for color image compression'', SPIE/IS\&T 1992 Symposium on Electronic Imaging: Science and Technology, 365—374,1992

[13] Nadenau, Marcus J and Reichel, Julien and Kunt, Murat,"Wavelet-based color image compression: exploiting the contrast sensitivity function", IEEE Transactions on Image Processing, 58—70, 2003

[14] Sathappan, S., and P. Suresh Babu. "Block-based prediction with Modified Hierarchical Prediction image coding scheme for Lossless color image compression", *Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on*. IEEE, 2016.

[15] Rafeal C. Gonzalez, Richard E. Woods "Digital Image Processing"

International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 6, Issue 4, July- August 2017 **ISSN 2278-6856**

Video Processing and its Applications: A survey

Neetish Kumar[1], Dr. Deepa Raj[2]

[1]Department of Computer Science, BBA University, Lucknow, India

[2]Department of Computer Science, BBA University, Lucknow, India

Abstract

A video is considered as 3D/4D spatiotemporal intensity pattern, i.e. a spatial intensity pattern that changes with time. In other word, video is termed as image sequence, represented by a time sequence of still images. Digital video is an illustration of moving visual images in the form of encoded digital data. Digital video standards are required for exchanging of digital video among different products, devices and applications. Fundamental consumer applications for digital video comprise digital TV broadcasts, video playback from DVD, digital cinema, as well as videoconferencing and video streaming over the Internet. In this paper, various literatures related to the video processing are exploited and analysis of the various aspects like compression, enhancement of video are performed. There are many issues and challenges still exist related to the video processing inclusive of security issues, and those are briefly discussed in this paper.

Keywords: Video processing, compression, recognition, security

1. INTRODUCTION

The first development of video technology is for mechanical television systems that were soon replaced by cathode ray television systems (CRT). After the advancement of new display devices, the technological aspect of video processing gets broader. With the emergence of new media, mobile media and social media, micro video as a new form of network video, the transmission case came into being. Transmission of videos in different network mode became very popular nowadays. For transmission purpose, it is desirable to occupy less bandwidth of the medium and less time for transmission. Video communication using video phone, video streaming, and video broadcasting has become more and more attractive with the fast pace of development in the wireless industry in present era. In order to transmit such rich multimedia content, it is desirable to occupy less bandwidth, the size of the original video signal must be reduced by some compression technique, without degrading video quality or data loss. To get rid of such hurdles, video compression techniques provide efficient solutions to represent video data in a more compact and robust manner so that the storage and transmission issues of video can be realized in cost effective way. Digital video has become a necessary part of everyday life. A part from compression it is well known that video enhancement as an alive topic in video processing has received much concentration in recent time. The aim behind it, is to improve the visual appearance of the video, or to provide a better transform representation. Accomplishing video enhancement understands the cause of low quality video. The common cause of the degradation is a challenging problem because of the various reasons like low contrast, signal to noise ratio is usually very low etc. The application of image processing and video processing techniques is to the analyze the video sequences in traffic flow, traffic data collection and road traffic monitoring. Various methods are present including the inductive loop, the sonar and microwave detectors has their own Pros and Cons. Video sensors are relatively cost effective installation with little traffic interruption during maintenance. Furthermore, they administer wide area monitoring granting analysis of traffic flows and turning movements, speed measurement, multiple point vehicle counts, vehicle classification and highway state assessment [5].

A video signal is a sequence of 2D images, captured onto the image plane of a video camera from a dynamic 3D scene. Basically in color scene, the color value obtained at any point in a video frame records is the reflected light at a particular 3D point in the observed scene. It is well known that light abide of an electromagnetic having the wavelength range between 380nm to 780nm in which the human eye is sensitive. The grasped color of light lean on its spectral content (i.e. its wavelength composition) e.g. “Red” light has its energy concentrated near 700nm while “White” appears if the light has equal energy across the entire visible band. In general, “Spectral color” is nothing but light having very narrow bandwidth. On the other hand, “Achromatic” light is synonymous to the white light [12]. Illuminating sources radiate an electromagnetic wave including the sunlight, TV monitors, bulb & so on [2]. The sensed color of an illuminating light source lean on the wavelength range in which the energy is emitted. Additive rule is followed by the Illuminating: i.e., the sensed color of several mixed illuminating light sources depends on the sum of the spectra of all the light sources.

2. LITERATURE SURVEY

V. Kastrinaki et al. suggested the applications of video sensors in traffic management and monitoring with the help of video processing and analysis methods. They recommended an overview of image processing and analysis tools used in these applications and relate these tools with complete systems for specific traffic applications. It is based on the automatic lane finding methodology. There are basically three solutions for their three purposes, first, to classify image-processing methods used in traffic applications. Second, to provide the pros and cons of these algorithms. Third, from this unified consideration, struggling towards evaluation of shortcomings and general needs in this field of active research [5]. Alain Traimu et al. worked on color image and video processing. Multiple area of present scopes like color vision, perception, and interpretation, acquisition systems, consumer imaging applications, and medical imaging applications, issues, controversies, and problems of color image science have analyzed by them. The science of color imaging may be defined as the study of color images and the application of scientific methods to their measurement, generation, analysis, and representation. It includes all types of image processing, including optical image production, sensing, digitalization, electronic protection, encoding, processing, and transmission over communications channels. They presented an extensive overview of the up-to-date techniques for color image analysis and processing [6].

Yunbo Rao and Leiting Chen [1] described about the Video enhancement. According to them video enhancement is the most important and difficult components in video research. The motive behind the video enhancement is to improve the visual appearance of the video, or to provide a better transform representation compared to earlier for future automated video processing, such as analysis, recognition, detection, segmentation,, surveillance, automatic traffic control systems, criminal justice systems. Brief survey of the existing techniques of video enhancement, which can be classified into two broad categories: (i) Self-enhancement and (ii) Frame-based fusion enhancement have performed by them. Finally they are able to describe the advantages and disadvantages of algorithms of image/video enhancement. They have described recent developments, methods of video enhancement and point out promising directions on research for video enhancement for future research. S. Ponlatha and R. S. Sabeenian compared the different video standard and concluded that there is a constant improvement in video compression factors, new techniques and technology, and some new formats in the horizon of H.265 and VP8: H.265 compression is still in the process of being formulation, and motive to gain a 25% improvement in the compression factor while lowering computational overhead by 50%: for the same perceived video quality. VP8 is a codec from On2 Technologies, which claims that the codec brings bandwidth savings and uses less data than H.264: to the extent of 40%. There is currently a fight over the standard to be chosen for Web video, and VP8 is bashing it out with H.264 [4]. W. Puech et al. briefly described about the secure signal processing and the recent cryptographic challenges in video processing. they has given an overview of approaches and challenges that exist in applying cryptographic primitives to important image and video processing problems, including content encryption, secure face recognition, and secure biometrics. basically their work aims to help the community in appreciating the utility and challenges of cryptographic techniques in image and video processing [15-16].

3. VIDEO COMPRESSION STRATEGIES

The straightforward form of video compression is to perform two dimensional images coding separately on each frame of a video temporal sequences. The JPEG baseline standard has widely adopted as a means of video compression. In this application, the video compression technique has become to be known as motion JPEG [3-5].

3.1 MPEG-1 video coding standard

The MPEG-1 video coding standard is universal in terms of functionality. The MPEG-1 video algorithm has been refined with respect to the JPEG and H.261 activities. Crucial features equipped by MPEG-1 constitute frame based random access of video, fast forward/fast reverse (FF/FR) searches through compressed bit streams, reverse playback of video and edit ability of the compressed bit stream. The basic MPEG-1 video compression technique is based on a Macroblock structure, motion compensation and the conditional replenishment of Macroblocks. The MPEG-1 algorithm processes the frames of a video sequence by block based methodology [7-8].

3.2 MPEG-2 video coding standard

MPEG goes on its typical attempt with a second phase (MPEG-2) in 1991, to provide a video coding solution for applications, which was originally lacking in the MPEG-1 standard. Specifically, MPEG-2 was given the charter to implement video quality not lower than NTSC/PAL and up to CCIR 601 quality. Digital cable TV distribution, networked database services via ATM, digital VTR applications and satellite and terrestrial digital broadcasting distribution were seen to benefit from the increased quality expected to result from the new MPEG-2 standardization phase.

3.3 MPEG-4 video coding standard

MPEG-4 Systems provides the technologies to interactively and concurrently represent and conveyed audio-visual contents comprises of various objects including audio, visual, 2D/3D vector graphics and etc. MPEG-4 inherits many of the features of MPEG-1 and MPEG-2.

3.4 H.264/AVC

The ongoing H.264/AVC compression standard is based on the picture wise processing and waveform based coding of video signals [13]. The technology now being expressed for the new standardization project on high efficiency video coding (HEVC) is a generalization of this approach which grants significant gains through modernization such as improved intra prediction, larger block sizes, more flexible ways of decomposing blocks for inter and intra coding and better exploitation of long-term correlations and picture dependencies[9][13]. It will support a wide range of encoder modes, which are typically optimized using mean-squared-error-based or related distortion measures.

4. VIDEO GENERATION TOOLS

There are various video making tools are available in present. Some of these are as follows e.g Nutshell, Videoshop, iMovie App, iMovie for Macs, Magisto, Animoto. All above mentioned tools works on different platform but the basic principle of all are same. Animito takes different camera clicked pictures in its primary memory and processed for adding filters to and embellishment for making the video. It has also the functionality to adding text and sound tracks during the processing of frames as desirable. Animoto offers advanced graphics that exceed either of the previous technologies developed. Like Kizoa, Animoto is cloud based, offering users the ability to create online slideshows that can be linked by URL or embed, but it also allows for a low resolution download in mp4 format. Animoto's main marketing message is its ease of use: users upload photos, add captions, select music, and the program will make its own decisions about effects and transitions. The resulting presentation has glossy, advanced graphics such as fluttering butterflies that look real.

5. ISSUES AND CHALLENGES

5.1 Challenges related to compression: As in modern trend we are dealing with tremendous amount of storage and need to transmit voluminous data effectively. Compression performance can be enhanced by employing motion-compensated prediction, which predicts each frame block wise from the previous frame. The prediction error can be more effectively compressed than the original frame data. In video conferencing application we somehow, lagging to produce the satisfactory result in terms of bit rate representation and still a big challenge to deal with.

5.2 Challenges related to enhancement:
In enhancement, one aims to process image to improve their quality. An image may be of poor quality because of its contrast is low, or it is noisy, or it is blurred etc. Many algorithms are proposed to remove the degradation but not any algorithm yet developed that doesn't hurt the signal. For ex. Noise reduction algorithms typically involve local averaging or smoothing which unfortunately blur the sides of the frames. It needs to look forward in this area.

5.3 Challenges related to recognition: An effective recognition system is not yet developed to detect the generic object lying or either a person having and behaves according if found suspicious. The area has vast scope to deal with.

5.4 Challenges related to security: Secure signal processing is an emerging technology to legitimate image/video processing duties in a secure and privacy preserving fashion. It has brought a great amount of research attention due to the boost demand enabling prosperous functionalities for individuals data reserved online. But treating multimedia as ordinary data and applying cryptographic ciphers such as RSA and AES, information leak is minimized so such an approach is inefficient for practical video processing applications. Cryptographic operations are often computationally expensive. Efficient usage of cryptographic protocols is therefore imperative. Another challenge may occur due to the certain cryptographic techniques that cause cipher text expansion of two orders of magnitude, such as public key encryption of image pixels. Success in this direction will depend on future research efforts to address the question of how to properly define security and the proper level of protection for multimedia data.

REFERENCES

[1]. Rao, Yunbo and Chen, Leiting, "A survey of video enhancement techniques", Journal of Information Hiding and Multimedia Signal Processing. 71--99, (2012).
[2]. Ranganathan, Parthasarathy and Adve, Sarita and Jouppi, Norman P"Performance of image and video processing with general-purpose processors and media ISA extensions" IEEE Computer Society. 124--135 (1999).
[3]. Singh, Lalit and Kaushik, Sandeep and others "Video Compression Technique", International journal of scientific \& technology research, 114--119, 2012.
[4]. Ponlatha, S and Sabeenian, RS "Comparison of video compression standards", IACSIT Press (2013).
[5]. Kastrinaki, V and Zervakis, Michalis and Kalaitzakis, Kostas, "A survey of video processing techniques for traffic applications" Elsevier. (2003).
[6]. Alain and Tominaga, Shoji and Plataniotis, Konstantinos N, "Color in image and video processing: most recent trends and future research directions", Journal on Image and Video Processing. (2008).
[7]. Girod, Bernd and Younes, K Ben and Bernstein, Reinhard and Eisert, Peter and Farber, Niko and Hartung, Frank and Horn, Uwe and Steinbach, Eckehard and Stuhlmuller, Klaus and Wiegand, Thomas, "Recent advances in video compression", IEEE 580-583 (1996).
[8]. Mantiuk, Efremov, Alexander and Myszkowski, Karol and Seidel, Hans-Peter, "Backward compatible high dynamic range MPEG video compression", ACM 713—723 (2006).
[9]. Rosa, Vagner S and Susin, Altamiro A and Bampi, Sergio, "An HDTV H. 264 deblocking filter in FPGA with RGB video output", IEEE 308--311 (2007).
[10]. Clemens C and Steffens, Liesbeth and Verhaegh, Wim FJ and Bril, Reinder J and Hentschel, Christian, "Qos control strategies for high-quality video processing", Springer. 7-29 (2007)
[11]. Yang, Kai-Chieh and Dane, Gokce and El-Maleh, Khaled. "Temporal quality evaluation for enhancing compressed video", IEEE 14.11 1219-1235 (2007):.
[12]. Ahumada Jr, Albert J and Peterson, Heidi A. "Luminance-model-based DCT quantization for color image compression'', SPIE/IS\&T 1992 Symposium on Electronic Imaging: Science and Technology, 365—374,(1992)

[13]. Luthra, Ajay and Sullivan, Gary J and Wiegand, Thomas, Murat, “Introduction to the special issue on the H. 264/AVC video coding standard”, IEEE Transactions on Image Processing, 557--559, (2003)
[14]. Sathappan, S., and P. Suresh Babu. "Block based prediction with Modified Hierarchical Prediction image coding scheme for Lossless color image compression", Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on. IEEE, (2016).
[15]. Wenjun Lu, Avinash Varna, and Min Wu, “secure video processing: problems and challenges” IEEE ICASSP 978-1-4577-0539-7/11(2011)
[16]. W. Puech1, Z. Erkin2, M. Barni3, S. Rane4, and R. L. Lagendijk2, “emerging cryptographic challenges in image and video processing” IEEE ICIP , 978-1-4673-2533-2/12 (2012)

IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-11, Nov 2018 E-ISSN: 2347-2693

Video Generation using Still Images

Neetish Kumar [1*], Deepa Raj[2]

[1]Dept. of Computer Science, Babasaheb Bhimrao Ambedkar University (A Central Univ.), Lucknow, 226025, India
[2] Dept. of Computer Science, Babasaheb Bhimrao Ambedkar University (A Central Univ.), Lucknow, 226025, India

[*]*Corresponding Author: neetish08537@gmail.com, Tel.: +91-9006167094*

Available online at: www.ijcseonline.org

Accepted: 21/Nov/2018, Published: 30/Nov/2018

Abstract**—** A video is termed as the sequence of kinetic frames captured/ play backed into some sequences. The rapid development of video capturing systems changed the word from recording the still frames on tapes to the storing frame digitally through charge coupled devices. The Video recording system changes from past analog era to digital era which made editing and manipulation very easy. This paper explained how Matlab is used for making the video from the still images and its performance in the sense of storage and time. Different video generation tools are elaborated and compared with the Matlab generated tools.

Keywords**—** Matlab, frames, Video generation, Recording systems, CCD

I. INTRODUCTION

A video is the visual multimedia source that combines a sequence of images to form a moving picture. Any video recording systems do not actually record videos but the sequence of the still image at the desired frame rate. The human eye can't distinguish delay between the processing of still frames if it can be processed with above 16 frames per second due to the persistence of vision. It gives an illusion of moving videos. The Modulated light (such as a computer display) is perceived as stable when the rate is higher than 50 Hz through 90 Hz frequency. This perception of modulated light as steady is termed as the flicker fusion threshold. However, when the modulated light is non-uniform and contains an image, the flicker fusion threshold can be much higher, in the hundreds of Hertz. In image recognition context; it has been found a specific image in an unbroken series of different images, each of which lasts as little as 13 milliseconds is recognizable. Persistence of vision sometimes accounts for very short single-millisecond visual stimulus having a perceived duration of between 100 ms and 400 ms. The video recording systems change from past analog era to digital era very rapidly. The very first video recording system was accompanied by analog systems e.g. kinescope recording. The recordings generated by the kinescope were of poor quality and cost inefficient, so these were used by those stations that were beyond the reach of cable or microwave interconnections. In 50's era video, the consumer market is fully occupied with the audio tape recorders due to its popularity. This gives a big hint for the video recording that was logically related to the audio but the complexity of video frequency is much higher than audio that concerns problems to the video recordings.

Due to the transition from analog to the digital era, the broadcasters got a valuable tool for video recording. Previously electromechanical irregularities were ubiquitous in analog form of recording. While working with the analog signal, recordings are most vulnerable to noise, a combination of phase modulation, transport jitter, amplifier noise, improper machine line up. In the digital form of recording the above problem got a bit resolved.

The Camera evolution from very first to current scenario has is as follows
1814 – JoesephNiepce Photograph,1837 – The Daguerreotype Camera, 1867 – Zoopraxiscope System, 1895 – Lumiere Portable camera, 1908 – Lighter camera, 1950 – colored video camera, 2007– Flip video era[1]. R. Cutler et al. [1] came across with the new idea of recording techniques of the distributed meeting. H. Pun et al. [2] elaborated the structure of slow-motion replay segments and interpreted Hidden Markov Model for the structure of slow-motion sequences and proposed an inference algorithm to detect the boundaries of slow-motion replay segments. Thomas Meier and King N. Ngan [3] developed segmentation algorithm for video object plane generation that can automatically extract moving objects from a sequence. In 2015 Li Yaho et al. [4] worked with automatic video description generation using artificial intelligence and machine learning and helped to provide a solution for poorly tagged video with natural languages uploaded on YouTube. Kenneth A. et al. [5] studies are focused on high-performance digital color video camera. Canon EOS 1D-C is in working in the current scenario. The digital era begins with Kodak Digital camera [6]. A brief system overview was suggested to capture audio as well as video recording systems. Section II briefly

describes the basic architecture of video camera while Section III exposes the experimental result and analysis. Section IV is the concluding section of our proposed ideas.

