MARCH 2016 Ph.D in Electrical and Electronics Engineering SHAIMA SAFA ALDIN BAHA ALDIN

UNIVERSITY OF GAZİANTEP

GRADUATE SCHOOL OF

NATURAL & APPLIED SCIENCES

DESIGN AND IMPLEMENTATION OF A TELECONFERENCE SYSTEM USING AN IMPROVED HEVC CODEC

Ph.D THESIS

IN

ELECTRICAL AND ELECTRONICS ENGINEERING

BY

SHAIMA SAFA ALDIN BAHA ALDIN

MARCH 2016

Design and Implementation of a Teleconference System Using an Improved HEVC Codec

Ph.D Thesis

in

Electrical and Electronics Engineering

University of Gaziantep

Supervisor

Prof. Dr. Ergun ERÇELEBİ

by

Shaima Safa aldin BAHA ALDIN

March 2016

REPUBLIC OF TURKEY
UNIVERSITY OF GAZİANTEP
GRADUATE SCHOOL OF NATURAL & APPLIED SCIENCES
ELECTRICAL AND ELECTRONICS ENGINEERING

Name of the thesis: Design &Implementation of a Teleconference System using an Improved HEVC codec

Name of the student: Shaima Safa Aldin BAHA ALDIN

Exam date: 1/3/2016

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Metin BEDİR
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

E. Erçelebi

Prof. Dr. Ergun ERÇELEBİ
Head of Department

This is to certify that we have read this thesis and that in our consensus/majority opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

E. Erçelebi

Prof. Dr. Ergun ERÇELEBİ
Supervisor

Examining Committee Members:	Signature
Prof. Dr. Ergun ERÇELEBİ	E. Erçelebi
Prof. Dr. Ulus ÇEVİK	....
Assist. Prof. Dr. Sema KAYHAN	....
Assist. Prof. Dr. Şaban YILMAZ	....
Assist. Prof. Dr. A. Mete VURAL	....

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these r

ules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Shaima Safa aldin BAHA ALDIN

ABSTRACT

DESIGN AND IMPLEMENTATION OF A TELECONFERENCE SYSTEM USING AN IMPROVED HEVC CODEC

BAHA ALDIN, Shaima Safa Aldin
Ph.D. in Electrical and Electronics Engineering
Supervisor: Prof. Dr. Ergun ERÇELEBİ
March 2016
151 pages

Teleconferencing has become an indispensable element in any business system, because it offers the opportunity for collaborators to participate in a virtual group while remaining in divergent regions. Teleconferencing also increases productivity, minimizes travel expenses and saves travel time. This thesis presents a reliable Teleconference system that utilizes an improved high-efficiency video codec (HEVC) H.265 technology with a congestion control. The improvement is based on Coding Unit (CU) size decision and entropy coding, which provides an adequate approach to enhance the real-time video/IP technology in terms of improved video quality and increased compression ratio compared to the previous codec (H.264) and original HEVC. In the thesis, we proposed fast algorithms that reduce encoder complexity in a way more suitable for a real-time processing. Our system ensured that voice, video, text and other control packets are streamed efficiently to various individuals with smooth real-time communication. This thesis proposed a solution for a qualified server/client teleconference system. The system is an open, expandable platform and can work efficiently in an intranet/internet environment.

Key Words: H.264, H.265, video/IP, Entropy, Intra-Mode.

ÖZET

TASARIM VE TELEKONFERANS SISTEMI UYGULAMASI BİR GELİŞMİŞ HEVC CODEC KULLANMA

BAHA ALDIN, Shaima Safa Aldin
Doktora Tezi, Elektrik-Elektronik Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Ergun ERÇELEBİ
Mart 2016
151 sayfa

Telekonferans herhangi bir iş sisteminde vazgeçilmez unsur haline gelmiştir çünkü paydaşlar farklı bölgelerde bulunurken telekonferans paydaşlara sanal gruba katılmak için fırsat sunmaktadır. Ayrıca, telekonferans verimliliği artırır, seyahat masraflarını en aza indirir ve seyahat süresinden tasarruf sağlar. Bu tez, tıkanıklığının kontrolü ile geliştirilmiş yüksek verimli video kodek (HEVC) H.265 teknolojisi kullanan güvenilir bir telekonferans sistemi sunmaktadır. HEVC video kodlayıcıdaki geliştirme, kodlama ünitesi boyut kararına ve entropi kodlamaya dayalı olup önceki video kodlayıcı H.264 ve orijinal HEVC kodlayıcı ile kıyaslandığında iyileştirilmiş video kalitesi ve artırılmış sıkıştırma oranı açısından iyileştirilmiş gerçek zamanlı videoya/IP teknolojiye daha uygun yaklaşım sağlamaktadır. Tezde, gerçek zamanlı işlemler için daha uygun kodlayıcı karmaşıklığını azaltan hızlı algoritmalar önerdik. Sistemimiz düzgün gerçek zamanlı iletişim ile ses, video, metin ve diğer kontrol paketlerinin verimli şekilde farklı kişilere akışını sağlamıştır. Bu tez, kaliteli sunucu/istemci telekonferans sistemi için bir çözüm önermiştir. Önerilen sistem açık, genişletilebilir bir platformdur ve intranet / internet ortamında verimli olarak çalışabilir.

Anahtar Kelimeler: H.264, H.265 video / IP, Entropi, içi-Mod.

DEDICATION

The Thesis is dedicated to the sake of ALLAH (my Creator and Master), for giving me the strength to carry on this research and for blessing me with many great people who have been a greatest support in both my personal and practical life. To my great teacher and messenger Mohammed (peace be upon him), who taught us the purpose of life. To my parents, who have always loved me and trusted me unconditionally. Thanks for providing everything we need, for believing in us, giving the tools to succeed and whose good examples have taught me to work hard for the things I aspire to achieve. Thank you father for supporting me, even when I am wrong. Thanks for protecting me, and make me strong. Thank you mother for your guidance as we move along, for teaching us the right from wrong. To my husband Hassan, who leads me through the valley of darkness with lights of hope, kindness and support. You were always making my dreams your own, and ensured that I have never been sad or forlorn. I am truly thankful for having you in my life. To my beloved sisters, who represent the symbol of love and giving; particularly my dearest sister Noor, who stands by me when life look bleak. Thanks for watching out when things don't seem fine. Thanks for pulling me up, when I'm out of line.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor **Prof. Ergun Erçelebi** for the continuous support of my Ph.D research. Thanks for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of my research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBRIVATIONS

PSTN	Public Switched Telecommunications Network
LAN	Local Area Network
QoS	Quality of Service
SNR	Signal-to-Noise Ratio
RAS	Reliability, Availability, and Serviceability
IEEE	Institute of Electrical and Electronics Engineers
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
NVR	Network Video Recorder
DVR	Digital Video Recorder
fps	Frames Per Second
SAN	Storage Area Network
VoIP	Voice over IP
MCU	Multipoint Control Unit
IP	Internet Protocol
SMTP	Simple Mail Transfer Protocol
HTTP	Hypertext Transfer Protocol
URL	Uniform Resource Locator
DNS	Domain Name System
TTL	Time-To-Live
DHCP	Dynamic Host Configuration Protocol
NAT	Network Address Translation
CSMA/CD	Carrier Sense Multiple Access with Collision Detection
MAC	Media Access Control
IANA	Internet Assigned Numbers Authority
RGB	Red, Green and Blue
VBI	Vertical Blanking Interval
LPH	Lines per Picture Height

CRT	Cathode Ray Tube
LCD	Liquid Crystal Display
HVS	Human Visual System
DVD	Digital Versatile Disk
CD	Compact Disc
AES/EBU	Audio Engineering Society/European Broadcasting Union
HEVC	High Efficiency Video Codec
CTU	Coding Tree Unit
CTB	Coding Tree Block
CU	Coding Unit
CB	Coding Block
TU	Transform Unit
PU	Prediction Unit
LCU	Largest Coding Unit
DCT	Discrete Cosine Transform
IDCT	Inverse Discrete Cosine Transform
FDCT	forward Discrete Cosine Transform
CB	Chroma Block
CABAC	Context-Adaptive Binary Arithmetic Coding
RDO	Rate Distortion Optimization
PCM	Pulse Code Modulation
RAU	Resource Allocation Unit
OpenAL	Open Audio Library
OS	Operating System
GUI	Graphical User Interface
API	Application Program Interface
QP	Quantization Parameter (Factor)

CHAPTER 1

INTRODUCTION

1.1. Overview

Nowadays the telecommunication defined as the world's profitable industry. The number of subscribers to telecommunication services, including mobile and PC clients, are 1800 million. By the year 2000, the Annual expenses on telecommunications reached approximately to 900,000 million dollars. It had a great fleet of transportation, employees, and the highest income. The Western Electric Company has the 7^{th} rank on the Fortune 500 in 1982, which represents the arm of Bell System manufacturing [1].

Also, the Bell system was at the top of the list on the Fortune 100 Utilities. Thus, it is at the head of the list. The telecommunication is a great business. It is defined as "the communications at a distance" by the Webster, "the transference of information over a far distance" by the IEEE dictionary such as by telegraph, radio or TV, "the electrical communication" as a usual, descriptive definition

The local telephone company is one of the best examples for the telecommunication that deals only with a voice telephony system. Telecommunication involves the transmission of speech, text, and picture information at a distance. It consists of four main types of medium: coaxial, wire-pair, fibre optics and radio.

In Productive countries, the telephone is regarded as a way of life. It is linked to the PublicSwitchedTelecommunicationNetwork (PSTN) for voice communication locally, nationally, and internationally. Picture and data information may also be carried over these lines, such as TV. Recently, the PC has started to undertake a like function as the telephone. Nowadays they become a mutual and couple system.

Usually, the PC uses a telephone line to have an internet and e-mail utilities. The Radio has started, assisted by the PC, to offer alike services as an adjunct to the telephone such as internet and fax, in addition to the voice. It is defined as the wireless, the opposite of wire. There are many devices that operate on such a service, such as TV remote controls, CD player controllers, cordless telephones, door openers, remote radio pagers and VCR.

Months or years are the period that someone should be waiting for, in order to get a potential subscriber to a telephone in some countries. Cellular radio gave a solution to get rid of this issue. There is an increasing traffic of data communication within the PSTN, in which the network works as a data-channel. The dial-up links may utilize the PSTN circuits. Naturally, the Internet gave an extra stimulus to the usage of PSTN circuit. One of the main examples of PSTN usages is the traffic supplement of TV Conference.

In the recent years, the clients tend to bypass the PSTN either in a partial or in a complete manner, by using the satellite links, or reserving a limited capacity from some other provider; which could be a power company with surplus capacity on its fibre-optic or microwave system.

1.2. Telecommunication Concepts

The PSTN includes LANs connected by a far-distant net. Figure 1.1 illustrates the PSTN notion. Although, there is a clear inclination toward privatization, the PSTN is still managed by a government authority generally. It has become a commercial enterprise in the USA since its beginning.

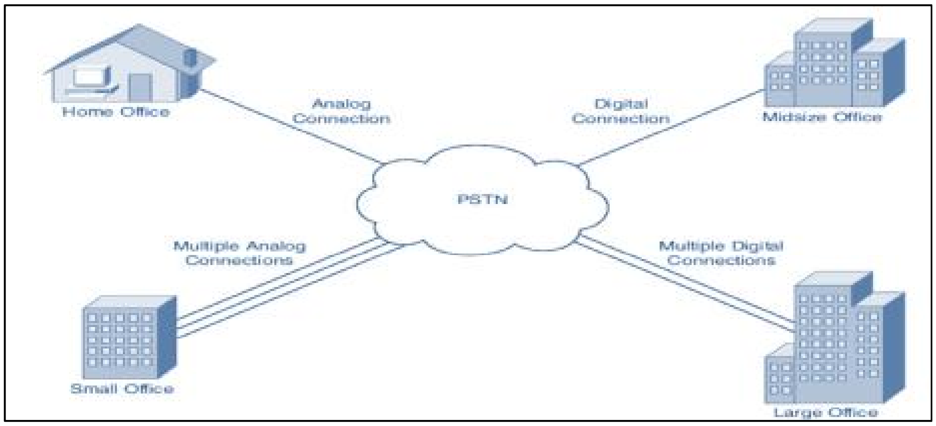


Figure 1.1 networks connected through the PSTN

1.2.1. End Users, Nodes, and Connectivity

End-users are the network inputs and receivers of network outputs. They represent the network I/O such as a PC, telephone, fax, or conference TV equipment. They usually connect to nodes.

The node defined as a point or connection-link in a communication system. It performs a switching function. While it is considered as a network interface unit in the case of a LAN; in which multiple end users connected together through it.

The way of connecting end users together through nodes is called the Connectivity. Figure 1.2 demonstrates these notions.

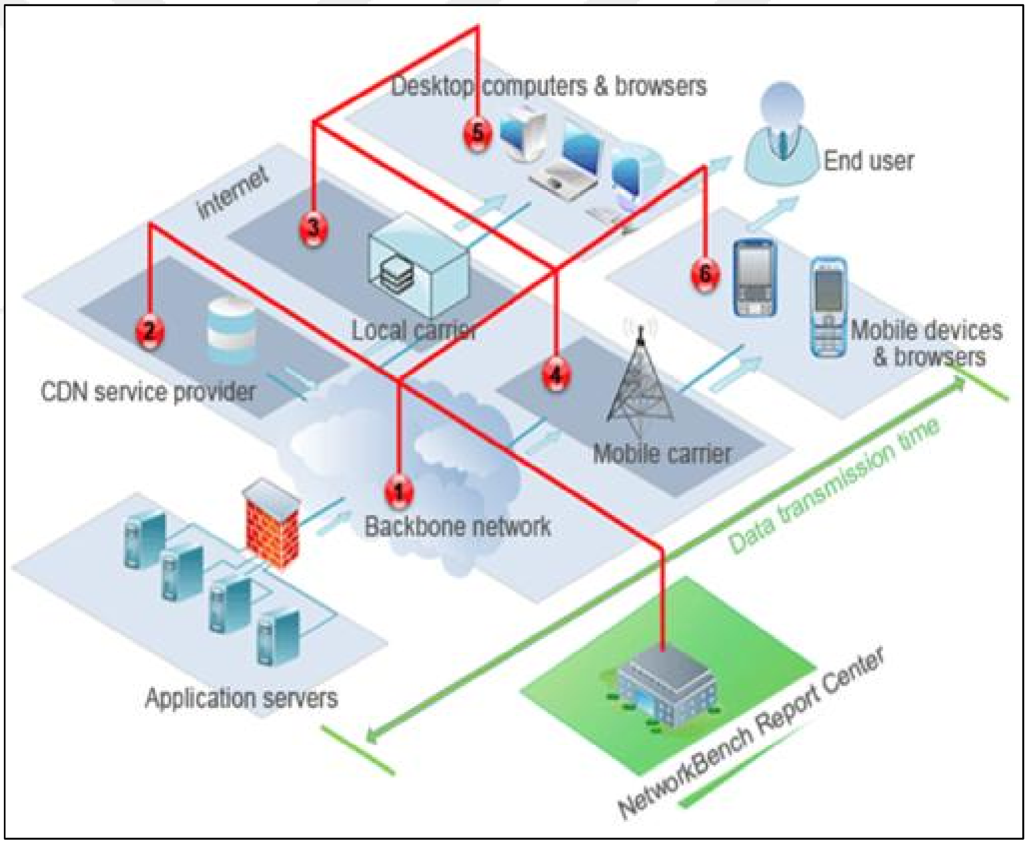


Figure 1.2 End-User, Node, a

nd Connectivity’s illustration

It is also defined as “the union of connections, switching circuits, and other active units prepared for a data transfer between multiple points in a telecommunications network” by IEEE.

1.2.2. Quality of Service (QoS)

The QoS may represent the amount of happiness of a telephone subscriber for using a specific service. System designers should always consider QoS in their design. QoS means how the carrier company keeps its customers happy. For example, the customer may spend his time on dialling; because of the bad call, no tone or cannot hear the other side. All the mentioned issues have an impact on the QoS. Thus the QoS is an influential factor in multiple fields of the telecommunication services. In the past days of tele-graphy, the amount of received service messages, at the switching centre, was the rough measure for the QoS, while it represents the observation of service in the recent telephony. From the perspective of a transmission engineer, it represents the satisfaction of a client, which is generally determined by how well the client can hear the others, as a loudness rating (measured in decibels (dB)). It also represents the percentage of missing calls, from the perspective of network and switching point. Therefore, this notion defines the service's grade. For example, a service has a high QoS, if it loses 1 call per 100 calls during one busy hour. There are many elements listed under QoS, such as:

- Dial-tone & post dialling delay.
- Availability of service tones.
- Billing’s Correctness.
- Acceptable service's cost to the client;
- The quickness in reacting to service demands;
- Operator's Courtesy;
- The time for setting a new telephone and additional services given by the telephone company.

In another word, the design of a telecommunication system depends on the QoS of its tools. In the transmission scope, the SignaltoNoiseRatio (S/ N or SNR) is the coefficient used for determining the signal's quality. SNR measures the incremental amount in dB of signal level to the noise level in a specified bandwidth. The transmission channel has a high quality as long as it has a minimum S/N ratio.

1.2.3. Reliability, Scalability and Availability

Reliability is a factor of any computer-related component that works regularly in accordance with its requirements. It is one of the three main factors (Reliability, Availability, and Serviceability/RAS) that are considered when buying, designing or using a computer product [1].

In theory, reliability means a free error system, while it means a product's quotient as a percentage in practice. The increasing reliability of evolutionary products returns to the solved issues in previous releases. For example, IBM's z/OS has a high reliability's reputation, due to the improvements of its products since the old versions.

IEEE sponsored the IEEE-ReliabilitySociety (IEEE-RS), which is a company dedicated to reliability in engineering. It confirms industry-wide approval of a systematic approach to assist in assuring reliable services. So, it insures the reliability in the analysis, engineering and maintenance as well. The sharing of information and mutual effort are supported by the members of the community, such as individuals and companies of all engineering fields. The perfect security supplier should have a reliability and capability of scoping, installing, and fixing systems. Because of the modular architecture (Expandable and Upgradable) of IP video systems, it is important to choose a supplier that will be available for a long period,

Scalability is the capability of a computer application or component to keep on working well as it is changed in size to satisfy a user need. In a telecommunication system, the provider/supplier should be capable of configuring expandable Teleconferences. Moreover, it must have a capability of configuring; in order to be compatible with other security techniques.

Availability is the ratio between the operational and elapsed time in a telephone circuit; while it is the accessibility of i/p and o/p ports on a network switching system.

1.2.4. Real-time vs. Non Real-time

When someone is downloading a file, the time plays an important role here. The same matter applies for opening a website; even that the delay does not stop downloading a file or loading a website, but it is an inconvenience factor and may become annoying. Waiting is the only thing that someone can do when the website is loading too slowly, especially when some website's images start to appear. These examples represent transfers, including non-real-time data. From a protocol viewpoint, the TransmissionControlProtocol (TCP) is used to control the connection by the sequence number inside each packet. The re-transmission is the used method for Lost or delayed data. Although, this process guarantees no packet loss, it adds an extra delay. With TCP, the transmission is precisely managed and will not continue without receiving the whole packets. Most applications that need security with no information lost are based on TCP [2].

Real-time data are only the reverse. In general the Real-time operation point to something that is time sensitive. If a Delay was an acceptable factor for non-real-time data, but it can be inconvenient and can effect on the overall performance for real-time data. Let's take the Voice call as an example. Imagine a delay in a telephone call between statements, like waiting more than two second to receive an answer. If, in the same conversation, the system were to lose a word here and there, the conversation becomes even more difficult.

However, it is not a matter to lose a word in a file transfer. The call would suffer an extra delay waiting for lost packets. Finally, if the packets arrived at a rate that changed (jitter), it might lead to unacceptable performance. Therefore, the real goal is to make latency, packet loss, and jitter as low as possible on real-time data connections. From a protocol viewpoint, the user datagram protocol (UDP) has usually spread; because we do not want retransmissions or the return of lost packets. UDP does not have a sequence number or acknowledgment values, thus has a low overhead. Therefore, it is used for applications in which the speed, but not security is considered as an important factor.

1.3. Teleconference Benefits

Teleconference system lets people meet together in various, diverse locations geographically in order to share information. As well as seeing and talking with each other you can:

• Display a close-up of images, maps, and small objects.

• Play a videocassette, DVD, CD

• Display a presentation slides or any other formats

• Record the open session

• Share the PC-information with the others

A Teleconference session can only have two sites and this is referred as a point-to-point session, or between multiple sites and this is referred as a Multipoint session. By contrast, a Teleconference system can grant a remarkable flexibility with cost-savings opportunities. Its advantages can be summarized into four key groups: IP technique; Lower Cost of Ownership, Enhanced Management Capabilities and Minimized Bandwidth for Compression, Transferring and Storage.

1.3.1. Lower Cost of Ownership

Teleconference systems can be an efficient trade solution for various reasons. For example, there is no more need for coaxial cabling, so no more power loss due to moisture, age or bending. Also, a Teleconference system based on a LAN or Ethernet-net is already available along with the Internet service. Thus, there will be no more need for an additional coaxial cable whenever a new hardware is installed. Since installing a new Ethernet cable is cheaper than a coaxial cable and uses newer techniques.

In addition to that, the modular architecture of a Teleconference system grants business advantages like flexibility and scalability. It is constructed on a modular grid that provides a multi-year expansion plans for a security system. Finally, using an IP technology can assist in minimizing the downtime of video network by permitting the choice of buying COTS (CommercialOffTheShelf) hardwares, such as servers, hard drives, and other components from local PC retailers.

1.3.2. “Future-Proof” IP technique

Buying a Teleconference system can be further “future-proof” than other systems. The IP technique is an upgradable system that allows both S/W and H/W components to be upgraded to fix any integration issue while the network is working. Such a capability gives an opportunity to use the system tools without any downtime and for a long period.

1.3.3. Enhanced Management Capabilities

The NetworkVideoRecorder (NVR) works as a heart of the system by making an extra management and changes with an easy way. Its controlling abilities involves the utilization of coherent digital technology, redirecting video feeds in an efficient way when a server fails, controlling high-traffic times and displaying the video feeds from any place by using the Internet connection.

NVR includes complete redundancy capabilities as a standard functionality, which redirect the security video feeds a new destination server whenever the original one fails. The expression "digital" commonly refers to both NVR and DVR. However, where to and how to apply the digitization, are the main decisions regarding the recording techniques.

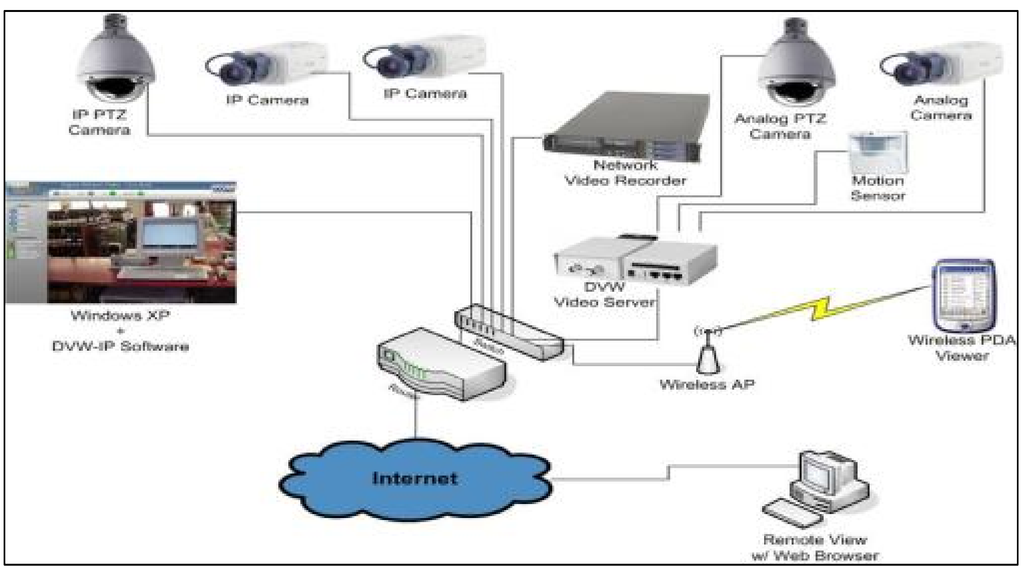


Figure 1.3 NVR Functionality

Digital video feeds are received from cameras by the NVR via Ethernet, then compressed digitally and stored on a hard drive, as shown in Figure 1.3. Whilst, the DVR performs the same function as NVR, except it receives analog video feeds through a coax cable, shown in Figure 1.4. Thus, the digital technology on a DVR is only used in the compressing and storing processes. Moreover, the whole system fails when a hard drive fails as long as the DVR utilizes only its internal hard drive. It is called a SPOF (SinglePointOfFailure) that is common in a DVR encoder of a single port. On the other side, the IP video system will not be vulnerable when the NVR fails because it is a distributed system. Also, a Teleconference system is an effective element in controlling the high-traffic on the network. It is also possible to view the networked cameras at one rate, in which they record at a different rate. So, they are programmed in a way that reduces the bandwidth of the network with no noticeable change in the quality of the displayed picture.

Figure 1.4 DVR Functionality

From the security management viewpoint, the Internet contributes in altering the way of coordinating the security functions by the administrators and owners. The utilization of IP technique permits the pictures to transfer throughout the Internet consistently, in spite of the travelled distance. So, they can be displayed with the same quality as if they were local. Teleconference systems allow the online video events and feeds to be displayed from any place via Internet. Therefore, it provides an opportunity for the control team to react faster and efficaciously to the security events [2].

1.3.4. Minimized Bandwidth for compression, Transferring and Storage

Nowadays, the Bandwidth is an important element for multimedia systems, image compression, transferring and storage. Compressing images particularly before transmission assists in minimizing the net-bandwidth requirements. Thus, it makes faster transmissions via the net with storing the video feeds efficaciously. Moreover, Teleconference clients can change the frame rates, compression settings and size of image without altering its quality, to accommodate to the high-traffics on the net. The network cameras are capable of displaying at one rate (30fps NTSC/25fps PAL), and recording at a different rate (1-30NTSC/1-25 PAL fps), when necessary. VoIP (Video over IP) system contains an effective compression protocol in comparison with the traditional systems. Where, the video is transformed digitally in the Traditional system at the DVR and then compressed for storage. Therefore, the DVR hard drive deals with all transformations, compressions and storing, while, the IP-based systems spread those jobs over the network with a minimized stress on the recorder. Moreover, some modern net-cameras can compress pictures before releasing video feeds to the network. The IP video network is more efficacious than traditional systems by its storing capability. The video feeds at NVR can be transferred automatically or manually to a storage area network (SAN) at specific times for extra space. Also, a traditional DVR hard drive represents a key storage component and has a limited capacity typically. Each Video clip required for prolonged periods should be transferred manually to an external hard drive, SAN, or burned to a DVD. Those storage options should be done by a person, therefore they are exhausting, consuming time, and may lead to a corruption in the files.

1.4. Literature Survey

- In 2014, Min, X., et. al. have dealt primarily with peer-to-peer (P2P) system performance based on modelling and simulation. They deduced that to transpose the spatial advantages of cellular P2P communication with superior throughput efficiency, in which a new P2P principle should be proposed. However, the streaming of mobile P2P over cellular networks is a modern and an innovative principle. In fact, there are no available simulations or real experimental results.

- In 2012, Harikrishnan N., et. al., implemented a VoIP Teleconference System that depends on MinimumAudibleAngle. They provided the basic model of a multi-speaker teleconferencing with autonomous users for simplifying the listening, talking or both. The system can receive, send and process the audio signals in real-time due to the available software stack on each client and the RTP (Real-timeTransportProtocol). It has been implemented using the JMF (JavaMediaFramework); in order to capture and play the audio signals.
- In 2008, C. Teck-Kuen and C. David designed a new efficient method for bandwidth utilization in a real-time VoIP teleconference. It has the same useful characteristics of the simple teleconference system, but it consumes less bandwidth than the normal one.
- In 2006, J. Yang and S. Shen suggested a multi-point video-conference using a visual C++. However, their coder based on the old H.263 codec technology.
- In 2003, T. Bengisu, et. al., designed an efficient SIP-based video conference that works on SUN's JMF and a commercial SIP stacks from Dynamicsoft. They implemented their model architecture on four SIP user agents.

1.5. Chapters Outlines

In this section, the contents of individual chapters of this thesis are briefly reviewed:

- **Chapter One**: This chapter gives an overview of the Telecommunication/Teleconference system with the literature survey.
- **Chapter Two**: The goal of this chapter is to providing the readers a basic knowledge of how Teleconferencing operates generally and the developments of specific technologies to support the Teleconferencing over IP network.
- **Chapter Three**: It gives a fundamental knowledge of video/audio signals to simplify the discussion of video transfer with examining the reason behind using a compression technique and the way of determining a suitable compression for an application.
- **Chapter Four:** The design and Implementation of the Teleconference system will be given in this chapter supported by the experimental results and the conclusion with the recommendations for future works.

CHAPTER 2

TELECONFERENCE OVER IP

2.1. Basic Concepts

The synchronized audio and video signals are what Teleconferencing uses in order to provide a two-way transmission among faraway places. Obviously, most of the calls that take the main role in the communication between peoples are non-verbal. So, when using a telephone, the facial expressions and gestures are not involved. Teleconference system tries to address these expressions by letting people both to hear and see as well.

The Teleconferencing over IP nets for many users is the perfect mixture of video technique and sophisticated networking. A lot of continual business travellers would prefer to make a cost saving by replacing some of their less significant travels with a Teleconference. As the network became popular in the business world; the need for Teleconference services increases significantly. Among the most important factors for the rapid growing of such a service is the requirement of cost-saving. Thus, it won’t be a surprise for any VideoOverIP participator to get an occasional demand for an IP Teleconferencing. Similar to the telephone calls, Teleconference systems can be installed in many ways. One of the simplest examples is the point-to-point communication; which can be held among two sites. Any call consists of more than 2 sites can utilize a multipoint Control Unit to switch the video and audio signals according to the user action. Or it can be done by a continuous presence, in which each participator can see and hear the others at the same time. Besides, the echo canceller equipment can also be used to support the teleconference system. We'll discuss the mentioned aspects of Teleconferencing in the coming sections [3].

2.1.1. Session Setup

In order to have a teleconference system, there are many tasks should be done before the startup process. It is actually not different from the precept of set up a real face-to-face conference. In order to have a right teleconferencing between multiple parties, then all of the Teleconference equipments should be ready and connected for each party before the startup. Occasionally, this process implements by a particular communication protocol within the devices or by a central management system. The main tasks that should be held before establishing a Teleconferencing are described as follows:

- The participator addresses should be known for each caller. The list of these Addresses is stored in a public or private directory. In contrast, this does not apply to IP addresses, since IP addresses may be modified periodically whenever a device joined to a network. The mobility option is available in some Teleconferencing protocols give the participator more flexibility and capability of calling from any location or using another device.
- All the teleconference devices and participators should be ready before the establishment with an adequate bandwidth to handle the teleconference streams. Most of the terminal devices provide multiple video speeds; it means a variety of data rates can be held based on the network characteristics.
- The documents sharing, Presentation slides, and a whiteboard for drawing can also be utilized as an alternative means of communications. Therefore, the teleconference system should be prepared in order to support all of the mentioned tools.
- Transferring the bills to participators and controlling the teleconference efficiency are some services that can be held during the session, so they should be prepared during the connection or reservation time.