II. ARCHITECTURE OF VIDEO CAMERA

The architecture of Video Camera is explained in the figure - 1 below which shows how video camera process pictures and create a video. It shows the different hardware components of a standard video camera that is capable of capturing the motion pictures. The outer perspective is captured by the lens and digitized by the charged coupled devices (CCD device). There are some SDRAMs and drivers that are attached to the playback side monitors for storing the captured sequence of frames. There are some features for setting up the desired frame rate for capturing a video by the user. Audio is captured by the separate cable line and compressed through the different audio codecs. Some software components are used in the video camera like JPEG coding and JPG decoding for image preparation, transformation quantization, and entropy coding. There are LCD and USB port, device drivers, Light flash and display device drivers interconnected that is needed to for the complete smooth capturing system [7][8].

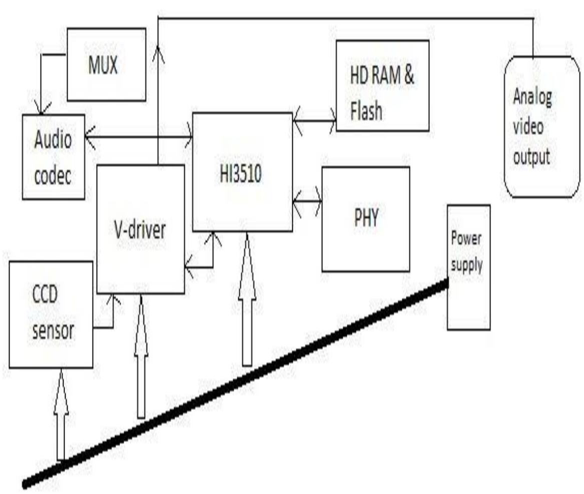


Figure. 1 Architecture of Video camera

III. EXPERIMENTAL RESULTS AND ANALYSIS

To generate a video from still images, following steps are followed in MATLAB environment. At first, the n numbers of still images are taken and desired memory space are allocated by using cell () function. All still images are read by imread function from the hard disk or from folders. The images taken are may be of different sizes so it is needed to resize all in similar size as 256*256. The reason behind it is to make the process computationally efficient. Then all images are converted into frames by im2frame function. After that, the decision of frame rate is decided according to one's need. With the function writeVideo (), frames are stored into a video file and implay() function is used to execute the video that gives human eye an illusion of motion pictures. Pictorial representation is depicted in Figure -2.

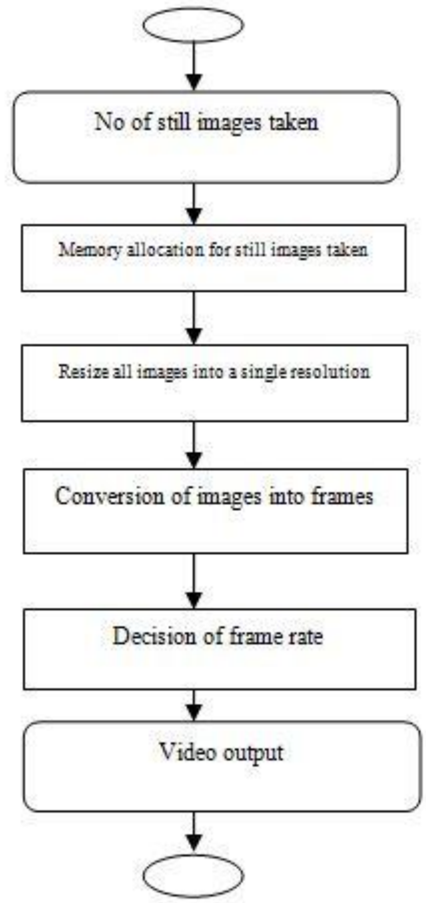


Figure. 2 Flow diagram of generation of Video from still images

Two video generation tools Movie maker and Animoto have been taken for performance measurement in terms of time and space in comparison with Matlab generated software. four test video is taken as Baby_video11, Baby_video22, Baby_video_grey11 and Baby_video_grey22. Still, images are depicted for Baby_video11 in Figure-3

Figure-3 Still images of Baby_Video11

For time consideration time elapsed for creating a video Baby_video11 having 11 frames, it takes 2.649 seconds without any additional effects. The size of the video generated is 591 KB. At this instant, it can be said that MATLAB provides more flexibility in terms of desired frame rate, so the video sequence at the playback side met guaranties of viewers need. For this video, Movie maker takes 11 seconds to execute it. The output video file is shown at the bottom rightmost corner of the above figure 3.

Figure-4 Still images of Baby_Video22

Another video Baby_video22, it contains 22 frames, still, images are depicted in figure-4. MATLAB takes 3.584 seconds to execute it. The size taken of the video is 1.03 Mb while movie maker software takes 22 seconds to execute it. The output video file is shown at the bottom rightmost corner of the above figure 4.

Figure-5 Still images of Baby_video_grey11

The grey video Baby_video_grey11 that have 11 frames takes slightly less time with respect to the color frames processed. 11 frames of the grey image video take 1.124 sec and size acquired is 494 KB. The grey image of another video Baby_video_grey22, having 22 frames takes 2.982 sec and occupies 923 KB of space. The size of the video generated by the grey image sequence is also less than the video generated by the color images. Figure-5 and figure-6 show the still images of baby_video_grey11 and Baby_video_grey22. The output video file with 11 frames and 22 frames is shown at the bottom rightmost corner of the above figure 5 and figure 6 respectively.

Figure-5 Still images of Baby_video_grey11

The moviemaker tool takes the total no of time equal to the no of frames processing i.e 22 seconds. Another software of video generation tool Animoto, for generating the video baby_movie11 takes a 31 second for the completion with the inclusion of special audio and video effects. Without Effect of video and audio, it is not working. After this study, we can say that Matlab is a good plate form for generating any video or still images of any sizes and any quantity very fast and space complexity is also less.Table-1 one shows the performance analysis of the Matlab generated tools with Animoto and movie maker and figure -3 shows the pictorial representation of their performance.

Table 1 Comparison analysis

Test videos	MATLAB software	Moviemaker	No of frames	Size
Baby_video11	2.649 sec	11 sec	11	591 KB
Baby_video22	3.584 sec	22 sec	22	1.03 MB
Baby_video_grey11	1.124 sec	11 sec	11	494 KB
Baby_video_grey22	2.982 sec	22 sec	22	889 KB

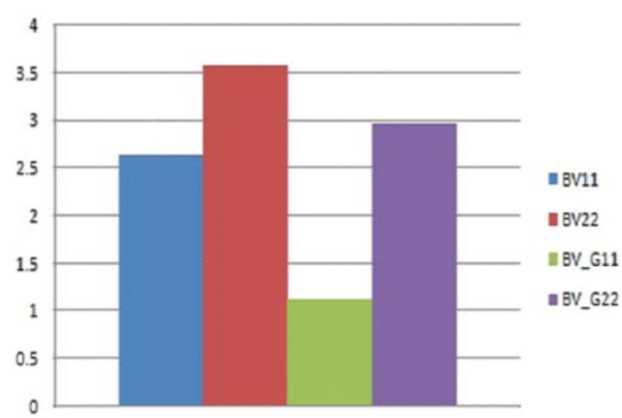


Figure-5 Total elapsed time for each video

IV. CONCLUSION

Matlab is a good platform for working on digital image processing application like compression, enhancement, and detection of an object in an image, segmentation and pattern recognition. From the above experimental study, it is concluded that Matlab is also very suitable for working in video processing application. Easily images can be read from file or folder and the frame is also generated by calling simple function im2frame().Video file is generated and all the frames are stored in the video file and execution of the video takes less time and acquire a less space in the memory compare with Animoto and movie maker.

REFERENCES

[1] R. Cutler, "The distributed meetings system", Acoustics, Speech, and Signal Processing, Proceedings (ICASSP '03) on IEEE International Conference on signal processing, 2003, pp. IV-756-9 vol.4.

[2] H. Pan, P. van Beek and M. I. Sezan, "Detection of slow-motion replay segments in sports video for highlights generation" ,2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings (Cat. No.01CH37221), Salt Lake City, UT, 2001, pp. 1649-1652 vol.3.

[3] T. Meier and K. N. Ngan, "Automatic segmentation of moving objects for video object plane generation", in IEEE Transactions on Circuits and Systems for Video Technology, vol. 8, no. 5, pp. 525-538, Sep 1998.

[4] LiYaho, AtousaTorabi , "Video Description Generation Incorporating Spatio- Temporal Features and a Soft-Attention Mechanism",Proceedings of the IEEE international conference on computer vision, 4507-4515, 2015

[5] Kenneth A. Paruiski, Lionel J. D'Luna. Brian L. Benamati,"High-performance digital color video camera", Journal of electronic imaging, 1992

[6] SamanNaderiparizi, Pengyu Zhang, MatthaiPhilipose,"Glimpse: A Programmable Early-Discard Camera Architecture for Continuous Mobile Vision", MobiSys, 2017, NY, USA

[7] Neetish Kumar, Dr Deepa Raj,"Video Processing and its Applications: A survey" International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), ISSN 2278-6856.

[8] Neetish Kumar, Dr Deepa Raj,"A Study and Analysis of Images in Different Color Models", International Journal Of Advanced Studies In Computer Science And Engineering IJASCSE Volume 7, Issue 1, 2018.

[9] Stephen J. Solari'Architecture of a digital video recorder'2000

[10] Luming Zhang, Peiguang Jing, Yuting Su, "SnapVideo: Personalized Video Generation for a Sightseeing Trip", IEEE transactions on cybernetics, 2016

Authors Profile

Neetish kumar, currently a research scholar of department of computer science of Babasaheb Bhim Rao Ambedkar University, Lucknow.

Dr. Deepa Raj, working as an Assistant Professor in the Department of Computer Science Babasaheb Bhim Rao Ambedkar University. She did her Post Graduation from J.K Institute of applied physics and technology, Allahabad University and Ph.D. from Babasaheb Bhim Rao Ambedkar University Lucknow in the field of software engineering. Her field of interest is Software Engineering, Computer Graphics, and Image processing. She has attended lots of National and International conference and numbers of research papers published in her field of interest.

A New Pentagon Search Algorithm for Fast Block-Matching Motion Estimation

Neetish Kumar, Deepa Raj

Abstract*: The impact of search pattern is a crucial part in the block-based motion estimation for finding the motion vector. An issue of distortion performance and search speed heavily depends upon the size and shape of search strategy applied. Performing the deep analysis for motion vector distribution on standard test videos, it is desirable to have such type of algorithm that meets the requirement of searching motion vector in less time. Hence, a new kind of Pentagon algorithm is proposed in this paper for fast block-matching motion estimation (BMME). It is an easy and efficient technique for finding motion vector. Experimental results expose the proposed Pentagon algorithm sparsely surpass the noted Diamond search (DS) algorithm. The new Pentagon search algorithm is examined with the previously proposed Diamond search algorithm in terms of performance measure; the proposed algorithm attains better performance with the less complexity. The experimental examination also depicts that the pentagon algorithm is better than the previously proposed Diamond search (DS) in terms of mean-square error performance and required the number of search points. The overall speed improvement rate (SIR) is about 31% with respect to the DS.

***Keywords* : Diamond Search, motion estimation, PSNR, search points .**

I. INTRODUCTION

Digital video is essentially a sequence of pictures displayed at the desired frame rate that makes an illusion of motion pictures for viewers. In those sequences of pictures, there are strong redundancies of different types e.g. spatial, temporal, perceptual and statistical. The compactness among the neighboring pixels is termed as spatial redundancy. There is a vast correlation among the neighboring pixel that needed to be removed for achieving the better compression ratio. On the other hand, there are also strong temporal redundancies between the frames in a video sequence [1][2]. Temporal redundancy is resolved by block matching motion estimation techniques in successive video frames. In similar way perceptual redundancy can be handled as it cites the detail of a scene that the human eye unable to perceive accurately [3][4]. It can be abandoned with no loss in quality of a picture. Block matching motion estimation techniques play a vital role in motion compensated video coding technique that is used in MPEG and H.26x series [9]. The idea behind this technique is to convert a frame in a video sequence into no of non-overlapping blocks and there is a motive to find the particular block in the successive reference block. Mean of absolute error/difference is mostly used as a Matching measure [13][15]. The mechanism behind finding the motion vector is to identify a candidate block in the current frame as it is in the reference frame. At first, the block is searched in a specified window size of (2W+1)*(2W+1) as shown in Figure 2.

Where W is the pixel value both in horizontal as well as vertical direction. Moving the candidate block in that window the best match block position is the motion vector as shown in Figure 1.

The reason behind the literature goes from heuristic search to fast search is the principle of local minima. The major drawback of heuristic search is that it is a time-consuming process. The candidate block is searched over the entire window [6]. Hence the search operation takes too much time for finding the motion vector. Although it is a time-consuming process but it guarantees the true motion vector that gives the optimal solution whereas through the fast search method we have to satisfy with only the suboptimal solution. As most real-world application many computationally efficient variants were developed, among which are typically the three-step search (TSS), New three-step search (NTSS) etc. NTSS is another variant of TSS that not only find the coarse motion vector but also the finer one. Centre biased property of motion vector leads to the introduction of NTSS.

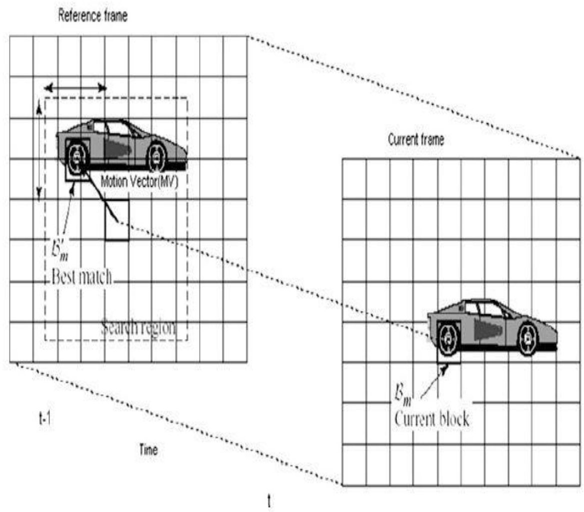


Fig. 1.**Motion vector between the two successive frames**

The various algorithms like Diamond search (DS)[12], Block-based gradient descent search (BBGDS) as well as Four-step search and [14] are developed to reduce the computational complexity less than $O(n^2)$ [16]. Searching pattern and the associated shape and size of various motion estimation algorithms varies.

Revised Manuscript Received on August 14, 2019.

***** Correspondence Author

Neetish Kumar*, Department of Computer Science, BBA University (A Central University), Lucknow, India. Email: neetish08537@gmail.com

Deepa Raj, Department of Computer Science, BBA University (A Central University), Lucknow, India. Email: deepa_raj200@yahoo.co.in

Some algorithms follow like square-shaped search patterns of different size (e.g. TSS, 4SS) while DS algorithm maintains diamond-shaped search structure and hexagonal search probes the six neighboring points variation.

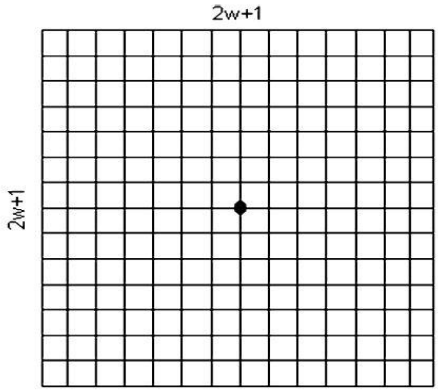


Fig. 2. A search window size of (2W+1)

This paper is organised as follows Section 2 deals with the various types of searching method for finding the motion vector. Section 3 elaborated on the algorithm proposed, followed by result and discussion in section 4. Section 5, which has the conclusion section that has its concluding remark and future work.

II. DIFFERENT SEARCH TECHNIQUE FOR FINDING THE MOTION VECTOR

The development of search pattern for estimation the motion vectors takes about tens of years. In the medical field, Tsai et al proposed a scheme for angiogram video sequence in 1994 that was fully relied upon a full frame DWT. To achieve high compression ratio this method is applied for using the frame difference method. Gibson et al. suggested about the possibilities for eliminating the artefacts from final image by adaptively examining prediction error and adjusting it appropriately for shrinking of digital angiogram videos that leads to the introduction of wavelet-based approach of lossy character. Later fast algorithms for searching are developed to evaluate motion estimation on real-time video [17]. Block searching techniques for the different type of I, P and B frames are developed. Some of the popular and efficient techniques are as follows.

A. Hexagon algorithm

Author This approach was introduced in 2002. The number of search points gets reduced after applying this algorithm compared to previously introduced algorithms, and hence results in declining the computational cost. HEXBS approach, summarized steps are follows:-

1) In The big hexagon with seven checking points is centred at (0,0) and all the seven points are tested for the block having minimum distortion. Two possibilities may arise in such case:

✓ If Minimum Block Difference (MBD) point is found to be at the centre of the hexagon, proceed to (3) (Ending);

✓ Otherwise, proceed to (2) (Searching).

2) the new MBD point chosen as centre and a new hexagon is formed. Again with this newly formed hexagon MBD point is identified by checking at all seven places. If it remain the point of newly hexagon at centre, terminates (3); otherwise, repeat this step continuously.

(3) the large hexagon is switched into a small hexagon. The four new points associated with it is compare to the present MBD point. Newly initialized MBD point leads to end stage of motion vector.

Algorithm can be explained as following on above discussed methodology. Firstly all the seven-point of the hexagon is searched if minimum distortion occurs at all six points except the centre then considering that point as centre the process is repeated [8]. If the MAD is found at the centre then the shape of the hexagon is reduced and then four extra points needed to be searched.

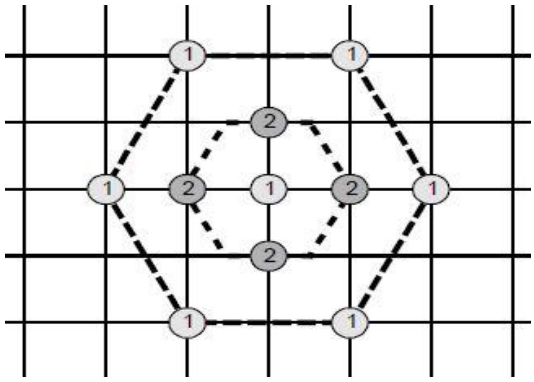


Fig. 3. Hexagon-based search (HEXBS): large (1) and small (2) hexagonal patterns

B. Diamond search algorithm

The Diamond Search algorithm illustrated as:

Step 1) let centre of LDSP assumed as origin, where we've 9 points to be checked on origin of search window. if the MBD points found at centre, then moved to (3) else proceed for (2).

Step 2) the new LDSP is created with the previous MBD point found, assuming the MBD as centre. According to previously described step s newly MBD point should be on found on centre, if it's not then moved to (3), else recursively repeat this step.

Step 3) LDSP is converted into SDSP (Small Diamond Search Pattern). The four-point of SDSP is tested. The new MBD point is the final solution of the motion vector i.e. best matching block.

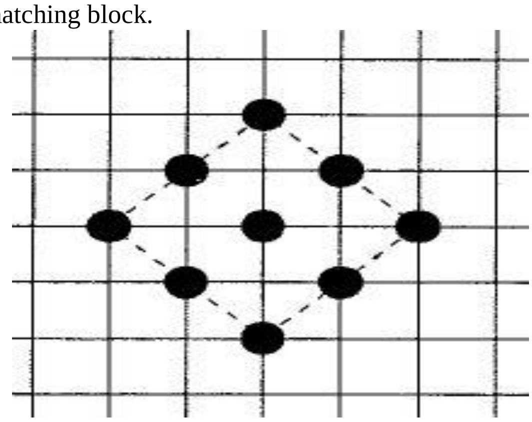


Fig. 4. Large Diamond Search Pattern

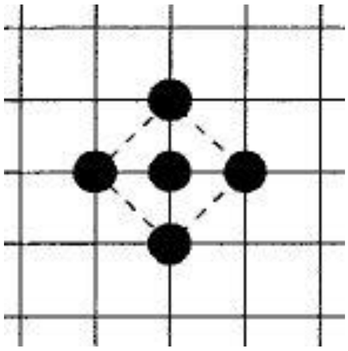


Fig. 5. Small Diamond Search Pattern

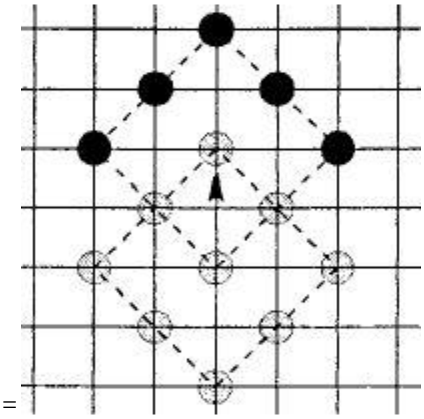


Fig. 6. Corner point

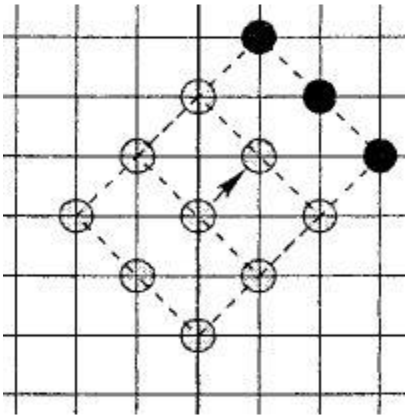


Fig. 7. Edge point

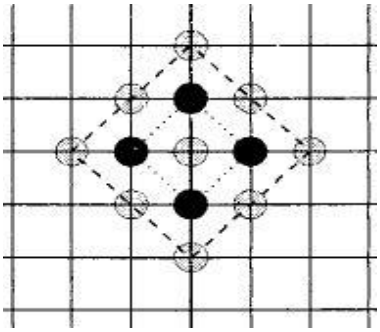


Fig. 8. Center point

In the above algorithm if the minimum found at the corner position of the LDSP then 5 extra points need to be checked and if the minimum found at the edges then 3 extra points needed to be checked [5] .

When the minimum found at the centre position than 4 extra points need to be checked and LDSP becomes SDSP [11][7].

C. Three step search

Koga et. al. described above mentioned algorithm in 1981. The performance evaluation is near optimal of this algorithm due to searching strategy of the best motion vectors detection in a coarse to the fine search pattern [12]. The algorithm steps are as follows:

Step 1) An initial step size is chosen. Eight blocks at a distance of step size from the centre (around the centre block) are picked for comparison.

Step 2) The step size is then reduced to halved. The centre is assumed to be at point with the minimum distortion. Both of the above steps are repeated till the step size becomes smaller than 1. A particular path for the convergence of this algorithm is shown below:

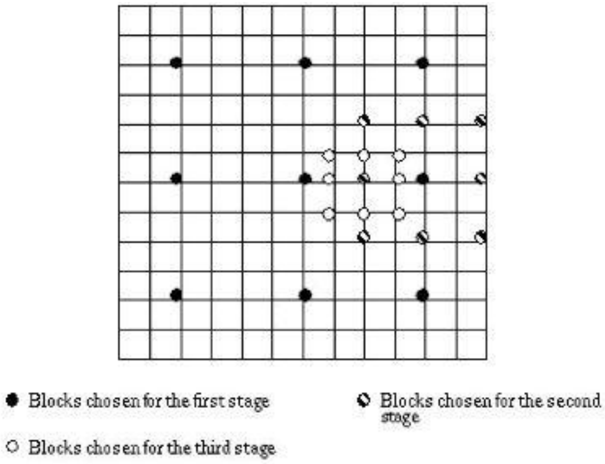


Fig. 9. TSS Algorithm steps

III. PROPOSED ALGORITHM

Based on the above observation a new fast pentagon algorithm is developed. The Figure for the Pentagon algorithm is shown in Figure 1. From the center point of the Pentagon all the distance is equal of 2 units except the lower corner ends that are $\sqrt{5}$. The Pentagon algorithm steps are as follows:

Step 1: The large pentagon with six checking points centered at (0,0), is tested for minimum block distortion with all six points. The MBD is tested and if, identified at the center of the Pentagon, proceed 3. (Ending); otherwise, go to step 2) (searching)

Step 2: Now a new large pentagon is created by assuming the MBD point of the previous search as the center. All new candidate points are checked and the MBD point is again identified. If the MBD point is still at the center point of a newly formed pentagon, then proceed to step 3) (ending); otherwise perform this continuous step repeatedly.

Step 3: Reduce the size of a search pattern from the large pentagon to four nearest neighbour points. The new four-point in the small pentagon tested for comparison with the present MBD point. The present MBD point termed as final solution for the motion vector.

In general term, the search strategy can be expressed as the following expression.

Npent (mx, my) = 6+3*n + 4
Where n = no of iteration the newer point exist
mx, my = motion vector position coordinates
Npent = total no of search point through proposed pentagon method

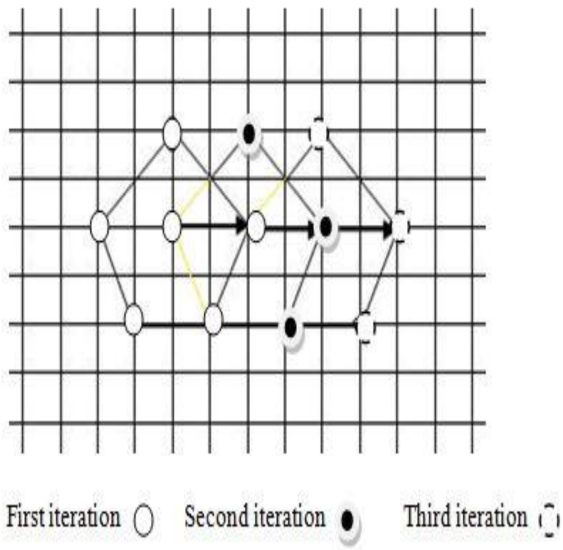


Fig. 10. Pentagon Search

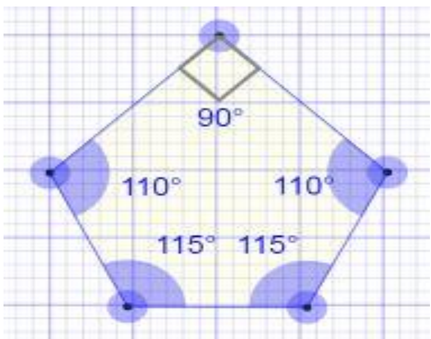


Fig. 11. The shape of the proposed Pentagon

The angles of this pentagon are 90, 110,110, 115,115 degree.

A. CASE 1

If the minimum MAD found at left most or right most corner then the new pentagon formed has only 3 new search points. The minimum should be checked at all three external points.

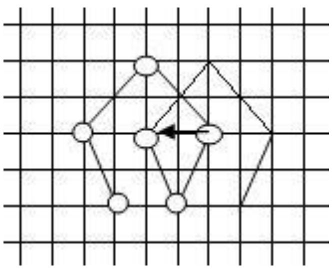


Fig. 12. minimum MAD found at left or right corner

B. CASE 2

If minimum MAD found at lower left or right corner then again 3 new points are searched. One point inside the previous Pentagon need not be checked because there is a correlation between neighbouring blocks [10].

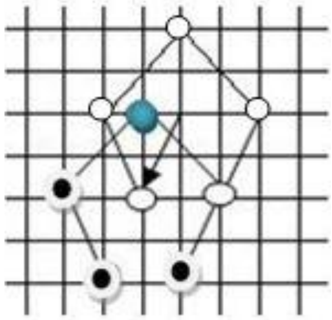


Fig. 13. Minimum MAD found at left position

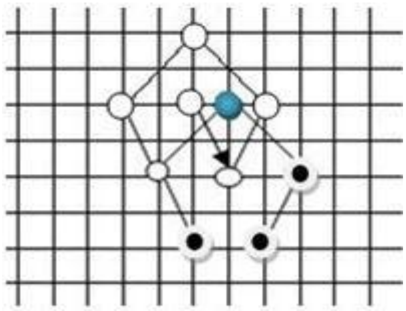


Fig. 14. Minimum MAD found at right position

C. CASE 3

Another case may arise if the minimum found at the top most corner of pentagon then two points of the newly formed pentagon

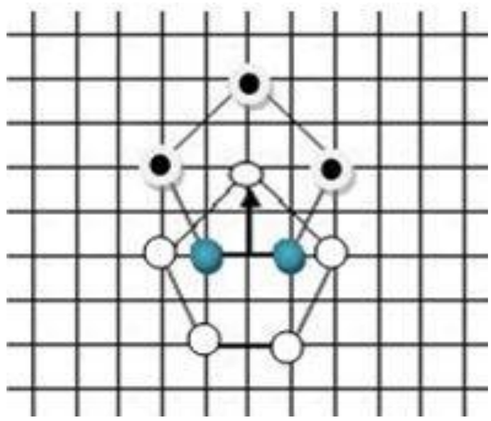


Fig. 15. Minimum found at the top position of the Pentagon

D. CASE 4

If minimum found at center position of the Pentagon then the size of the Pentagon is reduced. The new search position will be the closest position from the centre i.e. left right top and bottom.

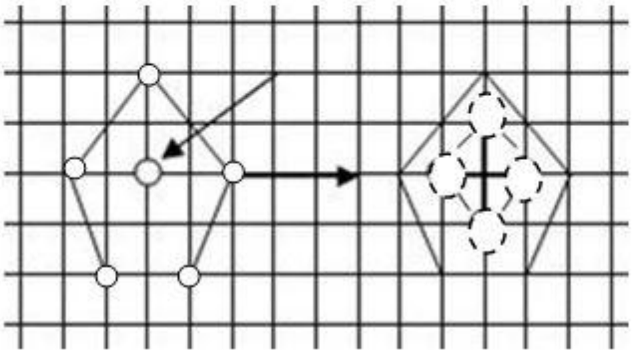


Fig. 16. Minimum found at centre position

IV. RESULT AND DISCUSSIONS

Two video sequences are taken for analysis purpose "missa" and "caltrain". In caltrain video sequence there are 32 frames of size 512*400. In missa video sequence there are 148 frames of 360*288 size.

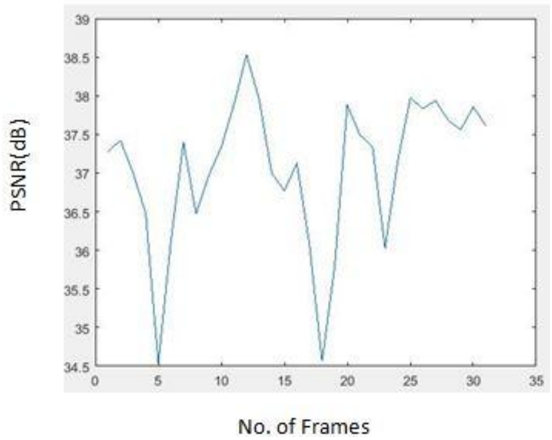


Fig. 17. PSNR value of 32 frames of "missa" sequence having total 149 frames

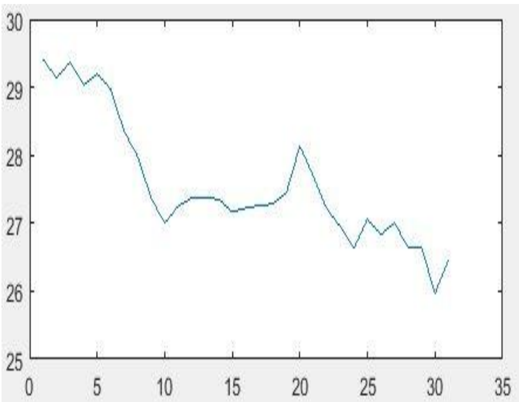


Fig. 18. PSNR value of "caltrain" sequence having 32 frames

$$SIR = \frac{(9 + M \times n + 4) - (6 + M \times n + 4)}{6 + 3 \times n + 4} \times 100\% \qquad (1)$$

$$SIR = \frac{MSE(DS) - MSE(PENT)}{MSE(PENT)} \times 100\% \qquad (2)$$

Case 1...(a) 6+3*1+4= 13(PENT) (b) 9+3*1+4= 16(DS)
SIR= ((16-13)/3)*100 = 23%
Case 2...(a) 6+3*2+4 = 16(PENT) (b) 9+3*2+4= 19(DS)
SIR= ((19-16)/16)*100 = 18.75%
Case 3.... SIR = ((9-6)/6)*100 = 50%(in both DS and PENT)
Avg SIR = (23+18+50)/ 3 = 31%

By analyzing the minimum possible search points of DS and Pentagon method, 3 points are saved on an average. Hence we see 31% increment in SIR performance over diamond search. Performance comparison for caltrain video sequence is given below

Table 1: Motion vector and total time elapsed for caltrain video sequence

	Motion vector	Frame size	Total time
DS	143	512*400	17.684
Hexagon	165	512*400	16.73
Pentagon	179	512*400	15.98

Table 2: Motion vector and total time elapsed for missa video sequence

	Motion vector	Frame size	Total time
DS	82	360*288	13.9419
Hexagon	91	360*288	13.8512
Pentagon	97	360*288	12.6734

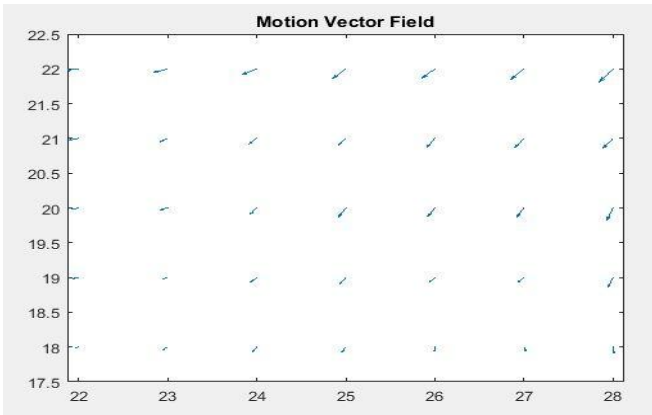


Fig. 19. Motion vector for caltrain video sequence

Table 3: Avg MAD per pixel for missa and caltrain video sequence at different frame rate

Mad per pixel		
	Frame diff1	Frame diff3
missa	4.809	4.984
caltrain	2.561	3.882

Table 4: Avg search per frame for missa and caltrain video sequence at different frame rate

Average search per frame		
	Frame diff1	Frame diff3
missa	0.604	0.843
caltrain	5.774	9.997

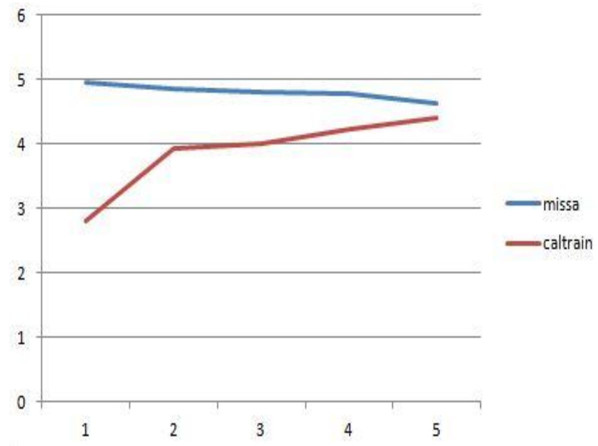


Fig. 20.Graph showing MAD for missa and caltrain sequence

V. CONCLUSION

In this paper, search pattern and search strategies of different block matching algorithm are analyzed. A new Pentagon block matching algorithm is proposed in this paper and simulation results shows that it has better performance over the existing hexagon and DS algorithm in terms of lesser no of search points. All the simulations are performed in MATLAB environment. The speed improvement rate is about 31% with respect to the DS algorithm. Like the DS and other algorithms it may be applied for the latest video compression codec.