Teleconferencing criteria handles the communication processes under different network settings. Although the Setup-Call is a complicated process, it has a higher flexibility in comparison with the setup of a voice-call that needs millions of lines to the complex phone company switches. A unique, fixed-limit bandwidth channel is frequently used for a teleconference set up to carry a specific protocol intended for the setting or tearing down a conference. Various signals go throughout this channel such as the teleconference requests, device commands, and acknowledgments. Ringing, connecting, and disconnecting are examples of teleconference signals. Therefore, the signalling channel is specified for transferring only the teleconference control commands and not for audio and video streams. Some of the IP Teleconference systems don’t utilize a handshake-protocol in the teleconference setup, because they depend on a central server to start and check connections. Clients start a conference reservation by choosing the devices that will participate in a session. Also, the gateways can be configured within a teleconference session to establish the conferencing between participators.

2.1.2. Multi-Point Conferencing

Multipoint conferencing has a great popularity nowadays in both video and audio communication. The term multipoint means any conference that is held between more than 2 endpoints. The number of endpoints does not rely on the number of the camera devices that are connected, because only the active one is counted in the teleconference session and so as for the audio devices. If the number of streams (audio/video) exceeds two, that means a Multipoint conference.

It's usually called a three-way calling or conference calling in telephony. Most of the government agencies or employees of a major corporation participate in a conference session. In a conference system, there is a bridge device to connect multipoint conversations. For each participator there is a mixture of all audio streams of participators without own audio. The amplification of all audio-participators to a similar level is what making a bridge of a conference a good one, as if it's a face to face meeting.

From the practical viewpoint, it isn't that easy. Since, there isn't any actual method to merge the pictures of participators into a one big single picture, as if the participators sit in the same room. Thus, for the sake of multipoint conferencing, there are two standard options; which are the MCU-switching and continuous presence.

2.1.3. MCU Switching

In multipoint control unit (MCU) switching, each terminal obtains a single video stream that is transferred (switching) during the progress of the conference. The MCU performs the switching function as a central device. Several modes are utilized for such a process and the following describes the most prevalent ones:

- Chair control: A single endpoint is specified as the conference's chair to manage the display of the others; which is useful for remote learning environments. There is flexibility in the selection of which endpoint will have the control chair of the conference system.
- Follow the speaker: The MCU tries to transmit the video stream from the active terminal to the others. Sometimes a little delay is included before a switch run, to avoid any switching results due to a single short noise.
- User control: Each endpoint has the ability to choose any of the participants for a call. It represents an extra load for the end user; which make it difficult to continue throughout a long conference. It is an annoying issue for some clients, which is similar to changing the TV channels by someone continuously.
- User-controlled broadcast: Any client has the ability to share his video stream in the conference session with the others by using the MCU. This operation can be useful for a presentation lecture.

Figure 2.1 demonstrates how a follow-the-speaker’s conference operates, in which it involves a central MCU and 4-endpoints (A – D). Each endpoint transmits a video stream into the MCU and obtains a single video stream in return. 'A' is speaking as the conference starts. The MCU transmits the video stream from 'A' to the others, and 'A' can receive a video stream in addition to that at the same time from any other endpoint. After a period, 'B' starts to speak. So, the MCU switches to transmit the video stream of 'B' to 'A', 'C', and 'D'. 'B' will keep going on receiving the video stream from the endpoint 'A'.

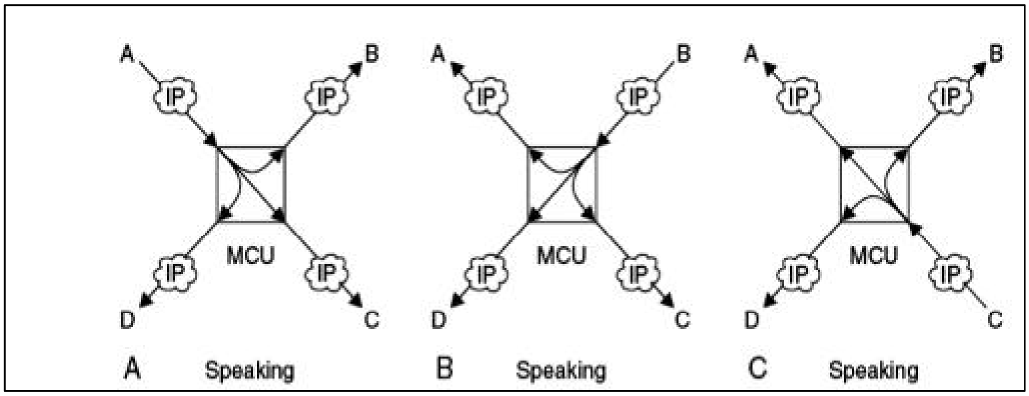


Figure 2.1 The MCU with the (A, B, C) Follow-the-Speaker conferencing

2.1.4. Continuous Presence

In this method, each participator can see the others on a continuous basis with no MCU. This method needs more net-resources than the previous method. In a four-way conference, for example, each endpoint should have a capability of handling the three received incoming video streams. Also, this method appears in a more natural way in such the video disruptions resulted from the switching operation are ignored.

Installing a continuous-presence conference can be much more difficult than the other methods. The MCU unit handles the entire request from each connected endpoint. Multiple copies should be created for each signal within this type, and be sent to a far site. Thus, Multitasking process is the best solution to handle such an issue; in which the network is in charge of creating all the required copies received from each one. If this process does not exist, then each video endpoint will be responsible for creating an individual unicast video stream for each destination, therefore there will be an increment in the bandwidth-overhead.

The four-way video-conference with the continuous-presence method is shown in Figure 2.2 on a Multicast net. It shows that any video stream coming from one location, such as A is going to the other locations throughout the network.

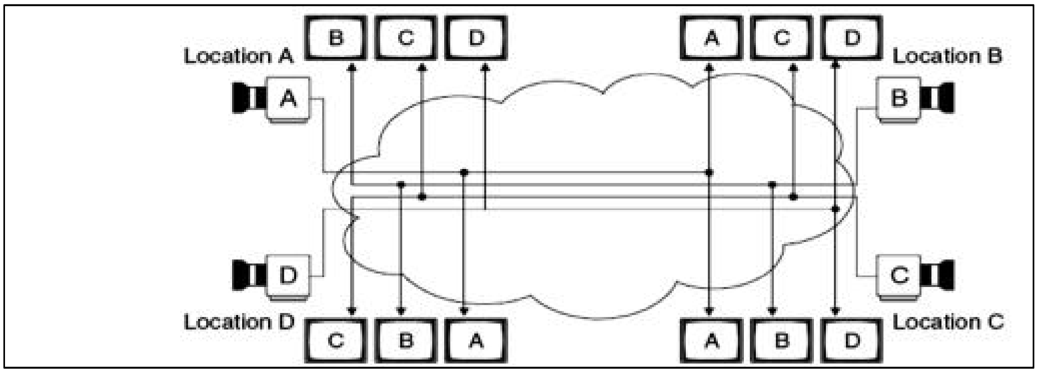


Figure 2.2 Continue-Presence conferencing

So every location has a video stream from the other 3 locations with no need for any switching mode such as follow-the-speaker and chair control; where each location can see all the others during the entire session.

2.1.5. Tele-Presence

This method represents a superior quality of video-conference setup by giving an impression to the users as if they are sitting within the same room. This is done by the usage of displays and HD-cameras linked in a continuous-presence mode with a good ergonomic architecture. Its design concentrates on how to display life-size images on each large monitor and how to place the cameras in a perfect manner; in order to give an eye-contact illusion between the participators.

The illusion of eye-contact is further improved by using a special design for the controlled lighting, conference tables, and a very low-delay codecs. Many organizations, such as HP, have a built-in Tele-presence system for their own utilization and selling to other clientele.

2.2. IP Video Transport

The discussion of video transport becomes much easier by understanding the basic criteria of the Internet Protocol (IP) networking and to put it into context. In this section, we will describe the Ethernet technology (the popular network technique in IP transport) [3].

2.2.1 How IP Fits IN

The IP gives a consistent addressing pattern to let the computer users communicate with each other throughout the net. It also provides a parallel application process; such that the browsing, e-mail, and video streaming operate parallelly on a single PC. It also has a flexibility in allowing various types of computers to work together like (PCs, mainframes, Macs, Linux machines) in spite of their physical communication method. Its links include spacious types of physical links like Ethernet (10BaseT or IEEE 802.3). The other techniques for supporting IP are the wireless link like Wi-Fi, dial-up modems, ATM and SONET circuits. It also can operate with a combination of different net technologies; like a CATV-system connected to a wireless home access-link to provide cable modem services, in order to send client data to the Internet by fibre-optic backbones. That's why the IP becomes the most popular technology nowadays. Simply, you can look at the IP address as a Telephone number; in which both of them are unique. As long as you know the phone number of someone, you can easily make a contact with him in spite of his real location and what kind of technology is currently in use, whether it is cordless, mobile phone, tone-dialed phone, or a fixed rotary phone and so as for the IP address in which each person can send his own data to the others depending on the known IP Address. In addition to that, knowing the telephone number of one person does not mean you can contact him so easily without his permission (i.e. Accepting the call). Also, there should be a common language between the participators in order to have a real communication. The same is true with IP networking.

In contrast, there is a different technology between the telephony and IP-networking; in which that the Telephony is a ConnectionOriented type and the IP is a ConnectionLess type. In the connection-less type, the information is broken into several subunits and each subunit is free to go on whatever path; while in the connection-oriented type, there should be a connection setup for only one route before any communication.

2.2.2 Limitations of IP

The IP technology relies on other software aware to complete its functions and also the other softwares depend on it. Figure 2.3 demonstrates how IP connects between the hardware and software part. Although, IP is not a protocol application or a user application, it is utilized within user-software to achieve their functions, such as playing a video, browsing, and transmitting an e-mail with the use of protocols that provide specific services to applications like HypertextTransferProtocol (HTTP) and SimpleMailTransferProtocol (SMTP) [2].

For example, the Hyper-Text-Transfer (HTTP) is a protocol service for providing the Uniform Resource Locator (URL); which is the location of resources on the Internet. As it’s known that the IP is a connection-less protocol; so it doesn’t provide a way for re-transmitting the lost or corrupted data, since this operation is the responsibility of other protocols; which involve IP. So, the data management is done by those protocols. Like the telephone operation, IP can establish a call without handling the control process if the call is interrupted.

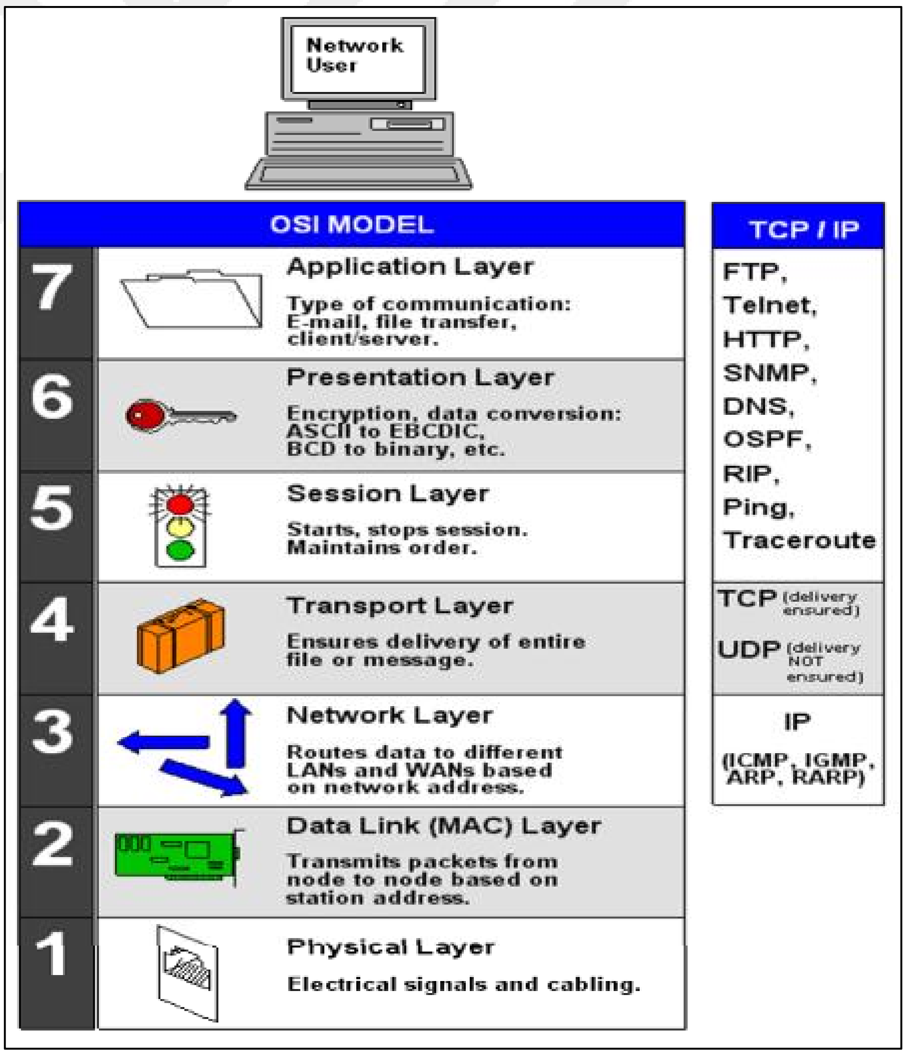


Figure 2.3 IP connects the upper and lower layer

2.2.3 IP Basics

The process of how the video traffic is transmitted over the IP-net can be more obvious if we understand some basic IP concepts. First of all, we will talk about the IP-Address as one of the most important features of the IP-protocol, and then we will discuss the IP design block and its packet with a quick look at some technologies that have an impact on the IP address and the video transmission [3].

- **IP Addresses**: It's the special format of IP technology; that is known as a dotted decimal consisting of four separated numbers by dots (or "periods"). 129.35.76.177, for example, represents an IP address for www.elsevier.com. The IP Address consists of a 32-bit number, separated by dots to four 8-bit numbers. In addition to that, any website has the ability to change its IP address by changing its ISP (InternetServiceProvider); which has a range of IP Addresses for their HTTP servers. As we know that the human brain can memorize words rather than numbers; thus the DNS (DomainNameSystem) was created. DNS holds a list of websites (domain-names) with their IP Addresses. Thus, anyone who types the URL name of a website is translated automatically to IP Address number by DNS. There are special tasks for some IP addresses. For example, the range (224.0.0.0-239.255.255.255) are used for multicasting.
- **IP Packets:** Simply, IP acts as a delivery service for IP-packet. A packet is a single message unit of a specific pattern that can be sent over IP. We will give a summarization for the structure of IP-Packet in Figure 2.4.

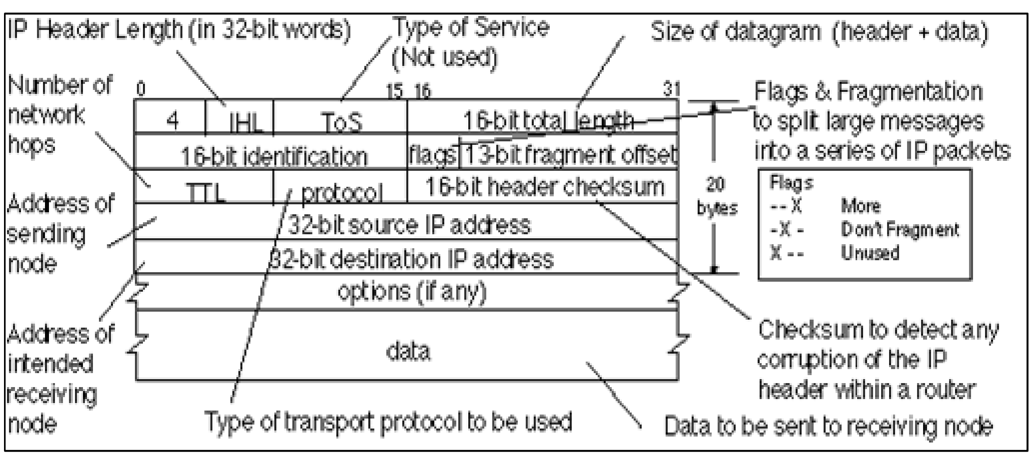


Figure 2.4 IP packet header

A various IP net links may be used in order to transmit a packet to its destination. There is always an active process at each step along the journey of the packet. This process includes the following steps:

a) A header checksum verification should be done in the header of the IP packet. If this verification is incorrect (like an error bit on the data link), then this packet will be ignored.
b) Guiding a packet to the right route typically by a router after determining its destination IP Address. The router device may send the packet to another to an outer network or deliver it locally depending on the IP address number.
c) On each passed node there will be a decrement in the TTL (TimeToLive) counter and a new checksum will be computed within the packet. When TTL becomes a zero value, then that means a distortion of the IP packet to avoid an endless circulation.

- **DHCP and NAT:** According to the structure of the IP packet, there are more than 4 billion unique IP addresses; which are not adequate for the whole humanities and not to mention how many IP addresses that one person might be used for each connected device such as a laptop, desktop, tablet, mobile, file server, net printer and so on. Thus, two technologies were created to handle this issue; which are DHCP (DynamicHostConfigurationProtocol) and NAT (NetworkAddressTranslation). The DHCP specifies a temporary IP address for each PC, while the NAT permits a sharing of a single IP address for multiple PCs. DHCP is an important service in ISPs (InternetServiceProviders) and shared networks. With DHCP, a PC utilizes a single IP address to access the Internet. It is a beneficial service for movable users, where they require a different IP address for each connection on a different Net. There should be a DHCP server to provide a DHCP service. As soon as the PC connects to the network, it broadcasts a DHCP request for an IP Address. Whenever a DHCP server obtains this demand; it chooses a free IP address from its list and keeps it to the requester PC. The greatest benefit of having a DHCP service is the Automatic IP addressing for all the computers connected to the network.

The converting from one IP address to another one is what the NAT service provides. It can isolate a stub network from the Internet. A common configuration is shown in Figure 2.5 for a home gateway with NAT; planned by a user to provide a broadband Internet link for multiple PCs.

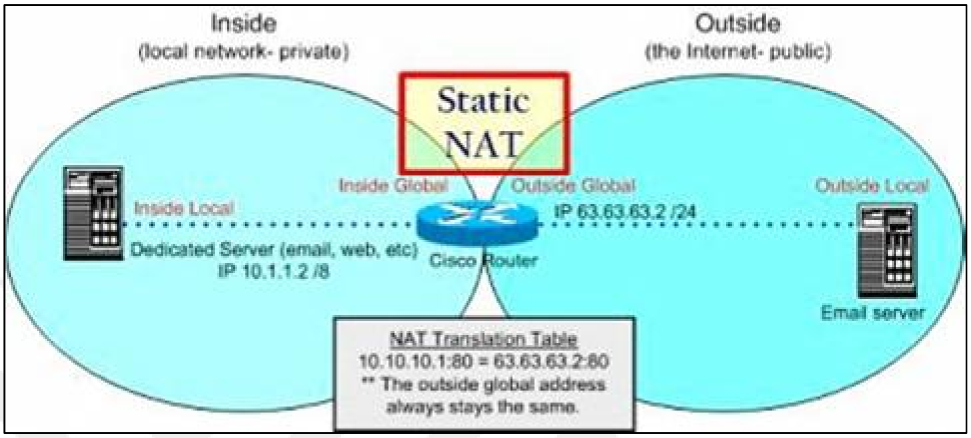


Figure 2.5 The NAT operation

The NAT device represents a single IP device from the Internet point of view, while it is a link to the Internet from a stub-net point of view. A set of private IP addresses is generated for all the PCs connected in the stub net. When a device on the stub net sends a packet outside its stub net, then the NAT device replaces the private IP address of the transmitted packet to its own public IP address in order to let it pass through to the outer nets with recording this operation in its own table. In contrast, the reverse operation is done, when a packet received through the NAT device from the outer nets to the stub net.

For video transmission, both NAT and DHCP can pose challenges. Usually, the servers and video encoders are designed to be data sources. Each PC should know the data source's IP address; in order to access its data and to communicate with the outer nets. In order to prevent any interfere in DHCP and NAT services, then the IP address of the data source (such as a video server, print server or any server) should be always a static IP (i.e. Does not change) and not a dynamic IP. However, it may be a useful-status for organizations to keep the data source with a dynamic IP address, in order not to share their content with external users.

2.2.4 Ethernet and IP

Ethernet and IP were related strongly to each other since a long time ago. Where, Ethernet is the most popular physical schemes for IP networks.

In the next sections, we will talk about the Ethernet addressing and the real correlation between it and IP. After that, we will discuss some Ethernet hardwares and their effect on video transport.

- **Classic Ethernet**: in the 1970s, the Ethernet was created as a mean of connecting PCs on a LAN. Originally, it utilizes a common group of wires linked to all computers. Keeping multiple computers from sending their data at the same time, is an important issue on such a network. Before the transmission, each computer investigates whether another computer is already transmitting data. If two computers send their data at the same time, they will start a technique to check whether a “collision” had happened. When detecting a collision, each computer at once ends the transmission, waits a random delay, and then listens to re-transmit again. The whole mentioned technique is called CSMA/CD (CarrierSenseMultipleAccess with CollisionDetection); replaced with a new system model of a specific cable to each Ethernet port on a PC. CAT5 UTP is the most popular cable type used nowadays (refers to a category 5 UnshieldedTwistedPair). CAT5 UTP can handle data transmission of speeds reach to 100 Mbps. One cable's terminal is directly connected to a router, switch, or a hub. We will give some description upon each one of these devices in the following sections.

- **Ethernet Addressing:** Ethernet utilizes the MediaAccessControl (MAC) addresses for each hardware device. Users who made some experiments on networking projects may have some familiarity upon MAC addresses; because MAC uses the letters A-F and numbers 0-9 for its pattern, divided into six groups of two characters each. The (00:01:02:6A:F3:81) represents a MAC-address for example on an Ethernet adapter within a computer. On each MAC number, the 1st six digits represent the Hardware manufacturer; which is a unique number. The contrast between IP and MAC can be demonstrated by a simple notion. When manufacturing the automobiles, a fixed, permanent serial number will be given to it. It is identical to assigning a MAC address to a hardware component. If a customer registers his own auto, then he will receive a license card. The license card's number is managed by the rules of the local jurisdiction. The auto may have various license cards' numbers during its lifetime, when it is registered by a different jurisdiction or have a new owner. Also the network hardware may contain several IP-addresses reserved at various times, as it moves from one net to another; but a fixed permanent MAC Address.

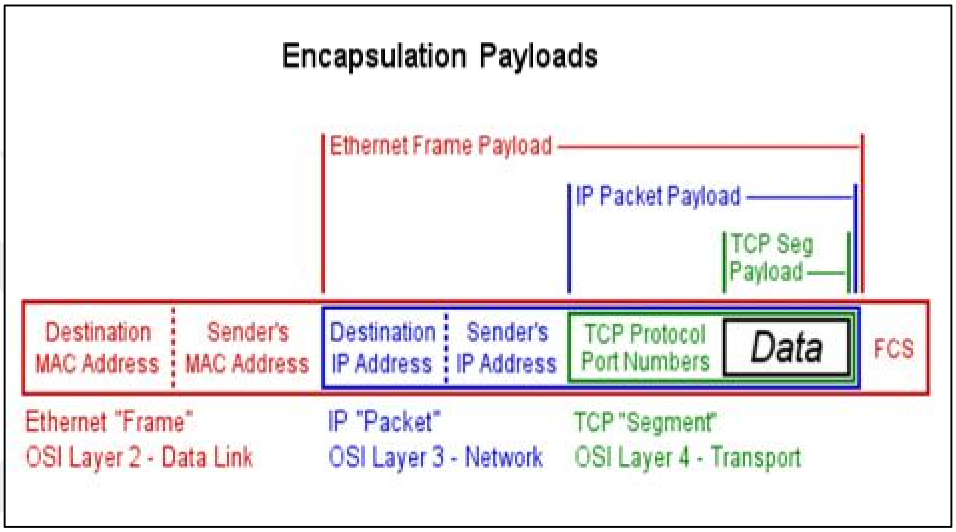


Figure 2.6 Ethernet Framing and IP

Talking further about the details, an auto’s serial no. is a private thing for the owner and the agency that releases the license cards. On the other side, the license card's number is a non-private that is conspicuously inscribed on the auto to be seen by the others. Similarly, a MAC address is a key number just for the LAN net, unlike the IP address; which is utilized globally by other PCs throughout the Internet communication with a specific device. The MAC is specified for the Ethernet transmission. Figure 2.6 shows the operation of data packaging to an IP header and then to an Ethernet frame.

- **Ethernet Hubs:** they are uncomplicated components used for providing an electrical ending and elongation to Ethernet cables. They can retransmit any Ethernet packet to all other connected devices, except the sender. They provide multiple key features in a network:

a) It works as a repeater, takes the input signal, amplifies it, and sends it to all ports connected to it instead the one that brings the input data.
b) It segregates the ports electrically for adding and removing links with no need to stop the network.
c) It has the ability to work as a combiner/splitter, by permitting to share a link between two devices with a third one. It does not provide a mechanism for preventing collisions; but it can re-send signals. Thus, each device should accomplish a CSMA/CD mechanism on its own.

- **Ethernet Switches:** It performs segmentation in its logical scheme, by isolating the physical and logical network on each interface. Figure 2.7 demonstrates a probable switch connection. Where, a common port is shared between Computer A and B (through a hub), so they should have to avoid packet collisions. Computer C will not suffer from collisions, because it has a dedicated port.

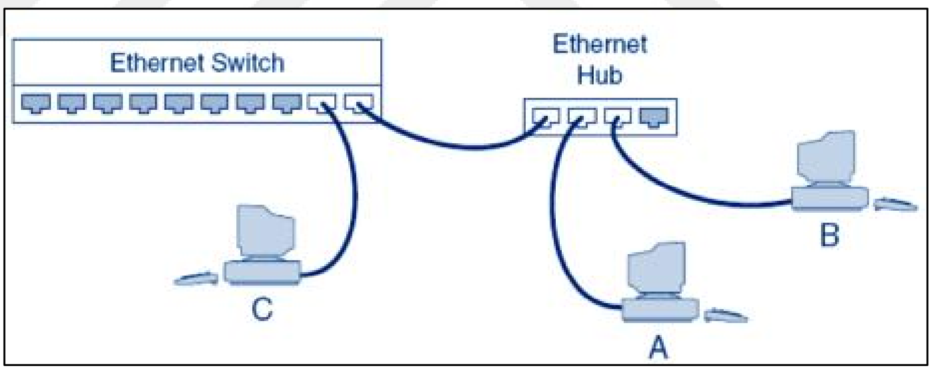


Figure 2.7 An Example for Ethernet Switch

Connecting each single device on a switch port gives three benefits:

a) There is no more worry about collisions with others. Therefore, the speed of transferring packets will rise.
b) No more broadcasting to all ports as the hub. It will send to a specified port of a destination address. Such a function will enhance the network reliability and security by prohibiting devices from seeing the addresses of packets not destined to them.
c) Specific devices can send and obtain packets at the same time; which is called a full-duplex mode. This feature can be advantageous for video-

streaming applications, because they use a huge amount of bandwidth. Note that a one way transmitting or receiving is called a half-duplex mode.

- **Routers:** Routers are called the intelligent device in the IP nets. Routers have variety configurations and can be small access routers that provide an interconnection to the internet for a small set of clients, or a series of enterprise routers; which can manage the requirements of a company network. Let us give a brief introduction to the router basic functions:

  a) It forwards the received IP-packets to their destination.
  b) For a Global transmission, it forwards the packets to a neighbour router.
  c) Keeping routing tables for multiple IP destinations to route each packet to its right path for the destination device.
  d) Bridging between various net technologies, such as telecom circuits and Ethernet LANs to transport packets over different types of network.
  e) Tracks the status of linked networks, to check whether they are available, or not, and routs the packets to prevent a congestion within the networks.
  f) Lines up the packets according to their priorities in private nets and gives a precedence to the highest priority packets.
  g) Broadcasting any modification in network configuration to all connected interfaces.
  h) Performing a lot of security and management tasks.

The Router differs from the switch in that it relies on the IP-address, while the switch relies on the MAC address. Such a basic phenomenon has some main effects, since IP is important for a global transmission, while MAC is important for a local transmission. Therefore, the switch focuses on the local devices that are connected to it. As a result, routers make some calculations and decisions before any transmission with a similar amount of bandwidth and port capacity as switches, and that’s why they are more expensive [3].

- **Ethernet Interfaces:** RJ-45 connectors (see Figure 2.8) are installed for utilization with Ethernet packets all over the world. So, it is an essential component on client-computers and servers. The RJ-45 wire consists of 8 pins organized in four pairs for CAT5 or CAT6 UTP cable. The coaxial cable and fiber are other useful Ethernet formats, but we will not talk about them since RJ-45 wire is our main subject. Ethernet UTP connections are installed in a physical star scheme; in order to give each terminal a straight, continuous link to one port on a Network-device that may be a router, switch, or a hub that receives the incoming packets from a terminal and re-transmits them to other devices. The most popular UTP interfaces include:
    a) 10BaseT, that works at a symbolic Bitrate of 10Mbps
    b) 100BaseT, that works at a symbolic Bitrate of 10^2Mbps
    c) 1000BaseT, known as a Gigabit Ethernet, that works at a symbolic Bitrate of 10^3Mbps.

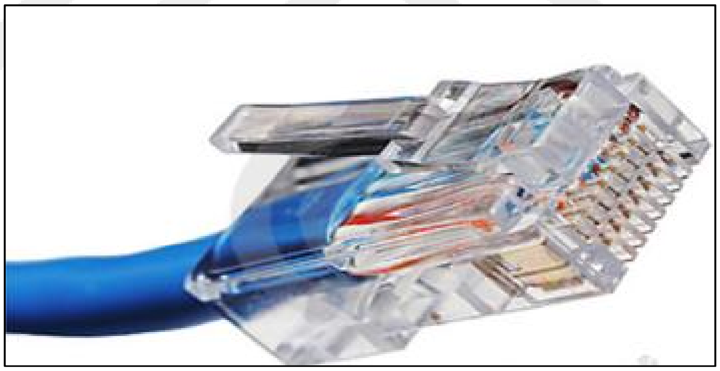


Figure 2.8 RJ-45 Connector

- **Wireless Ethernet Interfaces:** Its standards, like 802.11a, 802.11b, and 802.11g, provide a mobility and flexibility for users. There are several different types of low-cost wireless access devices for a variety of computers; which connect them to high-speed net-connections like the cable modems and DSL. The multimedia clients should be more cautious about the utilization of High-quality wireless links to stream a video. The quality and speed of a wireless connection rely on many factors, such as how far is the wireless receiver from the sender, the antenna configurations, and any local-sources of signal attenuation or disturbance. If the local circumstances change (for example, a metal door is closed or a laptop is moved), the error-rate of the wireless connection can also change quickly. 802.11 connections are created for dropping down to the minimum operating speed if the loss rate of packet increases enormously to lower the error rate. The transmission speed increases when the error rate goes down. In 802.11b system, for example, the Bitrate may vary at any time to another rate ranges from 1 Mbps to 11 Mbps without a caution, according to the efficiency of the radio channel between the two terminals. The services that operate on a wireless link can be affected with the instant decrease of a wireless Bitrate, such as downloading a file or streaming a video. If the wireless Bit rate decreases under the required value for a real-time streaming, then the contents become useless. In this manner, the mobility of wireless connections should be stabilized upon the possible issues that have a great impact on the video services.