REFERENCES

1. Vayalil, Niras Cheeckottu, and Yinan Kong.: 'VLSI Architecture of Full-Search Variable-Block-Size Motion Estimation for HEVC Video Encoding.' IET Circuits, Devices & Systems 11.6 2017: 543-548.
2. Paramkusam, A. V.: 'Efficient motion estimation algorithm on the layers.' Electronics Letters 53.7 2017: 467-469.
3. Jia, Hongjun, and Li Zhang.: 'Directional Cross Diamond Search Algorithm for Fast Block Motion Estimation.' arXiv preprint arXiv:0806.0689, 2008.
4. Chau, Lap-Pui, and Ce Zhu.: 'A fast octagon-based search algorithm for motion estimation.' Signal Processing 83.3 , 2003, 671-675.
5. Chauhan, Shahida. : 'Comparative study on diamond search algorithm for motion estimation.' , 2015.
6. Nie, Yao, and Kai-Kuang Ma.: 'Adaptive rood pattern search for fast block-matching motion estimation.' IEEE Transactions on image processing 11.12, 2002, 1442-1449.
7. Zhu, Shan, and Kai-Kuang Ma.: ' A new diamond search algorithm for fast block-matching motion estimation.' IEEE transactions on Image Processing 9.2, 2000, 287-290.
8. Zhu, Ce, Xiao Lin, and Lap-Pui Chau.: 'Hexagon-based search pattern for fast block motion estimation.' IEEE Trans. Circuits Syst. Video Techn. 12.5, 2002, 349-355.
9. P. T. Kovács, Z. Nagy, A. Barsi, et. al : 'Overview of the applicability of H.264/MVC for real-time light-field applications' 2014 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), Budapest, 2014, pp. 1-4,.
10. Neetish Kumar, Dr Deepa Raj, 'Video Processing and its Applications: A survey' International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 2017, ISSN 2278-6856,.
11. Neetish Kumar, Dr Deepa Raj, 'A Study and Analysis of Images in Different Color Models', International Journal Of Advanced Studies In Computer Science And Engineering IJASCSE 2018, Volume 7, Issue 1,.
12. Kulkarni, S. M., D. S. Bormane, and S. L. Nalbalwar. : 'Coding of video sequences using three step search algorithm.' Procedia Computer Science 49, 2015, 42-49.
13. Nguyen, Tung, Philipp Helle, Martin Winken, et al.: 'Transform coding techniques in HEVC.' IEEE Journal of Selected Topics in Signal Processing 7, no. 6, 2013, 978-989.
14. Ugur, K., Liu, H., Lainema, J., Gabbouj, et a : ' Parallel encoding-decoding operation for multiview video coding with high coding efficiency. 3DTV Conference, 2007 (pp. 1-4). IEEE.
15. Stankowski, J., Grajek, T., Wegner, K., et al.: ' Video quality in multiple HEVC encoding-decoding cycles.' 20th International Conference on Systems, Signals and Image Processing (IWSSIP),2013, (pp. 75-78). IEEE.
16. Xiong, Zixiang, Beong-Jo Kim, and William A. Pearlman.: 'Multiresolutional encoding and decoding in embedded image and video coders.' Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on. Vol. 6. IEEE, 1998
17. Yi, Kang, You-Han Lee, and Jeong-Hyun Joo. "A Fast Video Decoding Technique by Means of Converting Input Video Stream into Forward-Oriented Format Stream in Little-Endian Systems." 2015 7th International Conference on Multimedia, Computer Graphics and Broadcasting (MulGraB). 2015, IEEE

AUTHORS PROFILE

Neetish kumar, Research Scholar of department of computer science of Babasaheb Bhimrao Ambedkar University (A Central University), Lucknow. Currently working on video processing, image processing and pattern recognition. He completed his M.tech computer science degree from Central University of South Bihar, Patna

Dr. Deepa Raj, working as an Assistant Professorin the Department of Computer Science Babasaheb Bhim Rao Ambedkar University. She did her Post Graduation from J.K Institute of applied physics and technology, Allahabad University and Ph.D. from Babasaheb Bhim Rao Ambedkar University Lucknow in the field of software engineering. Her field of interest is Software Engineering, Computer phics, and Image processing. She has attended lots of National and International conference and numbers of research papers published in her field of interest.

International Journal of Grid and Distributed Computing
Vol. 13, No. 2, (2020), pp. 235 – 245

A Framework of Video Compression and Decompression

Neetish Kumar[1*], Deepa Raj[2]

[1]*Research Scholar, Department of Computer Science, BBAU(A Central Univ.), Lucknow,U.P , India. E-mail:neetish08537@gmail.com.*
[2]*Assistant Professor, Department of Computer Science, BBAU(A Central Univ.), Lucknow,U.P , India. E-mail:deepa_raj200@yahoo.co.in*

Abstract

As Multimedia and communication technology is on peak nowadays and keep growing day by day, the vast and voluminous content present on the web is video data. It needs to be compressed as it is of draconian in nature if not compressed. A lot of compression techniques are developed for handling the video data. If it is keenly observed then it is found that search techniques play a pivotal role in video compression. In Video compression, apart from the search techniques, some redundancies should be exploited. Various types of redundancies are the main contents of an uncompressed video. Spatial, temporal, Statistical redundancies occur based on their nature and properties and exploited accordingly without any loss of actual information in video. Block based motion estimation serves as the backbone of video compression and the search techniques used in the block based method play a pivotal role in detecting the motion vector. The encoding of motion vector and residual results in a compressed video form. This paper elaborates on a new kind of framework, having a slightly different type of technical strategy of video compression. All the search techniques used in this paper are pentagon search for block matching motion estimation.

Keywords: *Motion estimation, Motion compensation, Encoding, Decoding, Motion vector.*

1. Introduction

- **Introduction about the video:** Before going in depth of motion estimation and motion compensation, it must be understood the structure of a video properly. A normal video is composed of different types of shots. A shot mostly contains GOPs (group of frames) and those GOPs are core terms for compression because it contains different types of frames namely I, P and B frames. For analysis purpose a data structure is used for traversal of macro block, and then it comes to a matter of blocks and pixels that is the basic thing for a video. A semantic diagram of the normal video is shown in Figure 1.

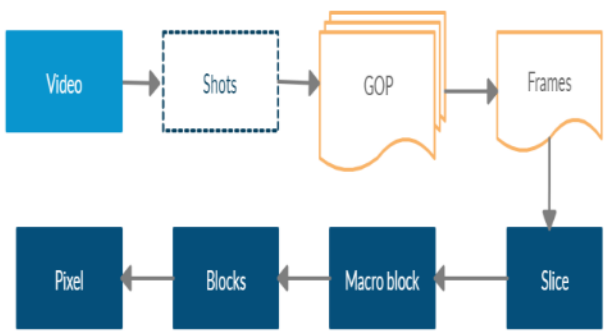


Figure 1. Organisation of video data

- **Motion estimation and motion compensation:** As motion estimation is the crucial part of video compression, a brief mechanism of it must be described. Firstly, the $\boldsymbol{Image_t(x,y)}$ is split in block $\boldsymbol{Blocks_t[n](x',y')}$. The function of motion compensation is to generate $\boldsymbol{Blocks_t[n](x',y')}$ from any part of $\boldsymbol{Image_{t-1}(x,y)}$. So, we have to find out the best possible matching block $\boldsymbol{Blocks_{t-1}[n](x'+mx,y'+my)}$ not strictly nearby of 16x16 boundary. Here, ***mx,my*** is called motion vectors. Basically, the encoder transmits the error and the motion vector for each block. The error can be calculated between the target and reference as $\boldsymbol{Err_t[n](x,y)=Blocks_t[n](x',y')-Blocks_{t-1}[n](x'+mx,y'+my)}$. The processor estimating $mx,my[n]$ for every n such that $\boldsymbol{Err_t[n](x,y)}$ is minimized is called Motion estimation. Whereas, the process of constructing $\boldsymbol{Blocks_t[n](x',y')}$ from It image pixels and ***(mx,my)[n]*** is called Motion compensation. The error image is what gets transmitted. Finally, decoder can redo motion compensation on its own using motion vectors and the error image to make the final re-construction of the image. The pentagon search is used for searching the motion vector throughout the paper for encoding and analysis. Search pattern of pentagon is shown below in Figure 2 [12].

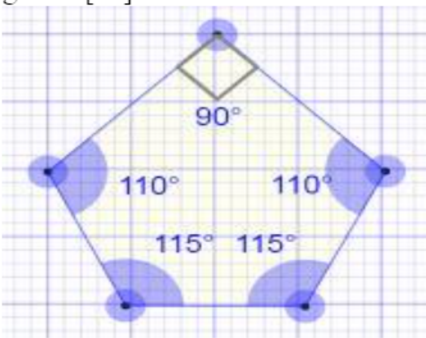


Figure 2. Shape of the pentagon

- **Compression strategy of videos**
  Steps in video compression
  - Motion estimation & Motion compensation: These two crucial parts are discussed in above section.
  - Transform and quantization: The basic transform coding process is almost similar to that of every codec mechanism. The process includes a forward transform and quantization followed by zigzag ordering and entropy coding.
  - Inverse transform: It is used in the reconstruction process. The reconstruction process includes an inverse quantization and inverse transform followed by motion compensation.
  - **Some codecs and their working principles**

There is some limitation on the certain simulator to analyze the result of video frames. In general these frames are of big size which takes too much time for analysis. As suggested in various literature, all the simulation for video compression is performed on QCIF video format due to prevent the heavy computational complexity. Time taken for the simulation of normal/ HD video takes through MATLAB usually takes much more time. Hence for analysis purpose CIF or QCIF are used. Spatial resolution and other properties for different formats are given in Table 1.

Table 1. Different formats of videos and its properties

Format	**Spatial Resolution**	**Frames/Second**	**Uncompressed Data Rate (bits/s)**
QCIF	176 × 144	15	4 561 920
CIF	352 × 240	30	30 412 800
ITU-R601	720 × 480	30	124 416 000

HDTV	1280 × 720	30	331 776 000
HDTV	1920 × 1080	30	746 496 000

MPEG 1:

Moving Picture Experts Group (MPEG) committee introduced the MPEG-1 as the first public standard for video compression. It was developed by ISO and was capable of compressing a video at the compression ratio of 26:1 and 6:1 for video and audio compression without excessive loss of quality. The standard consists of the following five parts.

MPEG 2:

The MPEG-2 Codec was developed in 1995 keeping, in mind to extend the compression strategy of MPEG-1. A step forward it not only compresses a video but also covers the scope of bandwidth usage. It is suitably designed for digital television broadcasting applications. It is capable of handling the bit rate of TV broadcasting that lies between the range of 4 and 15 Mbps (up to 100 Mbps), such as Digital High Definition TV (HDTV), Interactive Storage Media (ISM) and cable TV (CATV). Profiles and levels were introduced in MPEG-2.

MPEG 4:

MPEG-4 is a method of defining compression of audio and visual digital data. MPEG-4 is still an evolving standard and is divided into a number of parts for its simplicity and betterment of this codecs. Its supplimentry parts are MPEG-4 Part 2, MPEG-4 part 10 (MPEG-4 AVC/H.264 or Advanced Video Coding) etc. It was developed in October 1998 and it empowers audio as well as video in low bit-rate networks and allows the user to interact with the objects. Video object coding is one of the most prominennt features originated in MPEG-4 that enables arbitrary shape in spite of a rectangular frame for block matching. This type of compression incorporates the inclusion of shape, motion, and texture.

Table 2. Shows the different compression techniques of MPEG family

S.No	Year	Standard	Publisher	Popular Implementation	Bit Rate
1	1992	MPEG-1	ISO	Video-CD	1.5 Mb/s
2	1994	MPEG-2	ISO,ITU-T	DVD video, digital video broadcasting	>2Mb/s
3	1998	MPEG-4	ISO	Video in internet, DivX	Variable

H.261:

International Telecommunication Union (ITU) developed this standard in 1990. Motion compensated temporal prediction are easily handled by it. Due to the limitation of hadling voluminous visual information at that time, it supports two resolutions, namely, Common Interface Format (CIF) with a frame size of 352x288 and quarter CIF (QCIF) with a frame size of 172x144. It provides the coding and decoding services at the rate of p x 64 kbit/s, where p is in the range 1 to 30. As like of MPEG standard it uses discrete cosine transform(DCT) and performs both INTRA as well as ITNER frame encoding. The prominent features of H.261 is prediction, block transformation (spatial to frequency domain translation), quantization and entropy coding without degrading the video quality. Basically loop filtering serves as the backbone for improvement in video quality with extra processing power. It is mostly used in video conferencing and video communications systems. Its usage includes studio based video conferencing, desktop video conferencing, surveillance, monitoring, computer training, and telemedicine.

H.263:

With some limitations over the video conferencing application in H.261, H.263 is designed and developed by the ITU-T Video Coding Experts Group (VCEG). A step forward H.263 standard not only supports video conferencing but video telecommunication too. It was approved in early 1996. the compression core of the MPEG-4 standard lies in H.263. Variable block size compensation, overlapped block motion compensation are the prominent ideas of H.263 standard. While H.261 allows 18- 24

ISSN: 2005-4262 IJGDC

kbps for moderate quality over a regular phone line or wireless medium, H.263 allows 28-33 kbps for the high quality. It is used for various applications on the internet eg. Flash Video Content. It supports five resolutions: QCIF, CIF, SQCIF, 4CIF, and 16 CIF.

H.264:

H.264 or MPEG-4 Part 10, Advanced Video Coding (MPEG-4 AVC) was developed by the ITU-T Video Coding Experts Group (VCEG) together with ISO/IEC. The developer of MPEG series and H.26X series came together for this project and designed the best Codec. H.264 encoder provides better improvement in compression efficiency than the previous standards. It is basically a block-oriented motion-compensation-based video compression standard. H.264 is mostly used for encoding the content of Blue ray disc, pervelent in video conferencing, television broadcasting etc.

Table 3. H.26x family series

S.No	Year	Standard	Publisher	Popular Implementation	Bit Rate
1	1990	H.261	ITU-T	Video Conferencing, Video Telephony	P*64 kb/s
2	1995	H.263	ITU-T	Video Conferencing, Video Telephony, Video on mobile phones	<33.6 kb/s
3	2003	H.264	ISO,ITU-T	Blu-ray, DVD, Digital video broadcasting, HDTV, Apple TV	10's to 100's kb/s

2. Literature Survey

P. B. Penafiel and N. M. Namazi introduced a new framework for video compression that considers noise applicabitily in frame sequence directly in order to find the optimal compression ratio. By exploiting the temporal and spatial redundancies they found the compressed video. Processing is performed in blocks of N frames stored in a video buffer. Encoder and decoder are synchronized before the transmission of a new block [1]. K. Minoo and T. Nguyen suggested a framework for efficient entropy coding of data through parametric distribution model. Maximum a posteriori (MAP) or Maximum Likelihood (ML) and other statistical parameter estimation techniques are used in their modified framework that basically applies on entropy coding [2]. L. Xiaoli et. Al. Demonstrated a framework based on perspective to improve the degree of parallelism of the on-chip peripherals and the core for video compression [3]. M. Mody suggested a framework that doesn't need much more hardware support for video compression. The proposed model, VCF also defines simple primitive (e.g. threads, queues, semaphores, messaging, scheduling scheme etc) to enable it without using any hardware support [4]. S. Moni and S. Sista elaborated a framework for Images and video compression that are being extensively used in numerous areas such as video-conferencing, multimedia documentation, telemedicine, high definition television (HDTV) etc [5]. Xuguang Yang and K. Ramchandran discussed a novel region-based video compression framework based on morphology to efficiently capture motion correspondences between consecutive frames in an image sequence. They formed "clusters" based on thier idea of motion field associated with typical image sequences. These segmentation leads to the distinct objects or regions in the scene, that can be efficiently captured using morphological operators in a "backward" framework that avoids the need to send region shape information [6]. L. Herranz and J. M. Martínez worked on video summarization that provides compact representations of video sequences. It is basically based on the visualization, rather than the information conveyed [7]. J. Nightingale et.al. proposed an idea that focuses on improvements to video compression efficiency of

HEVC. In their work, they considerd all the barriers towards the development of HEVC framework. This study filled the gap at every stage of visual application, testing & evaluation, video streaming and packet loss in HEVC [8]. D. Schonberg et.al. presented a framework for compressing encrypted media, such as images and videos. They first develop statistical models for images before extending it to videos [9]. F. Kamisli demonstrated the different idea of video compression in which when spatial prediction is used, temporal information is ignored [10]. C. Zhang et.al. proposed a novel algorithm called spatially varying transform (SVT) is proposed to improve the coding efficiency of video coders. SVT enables video coders to vary the position of the transform block, unlike state-of-art video codecs where the position of the transform block is fixed [11].

3. Proposed framework for video Compression & Decompression

Till now there are a lot of video codecs developed, our proposed framework is a step forward in this regard. Basically at the back end i.e search strategy, a new pentagon search method is used that searches the true motion vector so much efficiently that further encoding method gets benefitted by it. Before knowing the complex process of compression it must be investigated first that what a video frame consists of. Basically video is a sequence of kinetic frames. These frames are bundled together by its properties, these bundles are known as GOP. In a GOP there are three types of frames I, P and B frames. All three types of frames are encoded in different ways. The first frame of the GOP is always an I frame that is encoded just like the still image. Figure 2 represents the diagramatic representation of GOP that deals with the coding of different frames.