2.3. IP Teleconference

The variety of different IP technologies gave a great strength for IP networks to transport packets over multiple distant nets. In this section we will take a brief view at the operation of preparing IP packets in a digital video stream. We'll cover the encapsulation method that chops up the video stream in a way to fit in IP packets. Then we will demonstrate the ways of multiplexing with some protocols work on top of IP to manage the data-flow, and we'll take a brief view at some used technologies for making video transport more reliable [1].

2.3.1 Encapsulation

It is the method of formatting a data stream to IP packets with putting headers and some important information to harmony with a particular protocol. The encapsulation method should be adequate for the performance requirements of various applications and nets. Each data should be encapsulated into IP packets before any transmission. This process works in real-time, just before the transmission of packets throughout the network. The encapsulation process is included in many software tools for many devices, such as file servers and desktop computers.

- **Packet Size:** Too many factors effect on the Performance of IP video streams like the packet size. Where, the packet length should be within the adequate limit in the IP specifications. There are some benefits and drawbacks of utilizing short packets as well as long packets. Figure 2.9 demonstrates how the overhead percentage varies according to packet size. Although, there is no basic criterion for selecting the length of a packet, we will describe the benefits of long and short packets:

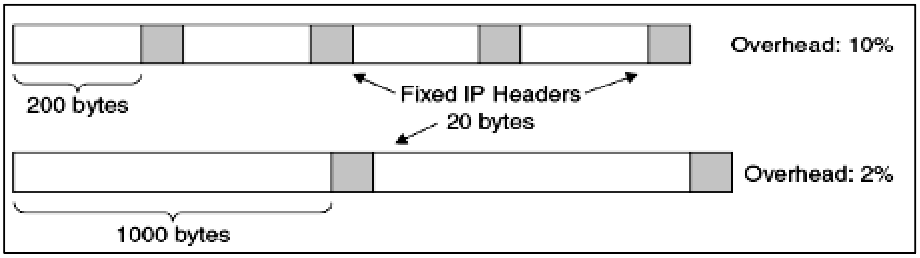


Figure 2.9 Long VS. Short Packet Overhead

a) <u>Long Packet Benefits</u>:

1. **Reduced packet-processing load:** Before any transmission, the packet header and checksum should be checked, with sending acknowledgments if necessary. These processes should be done on each workstation and network device for long-length or short-length packets. The processing time is reduced with larger packets, which can result in smoother net operation.

2. **Less overhead:** The number of bytes in the header is equal for both long and short packets to comply with IP. However, according to the large amount of real information in a long-length packet, a less header information and network bandwidth will be less than a short-length packet as a percentage of the whole packet size to transport a given video stream.

3. **Higher network loading:** In network technologies, there are some gaps between each sent data packet; but they are less for long-length packet, therefore a more reduced net loading.

b) <u>Short Packet Benefits</u>:

1. **Minimized latency:** Each packet cannot be transmitted till the reception of all its data when transmitting a video over IP. The amount of time for filling up all the required information in a packet for very low bit-rate signals (like an audio-only signal or a low frame rate web camera) can be added to the end-to-end delay of the transmitter. The accumulation process for some error-correction techniques can be faster as it processes multiple

short packets in one batch; therefore a more reduced latency, and that's why short-length packet is the perfect choice for Voice-over-IP system.

2. **Lost packets are less harmful:** Any bit error within a packet header means a packet discarding. Less overhead exists within short packets on each lost packet. A data recovery or error masking for the viewer can be done with some video streams in playback devices.

3. **Less need for fragmentation:** The fragmentation process establishes in a router whenever an incoming packet is larger than the maximum packet size for a specific network segment, so it is divided into small chunks called fragments to avoid network congestion.

Obviously, determining an optimum packet length is unknown. Usually, video stream prefers the usage of longest possible packet sizes, in order to ignore the need for a fragmentation process. An extra overhead can result from choosing the wrong packet size. Therefore, it is best to choose an optimal packet size according to the network conditions.

2.3.2. Multiplexing

It is important for a video distribution system to mix between multiple video signals coming from various sources; which are called a multiplexing. The multiplexer output can be a single or multi-signals, each one of them includes different video signals [1]. The reasons behind the usage of the Multiplexing process are described as follows:

- It is easier to transport and control a one large stream than multiple small streams. The peaks of one stream can harmonize to be valleys in another when combining variable-rate streams, resulting in an efficient usage of the overall system bandwidth.
- There is a good economic sense for using a fixed amount of bandwidth per channel system to mix as many signals as possible to entirely fill up the channel like the satellite and telecom links.

The multiplexer is either an autonomous box or part of another device. It may increase the cost on the network transmitter only, because a lot of video decoders and receivers have built-in De-multiplexers. There are two types of multiplexing, which are a statistical and time-division multiplexing. See Figure 2.10 for a comparison.

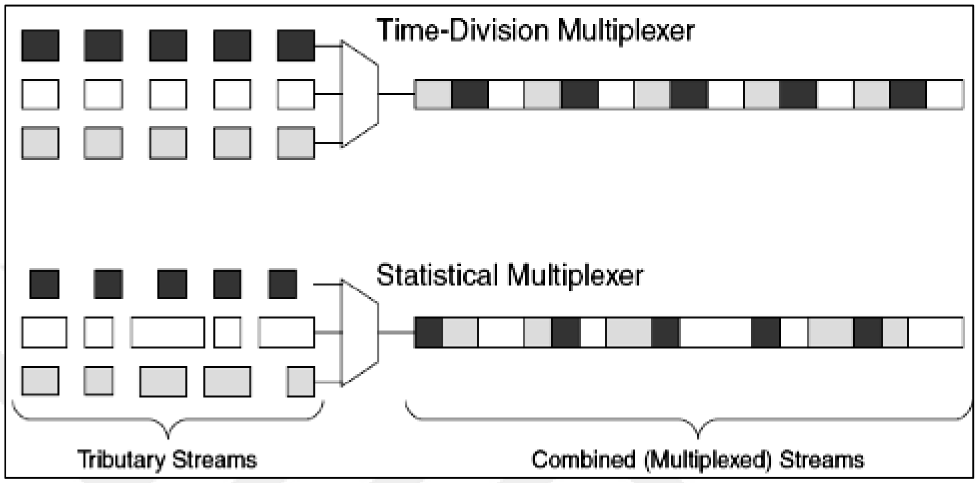


Figure 2.10 Statistical VS. Time-Division Multiplexing

- **Statistical multiplexing:** it gives a bandwidth to the arriving channels corresponding to their varying rates. The higher amounts of the net capacity are given dynamically to the streams of higher-speed streams. A max and min bit rate (rate limit) can be configured for each stream in many systems.
- **Time-division multiplexing:** it gives a constant bandwidth for each arriving stream that is placed into a single or multiple time-slots within the mixed-stream. This allocation can be controlled in most systems to satisfy streams that need less or more capacity.

2.3.3. Transport Protocols

They are useful for controlling the sending of packets in synchronism with IP. For transporting a real-time video, there are two major protocols:

- UserDatagramProtocol (UDP): It is the most older and simple IP protocol useful for video and other information that need speed.
- TransmissionControlProtocol (TCP): It is a firmly-established Internet protocol for data transmission. Many Internet devices can support TCP over IP (TCP/IP).

All the mentioned protocols work over the IP protocol due to their dependencies on IP datagram/packet transport services to transport data to other devices [2].

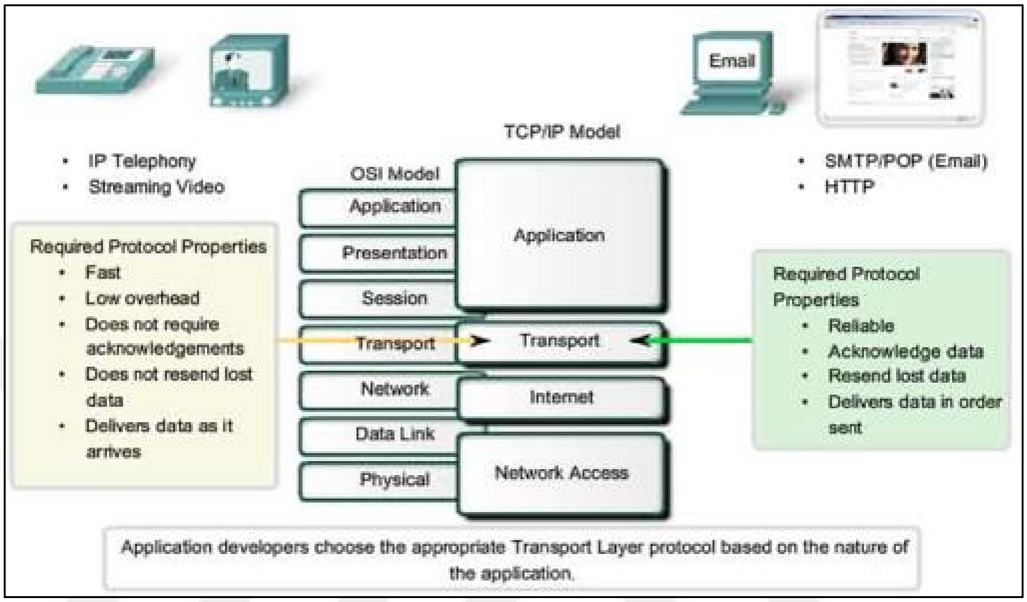


Figure 2.11 Transport Protocol Function

Figure 2.11 shows how Transport layer (TCP or UDP Protocol) fit into the networking hierarchy. We will take a quick look at some important definitions in the network technology:

a) **<u>Ports</u>**:

The common property between UDP and TCP is the usage of logical addresses called ports for packet transmissions for user-applications and high-level protocols. A different port numbers are assigned to packets that are fractions of an on-going file from the packets that include a net control data, despite their destination. The correct IP address with the right port number should be identified whenever a remote device gets its way in a certain software of another device. There are different set of port numbers such as well-known ports for common services, such as web browsing and e-mail, defined by the IANA (InternetAssignedNumbersAuthority). In a well-known port set, there are specific numbers for each well-known service such as port 80 reserved for Hypertext Transfer Protocol-HTTP. So any remote user can begin the web browsing by identifying the server IP address to send the proper packet to port 80.

b) **<u>Sockets</u>:**

The socket is a mix of port number and IP address, like 25.142.186.43:80. It is utilized in many ways by multiple transport protocols.

c) **<u>UDP:</u>**

It is a connectionless transport technique useful for applications that consider speed as the most important factor in their system, such as digital video and broadcasting. It is referred as a connectionless protocol; because there is no connection setup between a source and destination host. The UDP source puts the right destination socket in the datagram and gives it to IP for transmission with no checking mechanism for any lost or corrupted data during the transmission; that make it sometimes unfavourable for video transport due to the lost series of video images. The UDP characteristics are shown in Table 2.1.

Table 2.1 UDP VS. TCP Protocol

Characteristics	UDP	TCP
Description	Simple High speed low functionality "wrapper" that interface applications to the network layer and does little else	Full-featured protocol that allows applications to send data reliably without worrying about network layer issues.
Protocol connection setup	Connection less data is sent without setup	Connection-oriented, Connection must be established prior to transmission.
Data interface to application	Message base-based is sent in discrete package by the application.	Stream-based, data is sent by the application with no particular structure.
Reliability & Acknowledgments	Unreliable best-effort delivery without acknowledgments.	Reliable delivery of message all data is acknowledged.
Retransmissions	Not performed Application must detect lost data and retransmit if needed	Delivery of all data is managed and lost data is retransmitted automatically.
Features Provided to manage flow of Data	None	Flow control using sliding windows, window size adjustment heuristics, congestion avoidance algorithms.
Overhead	Very low	Low, but higher than UDP
Transmission speed	Very high	High, but not as high as UDP
Data Quantity Suitability	Small to moderate amounts of data	Small to very large amounts of data

However, the NTSC image of a video stream is displayed in just 33 milliseconds, so if some part of the image data is missing, then only the ideal receiver would recognize that to inform the sender to resend the image data again before displaying it. The 33 millisecond is not a long time, so it is not that easy to do all the mentioned process within this limited period. The detecting and correcting mechanisms are also included within many video formats, such as the Reed-Solomon forward error correction in the MPEG stream.

d) **<u>TCP</u>:**

It is a connection-oriented protocol that gives superior, reliable transmission. It is the most common protocol on the Internet and intranets. From its name, that means it needs a connection setup before any transmission. An acknowledgment message should be sent after each transmission according to a standard handshake sequence. The ability of handling the transmission errors is one of the most important characteristic in TCP. Depending on the sequence number field (Sequence Identifier) in a packet header, TCP can track each data byte that goes across a connection. Relaying on the sequence and acknowledgment identifiers, TCP can ensure that the receiver obtains every byte from the transmitter. Thus, TCP is suitable in error-free applications and security systems such as e-mails. Another characteristic of TCP is the capability of managing the data flow across a connection to avoid congestion using a window size field. Although, these characteristics are what made TCP a powerful protocol, but it may interfere with video transport for the following reasons:

1. If a packet arrives after a delay due to a retransmission, then it will restrict the receiver while examining the packet and then ignored.
2. The re-transmitting of packets can seize the network bandwidth for sending a new data.

In addition to that, the mechanism of controlling the data flow in a TCP may have an impact on video transmission. Whenever a packet is lost before its arrival to the destination, TCP can enter a speed reduction mode, except for real-time video streams. Where, ignoring the losses is the favourite way when dealing with the real-time streaming to synchronize the audio and video streams. TCP requires a different socket to manage multiple connections that share the same destination. By this way a single web server can manage many clients at the same time, which differs from

UDP that mix all the going data to a socket. The TCP characteristics are shown in Table 2.1.

2.3.4. Internet Video Streaming

It is the operation of transmitting video-streams to viewer equipment for a quick view. It is used in various systems such as training, communications, teaching, and customer support. Actual streaming returns to the TV origins and how to broadcast TV, CATV and satellite; since the video views quickly with no sorting management to have a real-time system and no storage that makes the viewing device less expensive. The actual streaming over IP net established through obtaining a digital stream and dividing it into IP-packets, which may be uncompressed, or compressed using some MPEG forms or other encoding formats (such as Windows Media or RealVideo) an then streamed out over the net with a data rate that is compatible with the video-rate. That means a video stream that lasts 21 min and 3 sec needs 21 min and 3 sec to stream. The player program receives the input packets and displays a picture on the viewer. The successful streaming means the video stream should reach to the viewer at the right instant; but many factors may have an impact on the delivery time. And that's why the servers use a specialized software or hardware to send the video stream at a fixed, and smooth-speed throughout a network. The network requires delivering streams to the viewers, without varying their timing or missing packets. The player-program requires accepting the input packet and addressing the defects caused by the net in the data-flow. In typical, it needs a little buffering capacity of the player equipment. Streaming is utilized successfully on multiple various network techniques, from low to high speed nets. The only demand is to make the exist net-bandwidth equal or higher than the stream data rate.

Table 2.2 Video Streaming Advantages and Disadvantages

No.	Advantages	Disadvantages
1.	The content is delivered when needed and not stored on the viewer device.	The network quality can have a major impact on the delivered video quality.
2.	True streaming can be used for live content, which is also used in Multicasting.	Firewalls and NATs can interfere with video streaming.
3.	Video streaming can be set up to allow viewers watching at any point in the stream.	Lost video packets are not retransmitted in true streaming.

The streaming advantages and disadvantages are shown in Table 2.2. Although, the public Internet became an effective technique for streaming a video, it is not ideal for video traffic. Table 2.3 lists some of the main characteristics of an ideal video net. With the care and right tool, it can be utilized for transferring high-bandwidth, high-quality video.

Table 2.3 Main characteristics of a video network

No.	Features
1.	High Bandwidth
2.	Low jitter
3.	Low delay
4.	Priority control
5.	Lossless transmission

The streaming needs an optimum amount of infrastructure over IP nets. The streaming server is one of the key equipment, which is responsible for streams delivering to each client device. Another one is the media player software; which produces a picture on the client viewer. In addition to that, the transport network and content preparation station between the server and the viewer are also important for streaming systems. Figure 2.12 shows how these main functions work together.

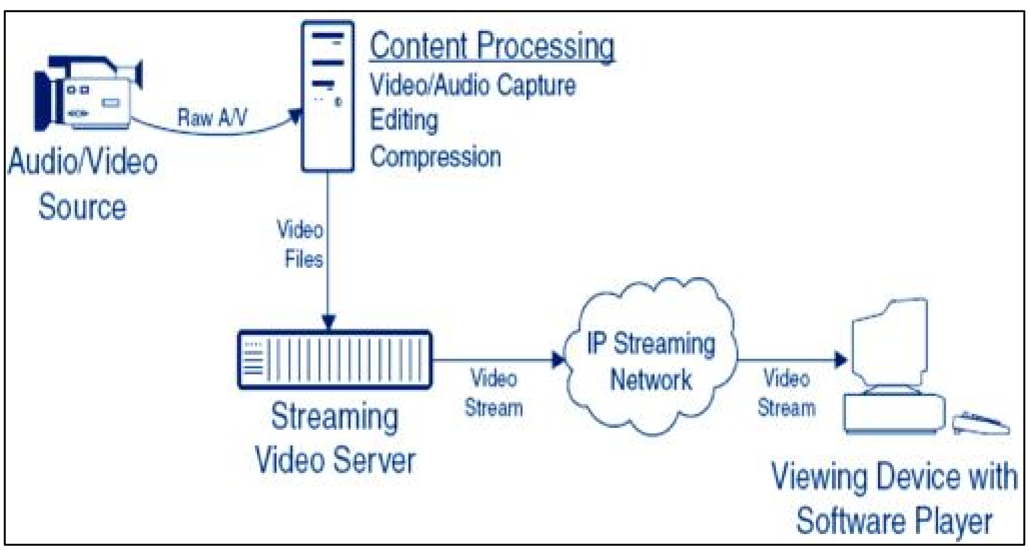


Figure 2.12 A typical Streaming System

- **Content Preparation:**

  The Raw video content, like an online camera picture or a recorded video on a cassette, is in general not suitable for streaming; because they need some processing for streaming such as video compression, format conversion, indexing, and publishing. The process of Capturing and content preparation for displaying can be a basic or a complicated operation, corresponding to the user purposes and his time and cost budget. In some organizations, catching footages with a VideoCam and putting a compressed VideoStreams on their web-site are quite enough; but for others like the expert production, it is not, since it requires more careful with editing pictures and graphics designed for web scene. Thus, a specific capacity is required for streaming a video content. Table 2.4 shows some of the performed key tasks throughout the content preparation.

Table 2.4 Streaming content preparation function

No.	Process	Function
1.	Capture	A process of gathering video content and placing it into the preparation system in a common format.
2.	Editing	A process of organizing the content into the form that viewer will see.
3.	Preprocessing	A process of conditioning the edited video prior to compression and can include color correction, noise removal and image resizing.
4.	Compression	A process of converting the video and audio streams in the formats will actually be streamed out to viewers regarding the stream rates, resolution or type of player software to be supported.
5.	Labeling & Indexing	A process of providing a content description and organizing it so that the viewers can locate it easily.
6.	Publishing	A process of transferring the content to the streaming server and creating web pages that will contain links to the media streams.

A two-pass encoding is used for non-real time streaming. Where the videos are encoded once, verified, and encoded again for the end result. The results of the 1^{st} pass can be utilized in controlling the second pass performance. For example, if the video of a lot of motion is to be encoded, then the 2^{nd}-pass can be prepared to fit with the motion; but the disadvantage of such a system is that it requires higher processing than the single-pass encoder. Although, it does not represent a problem to the saved contents on a server for later playback, it increases the required time for encoding each file. The sources of content can be video cameras, DVDs, pre-recorded cassettes, and others. They are all converted in many editing systems to a public standard before starting any process; such as a 32-bit RGB, in which each pixel per a frame has 8 data bits for each of the three primary colours red, green, and blue. Although, this format is so simple for processing and generating high-quality outputs, it occupies a large amount of disk space. Videos from a compressed standard like a DVD should be de-compressed to a public standard; by either a built-in adapter or autonomous equipment. Many modern video CAMs and camcorders have a capability of sending digital video-content directly to a computer by USB 2.0 or FireWire links. Usually, the process of content preparation is performed first to produce an adequate time for a complicated compression method. A smart operator can adapt the compression settings within the processing time for fast action sequences.

- **Streaming Server**:

  The streaming server has the responsibility of delivering media streams to viewers. It needs an internal-stored media content to create a stream according to each demand of a viewer; which may be a multi or uni-cast stream managed by different techniques. The storage and retrieval of the content are the key functions of a streaming server. The production will automatically start as soon as the content is available in different compression ratios and rates to be ready for many users with multiple net speeds. For an instant, there are many available choices of different playback speeds convenient for DSL-connection users, medium speed connection users, or dial-up connection users. It needs various versions of the content file for creation throughout the compression operation, in such 3 different files in a server may hold one piece of content, but with different playback speeds. If we consider several different types of players in a server to reach more audiences, then the situation will be more complex. Tallying this all up, a multitude of different copies of each content piece can be on a server, which means 9 different video files may be available in a server to handle three different connection speed choices and three different media player formats. The key function of a streaming server is the creation of real-time packets for each output stream. Also, the streaming server should prepare the headers with the encryption (if required) for each outgoing packet. Where, many new encryption methods utilize public key cryptography; which is a unique key shared among all users for the encrypted stream. The streaming server is capable of creating well-behaved streams with regular speeds in a real-time basis that also fits with the required data rate by the player software to avoid a buffer overflow or run of data. Sometimes, it may automatically alter the on-going stream-rate to a viewer according to the net condition reported from the software player; such as lowering the rate by switching to another lower speed version of recorded content (known as scaling a stream) to avoid congestion in a smooth way without noticing the switch over. One of the greatest advantages of a streaming technique is the capability of moving ahead/back within a video content by the player software support; since it should accept the user orders and send them to the server. The server should be able to change the output streams. The use of Gigabit Ethernet became common nowadays to handle large broadband streams. Also, multiple servers can have multiple copies of stream content with various physical locations through the Internet to support hundred or thousand users.

- **IP Streaming Network:**

  The streaming service is a great opportunity if it is available on any IP network and with controlling the main net performance variable to make it operate at a higher performance in reality. Table 2.5 lists some of the network factors that may have a great impact on the streaming efficiency.

Table 2.5 IP net-parameters that effect on Streaming

No.	Network Parameters
1.	Packet-loss ratio
2.	Packet-delay variation
3.	End-to-End delay

- **Player Software:**

  The Player software accepts the input stream and converts it to a displayed picture. At first, the content should be selected by the user before starting a streaming session; which resides inside a website on the streaming-server. The player-software will perform a decryption operation once the key is known, if the contents were encrypted by the StreamingServer. The keys may be straight-forward gathered from the StreamingServer or by communicating with a third-party AuthenticationService. The autonomous equipment like set-top boxes, have built-in decoders but with a fixed range of compressing algorithms (mostly in the MPEG standard). Also, the current PCs and some handheld devices are also able to run player software. Since, the processor speed, capacity of memory and number of other tasks that the processor is operating on, decides which device has the ability to decode a given stream. As a matter of fact, the capacity of disk drive is not a great parameter in comparison with the buffers held in the memory, which are required for video processing. Adding a hardware decoder into a PC will be a better choice for high-quality, full-screen video decoding [3].

CHAPTER 3

VIDEO/AUDIO CODEC TECHNIQUES

3.1 Video Basics

The teleconference systems have witnessed an extraordinary growth since 1990 and they have spread widely. It is the best choice to save the budget for many companies than paying the journey costs of transmitting managers to periodic meetings at a single place. The video and telecommunications technology has fully-developed to make teleconferencing more cost effective [3]. It will be easier to discuss the video transport if we have a basic understanding of video/audio signals. Each video stream is compressed before transmitting over an IP network. The reduction of the required bits to represent a video image is what we call it a compression. Even it is a selective process for each user, but sometimes it may be the separator between the failing and success a video net task. As an example, a compression process is used in digital mobile telephone for increasing its battery life and number of customers. Thus, it became an indispensable element to fit the contents into consumer delivery systems for Digital TV and HDTV systems, many images within a constant capacity of storage for Digital Camera, more audio files into the memory of MP3 players and Apple iPods, and more video channels for Satellite television users. In this chapter, we will demonstrate the main reasons behind using a compression method and how to choose the most suitable one for multimedia systems such as the Teleconferencing. Of course, that cannot be possible without understanding some video/audio basics [4].

3.1.1 Pixel, Luma, Scanning and Chroma

The talk about a video topic requires knowing some fundamental expressions. A basic definition of each term will be given here and the next sections will give the details [2]:

- ***Pixel:*** it is basis of a digital picture and the structure block of both still photography and video. The PC monitors, document scanners, digital cameras, digital video images and so many devices are all defined in terms of pixels. A pixel is the smallest element in a digital picture. A color TV has three various coloured dots (Blue, Green, and Blue), as shown in Figure 3.1. We can think of pixel as a one tiny dot of colour closer in representation to a one letter on a document or a byte within a PC hard drive. Although, the pixel is the key element of an image, it does not have so much data to create a realistic image without combining it with thousand or million pixels. The number of pixels increases with the increase of picture details. The intensity (saturation) and colour (hue) are what resides within a pixel. These values are fixed for a still image, but they are updated 30 (times/sec) for TVs in Japan, and North America, and 25 (times/sec) for the other systems. In addition to that, there is a location value for each pixel in both the vertical and the horizontal dimension; which should be displayed in a correct manner (from left to right and top to bottom) across the display device in a process called scanning.

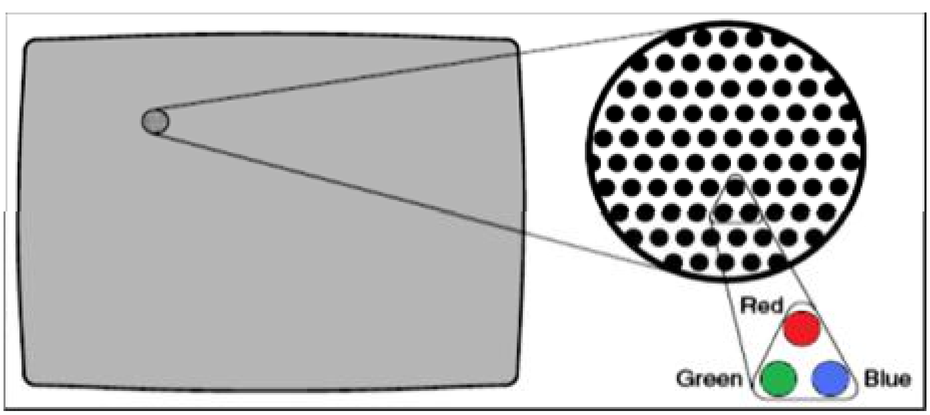


Figure 3.1 A Standard PhosphorDot Sample on a TV monitor

- ***Luma:*** it is the brightness of each pixel. Since, an All-White pixel represents a max Luma (Luminance) and the Black one represents a min Luma. It is a video signal portion that includes data regarding the video brightness (intensity) as noticed by a person eyes. Thus, more brightness in an image means there are pixels with high Luma and vice versa. Also, any change in the image brightness can be easily sensed by the human eye. Any degradation in the Luma means degradation in the quality of an image.

- ***Scanning:*** it is the process of capturing, saving, transmitting, and viewing the pixel Chroma and Luma values of video signal with ordering them in a specific manner useful for any video tool. It processes the image lines from left to right horizontally, in such that the scanning starts from the leftmost to the rightmost pixel of a picture with a short horizontal blanking interval between each line as shown in Figure 3.2. The horizontal scanned lines of a video image are sent one after another in a sequential order.

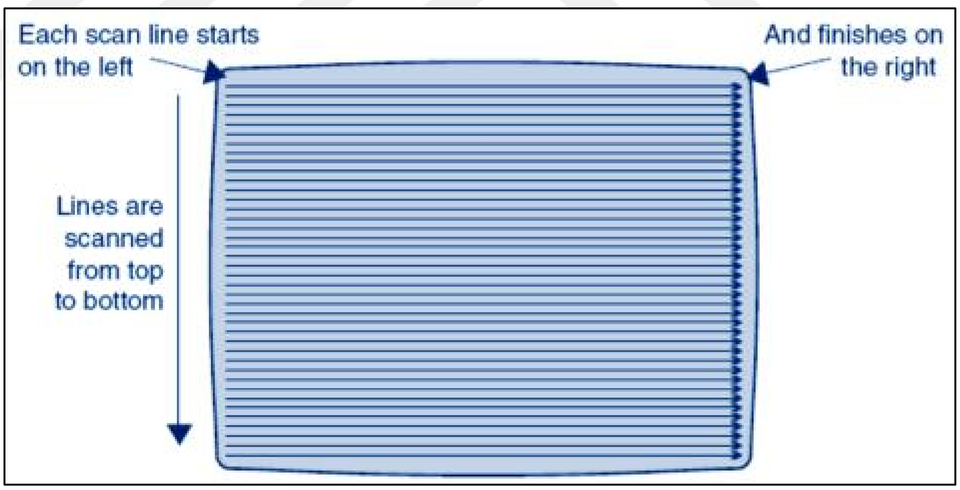


Figure 3.2 Scanning of a Video Monitor

- ***Chroma:*** it is pixel colour. Colours are organized over an entire spectrum of the human visual system. So, Chroma has the colour data of a video signal. Chroma runs in synchronization with Luma to create a normal, colorful, full-resolution image. The colorful monitor displays has three colored phosphor dots (Red, Green and Blue) in each pixel in comparison with a white phosphor dot per each pixel in a monochrome monitor. In fact, it is easy to send three distinct video color signals for short distances and that is the reason behind using a VGA video connector (or RGB connector); since it has multiple pairs of pins carry red, green, and blue video signals to link a PC to a monitor device. The Chroma is a combination between the red, green, and blue signals with different magnitudes. By having a great portion of green component; since the eye is highly sensitive to it, and little amounts of blue and red components leads to a new combination named a 'Y' signal. Based upon the human visual system, there are two other chroma signals used in both digital and analog systems called R-Y or PR and B-Y or PB and each one of them has lesser information than raw red and blue signals. That's why 3 wires are up to this time required for analogue video component (SD/HD).

3.1.2 Frames and Video Fields

A Picture frame is the basis unit of a motion imaging [5]. It is a one image from a sequence of images that creates a motion image. Also, it is formed by 2 fields, each has a half the horizontal lines of the whole picture in interlaced video.