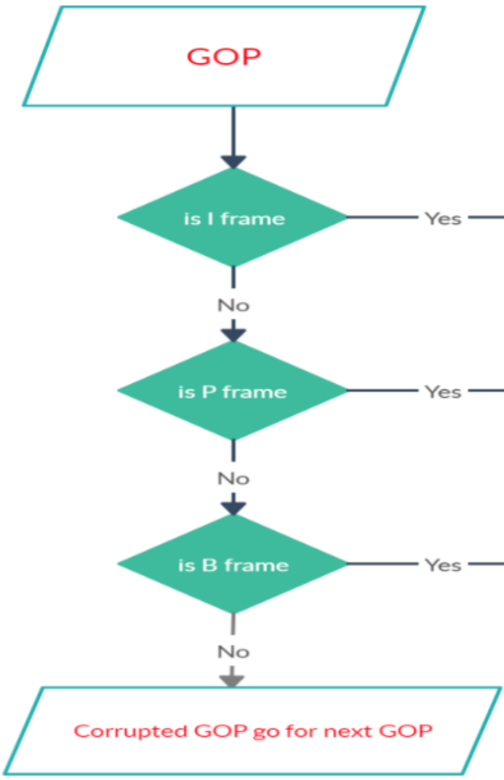


Figure 3. Encoding of different frame through GOP

3.1 Encoding of different Frames

a) **Encoding of I frame:** I frame is encoded just like the JPEG image. The first frame of every GOP is I frame. It is also known as INTRA frame as it has only the spatial redundancy. This is the only frame that is intracoded and also known as error-free frame because at this stage no prediction is needed for the encoding. It acts as a reference frame for the upcoming P and B frames. For encoding of this type of frame same process of DCT, Quantization and entropy coding is performed as like happens in JPEG. In DCT, the whole frame is subdivided into 8*8 blocks then DC coefficients are

DPCM coded and after that, the AC coefficients run-length coded. A neat diagram of encoding of I frame is shown in Figure 4.

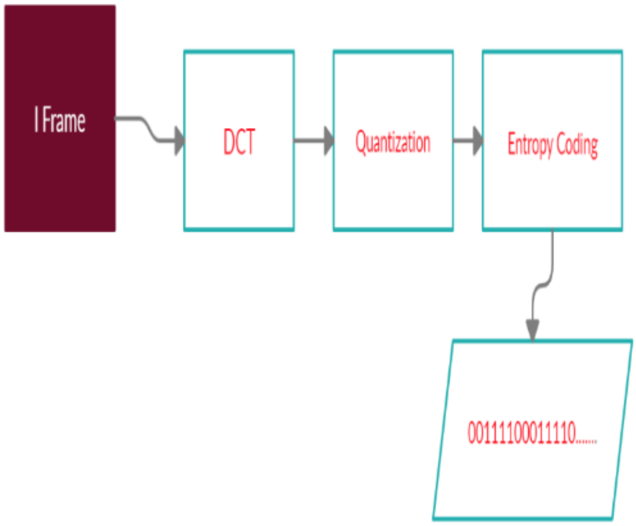


Figure 4. Encoding of I frame

b) **Encoding of P Frame:** P frame is also known as the forward predicted frame. Firstly the frame is predicted through the corresponding I frame. The search strategy used for finding the motion vector is pentagon search that gives the high performance for motion vector finding probability. Through motion estimation and motion compensation the motion vectors and the errors are basically encoded. To predict the motion of a particular block, a sufficiently larger size window is taken. The corresponding block is searched within the window. The exact match of the block in current, as well as reference frame, is calculated through the different metrics. The position shift of the corresponding block is known as the motion vector. A lot of searched techniques are developed for detecting the motion vector. In our paper pentagon search is used for this purpose that gives better motion vector finding probability than previously proposed search techniques. As fewer blocks movement and inclusion of new block information occurs at this stage. So new information is encoded and the remaining motion vector information is stored in the buffer. The same process of DCT, Quantization and entropy coding is performed for such blocks. Finally at the end of the encoding the visual information is converted into the bitstream. The compression of P frames gives a compression ratio of 20 to 30:1.

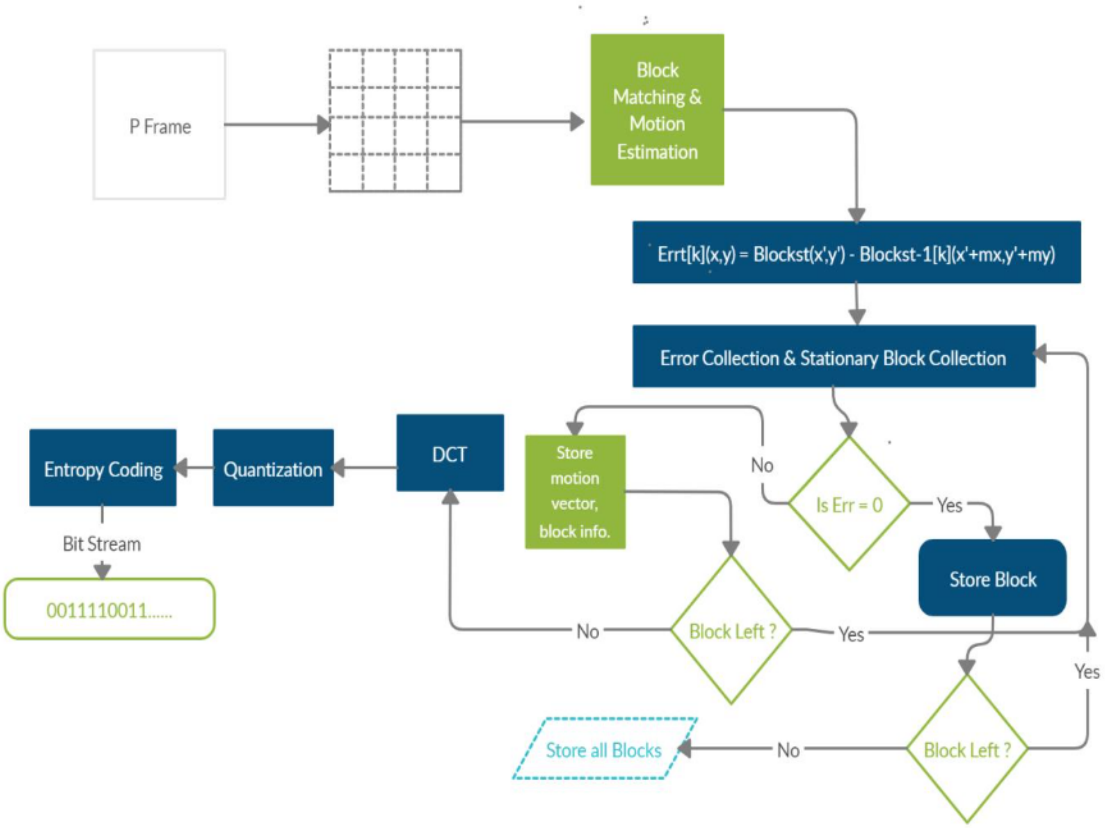


Figure 5. Encoding of P frame

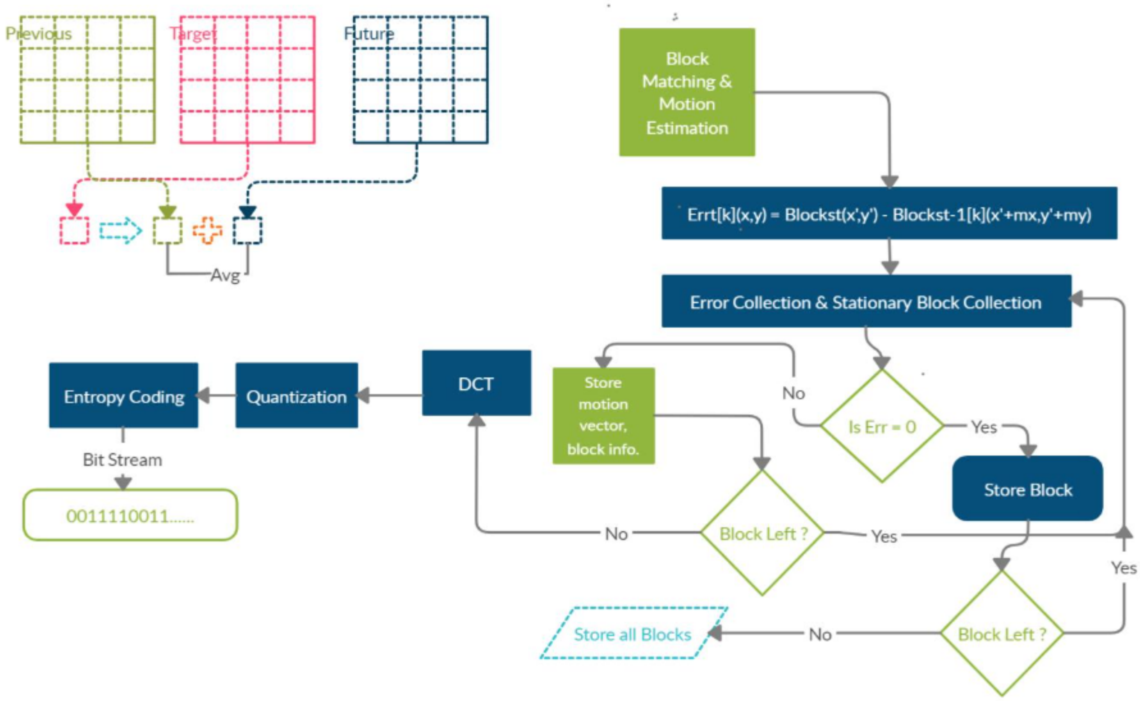


Figure 6. Encoding of B frame

c) **Encoding of B frame:** B frame is also known as the bi directional frame. It can be predicted through the past frame or future frame. All the process of block matching, motion estimation and motion compensation takes place just like as of P frames. At first, suffiently large window size is taken for identifying the motion vector. The search strategy used for finding the motion vector. Firstly the six blocks are chosen for comparison and the errors are calculated. We have already analyzed the pentagon search for finding the motion vector so that it can be used as the most effective search technique for determining the motion vector. There is a slight change in the block matching process in B frame. As in P frames, only past frame is considered as the reference but here past and future both frames are taken into consideration. So, in block matching phase average of these two is taken. Thus compression of B frame takes more space in the buffer. It typically requires very fewer bits than that of I and P frames. The compression ratio of B frames is very high and nearly about 30 to 50:1.

3.2 Decompression strategy of frames: In general the decoding of I frame is just like the decoding of the still jpeg image. For P and B frames there is a slight variation in this process. In the encoding of P and B frames there is a subtraction between input and predicted image, therefore at the decompression end, it is added to get the reconstructed frame. Exact reconstruction of a frame can't be possible because in all cases, at the encoder end, quantization is used that is an irreversible process. So, one has to satisfy at a faithful reconstruction level. Figure 7 shows the diagram of the video decoding. After getting the encoded bit stream, entropy decoding, dequantization, and inverse DCT are performed in order to get the error video. After the error is decoded the predicted video is added for getting the reconstructed frame. Decoding of I, P and B frames are discussed separately. Graphically the process of decoding of I, P, B frames are shown in Figure 8, Figure 9 and Figure 10 respectevily.

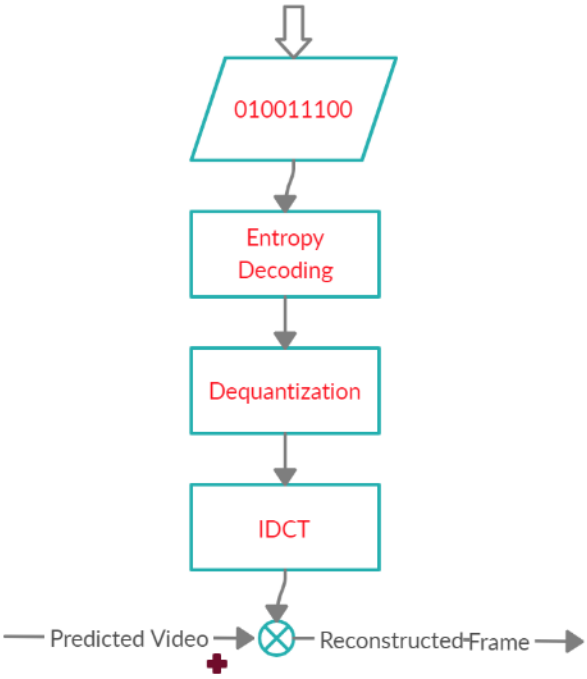


Figure 7. Decompression strategy of Frames

a) **Decoding of I frame:** I frame is decoded in the same manner as the still picture (JPEG) decoding technique. Simply the encoded bit stream gets decoded by entropy decoder, dequantizer and IDCT. The I frame is also known as the Error free frame because it is fully intra coded i.e. no past or future frame is required as the reference frame.

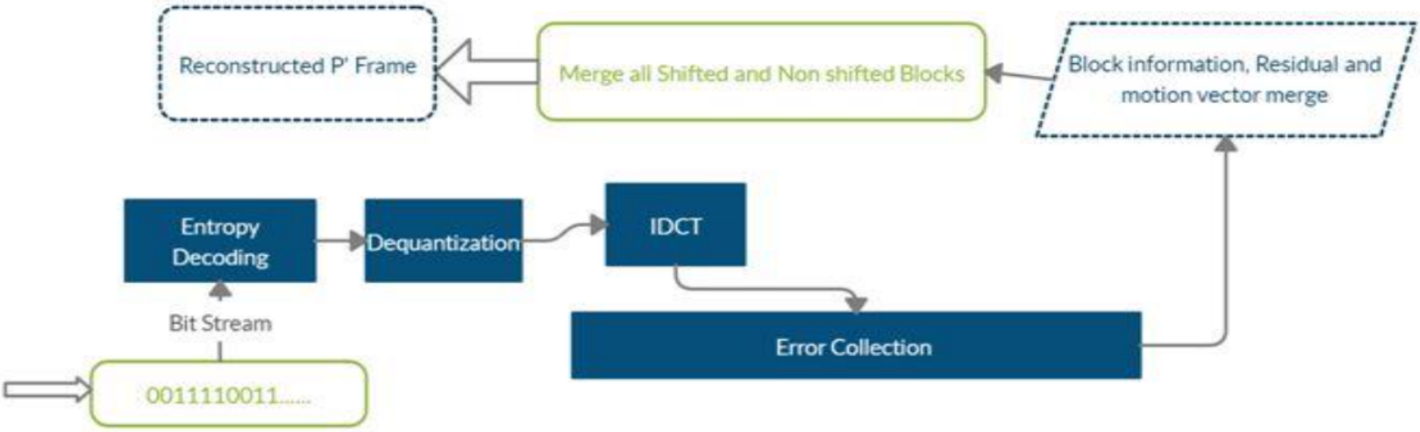


Figure 9. Decompression of P frame

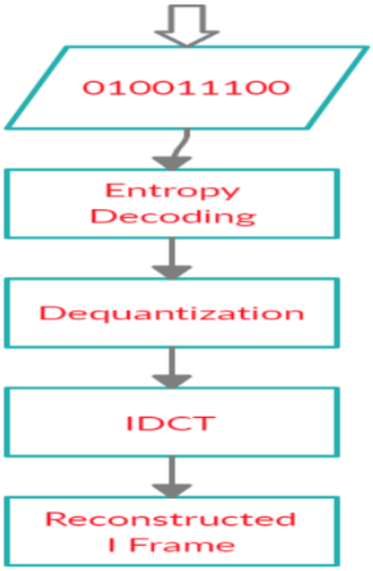


Figure 8. Decompression of I frame

b) Decoding of P frame:

At first the bit stream received at the decoder end, it is firstly entropy decoded followed by dequantization and inverse DCT. After all the above steps whatever received is only the error signal. For a faithful reconstruction, the corresponding motion vector and unchanged block must be appended to the error signal. After merging of the error signal, motion vector, and unchanged block the p frame is reconstructed at the decoder end. It is not actually the P frame but P' frame because at the encoder end there is a step of quantization which is a lossy process, so exact reconstruction is never possible. That's why one has to satisfy only with at an extent of faithful reconstruction. The decoding process of P frame is shown in Figure 9.

c) Decoding of B frame:

Same process of decoding takes place for B frame as that of P frame. A little bit difference occurs after the error accumulation, as there is only one anchor used as a reference while in the case of B frame two frames are used as a reference frame. The decoding process of B frame is shown in Figure 10.

There are some advantages of coding of B frames that reflects in coding efficiency. B frames has the less but quality information about the motion vector information. Rate of quality improves with accurate prediction and motion estimation. Backward prediction as well as forward prediction involves in the determination of motion vector so for decoding time the correct reference frame plays a major role. The level of accuracy may vary if the combination of forward and backward prediction wrongly used.

There are some disadvantages of handling the B frames. One is directly related to the memory. As it is the prediction of two reference frames, so it requires extra space in memory. In general the memory requirement gets doubled while handling the B frames, so it is not feasible. Another disadvantage appears in case of video conferencing. There is an unwanted delay occurs in this type of real time application.

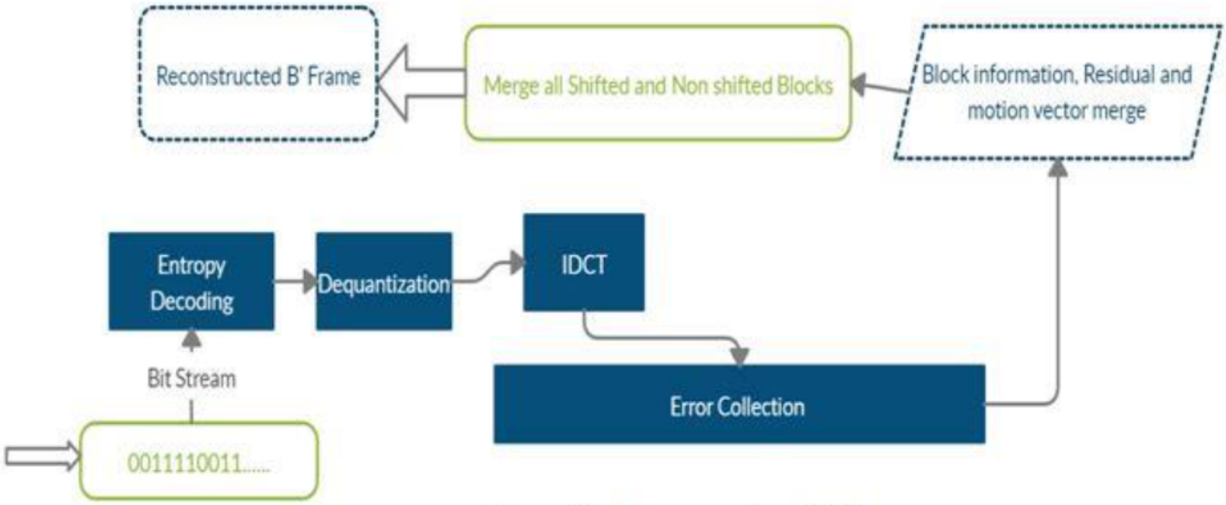


Figure 10. Decompression of B frame

4. Result and Discussion:

As like previous, the performance analysis can be performed through MAD. Our proposed framework efficiently works on P and B frames. We 'he already analyzed that the search pattern of the pentagon method for block matching motion estimation gives a satisfactory performance over different video frames. The search strategy used in this method is pentagon search that searches the motion vector in an efficient manner. Motion vector finding the probability through pentagon search is better than the previously proposed search strategy. Accumulating the error and process through this method is a different idea in both encoding and decoding. It also provides a better simulation result compared to others. The compression performance achieved for P and B frames through this method is larger than any other previously proposed framework. This framework is also applicable for compressing the real time video data.

5. Conclusion

This paper presents a more optimized framework of video compression using pentagon search for block matching motion estimation. As the search, for motion vector takes less time with more accuracy, it can be directly used for the optimized framework of video compression. A slight chane in the mechanism of compression is used in this paper. Whole bunch of blocks (changed and non changed) are processed parallely to reduce the effective time of compression. The same strategy of compression can be used in different domain e.g wavelet domain in future.

References

[1]. P. B. Penafiel and N. M. Namazi, "A new framework for noise-resistant video compression using motion-compensated prediction," *1995 International Conference on Acoustics, Speech, and Signal Processing*, Detroit, MI, USA, vol.4, pp. 2181-2184, 1995.

[2]. K. Minoo and T. Nguyen, "Entropy Coding via Parametric Source Model with Applications in Fast and Efficient Compression of Image and Video Data," *2009 Data Compression Conference*, Snowbird, UT, pp. 461-461, 2009.

[3]. L. Xiaoli, L. Chao, W. Qiang and D. Wenrui, "Design and Implementation of Real-time Video Compression System Control Framework," *2009 WRI International Conference on Communications and Mobile Computing*, Yunnan, pp. 581-585 2009.
[4]. M. Mody, "Video codec framework (VCF): Novel firmware architecture for video hardware," *2014 Twentieth National Conference on Communications (NCC)*, Kanpur, pp. 1-5, 2014.
[5]. S. Moni and S. Sista, "A framework for application specific image compression," *Proceedings DCC '97. Data Compression Conference*, Snowbird, UT, USA, pp. 456-1997.
[6]. Xuguang Yang and K. Ramchandran, "A low-complexity region-based video compression framework using morphology," *Proceedings of 3rd IEEE International Conference on Image Processing*, Lausanne, Switzerland, vol.2, pp. 485-488, 1996.
[7]. L. Herranz and J. M. Martínez, "A Framework for Scalable Summarization of Video," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 9, pp. 1265-1270, Sept. 2010.
[8]. J. Nightingale, Q. Wang and C. Grecos, "HEVStream: a framework for streaming and evaluation of high efficiency video coding (HEVC) content in loss-prone networks," in *IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, pp. 404-412, May 2012.
[9]. D. Schonberg, S. C. Draper, C. Yeo and K. Ramchandran, "Toward Compression of Encrypted Images and Video Sequences," in *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 4, pp. 749-762, Dec. 2008.
[10]. F. Kamisli, "Recursive Prediction for Joint Spatial and Temporal Prediction in Video Coding," in *IEEE Signal Processing Letters*, vol. 21, no. 6, pp. 732-736, June 2014.
[11]. C. Zhang, K. Ugur, J. Lainema, A. Hallapuro and M. Gabbouj, "Video Coding Using Spatially Varying Transform," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 2, pp. 127-140, Feb. 2011.
[12]. Neetish Kumar, and Deepa Raj. "A New Pentagon Search Algorithm for Fast Block-Matching Motion Estimation." *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* Volume-8 Issue-10, August 2019

ISSN: 2005-4262 IJGDC

ISSN- 2394-5125 VOL 7, ISSUE 09, 2020

ALGORITHMIC SOLUTIONS FOR HIGH QUALITY VIDEO EDITING SOFTWARE PROBLEMS

Neetish Kumar[1], Deepa Raj[2]

[1,2]Department of Computer Science, BBAU (A CENTRAL UNIVERSITY), Lucknow, India
Corresponding author Email- neetish08537@gmail.com

Received: 08.03.2020 Revised: 25.04.2020 Accepted: 14.05.2020

Abstract
It is said that paper has more patience than man, but a video has far more expression than a paper or a picture. As video is the sequence of kinetic images. The modern world accepted the video content more than words and touched the human life in almost every extent. Since it is growing faster there are various challenges related to it like enormous burden on the acquisition, storage, compression, and transmission of video data. Video editing is not only used in professional movie making but also in various areas like education field, marketing fields etc (ex. Since lockdown due to covid-19 the education world started classes online). Today we have more popular video cameras to manage the video data but algorithmic solution is needed to solve various problem related to the video editing. There are also more opportunities in quality degradation, video security, video copy detection etc. In this paper we are going to deal with the motion analysis, abrupt transition detection, gradual transition detection etc. Some algorithmic approaches are also used to cope with the modern video editing software problems.
Keywords: Video editing, Shot, Cut-point, Frames

1 INTRODUCTION

In the new digital era, the usage of multimedia has touched the every aspect of human life. Due to the world wide acceptance it draws the attention of researches to work in different domain of video processing [1]. For transmission of video over the wired or wireless channel, the compression aspects are kept in mind. There are some fields in which the videos are required to view with some modification or special effects. Special effects have countless features which may leads to further animation. In video editing process, frames that get corrupted during the capturing operation with camera are modified, extra noise from a certain frame is removed, unwanted objects are removed, etc [2]. These types of post-production activities are very much required in education field, film industry, online classes, news channels, YouTube channels etc. To understand the basic operation of video editing, firstly it is needed to know about video structure.

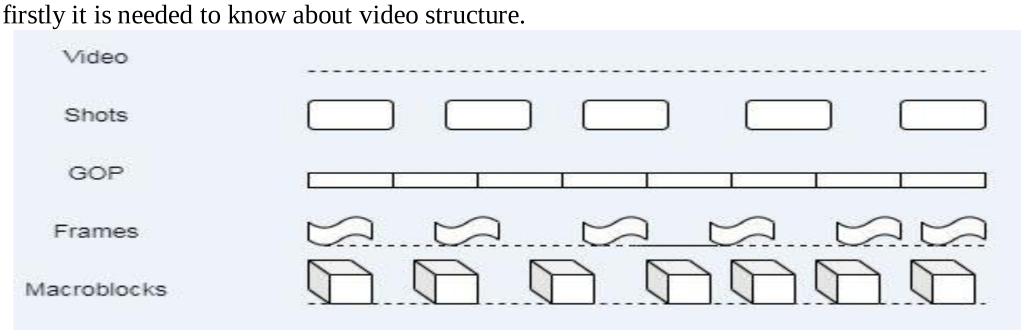


Figure 1: Video hierarchical structural

The above figure illustrates the structures of a video. Video is nothing but the collection of shots. Shots are the bundle of pictures commonly known as group of pictures (GOP). In GOP, there are 3 types of frames i.e. I, P and B frames [3]. To understand the basic idea of editing it needed to go to further on scale down which is known as Macroblocks. Matching, addition and deletion of frames are heavily depends upon the variation of blocks. The basic unit at which the editing operation performed is the pixel level. As the concept of neighbourhood similarity the pixels are clubbed to form a uniform block. Video editing has very broader aspect of addition deletion special effect creation on particular frame. The paper deals with the editing uses in videos and possible algorithmic solution for the same. This paper is organised as follows. Section 2 deals with the literature survey where as the 3 deals with the editing issues. Section 4 is the possible solution for that and lastly it has conclusion section.

ISSN- 2394-5125 VOL 7, ISSUE 09, 2020

2 LITERATURE SURVEY

ZHANG Songhai et. al. described a fast video structure analysis method that is based on image segmentation and matching. The decomposition of video further results in moving objects that depicts the information regarding color, position, shape, movement etc. The optimization techniques used by them is to solve size reduction of the object in last few frames are supported with greedy algorithm [4]. Chitru L. Mudhwuchutyulaet. al. attempted to bridge the gap between low level & high level multimedia signals through video editing framework. For that they developed a mechanism of video metadata editing which supports some functions that is required for standardized schema [5].Xian-Sheng Hua et. al. presented a home video editing system based on optimization techniques. The system automates the temporal structure and determines the segment of video of desired length of highlighted segments. In addition, with the help of the developed framework, best-matched music for a given video can be chosen for different editing styles [6].Asad Islam et. al. proposed an optimized, compressed domain video editing algorithms for common editing operations for mobile devices. The algorithm developed by them is time and space efficient [7].Carlos Cuevas and Narciso Garcia developed a novel shot detection technique which meets the requirement of real time video editing software. This technique detects both abrupt and gradual transitions between shot efficiently. In parallel, it also measures the variation of significant edge point of the image to identify the gradual transition [8].Zhen et. al. proposed a method to edit multiple video of different formats. They worked with DES (DirectShow Editing Services) to combine the MPEG, WMV, AVI etc formats effectively [9-10].

3 VIDEO EDITING OPERATIONS

As we know that the process of video editing is also known as the post production phenomena. According to the need visual effects are added, corrupted frames are removed and some new frames are added. Some of editing operations are discussed in this section

Condition for clear visualization of video frames

First of all, viewers of all kind want to watch any video without disturbance of frames. To see the video without any distortion there are some condition for visual legibility of a frame. The disturbance is monitored with respect to the spatial variation of pixels in frames.

i) Assume c_{min} = minimum pixel intensities in a frame

andc_{max}=maximum pixel intensities in a frame,

t' = threshold for spatial variation. Then,

$$|c_{max} - c_{min}| \geq t', \quad \text{for } 0 < t' \leq 255 \quad (1)$$

ii) again assume η = average pixel intensity of the frame

$\omega > 0$. Then,

$$(c_{min} + \omega) \leq \eta \leq (c_{max} - \omega) \quad (2)$$

Both of the above condition must be satisfied for clear visualization of the frame. The first condition checks about the meaning information about the intensity distribution where as the second condition depicts the smooth distribution of pixel intensity in order to ensure the maximum visibility. Literatures have suggested that the value of t' and ω nearly 20 for good results.

Frame and video clip Insertion: it is the very common editing process used when there is some fault during capturing side takes place. There are many way to add the certain frame in between. There is vast similarity between the successive frames in a video sequence. To insert a new frame these steps are followed

I)firstly the past and next frame is recorded.

II) Difference between both are calculated

III) on the basis of difference the average motion vector is added to the previous frame.

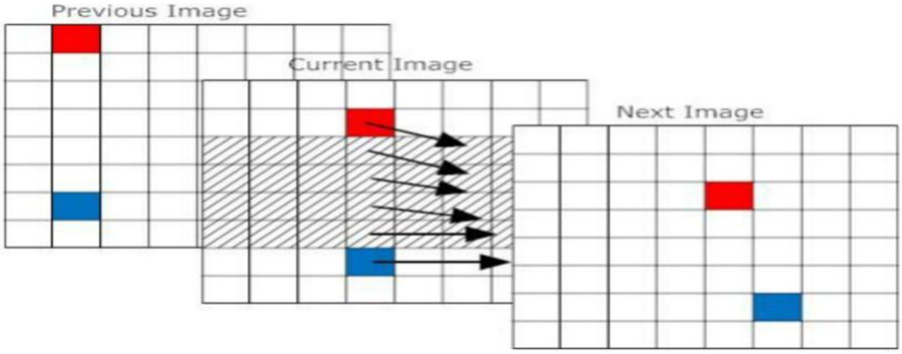


Figure 2:Frame insertion

ISSN- 2394-5125 VOL 7, ISSUE 09, 2020

To inset group of frames at a time it follows the same procedures at the past frame and next frame positions. There are three types of frames i.e I, P and B frames the junction point to be attached must have the I frame in order to make the clear illusion to the viewer. I frame contains most of the rich information that is not present in the P and B frames. So it needed to add the video clips that must have I frames.

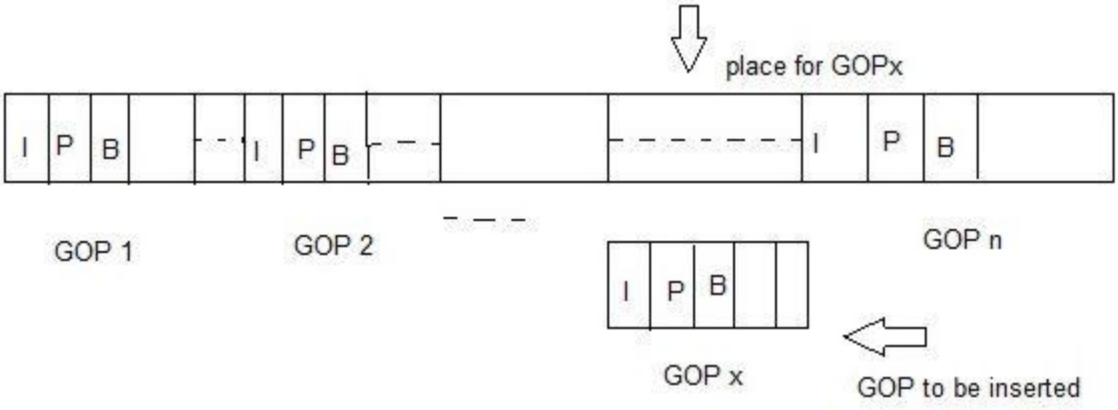


Figure 3: GOP insertion

Splicing: it is the process in which the two or more video clips are clubbed together. The prerequisite for the operation is to have the same format and same resolutions of the video clips. The fig below is showing the two video clips V0 and V1 are stitched at the splicing point. M is assumed to be the last frame of the V0 and n is the first frame of the V1 video clip. Stitching operation is said to be successful if the following conditions are satisfied.

Let t_m =time stamps of the last frame m in V0

t_n = time stamps of the first frame n in V1

$ŧ_{v0}$ = the duration of V0

$ŧ_m$ = the time duration of frame m

Then for perfect stitching point this condition must satisfy

$$tn = tm + ŧ_m = ŧ_{v0} \qquad (3)$$

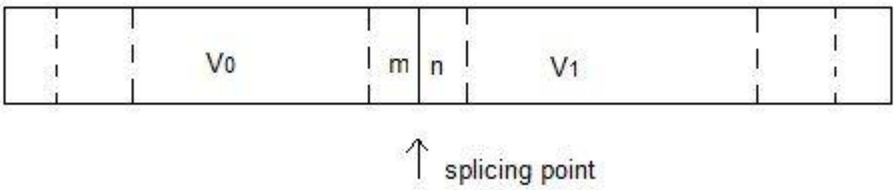


Figure 4: Splicing point & merging of two video clips

Cutting: Opposite to the splicing operation the result of the cut video clips are of same format and resolutions. Firstly the cut section is to be identified i.e. the cut in point and cut out point is detected. Secondly at the cut in point the previous P frame of the last GOP is converted into I frame. Then at the cut out point the clip length is needed to be removed after a certain time stamp. For a perfect cut the condition are as follows-Borrowing the above equation from the slicing part the equation is-

$$t_m = \sum_{k=0}^{m-1} T_k \qquad (4)$$

where,

t_m =time stamps of the last frame m in V0

t_n = time stamps of the first frame n in V1

T_0= time interval for first clip

T_1= time interval for second clip

T_k= time interval for k_{th} clip

ISSN- 2394-5125 VOL 7, ISSUE 09, 2020

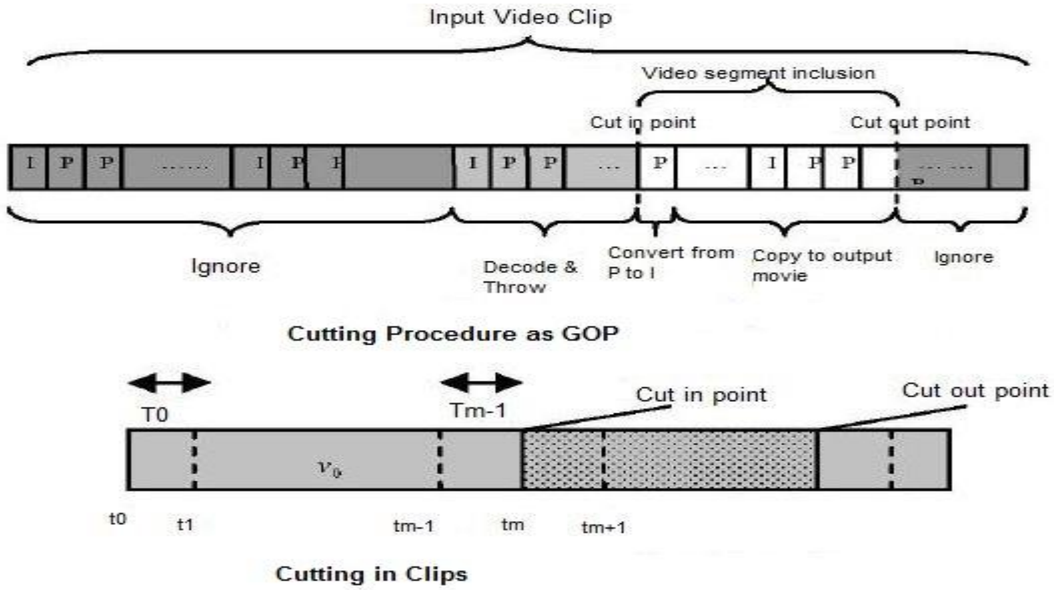


Figure 5:Cutting procedures

a) **Gray scale conversion:** It is the spatial effects of the video editing. The primary color that make the video colorful are the red, green and blue. The intensity combination of the three colors is used to gain certain color. It requires 24 bit to represent color image 8 for all three primary color. As we know that human eye is more sensitive towards intensity rather than color. The color scale can easily be converted into gray scale by simple averaging method.

Grayscale = (R + G + B / 3) (5)

b) **Slow Motion& Fast motion:** It is the temporal effect of the video clip that is done by modifying the frame rate. Frame rate is the rate at which the frames are played at the playback side. Frame rates are the essential thing used in both the sides of video processing i.e. at capturing side and playback side. If there is no change in the frame rate at the playback side the video clips runs normal but if there is a change in capturing and playback side then it results into the fast or slow motion videos.