- ***Frame Rate:***

If we take a look at the film utilized in a movie projector, we see a series of still pictures with a slight difference between a picture and its predecessor in a way to produce a motion by rapid viewing. The number of captured images (frames) per one second in a movie is 24, which is called a frame rate. A TV camera works in a like manner, but with an electronic picture sensor instead of a film. Each pixel, on a picture sensor, catches a signal on the light-intensity that strikes it. Capturing of a full picture means a full video frame is created. On TV, 25 and 29.97 fps are the two key frame rates [6].

- ***Frame Height:***

It represents the number of horizontal lines aligned in a vertical manner to create a video picture. The different frame heights are listed in Table 3-1. In common, the pictures of smaller-line height are defined by the overall line count, while HD images are defined by the picture's size. Moreover, the letter “i” represents the InterlacedScanning and “p” represents the ProgressiveScanning [7]. There are a number of lines known as VBI (VerticalBlankingInterval) within each video standard. Its name returns to their blanking or switching off, so they do not show as a portion of a visible picture.

Table 3.1 popular Video FrameHeights

Total Lines	Viewable Image	VBI	Common Name
525	480-486	39-45	525i/NISC
625	574-576	49-51	625i/PAL
750	720	30	720p
1125	1080	45	1080i

- ***Vertical Resolution:***

It represents the number of alternative white and black lines across a display vertically. Figure 3.3 shows a testing pattern of vertical image resolution. A 30% of the total lines can be resolved normally; which is lesser than the horizontal scan lines utilized by a camera. As an example, 576 visible horizontal lines within an image have about 400 lines of vertical resolution [8].

- ***Horizontal Resolution:***

It represents the number of alternative white and black lines across a display horizontally as shown in Figure 3.3. It's determined in terms of LPH (Lines per PictureHeight), which relates the image height to horizontal resolution, in lines.

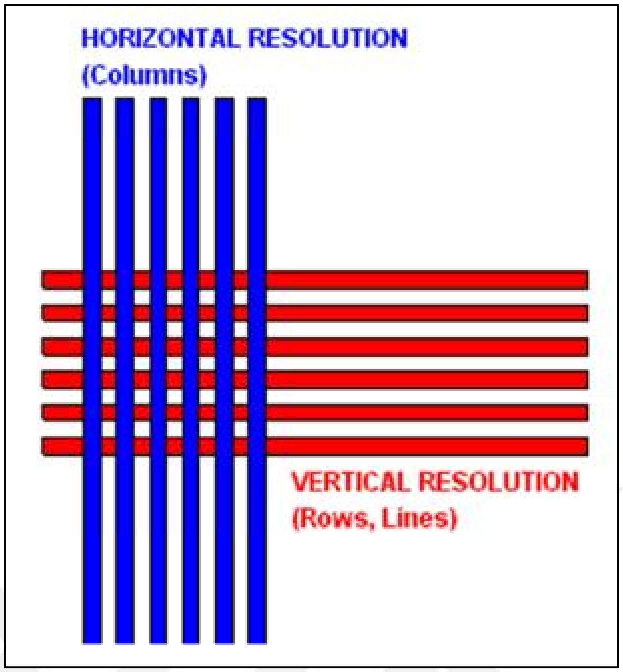


Figure 3.3 Vertical and Horizontal Resolution Test Patterns

There are many factors effect on the horizontal resolution of a picture [3]. For example, 720 pixels is the popular digital TV-format in compression and production technique for both 525 and 625- horizontal line systems, but it isn't the real resolution since it is limited by the signal processing circuits that minimizes the real horizontal resolution of signal to reach 455LPH. Table 3.2 shows the most common image sizes in video compression.

Table 3.2 Popular image sizes within a Video Compression
(Width in pixels x Height in lines)

1920x1080 High Definition
1280x720 High Definition
720x576 PAL (625 line) Full Resolution
720x480 NTSC (525 line) Full Resolution
704x576 4CIF[2]
352x576 PAL (625 line) Half-DI resolution
352x480 NTSC (525 line) half-DI resolution
352x288 CIF[3]
176x144 QCIF[4]

- ***Progressive and Interlaced Scanning:***

The progressive and interlaced scanning are the popular video scanning methods [4]. Each line within a picture is scanned horizontally in the progressive scanning (the “p” from Table 3-1), from top to bottom sequentially, as shown in Figure 3.4. While in the Interlaced scanning (the “i” from Table 3-6), just the image lines of odd-numbers (odd field) are first scanned sequentially and then the even ones (even field) as presented in Figure 3.5.

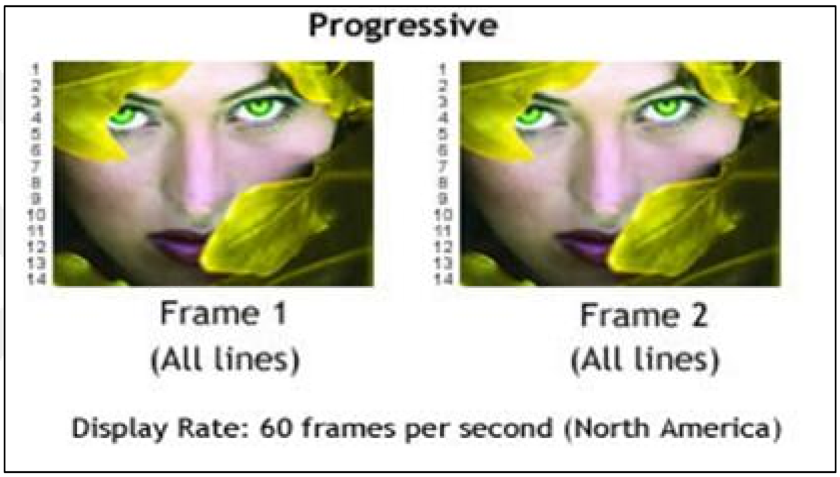


Figure 3.4 Progressive Scanning

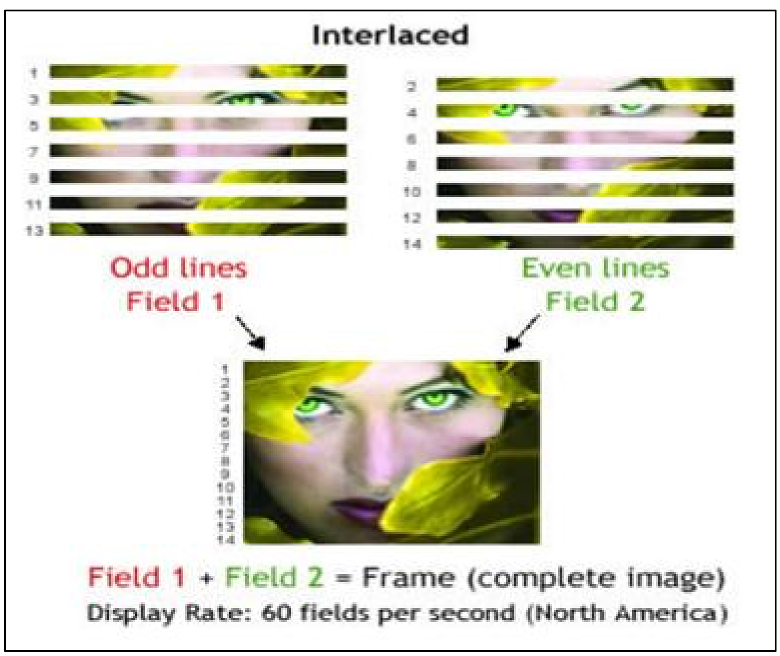


Figure 3.5 Interlaced Scanning

Because of the ease calculation of shapes and forming images, Progressive scanning is the popular method used for computer graphics displays, but the interlaced scanning offers a lower frame rate with no flickering and that is the reason behind using it for broadcast TV [6]. Flickering is the observation of little, quick variations in picture brightness, due to the low rate of updating an image than 50 times per second and that is why the interlaced scanning is used. Half of the display is updated in each field with a rate equals to 59.94 or 50 fps as in the 720p), so it is high enough to be used commonly in three video standards (1080i, 625i and 525i). Unfortunately, interlaced signals may suffer from some visual artifacts. The quick movement of an object horizontally within a picture makes the even-field displays this object in a various location from the odd field, producing a jagged edge.

3.1.3 Colour Spaces

Displaying a color video is what many digital video services rely on. Thus, it requires a technique for capturing and representing the color data. Colour images need 3 numbers for each pixel position, at least, to symbolize the color correctly [7]. The technique used for representing the brightness and colour is what we call it a colour space.

- ***RGB colour:*** In this space, 3 numbers are what define a color picture sample; which represent the relative correlations of Red, Green and Blue (3 main Light's colors). The combination of these colors, with variant amounts, produces a color [2]. Figure 3.6 shows a color picture.

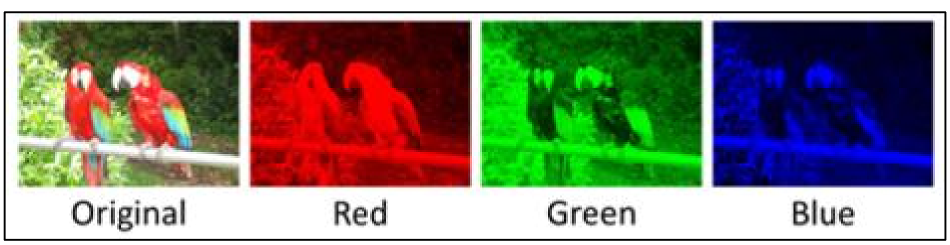


Figure 3.6 RGB components of Image

This colour space is suitable for capturing and displaying color pictures. The 3 colors should be filtered out from the scene to capture an RGB picture by a separate sensor array. Displaying an RGB picture is done by CRTs (Color CathodeRayTubes) and LCDs (LiquidCrystalDisplays) that illuminates the blue, red and green elements of each pixel in accordance to their intensities. Merging the seperate elements, from a normal viewing distance, gives the appearance of ‘true’ color.

- ***YCbCr:*** The sensitivity to color of the HVS (HumanVisualSystem) is less than to brightness (luminance). All the three colors of RGB space are usually saved with a similar resolution, because they have an equal importance [6]. However, the separation of color from the luminance information and with giving the Luma a higher resolution leads to an efficient representation of the color picture. The YCbCr (known also as YUV) color space is the most common method for representing a color picture efficiently. Y represents the Luma (luminance) element; which can be computed using the weighted averages of R, G and B:

$$Y = \mu b \times B + \mu g \times G + \mu r \times R \tag{3.1}$$

Where μb, μg, μr are weighting factors.

It is possible to represent the color information as *color difference* (chroma or chrominance) elements, in which each chroma element represents the difference between the luminance *Y and R*, *G* or *B*:

$$\begin{aligned} Cg &= -(Y) + G \\ Cb &= -(Y) + B \\ Cr &= -(Y) + R \end{aligned} \tag{3.2}$$

The *Y* element with the color differences *C*g, *Cr* and *Cb* (the differences between the mean luminance and color intensity of each picture sample) is used for giving an overall description to the color picture.

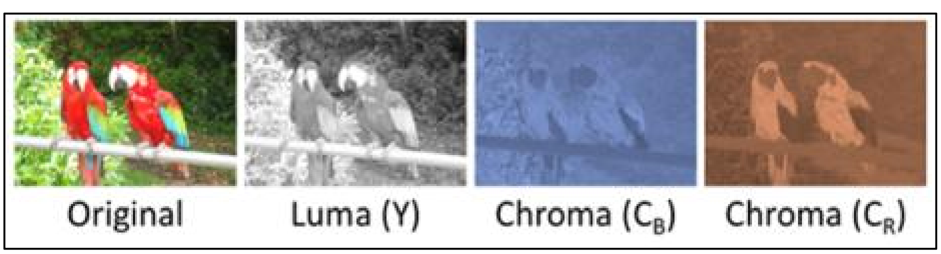


Figure 3.7 Chroma components of Image

Figure 3.7 demonstrates the chroma elements in correspondence with the RGB elements of Figure 3.6. Where, the dark grey means a negative difference, mid-grey means a zero difference, and the bright grey means a positive distinction. If there is a great distinction between a Luma picture and color element, then the chroma elements will get distinct values. However, the addition of *(Cr, Cb, Cg)* gives a fixed value, therefore just two of three chroma elements are required for storage or transmission; because it is easy to calculate the third element from the other two. As a result, only Cb, Cr and luma (*Y*) are used in YCbCr color space. This space is more superior to RGB, because it can represent the *Cb* and *Cr* elements of a *lower resolution* than *Y* since HVS has a higher sensitivity to luminance than color [3]. Therefore, such an opportunity contributes in reducing the data capacity for representing the chroma elements with no noticeable minimization in the visual quality.

Thus, YCbCr with reduced chrominance resolution is more simple and effective for image compression and transmission than RGB, due to its reduction in storage demands. Each picture should be reverted to RGB prior to its displaying. Transforming a picture from YCbCr to RGB and vice-versa is demonstrated in Equation 3.3 and Equation 3.4. Where, ($\mu r + \mu b + \mu g = 1$) and *G* can be computed based on YCbCr with no need for *Cg* component.

$$\begin{aligned} Y &= \mu b \times B + \mu r \times R + (1- \mu b - \mu r) \times G \\ Cb &= (B - Y) / (2 - 2\,\mu b) \\ Cr &= (R - Y) / (2 - 2\,\mu r) \end{aligned} \tag{3.3}$$

$$\begin{aligned} R &= 2 \times Cr + Y - 2 \times Cr \times \mu r \\ G &= Y + [(2 \times \mu r \times Cr - 2 \times \mu r^2 \times Cr) - (2 \times \mu b \times Cb - 2 \times \mu b^2 \times Cb)] / (1 - \mu r - \mu b) \\ B &= 2 \times Cb + Y - 2 \times Cb \times \mu b \end{aligned} \tag{3.4}$$

Also, μr and μb are equals to 0.299 and 0.114 by the ITU-R specification BT.601 [1]. So, by substituting the above mentioned values in the previous equations, we get the popular used conversion equations:

$$Y = 0.114 \times B + 0.299 \times R + 0.587 \times G$$
$$Cr = 0.713 \times R - 0.713 \times Y \quad (3.5)$$
$$Cb = 0.564 \times B - 0.564 \times Y$$

$$B = 1.772 \times Cb + Y$$
$$R = 1.402 \times Cr + Y \quad (3.6)$$
$$G = Y - 0.714 \times Cr - 0.344 \times Cb$$

- ***YCbCr Sampling Format:*** Three sampling patterns of Y, Cb, Cr are shown in Figure 3.8; which are supported by H.264 and MPEG-4. The sampling (4:4:4) has (Y, Cb and Cr) of equal resolutions. Thus, on each pixel position there is an available sample for each element. Its name refers to the relevant sampling rate of every element within the horizontal-line. That means there are four Cr and four Cb samples for every 4 Luma samples. While 4:2:2 sampling (known also as YUY2) has similar vertical resolution for both chroma and luma elements, and half horizontal resolution. The 4:2:2 means 2 Cr and 2 Cb samples in every 4 Luma samples within a horizontal-line. Such a sampling is useful in the representation of high-quality colors [1].

  The 4:2:0 (‘YV12’) is the most commonly used format. For each *Cr* and *Cb* element within this format, there is (1/2) the horizontal and vertical Y's resolution. It is popularly utilized in many customer applications like digital TV, video conferencing, and DVD storage. The 4:2:0 YCbCr video occupies half the RGB or 4:4:4 video, because each color difference element, within this format, has 1/4 samples in the Y element.

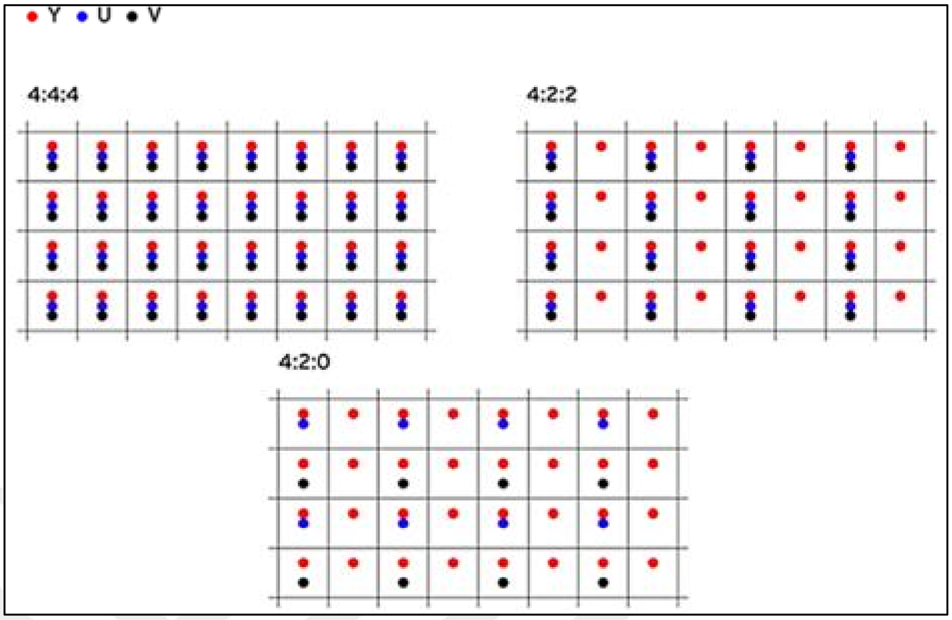


Figure 3.8 4:2:0, 4:1:1, 4:2:2 and 4:4:4 sampling formats (progressive)

3.2 Audio Types

In the multimedia world, the audio signals are necessary to associate with the video pictures. Because these signals are treated in a much different way than the video, it is important to put a specific section for them [2].

3.2.1. Analog Audio

It is the direct electrical input (i/p) signal to loudspeaker or a group of headphones to generate a sound. The videotape, satellite receiver, DVD and CD-player contain Analog audio outputs (o/p). While the TVs, stereo receivers, videotape/audiotape recorders contain Analog audio inputs. Also, the Connectors of Portable audio players and any laptop PC contain analog audio i/p and o/p of different types. An analog audio signal structure is just an electrical representation of the cone's position in a typical speaker. If a signal rises, the speaker's cone moves out and vice versa. The sound of High-pitches is a result from the quick variations in the level of signal; which in turn makes the speaker's cone shakes quickly. By increasing the audio signal level, the volume of the sound will also increase.

3.2.2. Digital Audio

Despite the conversion of many analog audio signals to digital audio formats, they will be with us forever since each system sound has some mechanisms for creating air vibrations in the audience's ears. Digital signals are better than the Analog signals in their immunity to noise signals. They became more popular by the invention of the CD (CompactDisc), mobile phones, MP3 players, recorders and so many digital video devices. There are two forms of Digital audio signals: compressed and uncompressed. Particularly, we will demonstrate the digital audio format of AES/EBU (AudioEngineeringSociety/EuropeanBroadcastingUnion), utilized in the professional video industry.

This type of format is formed by capturing the analog audio samples. The rate of the taken samples can vary in AES/EBU signals, 32,000-48,000 times/sec. A sampling of 96,000 times/sec exists in some modern advanced systems. The high-frequency sound means higher sampling rates. To capture a sound, the sampling rate should be at least double than highest-pitched frequency, according to the rule of thumb.

3.2.3. DVD Audio Formats

At least two audio codec formats are used within all DVD video disks of 525-line and 29.97fps format, which are the AC-3 Compressed and linear PCM audio. Linear PCM audio is a digital audio of the below characteristics:

- Sampling rate equal to 96,000 or 48,000samples/sec.
- Sample resolution equal to 24, 20, or 16 bits.
- Max audio channels equal to 8.
- A max total Bitrate=6.144Mbit/sec in single audio-stream (multi-audio streams can be utilized, but the overall should not be greater than 9.8 Mbit/sec-the max DVD-play out rate).

3.3 H.265-High Efficiency Video Codec (HEVC)

The increasing demands for HD video, and beyond, lead to a strong requirement for a coding technique superior to H.264. Such a demand became stronger when higher resolutions are associated with multi-view capture and display. In addition, the communication requirements for video-on-demand applications impose strict challenges on current networks [8]. Currently, HEVC (HighEfficiencyVideoCoding) is the latest VideoCoding technique of the ITU-T VCEG (VideoCodingExpertsGroup) and the ISO/IEC MPEG

(MovingPictureExpertsGroup). HEVC offers major commercial benefits as listed in Table 3.3.

Table 3.3 HEVC Applications and Benefits

Applications	Benefits
OTT	• Better User Experience • Lower CDN & Storage Cost.
IPTV Operators	• More streams to the home. • Increase reach.
Wireless Operators	• More efficient use of Bandwidth. • Better Video quality.
Ultra High Definition	• Improve HD user experience. • Utilize latest screen technology.
4k VoD Move Distribution	• Match cinematic quality. • Broaden reach of release.

The key purpose of the HEVC format is to improve the compression efficiency in comparison with the current formats [9], to reach about 50% bitrate reduction without a loss in the video quality, as shown in Figure 3.9.

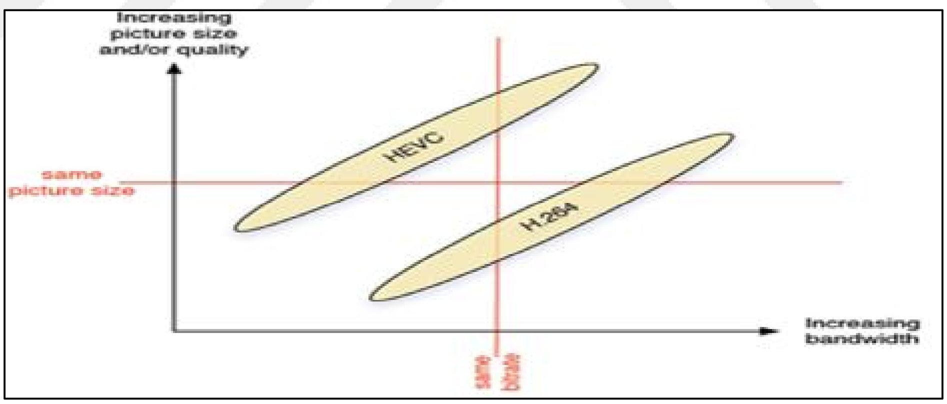


Figure 3.9 The potential gain of H.265 VS. H.264

In general, HEVC is similar to the earlier formats like H.264 and MPEG-2, in the basic structure, but it has numerous improvements like [10]:

- More flexible partitioning, from large to small partition sizes.
- Higher flexibility in the sizes of transform-block and prediction modes.
- More improved interpolation and de-blocking filter.
- More improved motion vectors, prediction and signaling of modes.
- Tools for effective parallel processing.

HEVC structure for encoding and decoding operation is shown in Figure 3.10.

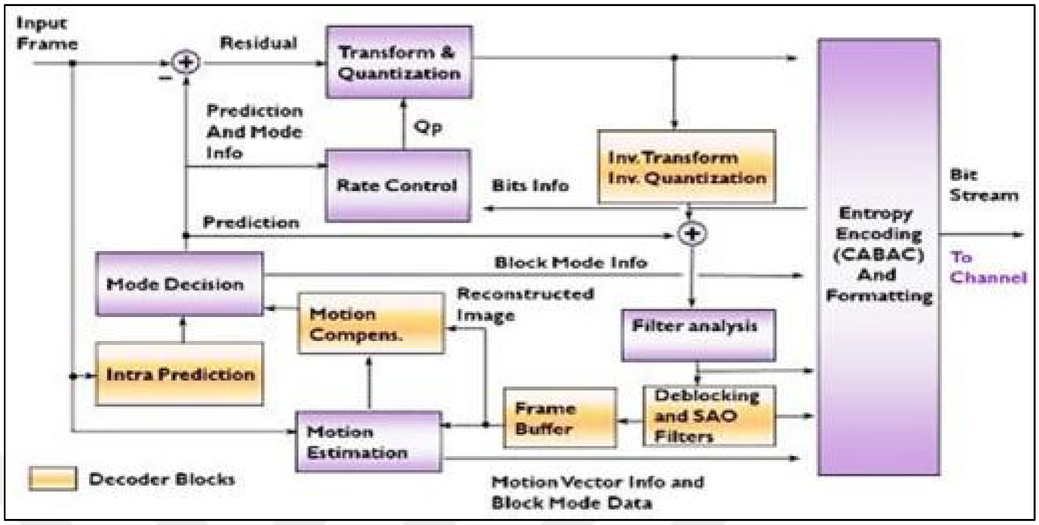


Figure 3.10 HEVC Codec structure

The HEVC design is similar to H.264/AVC, so it is the extension of the previous formats. The mutual characteristics with the others are represented by the superior coding model, many fundamentals of the high-level syntax, integer transforms, a like interpicture and intrapicture prediction algorithms and so on. Although, HEVC has many mutual algorithms with previous formats, it improved them greatly. This section will demonstrate the key dissimilarities with the H.264/AVC [11].

3.3.1 Partitioning:

HEVC uses a tree structure and quadtree-like signalling instead of 16x16 macroblocks. The image is segmented into CTUs (CodingTreeUnits), where each one of them has a luma CTB and the corresponding chroma CTBs with its syntax components [12]. The luma CTB's size can be either 16x16, 32x32 or 64x64 samples, as shown in Figure 3.11. Choosing a suitable size relies on compression quality and so many factors.

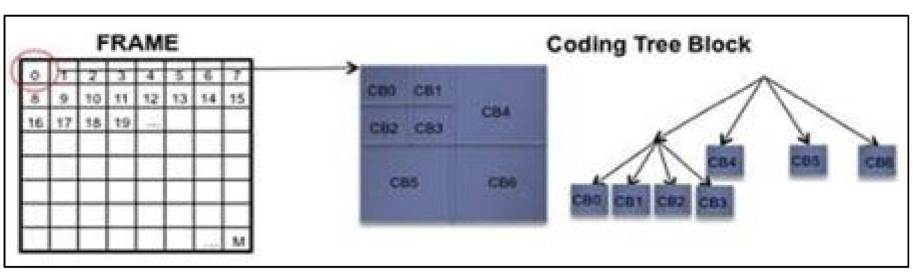
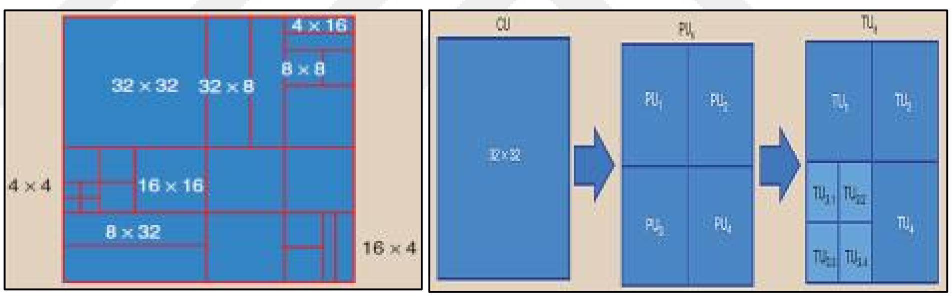


Figure 3.11 Picture partitioning

The CTUs can be divided into CUs (CodingUnits) that have 2 chroma CBs and single luma CB with the related syntax. A CTB may include many CBs of different sizes that reach to a minimum CB size of 8×8 luma samples in a consequence order or just single CB. Also, a CU can be split once into PUs (PredictionUnits) and TUs (TransformUnits) [13]. Deciding the suitable prediction's type is done at the CU level: interpicture or intrapicture. 64×64 to 4×4 samples are the available PB sizes. Moreover, each PU, in interpicture, has 2 or 1 motion vector for bi or uni-predictive coding respectively. The addition of larger block structures is one of the greatest contributions in the HEVC compression efficiency with the flexible sub-partitioning algorithms [14]. LCU (LargsetCodingUnit) is the fundamental block in HEVC that can be as large as 64x64 or recursively partitioned into smaller CUs that can be as small as 8x8 depending on the picture content as shown in Figure 3.12 (a). Also, it is possible to be partitioned further into tinny symmetric or non-symmetric PUs and TUs, as shown in Figure 3.12 (b).



(a) (b)

Figure 3.12 (a) Partitioning of a 64x64 LCU (b) partitioning a 32x32 CU

Single or multi-PUs are contained within each CU, which forms the basis of production. The size of PU can be the same as the root CU or has a 4 x 4 in Luma block sizes. HEVC implements a DCT-like (DiscreteCosineTransform) transformation, as the previous codecs, to the residuals for de-correlating data. In HEVC quantization and transformation operations, a TU is the fundamental unit used in them. The TU's shape and size relates to PU's size. One or multi-TUs may be included within each square CU; which can be partitioned into smaller TUs within a quad-tree partition scheme.

3.3.2 Prediction:

HEVC has 33 directional Intrapicture modes, in comparison with H.264/AVC that has only 8 modes [15]. There are 3 types of block CU' sizes (16x16, 8x8 and 4x4). For example, the samples are predicted from the above and left neighbours that are obtained from previously encoded and re-constructed blocks [16], as shown in Figure 3.13 and Figure 3.14 for 4x4 size.

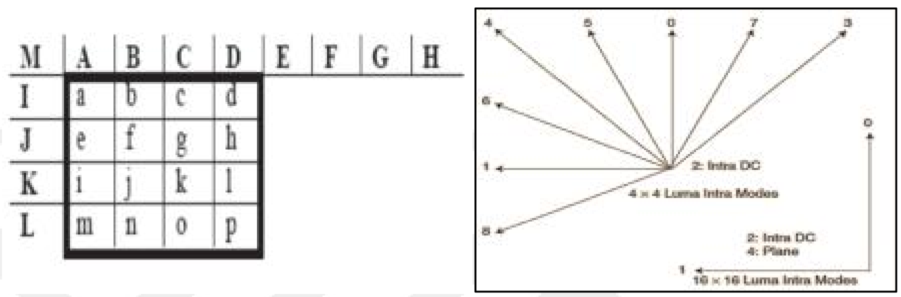


Figure 3.13. A 4x4luma block to be predicted and Intra modes [5]

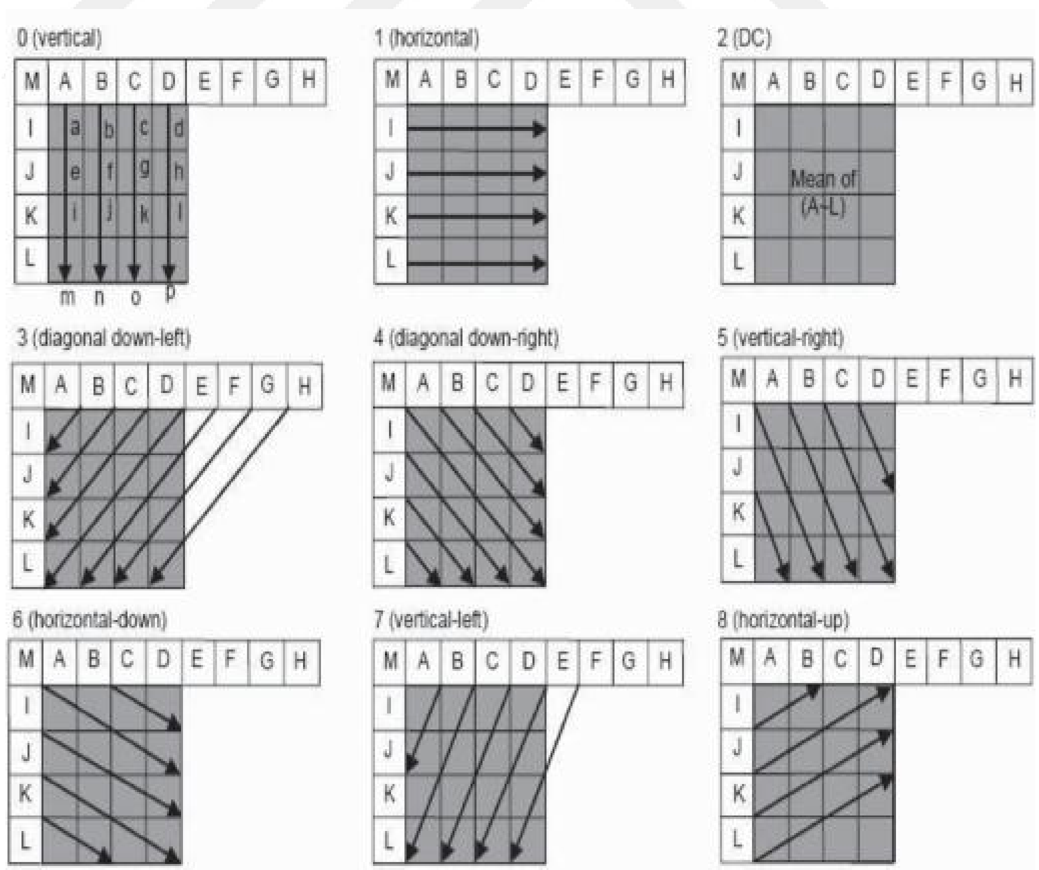


Figure 3.14 Modes from 0-8 [7]

For all 9 modes, the values of a-p pixels are predicted as shown in Table 3.4. In HEVC, there are variety sizes for Intrapredicted PUs ranges from 4x4 to 32x32 [17]. A Single intra-mode is chosen for two chroma CBs; which is determined by a vertical, horizontal, planner, diagonal, DC, or by the same prediction mode used in luma. For each TB (TransformBlock), the intra-prediction mode is implemented independently.