4 ISSUES IN VIDEO EDITING AND ITS SOLUTIONS

Shot Boundary detection: Shot boundaries can be detected in various ways like feature visual content based detection, by classification of continuity signal etc. there are some issues too in detection of shot boundary.

In pixel based approach, it is difficult to predict the threshold that detects thetransition stage to find out the shot boundary. It is very crucial to identify the right threshold value otherwise the frame difference between the two successive frames exceeds the threshold and gives wrong information about the shot boundary. Another method i.e. histogram based method is invariant to local or global motion as compared to pixel based approach so it gives better result. But there are another issue related to it i.e. color histogram sensitivity towards camera motion such as panning and zooming.

$$D(fn, fn+1) = \frac{1}{NxNy} \sum_{x=1}^{Nx} \sum_{y=1}^{Ny} P\left|(fn, x, y) - P(fn+1, x, y)\right| \quad (6)$$

The above equation is used for shot boundary detection in gray images. It uses sum of absolute difference method for that.

fn, fn+1 are the current and the next frame respectively,

x,y are the pixel coordinates in frames

Nx, Ny are width and height of the frame

Quality degradation: It is said to be the main aim of video editing to prevent the video from degradation. If the quality of the video is degraded at the capturing side there it more vulnerable to degrade over the channel while transmission. It is needed to preserve the quality of video without any compromise. The main factor for degradation is resolution factor and the addition of extra noise. When and unwanted noise come into the picture while capturing then it must be needed to remove the noise at the playback side.

The above is solved with noise reduction and filtering techniques. There are various types of noise and different types of noise are handled by different techniques. For example Gaussian noise, salt & pepper noise, shot noise, quantization noise speckle noise are handled by linear, nonlinear and adaptive filters

Abrupt transition detection: Basically in long length video clips, there is chance of two unrelated frames may come out in the picture at the playback side. Im more cases of shots the last frame of previous shots are unrelated with first frame of next shot. In this case the sudden change in intensity may cause the ignorance of frame because one may can't correlate the things between the frames. A large effort has been made by the research community to make an illusion between these types of unrelated shots.

$$\left(I^n, I^{n-1}\right)$$

$$\left(p^n, p^{n-1}\right)$$

$$M_q(I^n) = \frac{1}{HW}\sum_{h,w} p^n{}_{h,w} \quad (7)$$

The above equation is the measure of illumination changes denoted by M.

I^n, I^{n-1} -are the intensity variations in past and next frames

P^n,$^{pn-1}$ - are the mean intensities

H- height of the image

W- width of the frame or image

Video security: In today's advanced world, security is main concern in every fields. In education world it is much more needed to prevent the data from any kind of tempering present in video. One can change the frame content if it is not locked. Still a good cryptographic based algorithm is required to protect the video content at highest level.

There are many algorithms for video security that allows cryptographic operation on it. It may be based on symmetric key based or asymmetric key based.

CONCLUSION

This paper clearly describes the editing issues and possible solution for that. Video editing is much more required in every field of human life like journalism, movie making, online classes etc. It gives not only the descent representation among viewers but also allows the editor to add more effects according to the need of their viewers. There is certain limitation at the capturing side. If any mistake may occur then it will be solved by the editing process. Film industry, animation industries are fully dependent upon the video editing process, whereas, some sectors require mild editing. For example, during the pandemic of COVID-19, the education industry is not in standstill mode. Video are captured and further passes through the editing process before reaching to the student's desk by their schools or college authorities. Editing process can be applied in multiple dimensions.Over all it can be said that video editing is the blessing for the mistakes done at the capturing end.

REFERENCES

1. Neetish Kumar, Deepa Raj. A New Pentagon Search Algorithm for Fast Block-Matching Motion Estimation. International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, 2019, Volume-8 Issue-10.
2. FilliaMakedon, James W. Matthews, Charles B. Owen, Samuel A. Rebelsky. Multimedia authoring, development environments, and digital video editing. 2001 Dartmouth College, Hanover, NH.
3. Abdulhussain SH, Ramli AR, Saripan MI, Mahmmod BM, Al-Haddad SA, Jassim WA. Methods and challenges in shot boundary detection: a review. Entropy. 2018 Apr; 20(4):214.
4. Zhang S, Zhang Y, Chen T, Hall PM, Martin R. Video structure analysis. Tsinghua Science and Technology. 2007 Dec;12(6):714-8.
5. Madhwacharyula CL, Kankanhalli MS, Millhem P. Content based editing of semantic video metadata. In2004 IEEE International Conference on Multimedia and Expo (ICME)(IEEE Cat. No. 04TH8763) 2004 Jun 27 (Vol. 1, pp. 33-36). IEEE.
6. Hua XS, et al. Optimization-based automated home video editing system. IEEE Transactions on circuits and systems for video technology. 2004 May 4; 14(5):572-83.
7. Islam A, Chebil F, Hourunranta A. Efficient algorithms for editing H. 263 and MPEG-4 videos on mobile terminals. In2006 International Conference on Image Processing 2006 Oct 8 (pp. 3181-3184). IEEE.
8. Cuevas C, García N. Temporal segmentation tool for high-quality real-time video editing software. IEEE Transactions on Consumer Electronics. 2012 Sep 27;58(3):917-25.
9. Wang G, Yuan X. The realization of video mixing edit method based on DES. In2010 IEEE International Conference on Software Engineering and Service Sciences 2010 Jul 16 (pp. 419-422). IEEE.
10. Avesh M and Srivastava R (2016) Parametric study on the performance of active suspension system for variable passenger size and repeated road bumps. In: 2016 10th International Conference on Intelligent

ISSN- 2394-5125 VOL 7, ISSUE 09, 2020

Systems and Control (ISCO), Coimbatore, India: pp. 1–6: 7-8 January 2016. DOI: 10.1109/ISCO.2016.7726894. IEEE

Shot Boundary Detection Framework For Video Editing Via Adaptive Thresholds And Gradual Curve Point

Neetish Kumar [a], Deepa Raj [b]
[a]Research Scholar, Department of Computer Science, BBA University (A Central University), Lucknow, India.
[b]Associate Professor, Department of Computer Science, BBA University (A Central University), Lucknow, India.

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 10 May 2021

Abstract: Day to day huge volumes of extended videos entrained from documentaries, cinemas, athletics and surveillance cameras are evolving over video databases and in internet. Processing these videos manually is hard, costly and time-consuming. For extracting these long-duration videos an automatic procedure is desperately needed. As a vital factor the Shot boundary detection (SBD) is considered for lot of video analysis tasks, for example video editing, indexing, summarization and action recognition. In the analysis of video content SBD is considered to be one of the vital task. Based on this, we have presented an effective SBD approach. We have used the gradient and color information for abrupt transition detection. For Gradual transition detection the average edge information of the gradual curves in the sequence of frames are obtained. From the optimal edge detector an average edge frame is gained. The computational complexity is reduced by this approach by processing only the transition regions. The proposed approach when compared to the exiting work done have achieved improved results in terms of precision, recall and F1.

Keywords: Shot boundary detection, Gradient, Luminance, Abrupt, Gradual, Average edge.

1. Introduction

These days, with the rise in digitalization there is a rise in the distribution of multimedia data (particularly video) over the Internet. Owing to this surplus growth, an effective tool is desired for video retrieval and indexing. Properly recognition of the video contends are needed for designing an effective tool therefore the temporal video segmentation is required. SBD of Temporal video is task of video segmentation into significant shots by identifying the transition among sequential frames, and the boundary between two consecutive shots are marked by the transition **(Rashmi, B. S., and H. S. Nagendraswamy 2020)**. Abrupt transition and gradual transition are the two types of transitions. The formerly one is said as an amendment in the video contents in which there is no connection of frames among the margins of two shots. There is a slow change in the content of the frames in the gradual transition, some of the frames intersect in the consecutive shots and the duration of these frames is said as a gradual transition period. It is not easy to detect abrupt transition as it includes lot of complexities such as OCM and fast illumination causing extreme false positive results. The surplus volume of video contends have triggered a considerable desire for effectual schemes that can handle, store and deploy the entire contends **(Liu et al. 2017)**. For attained this, video content are added with their storage (indexing) and then with the help of video structure their basic units are examined. Because of certain video attributes the evaluation of video structure becomes a difficult task. The videos are split into simple elements as per the video structure analysis. There are four levels of video structure i.e. i) frames, ii) shots, iii) scenes, and iv) stories. A shot is considered to be the basic element of video which is believed to be the apt point for content based video indexing and retrieval (CBVIR) **(Youssef et al 2017)**. A hard transition (HT) is formed when two shots are attached directly. In contrast, employing the editing sequences (indirect concatenation) the soft transition is formed.

In videos, HTs are the most prevalent than ST **(Loukas et al. 2016)**. In the past two decades SBD has attracted the attention of many researchers. Which has classes: compressed and uncompressed domains. When compared to compressed domain the latter has gained lot of interest for valuable and fabulous visual information. Though, additional processing time is required by the uncompressed domain-based algorithms owing to the video frames decoding practice **(Singh, Alok et al. 2019)**. Some of the methods employed for feature extraction are Pixel-based, edge-based, histogram-based, and motion-based. Than the time domain, the transform domain is considered for the analysis of SBD. Because the signals are viewed in different domain which is allowed in the transform domain moreover provides a great shift concerning its potent ability for evaluating the components of the signal **(Kar, Tejaswini et al. 2017)**. An image/signal with the help of orthogonal polynomials can be to convert into transform (moment) domain from the time/spatial domain. Scalar quantities such as orthogonal moments and transform coefficients are signified using visual information. The projection of signal are represented by these moments are represent on orthogonal polynomials. The capability of orthogonal polynomials (OPs) are considered by their energy compaction and the localization properties **(Chakraborty, Saptarshi et al. 2019)**. In this work we proposed a SBD in which abrupt transition can be detected using gradient and color information. Luminance distortion and the gradient similarity are evaluated to measure each frames contrast and structural changes. In the abrupt detection phase the adaptive threshold across the videos are used to extract the transition. The Gradual transition is obtained by distribution of average edge information considering the gradual curves. Using the from optimal edge detector the average edge frame is obtained.

The complete organization of the paper is done subsequently: A brief survey about SBD is presented in Sect. 2. Sect. 3 presents the background information of the feature extraction methods employed in this study. Sect. 4 presents the brief description on the proposed SBD system. Analysis and discussion is made in the Sect. 5 which is followed by the Sect. 6 which draws the conclusion.

2. Literature Survey

Over the past years lot of Video SBD methods have been developed. The intention of these SBD techniques is to attain precise results rapidly. For cut transition detection these techniques offers pleasing results but until the last decade the Gradual transition detection endured to be a hitch when certain vital inventions in the field of SBD were attained. The SBD or the temporal video segmentation is a wide topic which is examined by TRECVid **Smeaton et al. (2010)**. By two methods the SBD can be evolved (1) functioning in the compressed domain and (2) uncompressed domain. In compressed domain some of the Features such as motion vectors, DCT coefficients are used for SBD. Color Histograms were presented by **Mas J et al. (2013)** for assessing video frame representation. The consecutive frames pixel intensities for Video SBD. A Singular Value Decomposition (SVD) approach was proposed by **Cernekova et al. (2007)** which was used for major feature values extraction from the non- significant values. The execution time was minimized by this approach neglecting the non-significant feature values. **(Ren et al. 2009)** proposed a SBD approach in MPEG videos using the global and local features of the frames. A motion vector was employed to detect shot transitions in the form of motion prediction error. **Adjeroh et al. (2009)** in proposed an edge based method that used multiple multilevel features for SBD. The limitation of this work is the requirement of major computational power. A Candidate Segment selection approach was proposed by **Li et al. (2009)** as a pre-processing step at this time a main development was attained in 2009 in decreasing the execution time of SBD techniques.For SBD **Lu et al (2013)** used Pattern Matching and SVD. A similar candidate segment selection was employed in this technique but with abetter adaptive threshold calculation. For SBD **GG et al. (2014)** proposed a hybrid system using color histogram and Gist. A frame transition parameters was proposed by **Tippaya et al. (2015)** combining global feature and local feature in which to classify the type of transitions a neural network is used. **Mas et al. (2003)** proposed a frame skipping approach for SBD which used an adaptive thresholding adopting a preprocessing technique. Gradual transitions were detected deploying a triangle pattern matching approach. To find out the transitions **Liang et al. (2017)** proposed a SBD using foveation technique calculating the local attention consistency measure. **Bi, Chongke et al. (2018)** to detect SBD energy, used moment and contrast in which using co-occurrence matrix the texture features are computed. **Chawla et al. (2018)** used the entropy and SURF feature for SBD in which the frames entropy is computed using the intensity histogram.

3. Background Information

The proposed feature extraction process is explained in this section.

3.1 Gradient Similarity

Information from images can be extracted using Image gradients **(Liu, Anmin et al. 2011)**. The change in intensity is measured by every pixel of a gradient image of that same point in a given direction of the original image. The gradient images in the *x* and *y* directions are calculated to attain the absolute range of direction. The gradient similarity presented is well-defined in Eqn 1.

$$G(x,y)=\frac{2(1-R)+T}{1+(1-R)^2+T} \tag{1}$$

Where

$$T=\frac{T'}{MAX(G_x,G_y)}$$

$$R=\frac{|G_x-G_y|}{MAX(G_x,G_y)}$$

The gradient values of the image blocks *x* and *y*, are signified as G_x and G_y. *G(x, y)* is the gradient resemblance among *x* and *y* and its values in range of [0, 1]. The Gradient value G_x (similar for G_y) is found as the supreme weighted average of difference for every block in an image. The value of T is fixed at 200.

3.2 Luminance Similarity

The visible distortion occurs due to the luminance changes moreover the structure also changes they are not so frustrating **(Sharma et al. 2005)**. Using the Eqn 2 the Luminance similarity is described.

$$E(x_i, y_i) = 1 - \left(\frac{x_i - y_i}{L_u} \right)^2 \tag{2}$$

In image blocks x, y the pixels at position i are represented as x_i, y_i, the pixel values dynamic range is denoted as L_u. Among the image pixels x_i and y_i within the range of [0, 1] the luminance similarity is represented as $E(x_i, y_i)$.

In gradient and luminance similarities the general form of integration for an image pixel pair *xi* and *yi* to obtain the overall quality indicator $Q(x_i, y_i)$ which is defined as:

$$Q(x_i, y_i) = (1 - W(G, E)).G + W(G, E).E \tag{3}$$

The shortened forms of $Q(x_i, y_i)$, $G(x_i, y_i)$ and $E(x_i, y_i)$ are denoted as Q, G, and E respectively. The relative importance of the two components are adjusted using the weighting function W(G, E).Using Eqn 4 W(G, E)is computed.

$$W(G, E) = P.G \tag{4}$$

The positive weighting parameter is denoted as 'P'. Since P also has to lesser than 0.5 and Gis in the range of [0, 1]. In this paper the value of Pis taken as 0.1.

3.3 CIEDE 2000 Colour Variance

The CIELAB colour space is the main facet behind the colour-difference formula of CIEDE2000 [22]. Eqn 5 defines the CIEDE2000 colour difference.

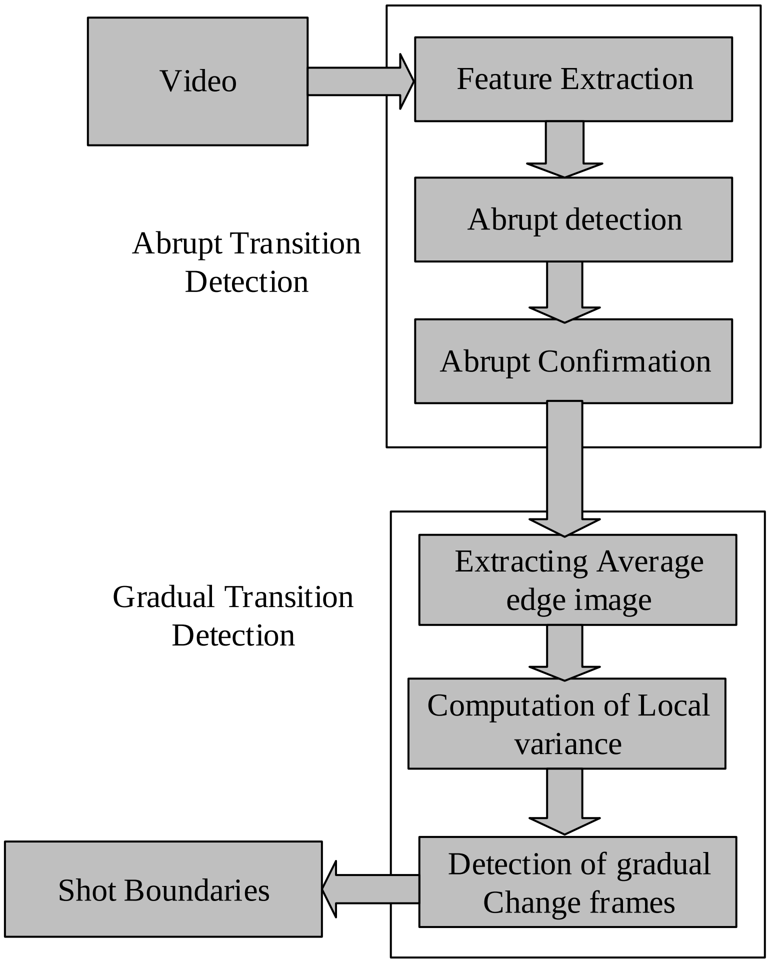
$$\Delta F = \Delta F(l_1^*, a_1^*, b_1^*, l_2^*, a_2^*, b_2^*) = \sqrt{\left(\frac{\Delta l'}{k_L s_L}\right)^2 + \left(\frac{\Delta c'}{k_C s_C}\right)^2 + \left(\frac{\Delta h'}{k_H s_H}\right)^2 + R_T \left(\frac{\Delta c'}{k_C s_C}\right)^2 + \left(\frac{\Delta h'}{k_H s_H}\right)^2} \tag{5}$$

In CIEDE2000 the differences in lightness, chroma, and hue are represented as l, c and h. The rotation function is denoted as R_T and the blue regionis responsible for the interaction between chroma and hue differences. In CIEDE2000 five corrections on CIELAB have been presented and they are lightness (s_L), chroma (s_c), the hue (s_H) weighting functions. The CIEDE 2000 Colour difference is applied in a Video Sequence as a rate to find the relationship across the frames.

4. Proposed Methodology

In this section the proposed system is explained in detail. Fig. 1illustrates the steps of the proposed system. The proposed system comprises of the feature extraction, thresholding, gradual and abrupt transition detection.

Figure.1 The proposed block Diagram



4.1 Feature extraction

It is the initial procedure of the proposed system in which the features are extracted from the equivalent frames such as Luminance Similarity Comparison (*E*) and Gradient Similarity Measurement (*G*) using Eqn (1) and (4). Combining these two features the equivalent frames quality similarity Measure (*q*) is found using (3). The features *Q* and *E* are respectively used to find the possible abrupt transition detection stage. The abrupt sections Lab colour difference is used in the confirmation stage.

4.2 Thresholding

Among the sequential frames a threshold is used for clarifying with the abrupt or gradual changes. Beyond a certain threshold when there is a distance concerning the sequential frames then a transition is declared. It is vital to choose an apt threshold for ensuring high accuracy and for finding the both changes. A possible abrupt threshold (α) and final threshold (λ) are used for identifying transitions in a video and they are represented in Eqn (6) and (7).

$$\alpha = \mu_q + \left(C_1 \times \sigma_q\right) \tag{6}$$

$$\lambda = \mu_{\Delta E} - \left(C_2 \times \sigma_{\Delta E}\right) \tag{7}$$

The possible abrupt transition is detected using the threshold α and in the abrupt confirmation stage λ is used. C_1 used in Eqn 6 is a constant and the value lying in between the range of [-3.2, -2.8] is selected likewise in Eqn 7 the constant C_2 having a range of [1.6, 2].

4.3 Abrupt section

Generally this section is categorized into two phases i.e., Probable Abrupt Detection Stage and Abrupt Affirmation Stage.

4.3.1 Probable abrupt detection Phase

All the frames in a video of this phase is classed into two types i.e., probable transition frames and non-transition frames. This phase is involved to confirm that the transition like behaving frames are retained and rest are rejected. The quality indicator (Q) is used for the classification which is the combination of gradient and luminance resemblance and is computed using Eqn 3. As given in Eqn 10 the probable abrupt frames (PA) are found using the probable abrupt threshold (α).

The irregular and the former frames of fPA_i are given by $fPA_{i\pm\eta}$. In Lab colour space fPA_i is the probable transition frame. Experimentally η is set as 2 which is a user defined constant.

$$PA = \begin{cases} \text{Pr}\textit{obable} \text{ tran., if } q_i \geq \alpha \\ non-tran., otherwise \end{cases} \tag{8}$$

4.3.2 Abrupt confirmation phase

The probable abrupt transition frames (fPA_i) in this stage is stated as actual abrupt transition (A) based on the Eqn 9.

$$A = \begin{cases} True,\ if\ \Delta E\left(\text{fPA}_{i-\eta}, fPA_{i+\eta}\right) \geq \lambda \\ False, \qquad otherwise \end{cases} \tag{9}$$

Among the frames fPA(i-η) and fPA(i+η) the CIEDE2000 colour difference is E(fPA(i-η), fPA(i+η)). The $\pm\eta^{th}$ frames from the i^{th} frame is indicated by *η which is a* constant. For evaluation the value of *η* is considered as 2.

4.4 Gradual transition detection phase

In this phase using the gradual curves which are considered by the dissemination of average edge information the Gradual transition detection is done

4.4.1 Average edge image extraction

An average edge image was built for each frame. It is found that than the average intensity of the pixels contain more intensities. Than the original image these images are smoother and distinct. The average edge image found with some changes was stated from effect average gradient (EAG). The average edge of image is obtained using the steps mentioned below.

1. Conversion of a color image into a gray image.
2. with the threshold 100 attain the edge image using optimal edge detector
3. Using equation (10) calculate the average gradient (AG)

$$A_G = \sum_{x,y} F(Gradient)(X,Y) / \sum_{x,y} R(X,Y) \tag{10}$$

Where $R(X,Y) = 1, if\ \text{F(Gradient)(X,Y)} > 0$

$R(X,Y) = 0, if\ \text{F(Gradient)(X,Y)} = 0$

4. Based on the average gray (A_G) value Obtain an average edge image; using equation (11) consider a threshold value which is new.

$$F(X,Y) = \begin{cases} F\ \text{Gradient(X,Y)}\ if\ \text{F Gradient(X,Y)} < A_G \\ 0, \qquad \text{if F Gradient(X,Y)} \leq A_G \end{cases} \tag{11}$$

4.4.2 Computation of gradual point

Based on the average gradient image detection of Gradual point is done. Initially in every 20 frames the variance was computed. To analyze the gradual change twenty frames were considered. For gradual sequence the variance is almost the same if it resembles as shot sequence. Based on equations (12) and (13) the variance is calculated.

$$VAR(x) = \frac{1}{T-1} \sum_{k=x}^{x+T-1} \left(AGI(k) - mean(k)\right)^2 \tag{12}$$

$$Mean(x) = \frac{1}{T}\sum_{k=i}^{x+T-1} AGI(k) \quad (13)$$

Where x = 1, 2 ...n – T+1

In a window the total number of frames is denoted as T

In the kth window AGI (k) is the variance of the AGI.

4.4.3 Detection of the gradual change point

One of the frames in the local parabolic sequence has a least value. The gradual change point was attained by analysing the width and depth of the sequence. A gradual change point is confirmed once the equations (14) and (15) are fulfilled.

$$D_{VAR} = |\omega LocMax[i \pm 1] - \omega LocMin[i] > 0 \quad (14)$$

$$D_{FRAME} = |FrmLocMax[i \pm 1] - FrmLocMax[i] < 20 \quad (15)$$

Where, i = 1, 2, 3...n

The variances of ith frame are denoted as $\omega LocalMin[i]$ and $\omega LocalMax[i]$ having the local minimum and maximum. The has local maximum frame number id denoted as $FrmLocMax[i]$

5. Experimental results and discussion

This section provides a detailed explanation regarding the database used and the experimentation of the employed approach.

5.1 Dataset

The dataset used for experimentation is the TRECvid 2007 video databases for temporal video segmentation. This benchmark dataset which is provided by *Netherlands Institute for Sound and Vision*. Compressed MPEG videos are used for this work.

5.2 Performance analysis

The parameters considered for the analysis of the proposed approach are Precision (P_{re}), Recall (R_{ec}) and F1 Score (F_1) factors, which are computed by means of Equations (16), (17) and (18).

$$R_{ec} = \frac{NC}{NC + NM} \times 100 \quad (16)$$

$$P_{re} = \frac{NC}{NC + NF} \times 100 \quad (17)$$

$$F_1 = \frac{2 * R_{ec} * P_{re}}{R_{ec} + P_{re}} \quad (18)$$

In a video the transitions which are correctly, missed and falsely detected are represented as *NC*, *NM* and *NF*. *Using the* chosen videos of TRECVid 2007 experimentation is conducted.

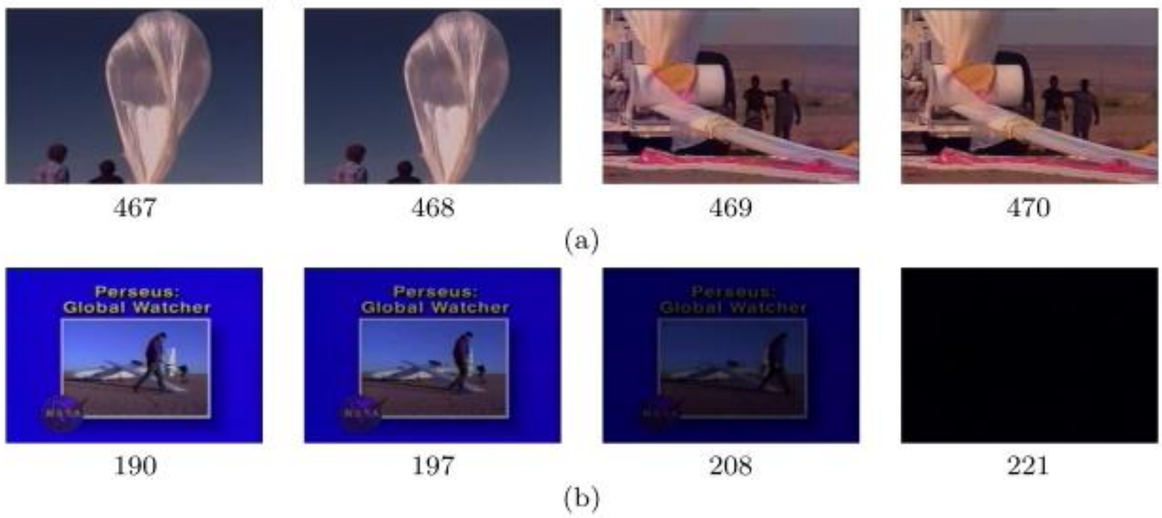
Table 1 provides the performance of some of the videos that are collected. Each videos computation time is provided in the Table 1. The correctly detected transitions from the *NAD*58.*mpg* video is represented in Figure 2.Using the videos the proposed systems average F_1 scores are 94.5% and 80.9% for both the transitionswith an average overall performance of 91.6%

Table. 1. Results of the Proposed System for TRECVid 2007 Database

Video	Computation time(sec)	Abrupt			Gradual			Overall		
		Rec	**Pre**	**F1**	**Rec**	**Pre**	**F1**	**Rec**	**Pre**	**F1**
BG 3027	1630	84.3	100.0	92.2	100.0	51.0	65.7	95.9	92.2	95.0
BG 3097	1535	86.9	100.0	92.7	-	-	-	88.9	100.0	92.6
BG 3314	1155	84	100.0	89.9				84	100.0	90.0
BG 16336	84	92.0	100.0	93.7				96.0	100.0	96.4

BG 37309	305	100.0	100.0	100.0	85.5	62.6	74.5	95.5	82	88
BG 37770	506	100.0	100.0	100.0	91	78.4	87.3	94	84	89
BG 22677	507	97.2	100.0	99	84.4	87.3	88.8	92.3	95	92
BG 36658	871	94.2	97.2	96	86	89	87	92.7	95.4	94.0
BG 8947	545	94.6	97.3	95.9	72.2	100.0	83.9	90.3	97.7	92.8
BG 4455	772	95.2	98.6	96.9	85.3	90	88	93	96	94
BG 35153	872	92.0	100.0	94.2	87	82.0	81.6	87.2	91	91
Clip	198	89.5	100.0	94.2				87.5	100.0	94
Average	**664.0**	**93**	**97**	**96**	**89.2**	**81.3**	**83.2**	**91.5**	**92.6**	**92.8**

Figure.2 illustration of the identified **a** abrupt transition and **b** gradual transition



To validate proposed system, three SBD techniques proposed by the authors in eigen value decomposition and Gaussian transition detection method **(Amiri et al. 2012)**, temporal segmentation method **(E Santos et al. 2017)** and 3D convolutional networks method (**Liu et al. 2017)** are considered.

Table. 2. Summary of the assessment of the proposed scheme with the existing works

Process	**Evaluation parameter**	**videos**				**Average**
		Anni006	**Anni009**	**Anni010**	**NAD58**	
Proposed	R_{ec}	83.9	85.5	87.3	90.2	86.5
	P_{re}	77.3	90.2	93.6	94.1	91.7
	F_1	83.6	87.5	92.4	92.5	89.6
Eigen value decomposition and Gaussian transition detection method	R_{ec}	93.8	84.5	91.2	93	92.3

	P_{re}	85.2	82.1	78.4	92	84.5
	F_1	89	81.5	83.5	91.4	87
Temporal segmentation method	R_{ec}	84.3	87.9	88.5	92.8	89.1
	P_{re}	92.2	86	84.2	91.5	86.7
	F_1	91.1	84.2	85.0	93.7	89
3D convolutional networks method	R_{ec}	92.8	94.3	84.5	90.2	91.7
	P_{re}	95.5	80.6	85.6	90.6	87.3
	F_1	90.2	88.2	88.9	92.1	89.7

Moreover, for gradual detection the obtained percentage of **85.7%** is high. The F_1 score as per the table 1 endure as high as **91.2%.** Table 2 illustrates the assessment made among the proposed system and the existing methods.

6. Conclusion and future suggestions

A SBD approach is employed in this study for the detection of abrupt and gradual transitions. The abrupt transition is detected using adaptive threshold (α).In the confirmation phase, the similarity features (gradient and luminance) are used. Along with the adaptive threshold (λ), the lab features (λE) are used to analyse the frames. For detecting the gradual transition the gradual curve was obtained. For this initially the average edge image was found. Amongst sequences and local variance the minimum difference were considered. The gradual change point finally was identified. The presented high-quality real-time results proves that our proposed system is more enhanced that the Existing shot detection approaches for video editing based SBD which needs quality and where speed.

References

1. Adjeroh, Don, M. C. Lee, Nagamani Banda, and Uma Kandaswamy. (2009). Adaptive edge-oriented shot boundary detection. EURASIP Journal on Image and Video Processing 2009, no. 1. 859371.
2. Amiri, Ali, and Mahmood Fathy. (2012). Video shot boundary detection using generalized eigenvalue decomposition and Gaussian transition detection. Computing and informatics 30, no. 3: 595-619.
3. Bi, Chongke, Ye Yuan, Jiawan Zhang, Yun Shi, Yiqing Xiang, Yuehuan Wang, and RongHui Zhang. (2018) Dynamic mode decomposition based video shot detection. IEEE Access 6: 21397-21407.
4. Chakraborty, Saptarshi, and Dalton Meitei Thounaojam. (2019). A novel shot boundary detection system using hybrid optimization technique. Applied Intelligence 49, no. 9: 3207-3220.
5. Chawla, Rashmi, Poonam Singal, and Amit Kumar Garg. (2018). A Mamdani fuzzy logic system to enhance solar cell micro-cracks image processing. 3D Research 9, no. 3: 34.
6. E Santos, Anderson Carlos Sousa, and Helio Pedrini. (2017) Shot boundary detection for video temporal segmentation based on the weber local descriptor. In 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1310-1315. IEEE, 2017.
7. GG, Lakshmi Priya, and S. Domnic. (2014). Walsh–Hadamard transform kernel-based feature vector for shot boundary detection. IEEE Transactions on Image Processing 23, no. 12 : 5187-5197.
8. Kar, Tejaswini, and Priyadarshi Kanungo. (2017). A motion and illumination resilient framework for automatic shot boundary detection." *Signal, Image and Video Processing* 11, no. 7: 1237-1244.
9. Li, Y-N., Z-M. Lu, and X-M. Niu. (2009). Fast video shot boundary detection framework employing pre-processing techniques. IET image processing 3, no. 3: 121-134.
10. Liang, Rui, Qingxin Zhu, Honglei Wei, and Shujiao Liao. (2017). A video shot boundary detection approach based on CNN feature. In 2017 IEEE International Symposium on Multimedia (ISM), pp. 489-494. IEEE.
11. Liu, Anmin, Weisi Lin, and Manish Narwaria. (2011).Image quality assessment based on gradient similarity. IEEE Transactions on Image Processing 21, no. 4 : 1500-1512.

12. Liu, Tingxi, Yao Lu, Xiaoyu Lei, Lijing Zhang, Haoyu Wang, Wei Huang, and Zijian Wang. (2017). Soccer video event detection using 3D convolutional networks and shot boundary detection via deep feature distance. In International Conference on Neural Information Processing, pp. 440-449. Springer, Cham.
13. Loukas, Constantinos, Nikolaos Nikiteas, Dimitrios Schizas, and Evangelos Georgiou. (2016). Shot boundary detection in endoscopic surgery videos using a variational Bayesian framework. International journal of computer assisted radiology and surgery 11, no. 11, 1937-1949.
14. Lu, Zhe-Ming, and Yong Shi. (2013). Fast video shot boundary detection based on SVD and pattern matching. IEEE Transactions on Image processing 22, no. 12: 5136-5145.
15. Mas J, Fernandez G (2003) Video shot boundary detection based on color histogram, in Proc. TRECVIDCernekova, Zuzana, Constantine L. Kotropoulos, and Ioannis Pitas. Video shot-boundary detection using singular-value decomposition and statistical tests. Journal of Electronic Imaging 16, no. 4 (2007): 043012.
16. Mas, Jordi, and Gabriel Fernandez. "Video Shot Boundary Detection Based on Color Histogram. 2003. In *TRECVID*. .
17. Rashmi, B. S., and H. S. Nagendraswamy. (2020). Video shot boundary detection using block based cumulative approach. Multimedia Tools and Applications : 1-24.
18. Ren, Jinchang, Jianmin Jiang, and Juan Chen. (2009) Shot boundary detection in MPEG videos using local and global indicators. IEEE Transactions on Circuits and Systems for Video Technology 19, no. 8: 1234- 1238.
19. Sharma, G., Wu, W., & Dalal, E. N. (2005). The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur, 30(1), 21-30.
20. Singh, Alok, Dalton Meitei Thounaojam, and Saptarshi Chakraborty. (2019). A novel automatic shot boundary detection algorithm: robust to illumination and motion effect. Signal, Image and Video Processing : 1-9.
21. Smeaton, Alan F., Paul Over, and Aiden R. Doherty. (2010). Video shot boundary detection: Seven years of TRECVid activity. Computer Vision and Image Understanding 114, no. 4: 411-418.
22. Tippaya, Sawitchaya, Suchada Sitjongsataporn, Tele Tan, Kosin Chamnongthai, and Masood Khan. (2015). Video shot boundary detection based on candidate segment selection and transition pattern analysis. In 2015 IEEE International Conference on Digital Signal Processing (DSP), pp. 1025-1029. IEEE.
23. Youssef, Bendraou, Essannouni Fedwa, Aboutajdine Driss, and Salam Ahmed. (2017). Shot boundary detection via adaptive low rank and svd-updating. Computer Vision and Image Understanding 161: 20-28.