Table 3.4 H.264/AVCModes for 4x4 block

Mode 0 (vertical)	Mode 1 (horizontal)	Mode 2 (DC)
☐ a,e,i,m = A ☐ b,f,j,n = B ☐ c,g,k,o = C ☐ d,h,l,p = D	☐ a,b,c,d = I ☐ e,f,g,h = J ☐ i,j,k,l = K ☐ m,n,o,p = L	☐ All the predicted pixels are equal to mean value of all references (A, B...L)
Mode 3 (diagonal down-left)	**Mode 4 (diagonal down-right)**	**Mode 5 (vertical-right)**
☐ a = (A+2B+C+2)/4 ☐ b,e = (B+2C+D+2)/4 ☐ c,f,i = (C+2D+E+2)/4 ☐ d,g,j,m = (D+2E+F+2)/4 ☐ h,k,n=(E+2F+G+2)/4 ☐ l,o = (F+2G+H+2)/4 ☐ p = (G+3H+2)/4	☐ d = (B+2C+D+2)/4 ☐ c,h = (A+2B+C+2)/4 ☐ b,g,l = (Q+2A+B+2)/4 ☐ a,f,k,p=(A+2Q+I+2)/4 ☐ e,j,o = (Q+2I+J+2)/4 ☐ i,n = (I+2J+K+2)/4 ☐ m = (J+2K+L+2)/4	☐ a,j = (E+A+1)/2 ☐ b,k = (A+B+1)/2 ☐ c,l = (B+C+1)/2 ☐ d = (C+D+1)/2 ☐ e,n = (I+2E+A+2)/4 ☐ f,o = (E+2A+B+2)/4 ☐ g,p = (A+2B+C+2)/4 ☐ h = (B+2C+D+2)/4 ☐ i = (E+2I+J+1)/4 ☐ m = (I+2J+K+2)/4
Mode 6 (horizontal-down)	**Mode 7 (vertical-left)**	**Mode 8 (horizontal-up)**
☐ a,g = (E+I+1)/2 ☐ b,h = (I+2E+A+2)/4 ☐ c = (E+2A+B+2)/4 ☐ d = (A+2B+C+2)/4 ☐ e,k = (I+J+1)/2 ☐ f,l = (E+2I+J+2)/4 ☐ o,i = (J+K+1)/2 ☐ j,p = (I+2J+K+2)/4 ☐ m = (K+L+1)/2 ☐ n = (J+2K+L+2)/4	☐ a = (A+B+1)/2 ☐ b,i = (B+C+1)/2 ☐ c,j = (C+D+1)/2 ☐ d,k = (D+E+1)/2 ☐ e = (A+2B+C+2)/4 ☐ f,m = (B+2C+D+2)/4 ☐ g,n = (C+2D+E+2)/4 ☐ h,o = (D+2E+F+2)/4 ☐ l = (E+F+1)/2 ☐ p=(E+2F+G+2)/4	☐ a = (I+J+1)/4 ☐ b = (I+2J+K+2)/4 ☐ c,e = (J+K+1)/2 ☐ d,f = (J+2k+L+2)/4 ☐ g,i = (K+L+1)/2 ☐ h,j = (K+2L+L+1)/4 ☐ k,m,l,n,o,p = L

A multi-directional spatial prediction is used for reducing the spatial-redundancy through the neighbour-samples as shown in Fig.3.15.

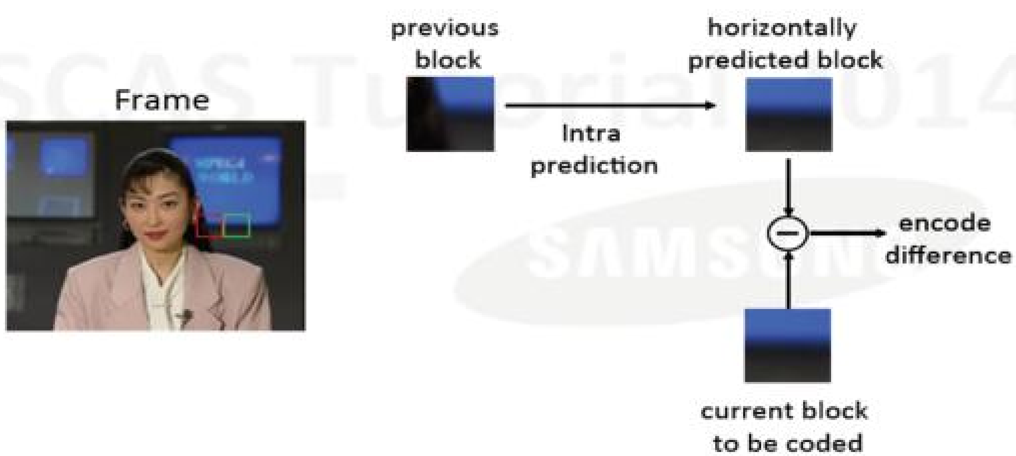


Figure 3.15 The process of Intra-prediction

Similar to AVC, HEVC uses the left and top pixels thar are noted by R(x,y) for predicting a pixel (P) located at (i,j). AVC utilizes the top, left and left-top sides only, unlike HEVC that benefits from extra samples locates in left-below side ($R_{0,N+1}$,...$R_{0,2N+1}$) as shown in Figure 3.16.

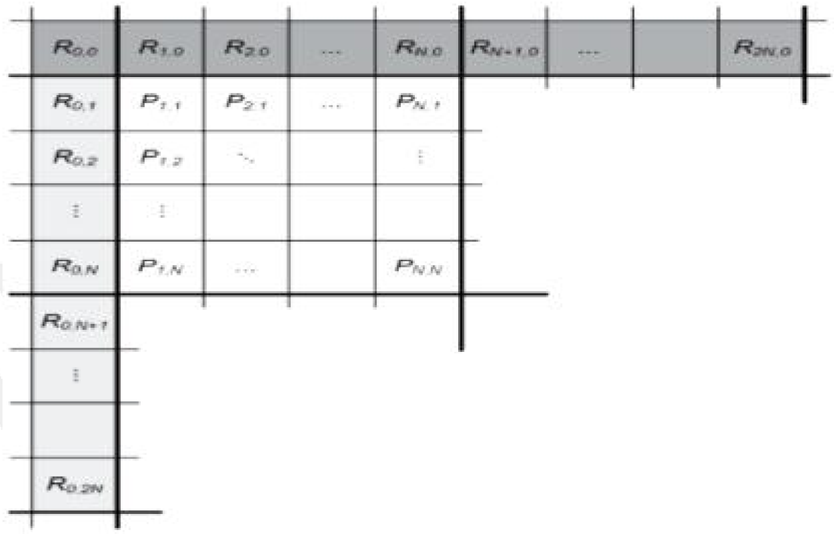


Figure 3.16 The way of predicting P(i,j) by R(i,j) in NxN block size

As mentioned before, HEVC consists of 33 angular predictions as shown in Figure 3.17, where V and H represent the Vertical and Horizontal directions respectively and the numbers beside each one of them represent the displacement as 1/32 pixel fraction [17].

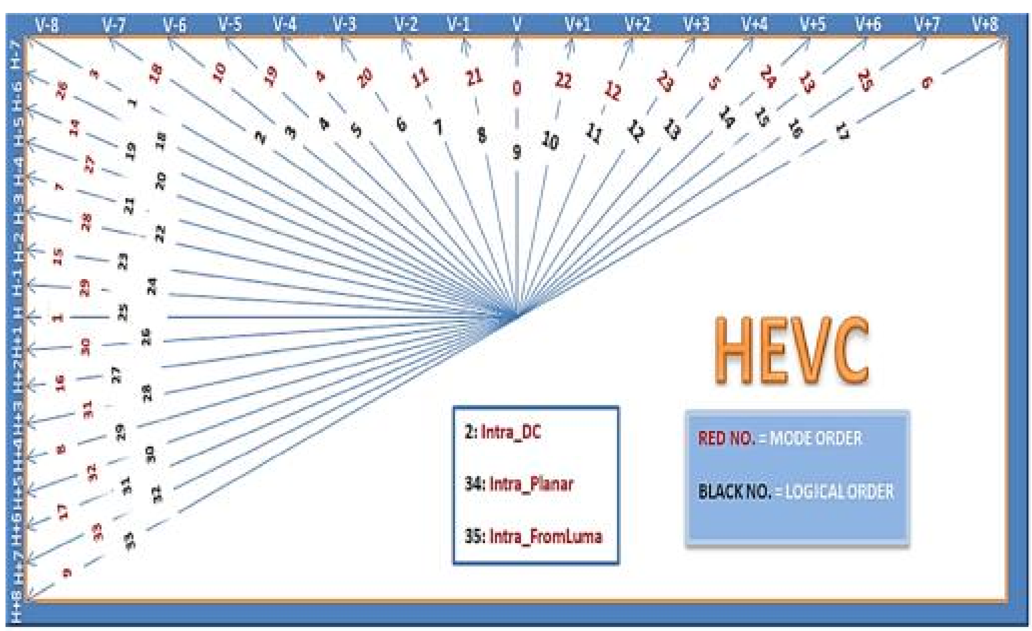


Figure 3.17 Luma intra-prediction modes of HEVC.

All the sample's values are projected using a single reference either column or row; in order to simplify the handling of references [18]. Where, modes 2-17 utilize the Left reference-samples and modes 18-34 utilize the above reference rows. Predicting the sample's value of P(i,j) is done using the linear interpolation as follows:

$$P(i,j) = ((32 - K_j) \cdot R(x,0) + K_j \cdot R(i+1,0) + 16) >> 5 \tag{3.7}$$

$$F_j = (j \cdot s) >> 5 \tag{3.8}$$

$$K_j = (j \cdot s) \,\&31 \tag{3.9}$$

$$x = i + F_j \tag{3.10}$$

Where 'R(x,0)' represents the reference sample, 'K_j' results from weighting two reference sample in the prediction mode's direction, , >> denoted a right bit-shift or in other word a division by two (i.e. 32) and then flooring results, 's' is the displacement tangent with relative to 1/32 pixel fraction that ranges from -32 to +32, '&' refers to the bitwise AND. The vertical intra-frame prediction of modes (18-34) is shown in Equations (3.7-3.10); in which the values of 'i', 'j' and the index are just swapped for the horizontal cases.

There are other two modes in addition to the angular modes, which are the DC mode (it computes the mean values of reference samples) and the planar mode (it computes the average of linear prediction at four reference-corners), as follows:

$$P^V(i,j) = (N-j) \cdot R(i,0) + j \cdot R(0,N+1) \tag{3.11}$$

$$P^H(i,j) = (N-i) \cdot R(0,j) + i \cdot R(N+1,0) \tag{3.12}$$

$$P(i,j) = P^V(i,j) + P^H(i,j) + N >> (\log_2(N) + 1) \tag{3.13}$$

Where, N represents the block size.

3.3.3 Transform and Deblocking Filter:

A square transform is utilized for coding the residual after implementing the prediction process. The transform size can be the same as the CB or it may be divided into 4 tiny TBs. HEVC can utilize a 4x4 to 32x32 transform sizes for the DCT-like process. Typically, the new coding techniques encode the video at bit rates that work out to substantially less than 1 bpp (BitPerPixel) [19]. Typically, such a process is accomplished by gathering pixels together, into 4x4, 8x8, 16x16 or 32 x 32 blocks. After that a two-dimensional *DCT* transformation process takes it role with those blocks. The DCT works on **'X'**, a set of $N \times N$ samples (either residual values after prediction or picture samples) and produces **'Y'**, an $N \times N$ set of elements. The DCT process (and its inverse, IDCT) are defined according to a transform matrix **'A'**. The FDCT (ForwardDCT) of an $N \times N$ block sample is as follows:

$$\mathbf{Y} = \mathbf{A}\,\mathbf{X}\,\mathbf{A}^{\mathrm{T}} \tag{3.14}$$

and the IDCT by:

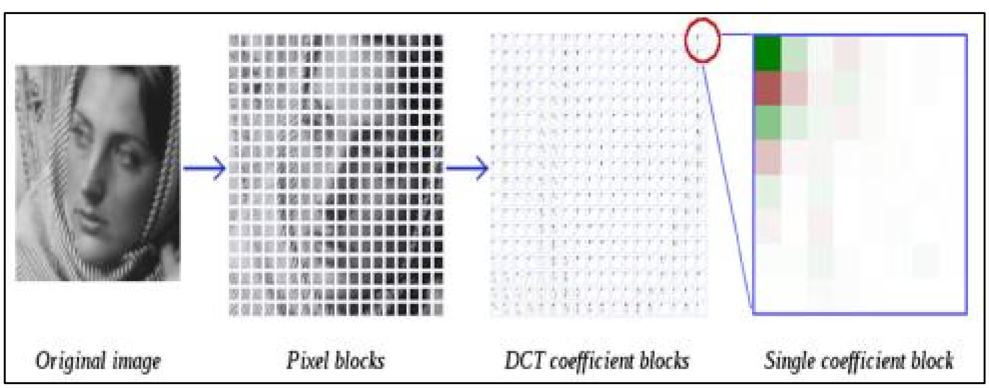
$$\mathbf{X} = \mathbf{A}^{\mathrm{T}}\,\mathbf{Y}\,\mathbf{A} \tag{3.15}$$

Where **Y** is a coefficients matrix, **X** is a samples matrix, and **A** is an $N \times N$ transform matrix. The **A's** components are:

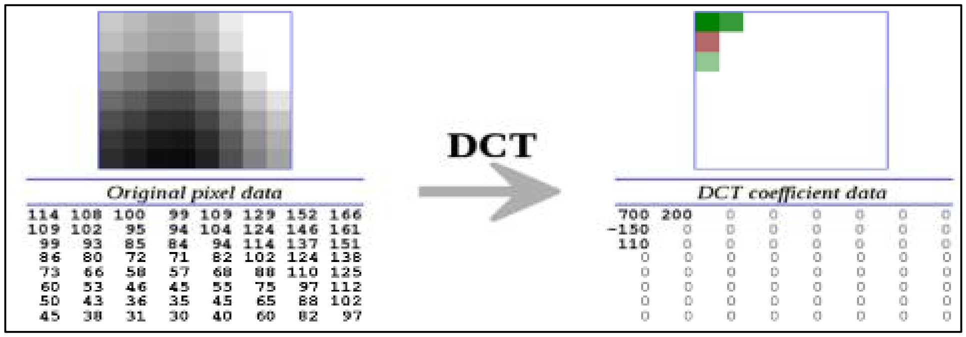
$$A_{ij} = C_i \cos\frac{(2j+1)i\pi}{2N} \qquad \text{where } C_i = \sqrt{\frac{1}{N}}\ (i = 0), \qquad C_i = \sqrt{\frac{2}{N}}\ (i > 0) \tag{3.16}$$

Equation 3.14 and equation 3.15 can be in a summation format:

$$\begin{aligned} Y_{xy} &= C_x C_y \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} X_{ij} \cos\frac{(2j+1)y\pi}{2N} \cos\frac{(2i+1)x\pi}{2N} \\ X_{ij} &= \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} C_x C_y Y_{xy} \cos\frac{(2j+1)y\pi}{2N} \cos\frac{(2i+1)x\pi}{2N} \end{aligned} \tag{3.17}$$



(a)



(b)

Figure 3.18 Illustration of DCT operation on 8x8 block

Figure 3.18 illustrates a DCT transformation process for 8x8 grayscale pixels block to an 8x8 frequency coefficients block. Although the number of frequency coefficients in DCT is the same as the input pixels, it typically represents the same data with a further compact shape; which can be saved in lesser bits than the spatial domain.

The DCT coefficients are then quantized by the encoder to a lower precision. An inverse DCT-2D is then performed by the decoder to re-create every block of the source picture.

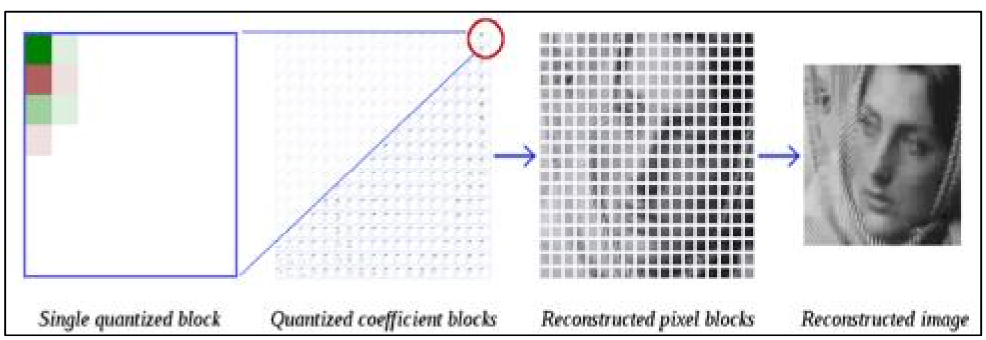


Figure 3.19 Operation of IDCT of video decoder

Reconstructing a source picture from sets of quantized DCT coefficients is the fundamental operation of the decoder [3]; which is illustrated in Figure 3.19. The IDCT reconvert every coefficient to a set of pixels by an inverse DCT. Then, the sets of pixels are reconstructed to form a full picture.

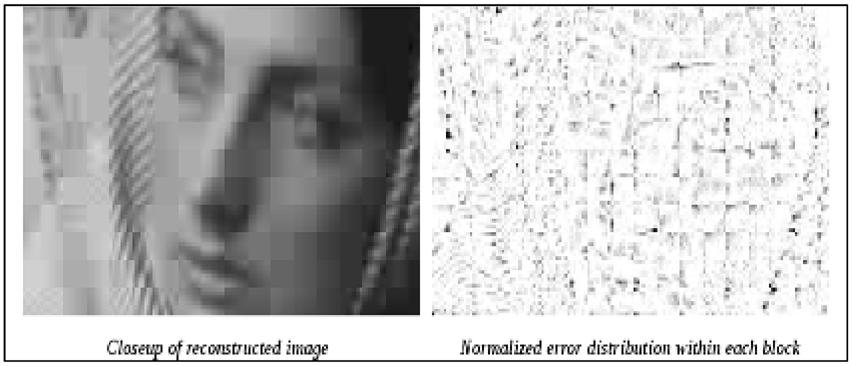


Figure 3.20 Screen Door Effect

The effect of 'screen door' is shown in Figure 3.20, in which a clear grid of borders is shown in the reassembled blocks. Such an issue is particularly obvious in video, in which the grid remains constant as the objects moving.

That's why a deblocking-filter is used to solve this issue by smoothing the edges between blocks. It also called a *loop filter* that may be available inside the encoding loop or it may be a post-processing stage [20].

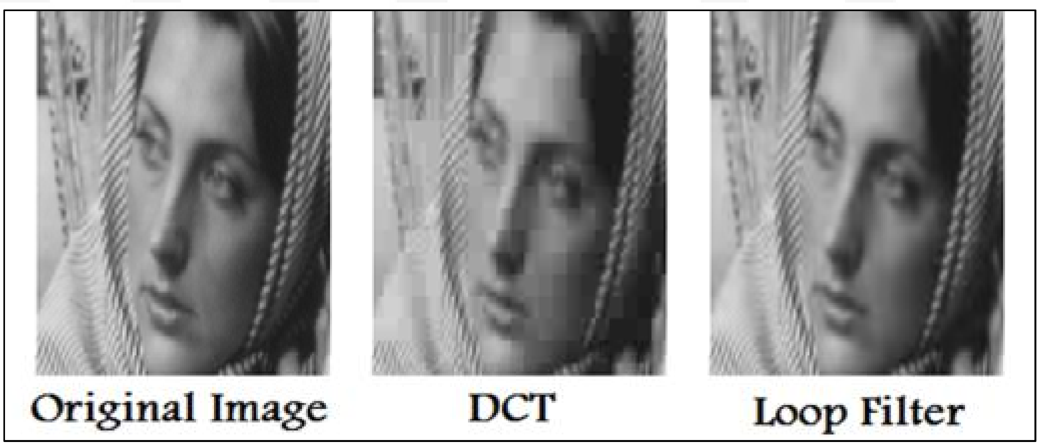


Figure 3.21 Effect of DCT and Loop Filter Transform

A Comparison between source pictures and transformed/quantized pictures by the loop filter and DCT is shown in Figure 3.21.

3.3.4 Entropy Coding:

The CABAC (ContextAdaptiveBinaryArithmeticCoding) is utilized for entropy coding like the H.264; but with improved coefficient that leads to a superior quality and throughput for larger transform sizes [21].

Entropy coding is performed after transformation to code all quantized transform coefficients and syntax elements. CABAC is the entropy encoder of HEVC that offer higher coding performance than CAVLC of H.264, due to its higher sophisticated context modelling and arithmetic coding engine [22]. Although CABAC enhances the coding performance, it raises the coding complexity. It is more obvious at higher bitrates [small QPs (QuantizationParameters)], in which the transform coefficient has a master role in encoded streams.

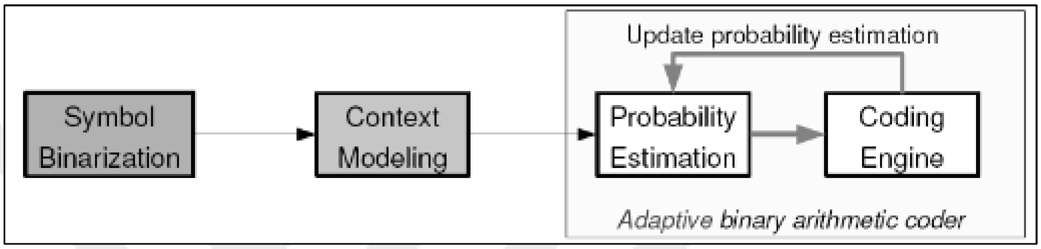


Figure 3.22 CABAC structure

There are main components within the CABAC design; which are the binarization, context modeling, and arithmetic coding as shown in Figure 3.22. These components are demonstrated by the primary building sets of the CABAC encoding block layout, as shown above [23]:

1. ***Binarization:*** a Binary Arithmetic Coding is used for encoding only binary decisions.
2. ***Context model selection***: It is a probability model for single/multi-bins of the binarized -symbol represents the Context Model; which is taken from a set of available models based on the statistics of latest-coded symbols. It saves the probability of each bin (‘0’ or ‘1’).
3. ***Arithmetic encoding:*** It encodes each bin on the chosen probability model.
4. ***Probability update:*** The chosen context model is updated upon the real-coded values (e.g. when the bin = ‘1’, the frequency count of ‘1’ increases).

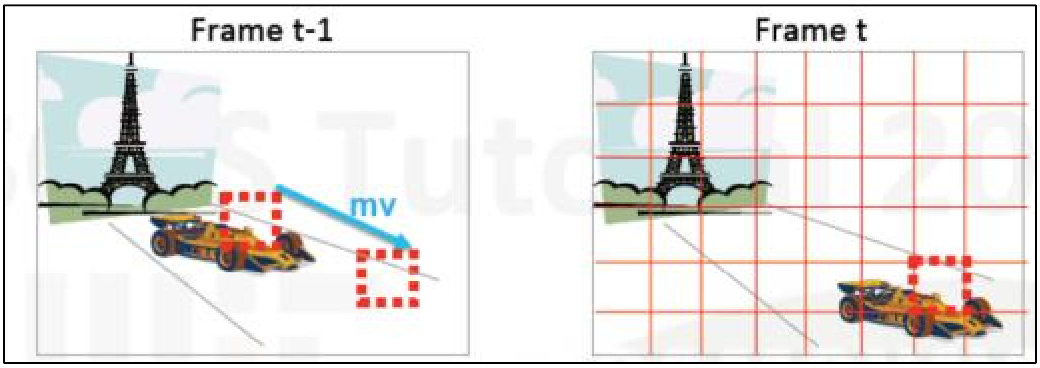


Figure 3.23 Motion Estimation

3.3.5 Motion Estimation (ME) and Compensation:

The popular algorithm of Motion compensation is compensating the movements of rectangular 'blocks' of the current frame, as shown in Figure 3.23. The next steps are applied for every set of $M \times N$ samples within the current frame [24]:

1. Searching a region in the reference frame (a previous or next frame that is encoded and sent previously) to obtain a best match of M × N-sample block. The energy in the residual is the commonly matching standard; which is resulted from subtracting the candidate area from the current $M \times N$ block. Where, the best match is the candidate area that reduces the residual energy. The demonstrated process is called Motion estimation.
2. The selected candidate area will be the predictor for the current $M \times N$ block; which is subtracted from the current block to produce Motion *Compensation* (residual $M \times N$ block).
3. The residual block will be coded and sent. The offset between the location of the candidate area (*motion vector*) and the current block will be also sent. The decoder is utilizing the obtained motion vector for re-producing the predictor area, then decoding the residual block, and adding it to the predictor for reforming the source block.

3.3.6 Mode Decision:

Choosing the best mode for an input set of image is one of the real-time encoder demands [25]. It requires the estimation of various choices to select the suitable mode that provides a lower cost, to obtain lowest rate and distortion. Such a process is more complicated in HEVC since it need to select either intra or inter modes for an input CU; determines the intra prediction mode, the split type of inter prediction, and the transform size and the CU partitioning [26]. All the mentioned tasks are preformed within an RDO process. Moreover, the encoder should be complied with the bit-rate restrictions; which in turn produces a new level of complexity.

3.4 HEVC Complexity Aspects

In HEVC, the complexity calculations of the main modules are superior to H.264. Since the most complex section is the encoder, because of the various options [27]. However, the increment of intra-modes has increased the flexibility. Thus, it's fairly reasonable that HEVC added extra complexity regarding prior codecs, about (1.6%) and (4.5%) increment in the decoder's and encoder's overhead. Moreover, the reservations of memory-buffer are greater than the prior codecs; which is not proper for real-time operations [28].

The encoder and decoder complexity is big as there is a very tight feedback loop between the arithmetic coder and context modeller CABAC. At the decoder, the feedback loop is stricter than at the encoder side, because a model is required for decoding a symbol that is necessary in the calculation of the following model [29]. The main bottleneck is the arithmetic decoder since it processes all the encoded data consecutively. And there is no way around the arithmetic decoder, all the data in the bitstream has to pass the entropy decoder in the H.265 scheme before other blocks begins the decoding such as IDCT and motion compensation.

The issue increases in the higher bitrate and resolution [2]. A video of resolution 352x288 pixels at 30 fps produces 1.5 million decoded symbols/sec, while 50 million decoded symbols/sec will be generated from the HDTV resolution of 1280 x 720 at the same rate.

Determining the best mode of the lowest distortion and rate within the specified block is the main task of a real-time encoder within an image; which is implemented through the estimation of various options [30]. This operation is somehow difficult as the HEVC chooses either inter or intra modes for each CU, decides the CU-transform and partitioning-size, adapted with the bitrate limits, etc.

There are 35 modes depending on the PU-size, which resulted in an extra overhead regarding the prior codecs. The HEVC-angular modes are further complicated in calculation than the directional H.264-modes [31]. Also, the reference patterns may be refined before any prediction, as in AVC/ H.264, but in a more selective way, relying on the prediction mode. The increment in the prediction-modes will need a proper mode selection method to preserve an acceptable search complexity.

The RDO process determines the most effective CTU-partitioning in the mode decision [32]. Where, this process depends on the estimation of the compression Distortion (D) and Compression Rate (R). Although, it estimates the trade-off between (R) and (D) to determine the criterion of selecting the best coding option, it's a performance bottleneck due to the diverse coding options in HEVC than AVC. So, 341 possible subblocks must be computed for selecting the optimal coding-option, which causes poor compression efficiency [33].

There are 11,935 compressions (341 sub-block x 35 intra-modes) in every CTU that are utilized in RD-cost's estimation. Therefore, a full RDO is not utilized in any company because of its complexity. As a result, finding an efficient algorithm for the RDO mode becomes an essential matter [34].

In addition to the mentioned above, HEVC has a capability of simultaneous processing of LCU multiple parts. The HEVC-CU includes different sizes (128×128, 64×64, 32×32, 16×16, 8×8, and 4x4). For example, four equivalent-sizes of CUs are resulted from splitting an LCU that may be encoded directly or divided to another 4 CUs based on the RDO value as demonstrated in Fig. 3.24. Terminating a split process depends on the size of resulted CU (smallest size) [35]. Thus, an efficient encoder has a more flexible quad-tree structure that results in a high coding' gain based on the varying of picture's content effectively [36]. As a result, a new tree-structure of 64x64 LargestCodingUnit (LCU) results in a 12% average reduction in the bitrate compared to 16x16 LCU.

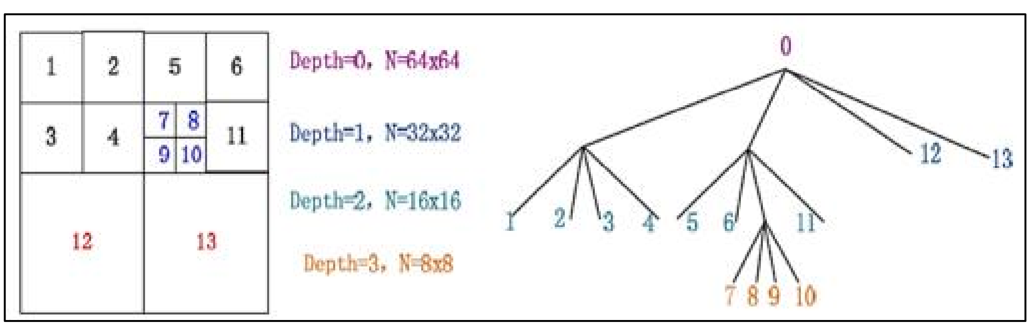


Figure 3.24. HEVC Quad-Tree structure.

A comparison between H.256 (HEVC) partitioning and H.264 is shown in Figure 3.25. Note that, LCUs define regions of lesser diverse contents than others and vice versa [37].

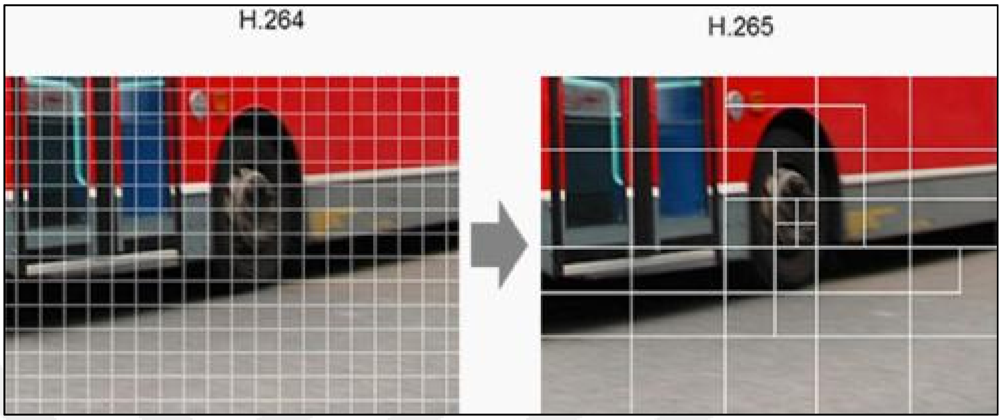


Figure 3.25 A partitioning-comparison between HEVC and H.264

Although, a significant bitrate's minimization can be achieved with the LCUs, the HEVC encoder should compute every possible CU for determining the optimal one, which results in a further computations' complexity [38]. Where, finding the optimal CUs is like implementing the encoder process 4-times on an LCU because of the multi CU's sizes (64, 32, 16, and 8). Therefore, a 75% of the computation resource is dissipated because the encoder will select just one optimal size from the 4-sizes [39], [40]. This waste in the calculation's time and resources is not proposer for realtime services.

3.5 MPEG Audio Compression

Like the video compression, the MPEG standard has different audio compressing techniques. The Lossy compression is the type of all MPEG audio encoders. Moreover, they are all perceptual encoders, where the compression technique is working upon what can the human-hearing system detects and what can't. A significant amount of audio compression can be obtained through a perceptual coding. The human ear sensitivity is greater than the eye sensitivity to the small distortions obtained by a compression process. However, the human ear can't hear short streams of less than a millisecond, thus some data can be ignored from the audio stream with small or no penalty. Also, the quite audio stream may become hidden by a loud one that instantly precedes it or follows it, since the human ear is limited to a range of frequencies [6].

The MPEG technique takes into consideration all the human-hearing limits when compressing an audio stream. The Layer I of MPEG is the simplest compression method. On each compression process, it uses 384 input samples that relates to 8 msec audio stream using 48-kHz sampling. It can generate a fixed output bitrate at a 4:1 compression ratio, so it is possible to compress a 1.4-Mbps CD-quality audio stream to occupy a 384-kbps stream without a remarkable loss in quality, but it will be noticeable if the Compression beyond this to 192 or 128 kbps. For Applications that requires high fidelity, but not the highest quality, a lossless compression is the best choice. In MPEG-4, there are various audio objects as well as videos, which can be synthetic or natural objects. The assembling of all audio objects is the responsibility of the decoder to produce a final combined object. AAC signals are usually utilized in video/audio streaming and are supported on portable tools like the iPod. As an overall, the MPEG audio has more flexibility and doesn't consume much power as MPEG video. By increasing the number of layers, then the encoder/decoder complexity and the compression ratio increase. A summary of all the key characteristics of various audio levels is demonstrated in Table 3.4. The Software-only Layer III decoders can work smoothly within a various types of PCs. Also, the AAC decoders are becoming very widespread because of the Apple Company’s support. In order to choose a suitable audio encoder, then the total bandwidth should be high enough for carrying the video/audio stream with control signals. A higher compression ratio in an Audio-encoder will permit more bandwidth for video signals [3].

Table 3.4 Comparison of MPEG Audio Formats

Compression Method	**Samples Per Block**	**Compression Ratio**	**CD-Quality Bit Rate (kbps)**
Layer 1	384	4:1	384
Layer 2	1152	8:1	192
Layer 3	1152	12:1	128
AAC	1152	16:1	96
HE AACv2	2084	32:1	48

These compressed file formats allow your games to come with 10s of hours' worth of sound effects, vocal tracks, and music, all at a reasonably small space requirement. However, remember that at some point, the compressed data must be turned back into raw PCM data, for sending to the sound card. Figure 3.26 shows an example of our sound wave being transformed into PCM data.

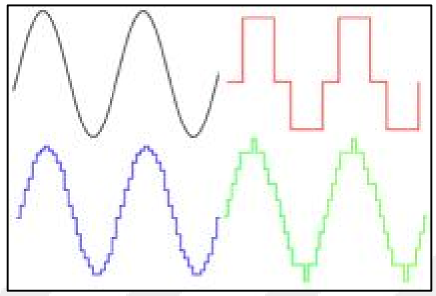


Figure 3.26 Original wave signal quantized at different bit depth and frequency

The PCM data for a single music track in CD quality will take up around 80MB - that's quite a lot, especially for consoles or mobile devices! It's therefore not feasible to store an entire track in memory at once; but what we can do, is streaming the data as we need it. When playing back the audio, we will very nearly always be playing back the PCM data from its beginning to end, so once a particular section of PCM data has been played, it doesn't really need to be in memory at all! In fact, we only need enough data from our PCM data stream to play back a short section at a time – the length of this section is determined by how quickly the next section of data can be loaded from disk and decompressed back into PCM data. If the section loaded up is too short, it may have finished playing before the next section has been loaded and decoded, resulting in playback artifacts (sections of sound repeating, pops and clicks, pauses in playback); conversely, if the section is too long, it will be wasting memory.

The simplest way to handle the streaming of audio is to have a number of short buffers, instead of one big buffer for your streamed audio. These buffers can then be filled up with some PCM audio data, and sent to the sound hardware, one after the other, via a queue. Then, as each buffer is popped of the queue, and its data consumed by the sound hardware, it can be filled up with new data for the next section of the audio track, and popped back on the queue as shown in Figure 3.27. Providing that the disk access and audio decoding does not take too long, this will result in seamless, streamed playback of an audio file of any size, without the large memory footprint.

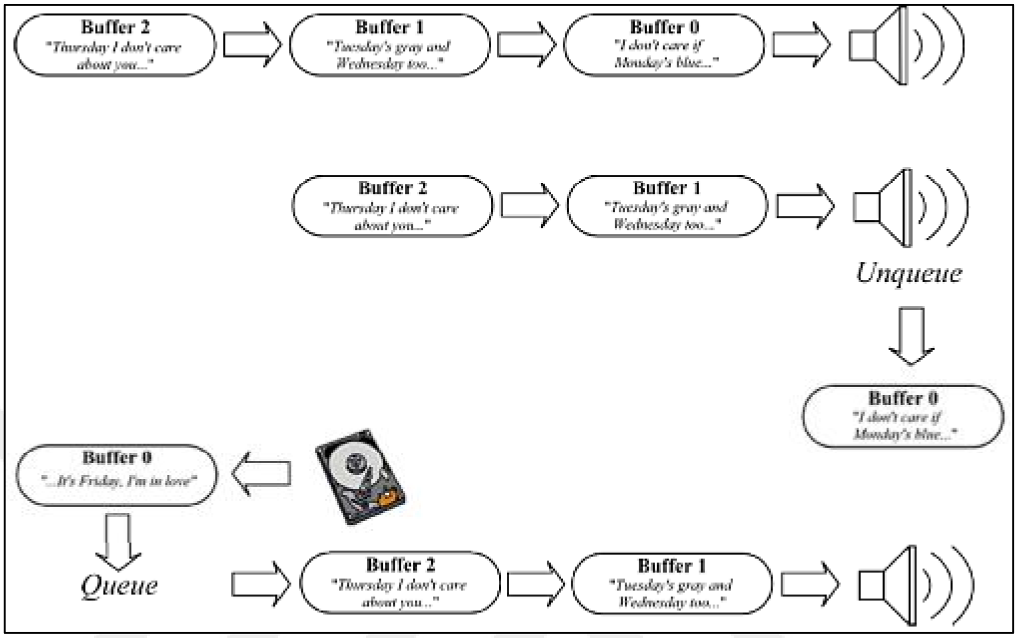


Figure 3.27 Example of Audio Buffer

CHAPTER 4

SYSTEM DESIGN, INTERFACE & RESULTS

4.1. Requirements

There are a lot of requirements when designing a Teleconference system. One of these requirements is the ease of use. Thus, our system has been designed in an easy way; so that even a person with fair computer knowledge can deal with it. It is self-explanatory and able to handle large customer base. Another requirement is the independence, since our system is a database independent and operating system independent. It has the ability for upgrading in the future and designed as a marketable product.

If the data reaches to the end-server in a reliable and high-quality form that what really makes a scalable and a distributed transmission network. Still the real-time streaming suffers from many issues because of the obtained errors when transmitting from the encoder to the end-server, such as packet loss, and buffer handling. Typically, there is a buffer of few seconds contained within the end-server for holding the arrived data that is appended at the end of the buffer for a streaming process. Where, the size of buffer effects on the recovering process of lost data. In order to have loss-free packets with a low computational overhead, we have implemented a reliable technique in our system; in which a TCP protocol has been utilized for reliable connections. Also, the data have been compressed to the smallest possible ratio using our proposed algorithm with an acceptable level of quality.

Moreover, in order to avoid Scalability and performance issues, a new encoding technique has been used which is HEVC. The videos can be encoded in HEVC at the lowest bitrate without losing the quality and the compression-ratio is doubled regarding H.264/AVC.

It has been developed in order to target the ultra-high resolution with higher frame rates regarding the prior codecs. It also ensures that the speech, images and other data are transferred effectively in a real-time communication. The system utilized the multi-core processors to have a better multitasking, improved video rendering with a higher privilege in server performance. Since building 2 cores or more within a single Die increases the processing power with maintaining the clock speeds at an efficient level. Also a processor of 2 cores processes instructions with a less energy than a single-core that works at a double clock-speed. Figure 4.1 shows a simple demonstration of our Teleconference system.

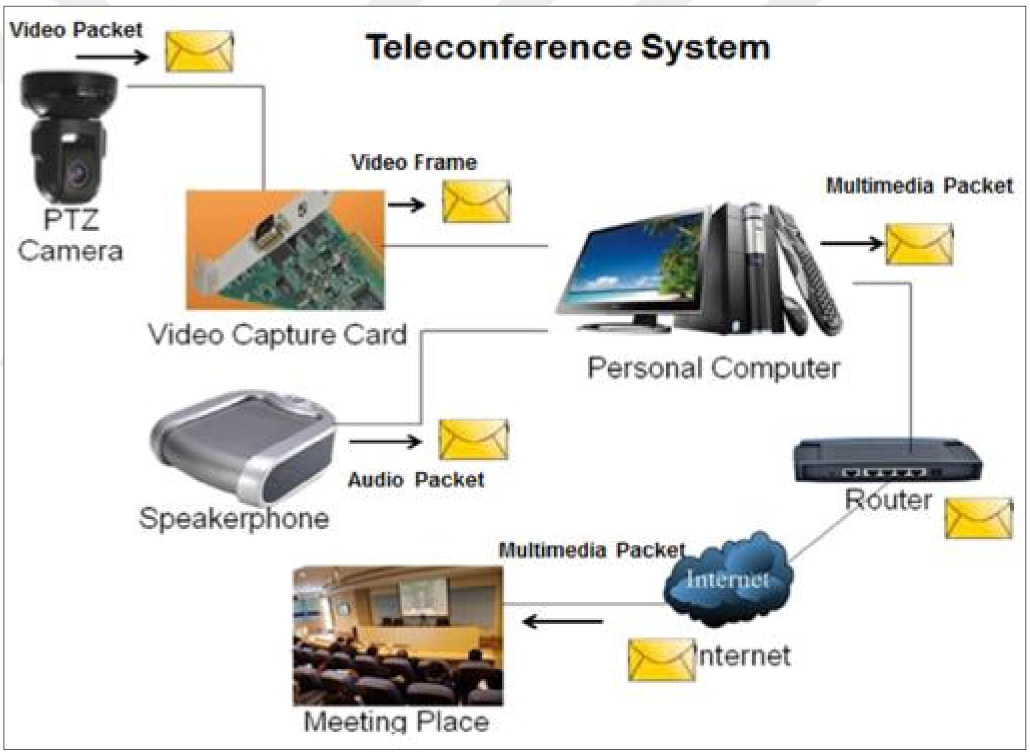


Figure 4.1 Our Teleconference System

4.2 System Flowcharts and Streaming Process

Our Application's environment is visual C++ 10.0, under the Windows Vista operating system. In our system, two major issues have been faced from the core of the real time processing. Therefore, the multi-thread issue will be discussed, and then the data buffering issues with the solutions.

4.2.1 Multi-Thread Processing & Packet Loss Recovery:

When requesting an online (live) stream from an end server, the server starts to encode the requested data for streaming it to the requester clients. Such a process occurs in a consequence form. Thus, on each datum-propagation, in order to make sure that each datum arrives successfully with no loss or a little data loss as possible as we can, we used a TCP-transport layer protocol. This protocol guarantees a reliable connection that handles the retransmission of garbled or dropped packets with the ACK signals on each arrived data as well as the data re-ordering and flow-control intercommunication system that deals with the network congestion in a manageable manner.

In the Audio part, the current-chunk is recorded during the encoding of the last-chunk. Therefore, they were implemented under multiple threads for a real-time processing, with the support of two flags, one for recording and another for the encoding as in Figure 4.2. Also for the video part, in order to save the memory a Quantization parameter of value equal to 32 has been chosen with the use of multi-thread tasks.

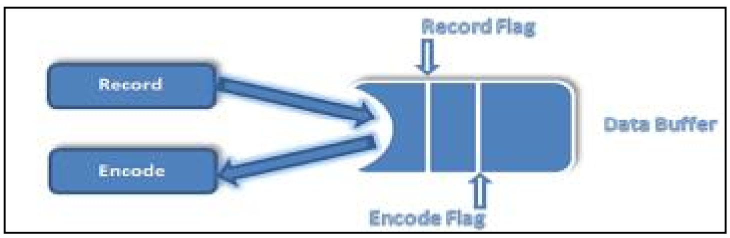


Figure 4.2 Multi-Thread Process

Therefore the system has been enhanced by using Multi-Core Processors to get a Better Multitasking, Improved Efficiency of Threaded performances, integrating with Modern Operating Systems, Improved Video Rendering, and Increased Server Performance.

4.2.2 Radio Link Buffer Management

A wireless Server/Client data-model has been implemented in our work. In total, multiple clients within a coverage region of a base-station will obtain the multimedia sequences from the server. The NALUs on each transmission are transmitted from the edge-server to the clients after encapsulating them to TCP-packets within the radio-link buffer.

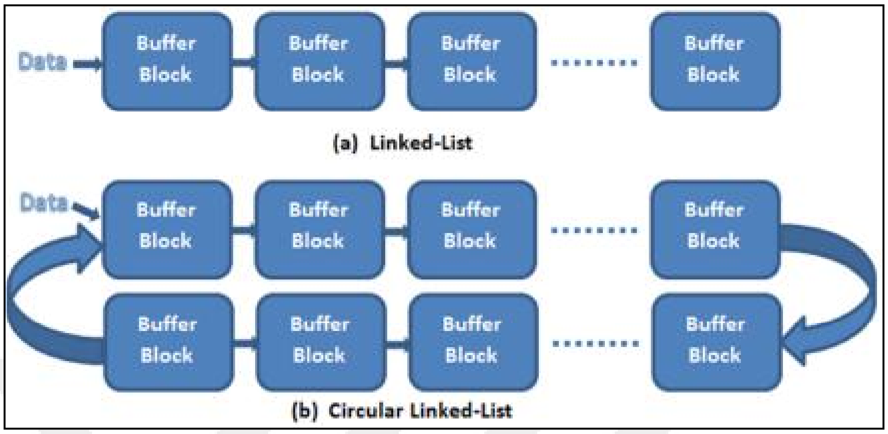


Figure 4.3 Two types of Linked-List

The scheduler operation gives a priority for each client to access the resources, while the RAU (ResourceAllocationUnit) schedules the client's transmission on the downlink/uplink. The causes of congestion/overload issues returns to the competing streams on the same resources or the bad scheduler that makes the radio link buffers not to be served as fast as enough. Thus, limiting the buffer's size and getting rid of the long waited data in the buffer are useful for real-time processing.

So, the buffer handling represents a key factor for real-time processing; which has two types (linear and circular linked list) as shown in Figure 4.3. The 1st one is simpler, but takes a lot of memory. The 2nd one is complex, but saves the memory. So we used the circular linked list.

4.2.3. Server-Side Interface

The server side will create a secure-reliable interconnection. Then it will address its connections according to its IP address and port number (Bind Operation). Then it will listen to any request from clients. If the authentication process is successful, then the send/receive process for packets will start immediately; but if the user has no record in the Teleconference system, then an Authentication Failure message will be displayed.

After checking the existing user record, an Authorization message will be shown to give the administrator the ability either to accept the user logon to the session or reject him/her. For each open session a list of joined clients will be shown in the server side. If the administrator has finished the teleconference session, then he will close the server program and then an alarm goes to each client informing him/her about the shutdown of the server to and then forced to close the open sessions. Figure 4.4 shows a Teleconference-Server Flowchart.

4.2.4. Client-Side Interface

A list of the login-users with their login\logoff time and the existed audio/video capture devices will be available on each client. After a successful login for the user then he will be able to use the Teleconference system. Figure 4.5 shows a Teleconference-Client Flowchart.

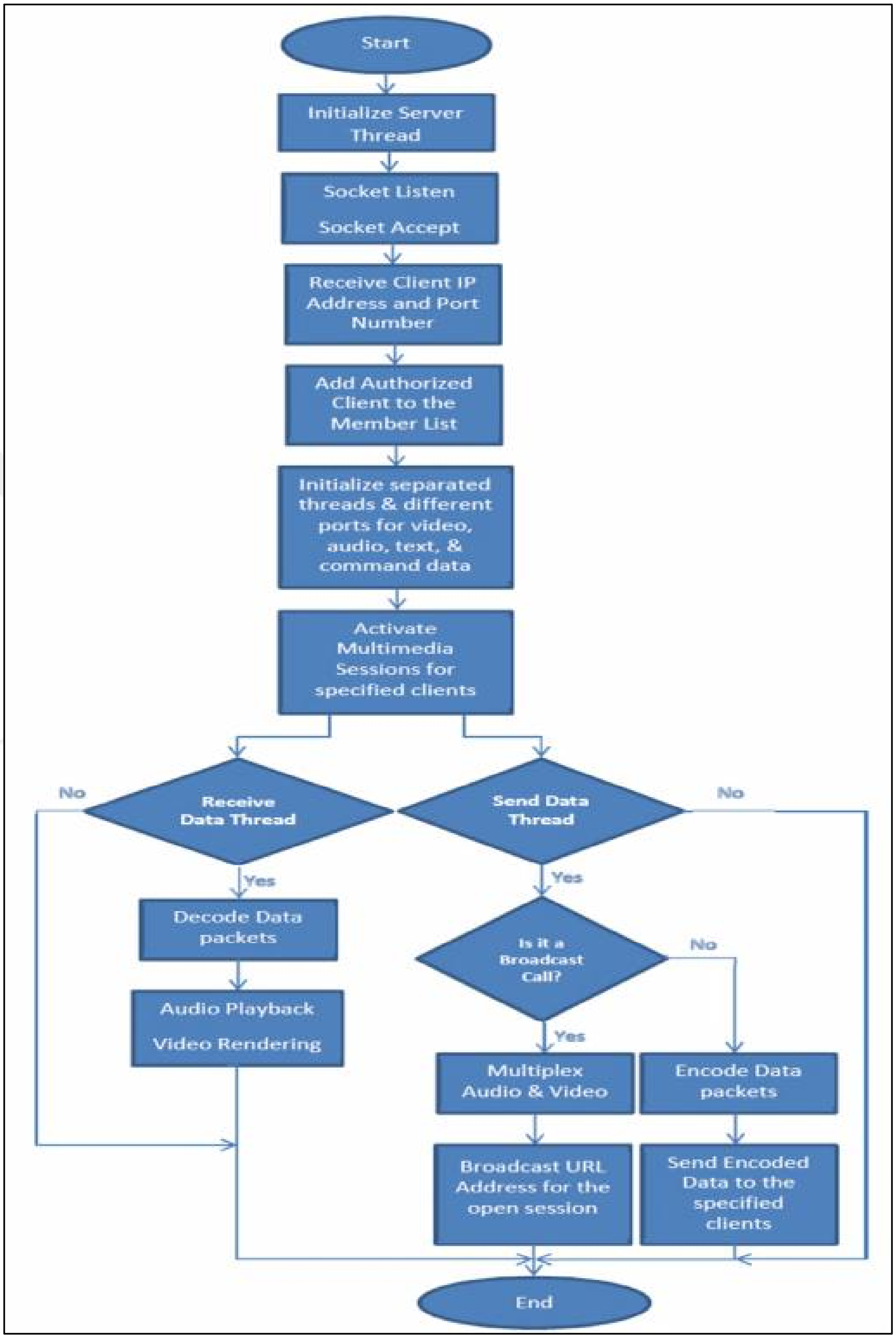


Figure 4.4 Teleconference-Server Flowchart

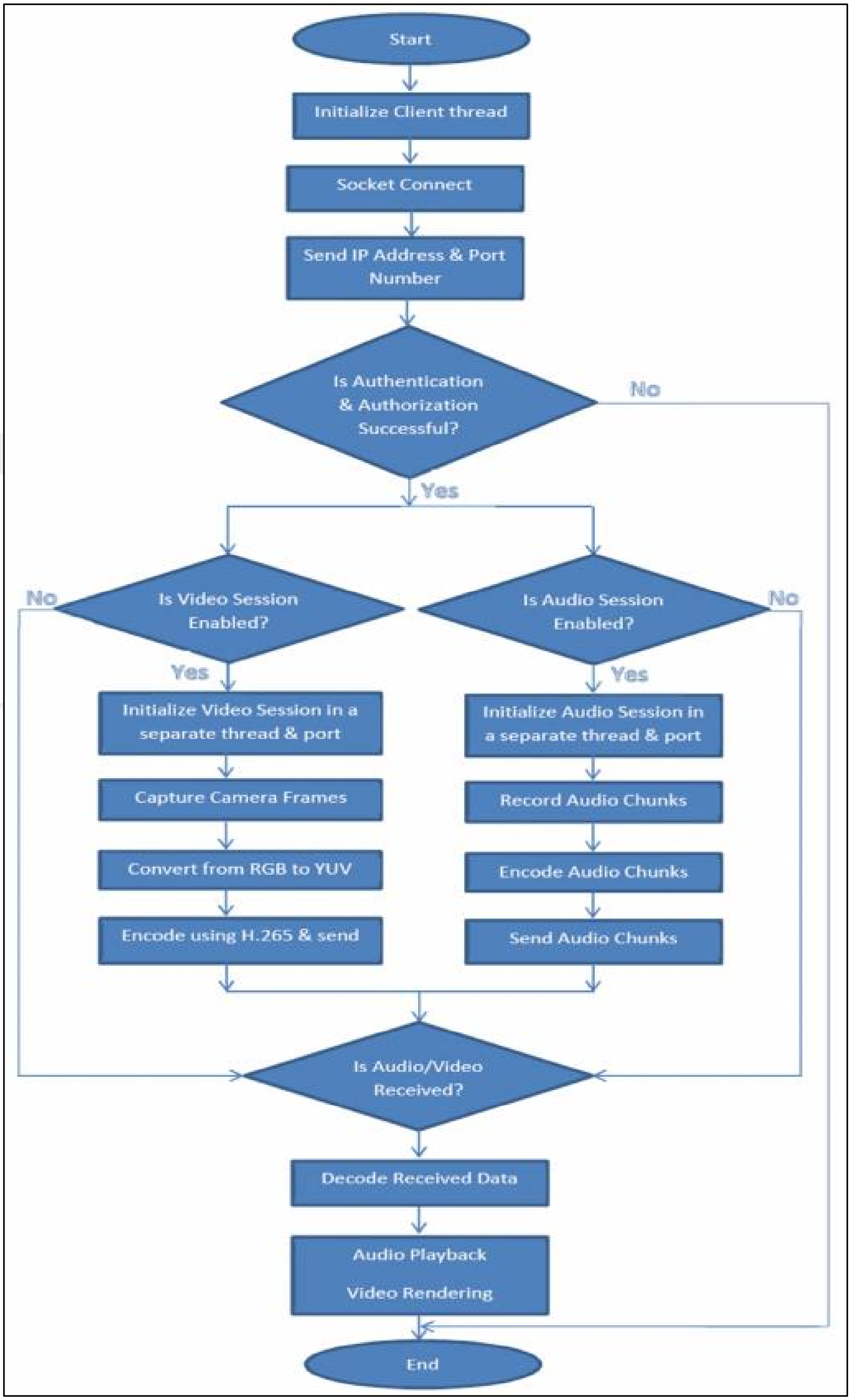


Figure 4.5 Teleconference-Client Flowchart

4.3 Key Components

In a standard Video/IP system, there are four main parts, video capturing/rendering, speech recording/playback, CODEC (encoder/decoder) and socket communications. Before transmitting the signal to IP nets at the transmitter, it is gathered and encoded; and vice versa at the receiver. The received data stream is decoded and recovered in its original form and then played back.

4.3.1 Video Capturing/Rendering:

Capturing the frames is done by the C++ DirectShow libraries and rendering them is done using the C++ SDL libraries by the Video Card Hardware. Figure 4.6 shows the video Capturing/Rendering diagram. SDL is a cross-platform software development library designated to grant a low level hardware abstraction layer to PC H/W devices. It can be used to implement high-privileged multimedia programs that can run on different OS.

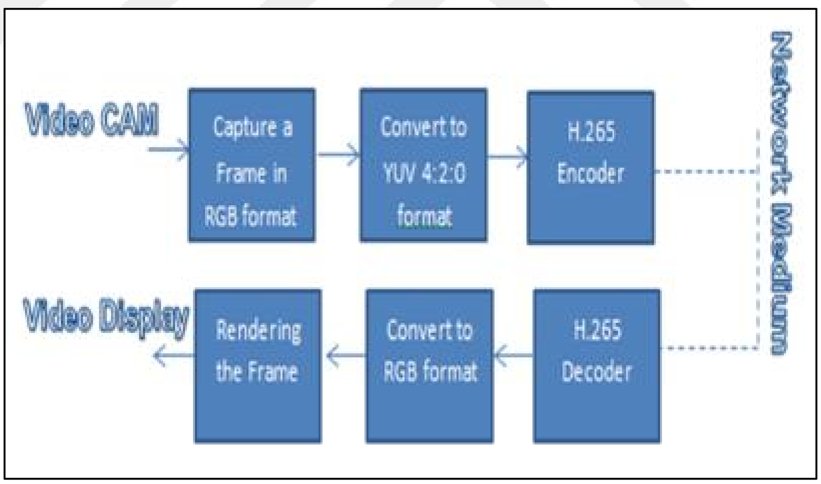


Figure 4.6 Video Capturing & Rendering Diagram

Each frame is captured using the OpenCV filter in an RGB format. A conversion from RGB to YUV 4:2:0 is done using a C++ function to make the frame ready for encoding as a raw format using the H.265 Encoder. After receiving the encoded frame throughout the socket implementation, an H.265 decoder operation starts to decode the received frame and generate another YUV format. But in order to display the reconstructed frame throughout the Graphic display, another conversion will be started from YUV to RGB format. Finally, the resulted frame will be displayed using the SDL Graphic display library.

4.3.2 Speech Recording/Playing:

Speech recording and playback processing is implemented by the software (C++ sound class of OpenAL library) and the hardware (sound card). We'll discuss about the sound buffer utilized in the implementation section. Figure 4.7 shows the speech part.

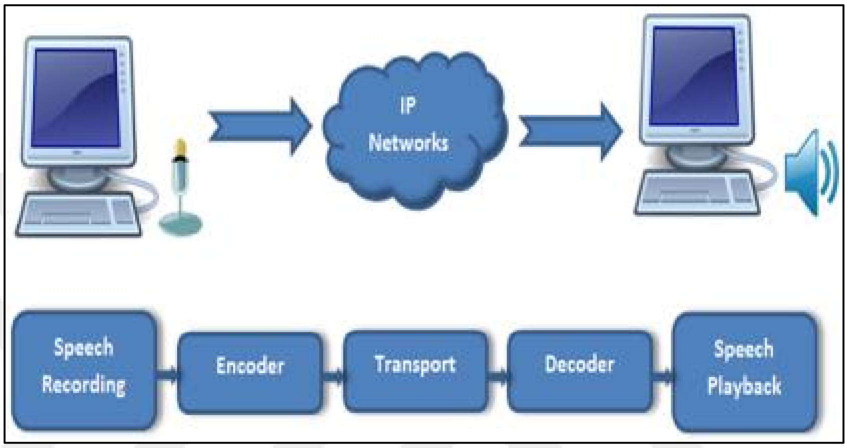


Figure 4.7 Speech System Structure

OpenAL (OPEN Audio Library) is A 3D audio programming interface; which is available for multiple operating systems (OS). It is an open source alternative to MS-DirectSound interface. Figure 4.8 shows the relationship between OpenAL and the sound programming interfaces (APIs) from Microsoft. Note that OpenAL API deals with the raw PCM format.

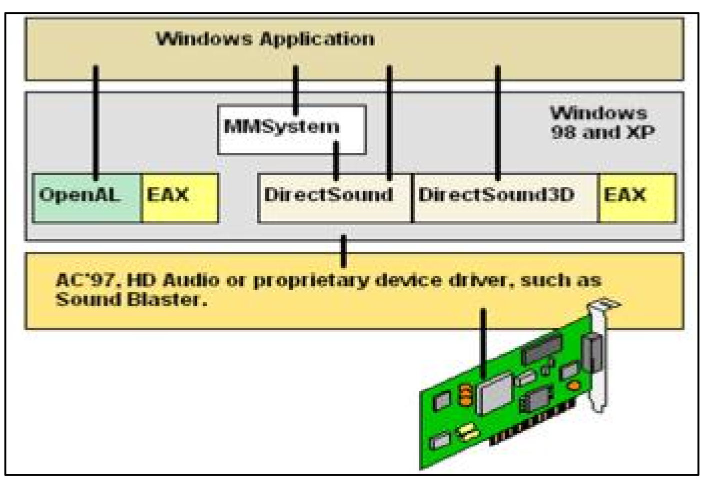


Figure 4.8 OpenAL & sound-API's relationship

4.3.3 CODEC

"Compressor/De-compressor“- makes the audio/video data "small enough" to be adequate for transferring over the expensive net connections. A codec receives the analogue signals, compresses them, then perform a digitizing operation, and finally transmits the digital signals over digital phone lines. The sound data is a continuous analog signal and that poses a problem in the PC; since computers store discrete digital values. Therefore, PCM is the solution to represent a digital format of sound. The sound is regularly sampled, and its amplitude stored as a value. PCM-data has two important components - sample rate (Hz); which determines how often a sample of the sound is taken, and bit depth, which determines the range of amplitude values that can be stored. As there is limited bit depth, and so only a finite number of values that can be stored, the sampled values of a sound wave are quantised - rounded to the nearest value that the bit depth can hold.

You should be able to see how both bit depth and sample rate determine the final quality of the digitised sound. To give you an idea of what the commonly used values for bit depth and sample rate are audio CDs, which encode their data at a sample rate of 44,056hz (over 44k samples a second!), and a bit rate of 16 (65,536 unique values). The audio hardware in your computer can natively handle PCM data, and can turn it back into analog data, for outputting via speakers or headphones.

The speech coder encodes the samples of speech into minimum No. of bits to avoid the connection errors, jittery nets, and burst-communication. At the receiver, the bits are decoded back to the PCM samples; in which it will be converted to the analogue waveform as shown in Figure 4.5 and its performance in our project is as shown in Figure 4.9.

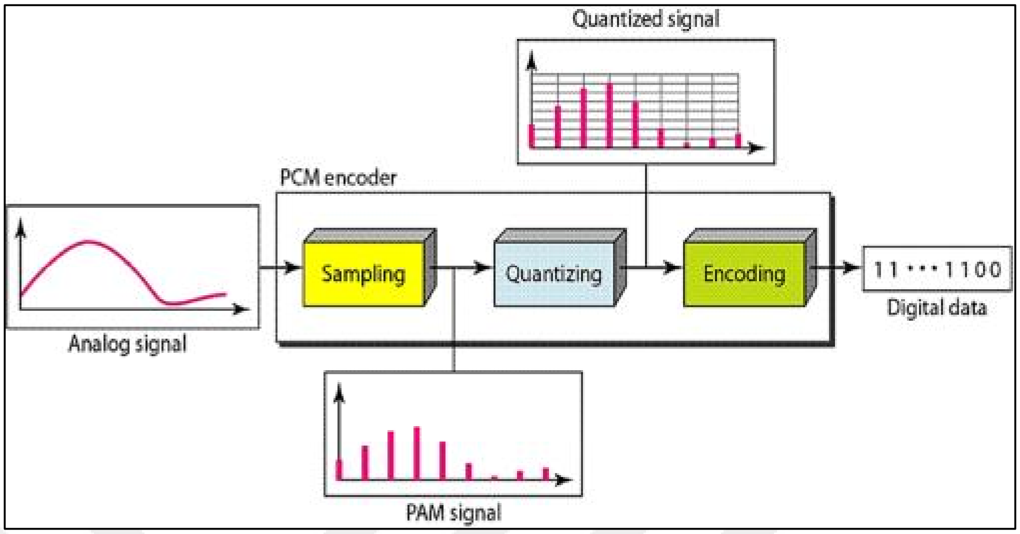


Figure 4.9 Components of PCM Codec

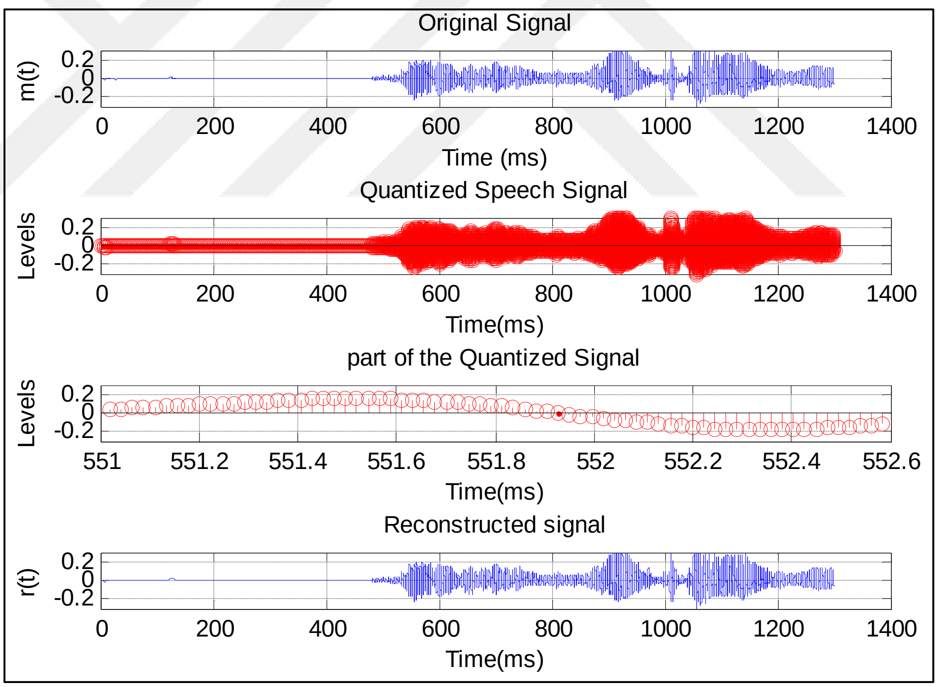


Figure 4.10 PCM Codec's performance

In our project we used the "Open AL" library, (short for Open Audio Library to create a virtual 3D world of sound. Its API model and convention deliberately resembles those of OpenGL. Figure 4.11 shows the OpenAL structure.

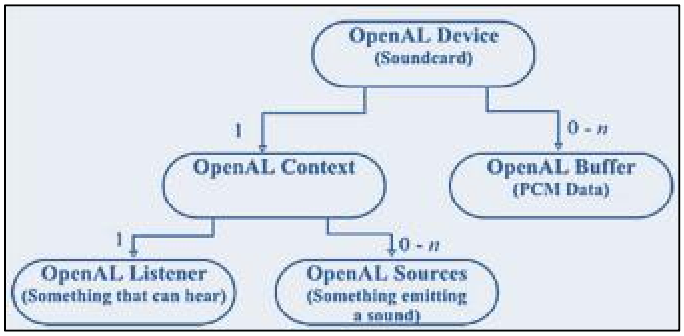


Figure 4.11 Structure of OpenAL

OpenAL-Buffers include audio chunks in PCM format, either eight or sixteen-bit, in either mono or stereo format. The rendering engine accomplished all required computations as far as distance attenuation, Doppler Effect, etc. While for the video part, we implement H.265 codec; which is a new video coding technology in order to enhance the real-time teleconference system. Figure 4.12 shows the H.265 codec. We have changed the encoder/decoder implementation to be more suitable for real time applications as discussed in the previous chapter.

The main target of the our H.265 standardization potential is to improve the compression rate significantly regarding current standards to a 50% bitrate of equal perceptual video-quality with preserving a high video quality.

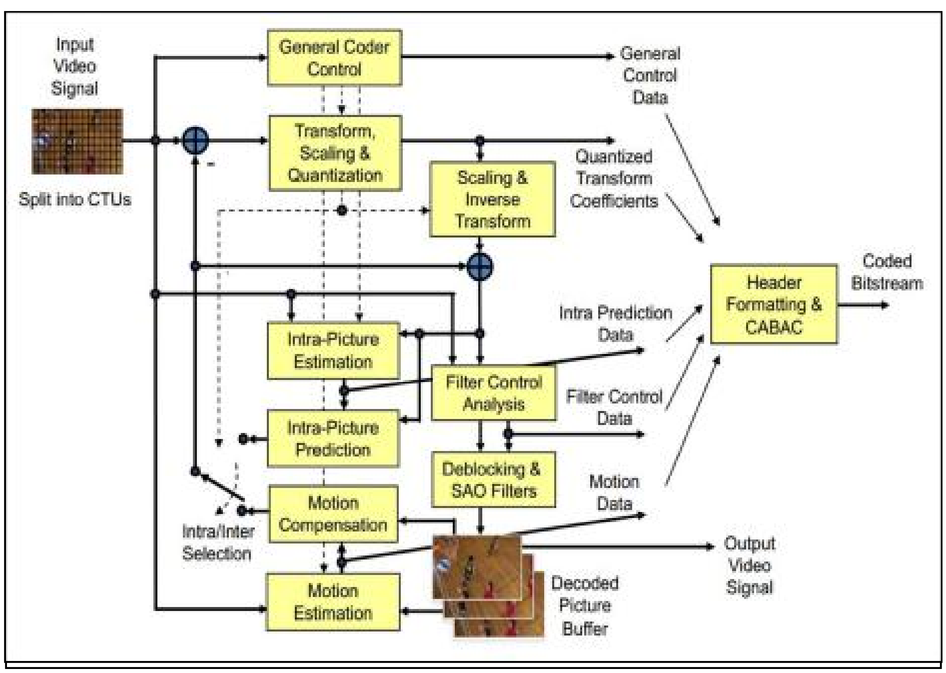


Figure 4.12 H.265 CODEC

4.3.4 Socket Communications:

The Socket is an endpoint-intercommunication of a medium controlled by the transport-services (Figure 4.13 shows the communication's structure of Socket).

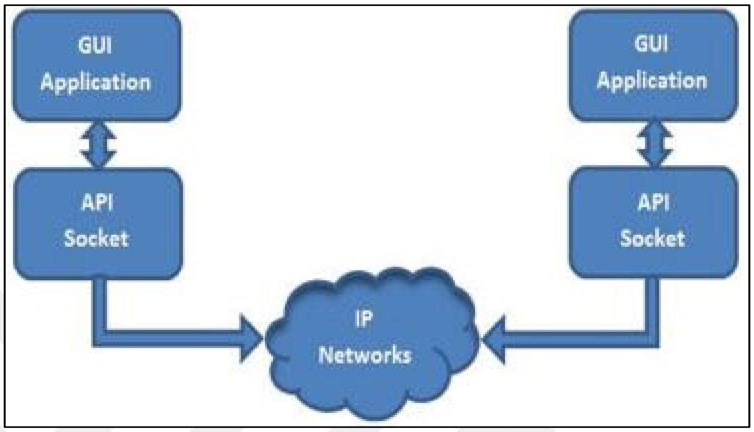


Figure 4.13 Socket-Communication's Structure

From the interconnection viewpoint, there are 2 socket's types; which are the oriented and connectionless. For the connection-oriented, the communication should be prepared before transferring data. It has more reliability than the connectionless. TCP represents a transport protocol of the connection-oriented type. While the connection-setup, packet-acknowledgement and re-transmission services do not exist within a Connection-less communication. UDP is the transport protocol for this type of socket. Thus, in order to have a more secure and reliable communication, we chose TCP as our transport protocol. Figure 4.14 shows the TCP connection setup.

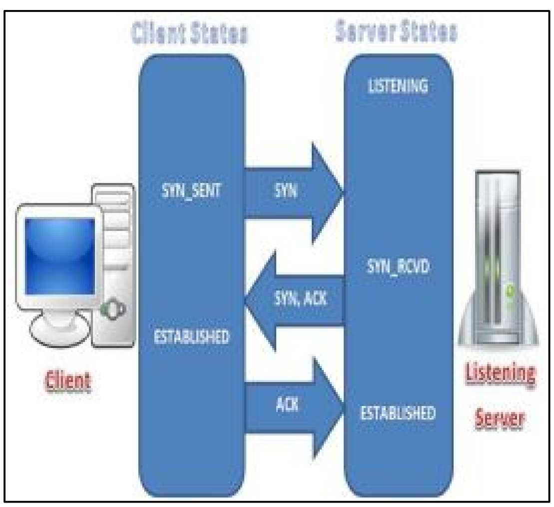


Figure 4.14 TCP Connection setup

From the synchronization viewpoint, 2 types of socket communications exist (asynchronous and synchronous). In the first type, there's no suspending in the client application's execution while waiting for the server's response. For the second type, the execution will be suspended until having a response from the server. Also 2 types of socket interfaces are available (Berkeley Socket Interface for UNIX system and WindowsSocket). As our system works under windows vista OS, thus a winsock2 has been chosen; which represents an interface and not a protocol. Figure 4.15 shows the Server/ Client WinSock2 Architecture.

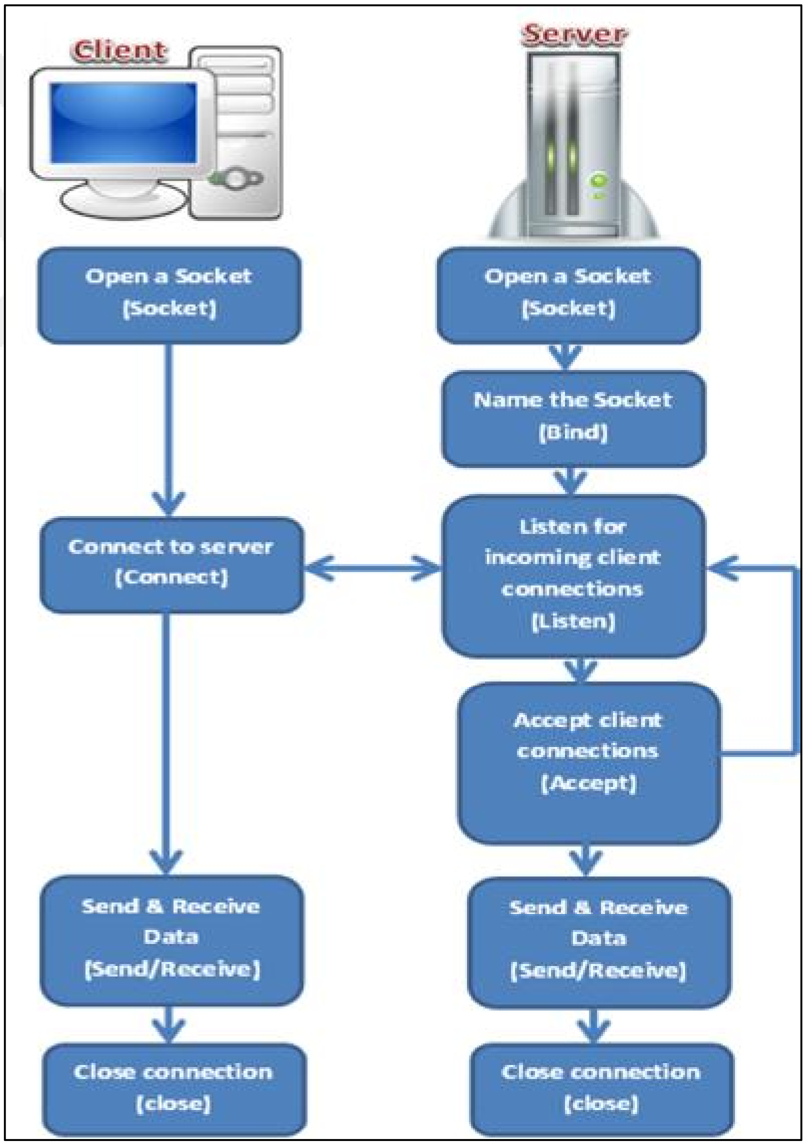


Figure 4.15 Winsock2 Architecture

4.4 Architecture Design Model

In order to get rid of the security and management issues, we have designed our Teleconference system over a Server/Client Model. The Server acts as a controller of the Teleconference sessions. As shown in Figure 4.16, the system model includes two components (Server and Client).

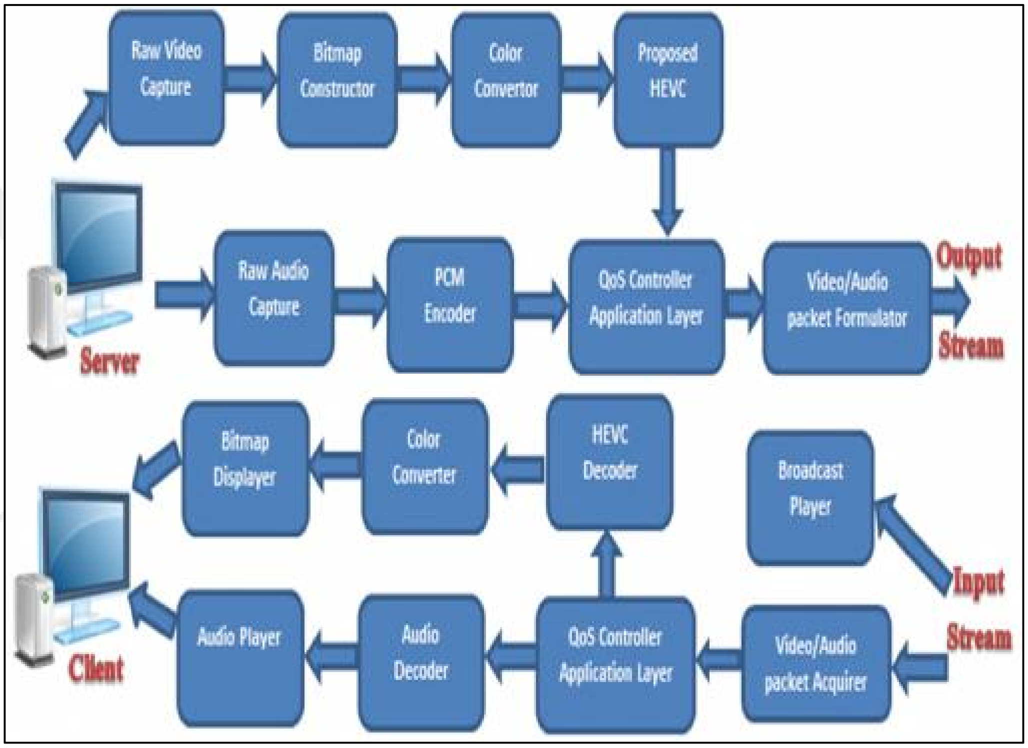


Figure 4.16 Teleconference System Model

4.4.1 System Model

One of the main roles in our server model is the conference's management, such as user registration, member authentication, and authorization. The session setup module is responsible in providing a reciprocal members' awareness in their teleconference system. It relies on a TCP protocol for a reliable communication. Therefore, when a user begins a new session, a status-record of this session will be saved within the teleconference server with all the members that attending this conference-session. And whenever a client attends this session, a status packet will be transferred from the server to this client informing him of all the online members with their IP-addresses.

In addition to that, a broadcast status-packet is transferred from the server to all attending members notifying them about the newly attending client and his IP address. As the notation (distributed system) defines the nature of the tele-conference-server; so it needs a control protocol to coordinate the teleconference-members on each modification of the teleconference parameters or network conditions. Any modification in the user's status represents a modification in teleconference parameters such as joining/leaving a session and starting/stopping an (audio/video) streaming. Both the end-to-end delay and the available bandwidth represent the network conditions.

A series of update-statuses will be triggered with each event with message transferring to other clients who are attending the conference. Each member will be notified about any modification that occurs at the server (Administrator) like activating and de-activating of his own audio/video calls. Moreover, the server has many other options such as Blackboard Sharing (some notes or lectures are shared between the Administrator and members), Screen Sharing (A tool for capturing the Administrator's Live Screen for a demonstration purpose to other members), Multimedia Broadcasting (Video/Audio Calls), file Sharing (Sending/Receiving Files) and Text chats.

Also, the administrator can broadcast a session for others who are not members within the teleconference; in order to watch and listen to the live lecture for example, using the HTTP protocol for sharing the streaming link of the teleconference. Each member has a list of the login members and the time for their login and logoff with a list of the available video/ audio capture devices. After a successful authentication, each client can start the Teleconference system.

4.4.2 Video/Audio Packet Formulation:

According to Figure 4.16, our system has been designed based on the server/client model. Each model transfers the data but with different functions as discussed before. Now we will demonstrate in detail the components of data sending and receiving. The transmission process contains seven sections such as:

1) *Video Capturing Process*: Capturing either a frame from a connected camera, or a window screen, or administrator's blackboard to a memory buffer.

2) *Audio Capturing Process*: Recording audio segments from a connected microphone to memory buffer.

3) *Bitmap Constructer*: Creating a DIB-Bitmap of RGB color 24-bit, pixel dimensions (640x480) from a grabbed frame in the buffer.

4) *Colour Converter*: Converting a grabbed frame from RGB to YUV4:2:0 format based on the matrix shown in (4.1); because the sensitivity of the human eye to colour is lesser than to brightness. Therefore, a more reduction can be obtained in the size of the image by under sampling the V and U components, which in turn increases the speed of the transmission and reserves more space.

$$\begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{U} \\ \boldsymbol{V} \end{bmatrix} = \begin{bmatrix} \mathbf{0.257} & \mathbf{0.504} & \mathbf{0.098} \\ \mathbf{-0.148} & \mathbf{-0.291} & \mathbf{0.439} \\ \mathbf{0.439} & \mathbf{-0.368} & \mathbf{-0.071} \end{bmatrix} \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{G} \\ \boldsymbol{B} \end{bmatrix} + \begin{bmatrix} \mathbf{16} \\ \mathbf{128} \\ \mathbf{128} \end{bmatrix} \tag{4.1}$$

5) *Proposed HEVC Encoder*: Compressing YUV420 raw images by our proposed algorithm into an HEVC/NALU streams.

6) *Video/Audio Packet Formulator*: Converting the video NAL units into fragmentation units, and then to TCP packets for the video transmission. Converting the 16-bit audio segments to TCP packets for audio transmission.

7) *Audio PCM Encoder*: Encoding the captured audio chunks to minimum number of bits for getting rid of the link errors, jittery nets, and burst transmission.

8) *QoS Controller*: controls the audio/video streams according to the network condition and QoS requirements.

The reception process consists of seven parts such as:

1) *QoS Controller*: Performs the same function as in the transmission process.

2) *Video Decoder*: Decodes the HEVC-NALU streams to YUV420 frames.

3) *Audio Decoder*: Decodes the audio streams to 16-bit PCM samples.

4) *Colour Converter*: Converts the YUV420 frames to RGB24bpp format using the matrix shown in (4.2).

$$\begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{G} \\ \boldsymbol{B} \end{bmatrix} = \begin{bmatrix} \mathbf{1.164} & \mathbf{0} & \mathbf{1.596} \\ \mathbf{1.164} & \mathbf{-0.391} & \mathbf{-0.813} \\ \mathbf{1.164} & \mathbf{2.018} & \mathbf{0} \end{bmatrix} \left(\begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{U} \\ \boldsymbol{V} \end{bmatrix} - \begin{bmatrix} \mathbf{16} \\ \mathbf{128} \\ \mathbf{128} \end{bmatrix} \right) \tag{4.2}$$

5) *Bitmap Displayer*: Displays the RGB24bpp bitmap image on the handle of the screen window.
6) *Audio Player*: Converts the 16-bit PCM samples into analogue form for Audio playback.
7) *Video/Audio packet Acquirer*: Acquires the video packet to HEVC-NALU streams and audio streams.
8) *Broadcast Player*: Displaying the HTTP online streaming from the Server.

4.5 Experimental Results and Discussions

Our Teleconference system is experimentally tested upon a sequence of frames that are captured using a digital's camera and speech segments using a sound recorder within two computers of 2.0 GHz-core. The Quantization-Factor of High-Level Efficiency codec parameters used in our H.265 encoder equals to 32.

4.5.1 First Method

Our algorithm features a faster method for Intra Prediction Mode Decision and partitioning, as shown in Figure 4.17. Although HEVC supports 35 directional modes, which leads to greater precision than H.246, it produces higher complexity and more computational overhead. HEVC uses the RDO process for choosing the candidate mode but this process consumes so much power and time due to its complexity. Therefore, we have proposed a faster and less complicated method for choosing the candidate mode, and then Shannon entropy have been implemented for the splitting process instead of the RDO entropy to reduce the computation load.

The basis of this algorithm depends on the CU's contents for deciding the candidate mode; which represents the most probable one. We found that both Luma & Chroma information have nearly identical values in the pixels along the local edge's direction. Therefore, we can obtain the prediction mode by predicting the pixels in accordance with their neighbours in the same edge direction. Where, the Sobel filter computes the gradient information of each pixel in a PU for selecting the candidate prediction mode.

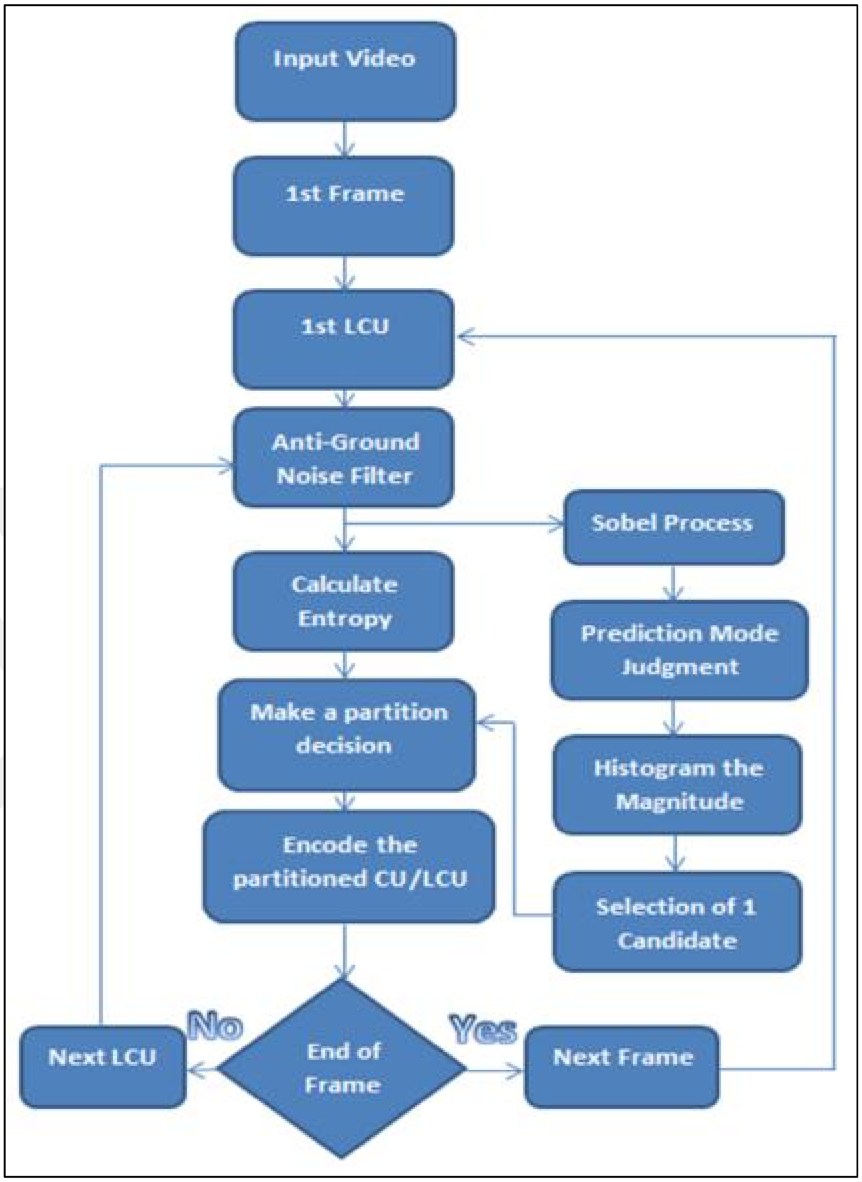


Figure 4.17 Flowchart of the proposed algorithm.

- ***Gradient generation engine***

The formula of Sobel operator is given by equation (4.3), and the dimension of matrix M is a 3×3 pixel unit shown in equation (4.4).

$$S_x=\begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix},\ S_y=\begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} \tag{4.3}$$

$$M=\begin{pmatrix} Px(i-1,j-1) & Px(i-1,j) & Px(i-1,j+1) \\ Px(i,j-1) & Px(i,j) & Px(i,j+1) \\ Px(i+1,j-1) & Px(i+1,j) & Px(i+1,j+1) \end{pmatrix} \tag{4.4}$$

The Sobel operator consists of two convolution kernels, which also represent the degree of difference between the vertical and horizontal directions. The corresponding directional vectors {T(xij), T(yij)} for a Luma pixel is defined in equations (4.5) and (4.6) respectively.

$$T(x_{ij}) = S_x \times M = \mathbf{2} \times \mathbf{Px}\,(\mathbf{i},\mathbf{j}+\mathbf{1}) - \mathbf{Px}\,(i,j-1) - \mathbf{Px}\,(i-1,j-1) - \mathbf{Px}\,(i+1,j-1) + \mathbf{Px}\,(i+1,j+1) + \mathbf{Px}\,(i-1,j+1) \quad (4.5)$$

$$T(y_{ij}) = S_y \times M = \mathbf{2} \times \mathbf{Px}\,(i+1,j) + \mathbf{Px}\,(i+1,j-1) - \mathbf{Px}\,(i-1,j-1) + \mathbf{Px}\,(i+1,j+1) - \mathbf{Px}\,(i-1,j) - \mathbf{Px}\,(i-1,j+1) \quad (4.6)$$

The vector's amplitude and direction; which is perpendicular to the edge direction, are computed using the equations below. An example of the Sobel operator and its calculations is shown in Figure 4.18.

$$\mathbf{A(T(x_{ij}))} = \sqrt{\mathbf{T\,(x_{ij})^2 + T\,(y_{ij})^2}} \quad (4.7)$$

$$\boldsymbol{\theta(V)} = \boldsymbol{tan^{-1}\,\frac{T(x_{ij})}{T(y_{ij})}} \quad (4.8)$$

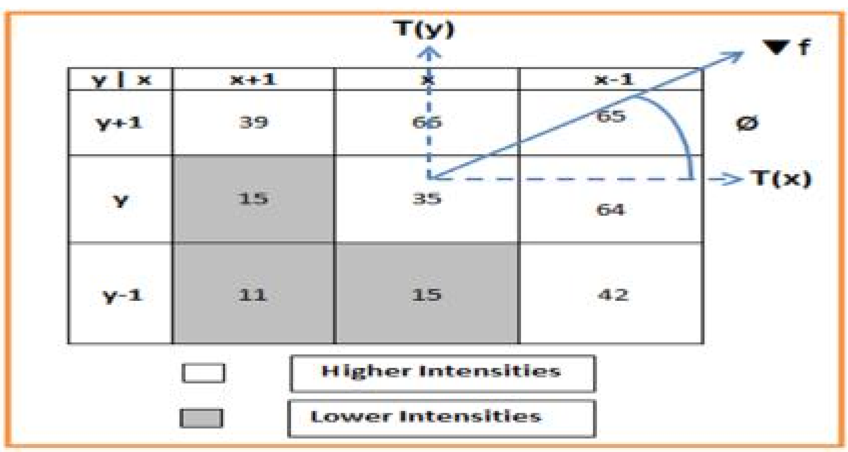


Figure 4.18 An example of Sobel Filter Operation

T(y) = ((39-11)/2 + (66-15)/2 + (65-42)/2)/3 = (14+25.5+11.5)/3=17

T(x) = ((65-39)/2 + (64-15)/2 + (42-11)/2)/3 = (13+24.5+15.5)/3=17.67

Ø = $\tan^{-1}(T(y)/T(x)) = \tan^{-1}(17/18) = 43.9^{o}$

|▼f | = $(17^2 + 17.67^2)^{1/2} = 24.52$

- ***Prediction mode judgment***

In the beginning, the vector of each pixel in the PU is computed, and the results are placed into two arrays (one array is for the magnitudes and the other one is for the angles). If the computed vector lies between the prediction vector and the Boundary line of judgment vector, then it belongs to the adjacent prediction mode. Where, the Boundary line vector represents the Bisectrix of the two adjacent prediction mode vectors, as shown in Figure 4.19. Referring to the mentioned Figure as an example, the vector we got lies in the area between the prediction mode 3 and the Boundary line of the judgment; therefore mode 3 is the obtained vector.

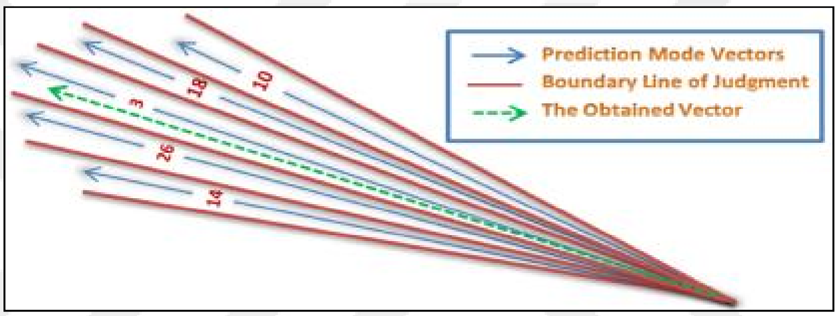


Figure 4.19 Example of the boundary line for prediction mode judgment

The next step is creating a histogram to count the sum of magnitudes for each prediction mode in a PU. The prediction modes lie on the X-axis, whereas the sum of magnitudes lies on the Y-axis. Finally, one candidate is chosen based on the highest prediction mode in the histogram. The prediction mode 23 is selected as the candidate mode referring to the example shown in Figure 4.20.

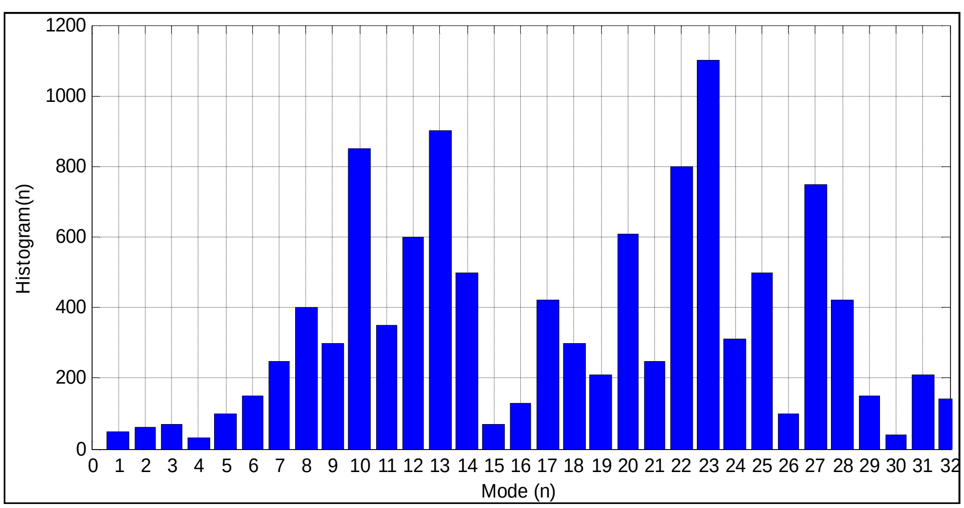


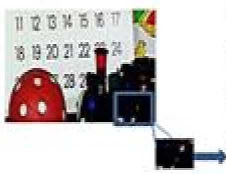
Figure 4.20 Histogram of the Edge direction

- ***Fast splitting process***

As mentioned previously, we have used a Shannon Entropy codec to encode the candidate mode ranks instead of the RDO Entropy. A stepper quantization process has been performed beforehand; in which each pixel is divided by (8) as shown in the equation 4.9, to get better values closer to the RDO results, as shown in Figure 4.21.

$$\mathbf{Q=Round(Px/8)} \tag{4.9}$$

Where: 'Px' is the pixel's value, and 'Q' is the quantized value. As a result, there will be a pixel's range of (0-31) instead (0-255) that gives a faster computation of entropy and a higher saving in memory.



106	105	106	105	116	108	119	120	114	111	112	109	106	107	108	110
105	104	105	105	106	108	119	119	113	111	110	108	107	107	108	110
105	104	106	105	107	108	120	119	113	112	112	108	106	107	107	111
106	104	106	105	116	108	119	120	114	111	112	109	106	107	108	110
108	104	106	105	116	107	119	120	114	111	112	109	106	107	108	110
106	104	106	105	116	108	119	120	114	111	112	109	106	107	108	110
106	104	106	104	116	108	119	120	114	112	112	109	106	107	108	110
108	103	106	105	116	106	119	120	114	111	112	109	106	107	108	110
108	102	105	103	113	108	119	120	114	111	112	109	106	107	108	110
107	104	103	103	111	108	117	118	120	110	112	109	106	107	108	110
106	104	104	105	111	106	119	120	114	111	112	109	106	107	108	110
107	104	102	104	113	105	119	120	114	109	111	109	106	107	108	110
106	104	106	105	115	108	119	120	114	109	110	109	106	107	108	110
104	104	106	105	116	108	119	120	114	110	112	109	106	107	108	110
102	104	107	105	116	108	119	120	114	111	112	109	106	107	108	110
103	104	107	104	116	108	119	120	114	111	112	109	106	107	108	110
102	104	106	105	106	104	119	120	114	112	111	108	106	107	108	110

13	13	13	13	14	13	14	15	14	13	14	13	13	13	13	13
13	13	13	13	13	13	14	14	14	13	13	13	13	13	13	13
13	13	13	13	13	13	14	14	14	14	14	13	13	13	13	13
13	13	13	13	14	13	14	15	14	13	14	13	13	13	13	13
13	13	13	13	14	13	14	15	14	13	14	13	13	13	13	13
13	13	13	13	14	13	14	15	14	13	14	13	13	13	13	13
13	13	13	13	14	13	14	15	14	14	14	13	13	13	13	13
13	12	13	13	14	13	14	15	14	13	14	13	13	13	13	13
13	12	12	13	14	13	14	15	14	13	14	13	13	13	13	13
13	13	13	13	13	13	14	14	15	13	14	13	13	13	13	13
13	13	13	13	13	13	14	15	14	13	14	13	13	13	13	13
13	13	13	13	14	13	14	15	14	13	14	13	13	13	13	13
13	13	13	13	14	13	14	15	14	13	14	13	13	13	13	13
13	13	13	13	14	13	14	15	14	13	14	13	13	13	13	13
12	13	13	13	14	13	14	15	14	13	14	13	13	13	13	13
12	13	13	13	14	13	14	15	14	13	14	13	13	13	13	13
12	13	13	13	13	13	14	15	14	14	14	13	13	13	13	13

Figure 4.21 Stepper Quantization Example

The Entropies of CUs per each LCU have been determined using the following equation:

$$\mathrm{En}(x) = -\sum_{j=0}^{k} \boldsymbol{P_j} \times \boldsymbol{log_2}(\boldsymbol{P_j}) \qquad (4.10)$$

Where 'En(x)' is the CU's entropy, 'Pj' is the probability of pixel that is computed using the following equation:

$$P_j = \frac{\boldsymbol{n_j}}{\boldsymbol{T}} \qquad (j=0, 1, 2,..., 31) \qquad (4.11)$$

Where 'n_j' defines the number of pixels whose value='j' and 'T' defines the total pixels. Partitioning a CU or not is based on the relationship between each CU and its entropy. The CU will be split if its entropy is larger than 3.5. The principle of determining the mentioned threshold depends on the maximum likeness between the original and the proposed CUs, taking into consideration many different video sequences. The proposed algorithm for Fast LCU partitioning is shown in Algorithm.

Algorithm 1: Fast LCU Partitioning

Input: Image A

Output: Encoded Image A

for LCU=1 $\longrightarrow$ P **do**

  -Perform Quantization with a step size=8 using Eq. (4.9)

  for CU=1 $\longrightarrow$ N **do**

    -Compute the probability of pixel occurrence per CU using Eq. (4.21)

    -Compute the Entropy (En) of CU using Eq. (4.20)

    If En> 3.5 **Then**

      -Perform Partitioning for CU

    else

      -Perform no Partitioning for CU

    end If

    -Perform encoding process for CU

  end for

end for

A bitrate comparison between our proposed HEVC, original HEVC and AVC is hown in Figure 4.22 versus different QPs; in which 2.1(kbits/s) saving in the bitrate occurs using the proposed HEVC in average. Therefore, to determine the reduction in time, we used the following equation:

$$\blacktriangledown T = (T_{HM12}-T_{proposed}) / T_{HM12} \tag{4.12}$$

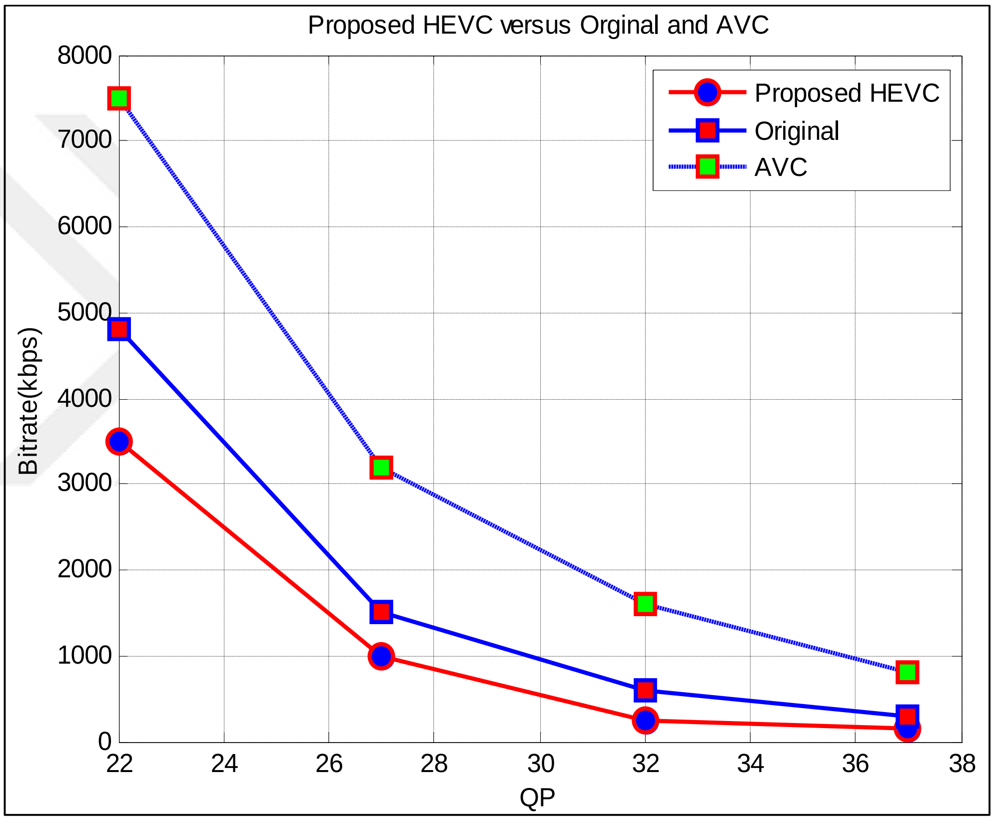


Figure 4.22 Bitrate vs. QP.

Where: '$T_{proposed}$' represents the evaluated coding time using our proposed algorithm, 'T_{HM12}' represents a coding time of HM12 with RDO, and '$\blacktriangledown T$' represents the reduction in coding time. There is an average similarity, regarding Table 4.1, of approximately 85.3% to the original algorithm, which is computed by equation below, where 'M_{MATCH}'= Number of matched CUs, and 'M_{CU}'= Total CUs.

$$\mathbf{S=M_{MATCH} / M_{CU}} \tag{4.13}$$

Table 4.1. Similarity percentages

Similarity (%)	**Video Sequence Name**
81.5%	ParkScene_1920x1080
84.3	RaceHorses_832x480
86.1%	BasketballPass_416x240
89.3%	Mobile_Cif_352x288
85.3%	**Total Average**

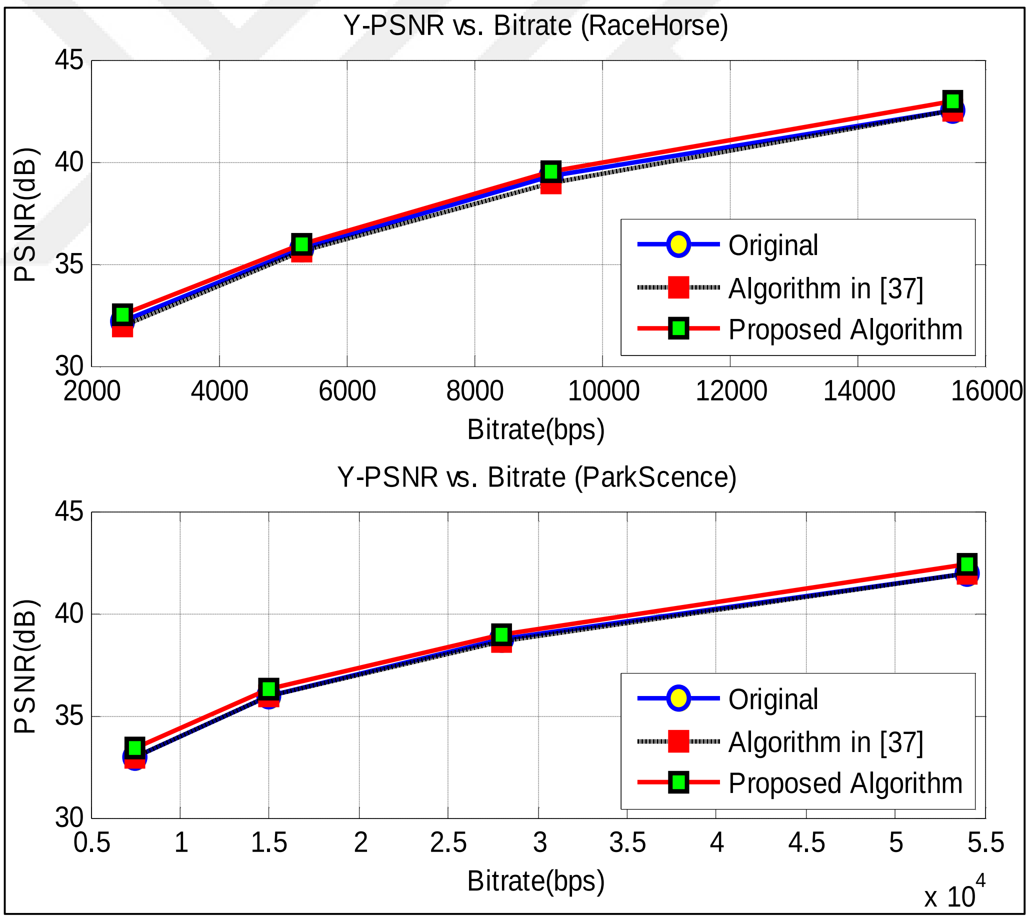


Figure 4.23 RD Curves of ParkScene (Class B) and RaceHorses (Class C)

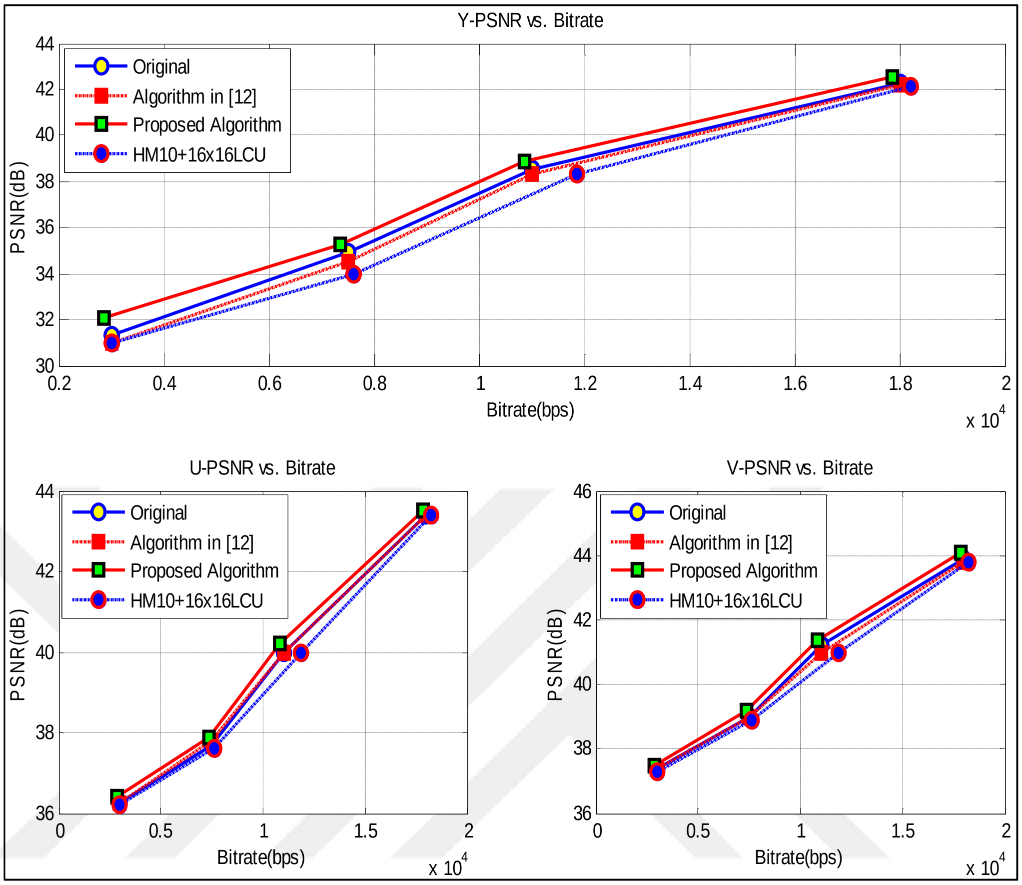


Figure 4.24 “RaceHorsesC” (832 × 480) under different QP

The proposed algorithm results in an increase in the PSNR rate by approximately 1.33% (for Y-PSNR) compared to the algorithm in [12] and 0.43% (for the park scene) compared to the algorithm in [14], as shown in Fig. 4.23 and Fig. 4.24, respectively. Table 4.2 shows a reduction in the average coding time by approximately 57.2% and the BD rate (Bjøntegaard delta) shows a 0.33% loss.

Table 4.2. The proposed Algorithm Results

Video Name	**Algorithm in [36]**		**Algorithm in [4]**		**Proposed Algorithm**	
	BD Rate (%)	**ΔT (%)**	**BD Rate (%)**	**ΔT (%)**	**BD Rate (%)**	**ΔT (%)**
ParkScene	0.27	36.0	0.51	34.2	0.42	56.0
RaceHorses	0.39	41.7	0.15	32.6	0.17	54.2
BasketballPass	1.84	37.5	0.61	47.4	0.57	58.3
Mobile_Cif	1.31	27.0	0.22	53.9	0.16	60.1
Average	**0.95**	**35.6**	**0.373**	**42.02**	**0.33**	**57.2**

4.5.2 Second Method

In our algorithm, we propose a fast algorithm for the decision of CU size. Where, it is inefficient to utilize all depth levels because the optimal depth level is highly content dependent. We can compute the CU depth range and ignore some particular depth levels that are rarely utilized in the previous frame and neighbouring CUs. Moreover, the proposed algorithm gives an early termination method depending on the Shannon entropy value instead of the RD cost used in the RDO process to reduce the computation complexity.

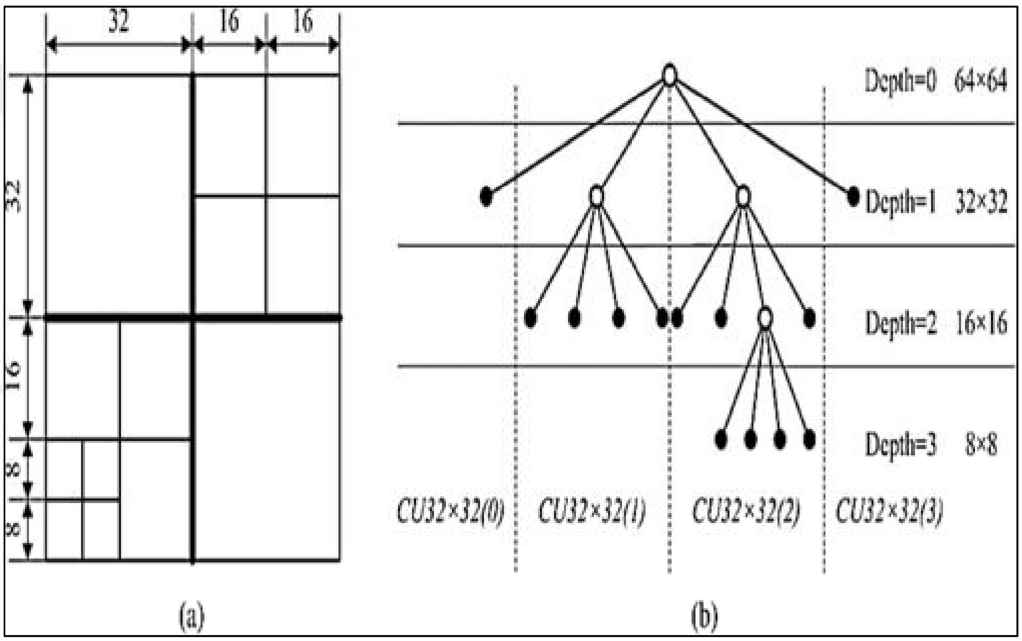


Figure 4.25 The Tree-block of depth-levels

In the original HEVC, the CU depth is fixed for any video sequence. The depth level is the criterion for splitting the tree block as shown in Figure 4.25. However, this criterion is unsuitable for teleconferencing video because of its slight difference between frames owing to the constant background. Thus, the computation of the depth level should be determined for each CU. Thus, our next step is to evaluate the depth levels and determine a threshold value based on the Shannon entropy of each CU to decide whether to split the specified CU or not.

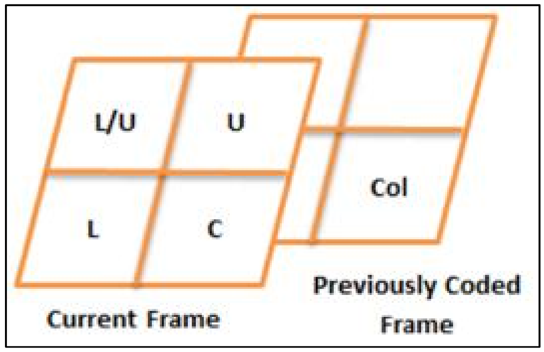


Figure 4.26 The Spatial and Temporal correlations of CUs

Predicting the optimal depth level of a tree block (Dp) is accomplished through the spatial neighbouring tree blocks (Upper (U), Lower (L), Left Upper (L/U)) and the collocated (Col) tree block that represents the previously coded frame as shown in Figure 4.26:

$$\mathrm{Dp} = \sum_{k=0}^{J-1} \mu(k) \cdot \varphi(k) \qquad (4.14)$$

Where 'J' represents the number of tree blocks = 4, '$\mu(k)$' represents the computed weight based on the correlation between the current tree block and the neighbouring tree block, and 'φ(k)' represents the depth level. The 4 weights are normalized to obtain $\sum_{k=0}^{3} \mu(k) = 1$. The neighbouring horizontal/vertical tree blocks and the collocated tree block in the previously coded frame have a higher correlation of the optimal CU size for the current tree block regarding the diagonal direction. '$\mu(k)$' of the L/U tree block is set to 0.1, whereas the weight values of the U, L, and Col tree blocks are set to 0.3.

Each tree block is segmented as one of five types (G4, G3, G2, G1, and G0) based on the predicted value of the optimal depth level; the segmentation steps are as follows:

1) The current tree block is located in the still or homogeneous motion region classified as type G0 if Dp in the current tree block equals 0 and its optimal prediction mode at depth level 0 is selected with SKIP mode.

2) The current block is located in the moderate motion region classified as type G1 if the value of Dp <<0.5. Note that the optimal depths of many neighbouring tree blocks are chosen to be at the depth level (0).

3) Most neighbouring tree blocks are classified as G2 because their optimal depth levels are equal to 1, which is located within the range 0.5 < Dp < 1.5.

4) Most neighbouring tree blocks have optimal depth levels equal to 2, so they are classified as type G3 because their values are within the range 1.5 < Dp < 2.5.

5) The current tree block is located in the fast motion region because most neighbouring tree blocks select optimal depth levels equal to 3, so they are classified as type G4.

Implementing ME (motion estimation) for type G4 tree blocks at depth levels "0" and "1" is inefficient. In G0 type, the optimal depth levels of temporal and spatial neighbouring tree blocks are all at the depth level (0), and the area covered by the current tree block and its neighbouring blocks have slow motion or motionless content. Therefore, the current tree block will need the ME only for the size of the tree block. As a result, for each tree block, the candidate depth levels that are tested later using Shannon entropy are summarized in Table 4.3. Two or three tested depth levels are skipped in most tree blocks.

Table 4.3 Candidate Depth-Levels of Each Tree-Block Type

Tree-block Type	**Candidate depth levels**
G0	0
G1	0,1
G2	0,1,2
G3	1,2,3
G4	2,3

As usual, the RD cost of temporal and spatial neighbouring tree blocks or CUs are identical within the RD cost distribution. Therefore, we can benefit from such a distribution by using a Shannon entropy value rather than the RD cost to determine the ET threshold. Determining the Shannon entropy (En) is the same as in fast splitting process of the first method.

The ME of the next depth level can be skipped when the Shannon entropy value of the current CU size is smaller than the computed threshold. The threshold value can be computed as follows:

$$\text{Thr} = \boldsymbol{\sigma}.\ \sum_{x=0}^{J-1} \boldsymbol{\mu(k)}\ .\ \boldsymbol{En(x)} \tag{4.15}$$

Where 'J' represents the number of neighbouring CUs (4), ' $\mu(k)$' represents the weights determined by the correlations between the current CU and its neighbouring CUs, 'σ' represents the regulation factor, and ' $En(k)$' represents the Shannon entropy of the CU. The weights of the CUs (L, U, and Col) are set to 0.3, and the weight of (L/U) CU is set to 0.1. The selection of the regulation factor should greatly reduce the complexity while maintaining high accuracy. Based on our experiments, the regulation factor is set to 0.38 that produces a good and consistent performance with teleconferencing videos. The overall algorithm is demonstrated in Figure 4.27. There is an average similarity shown in Table 4.4 reaches to 85.3% in average.

Table 4.4 A Similarity comparison

Video Sequence Name	**Similarity (%)**
ParkScene_1920x1080_24	81.5%
RaceHorses_832x480_30	84.3
BasketballPass_416x240_50	86.1%
Mobile_Cif_352x288_24	89.3%
Total Average	**85.3%**

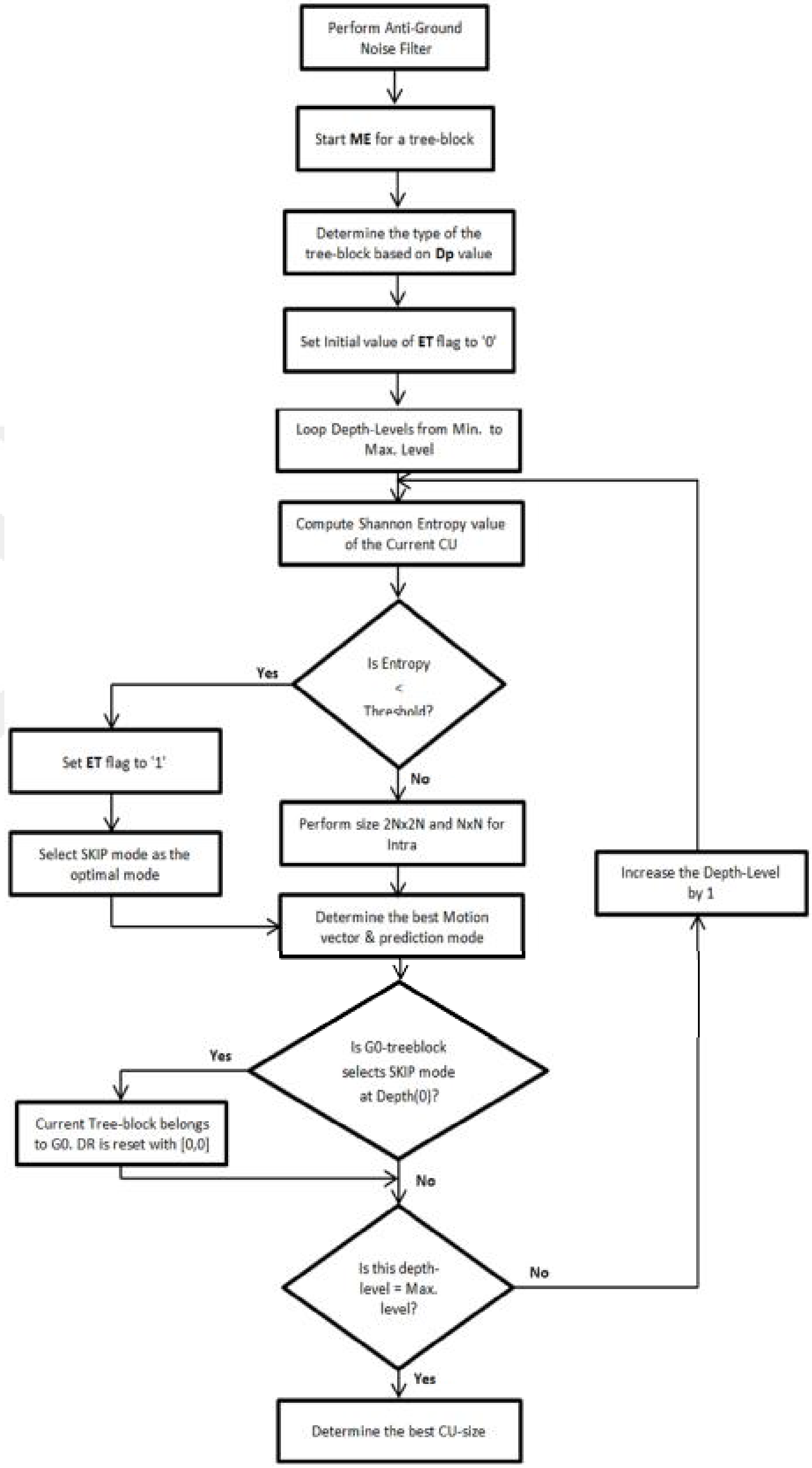


Figure 4.27 Flowchart of the proposed algorithm.

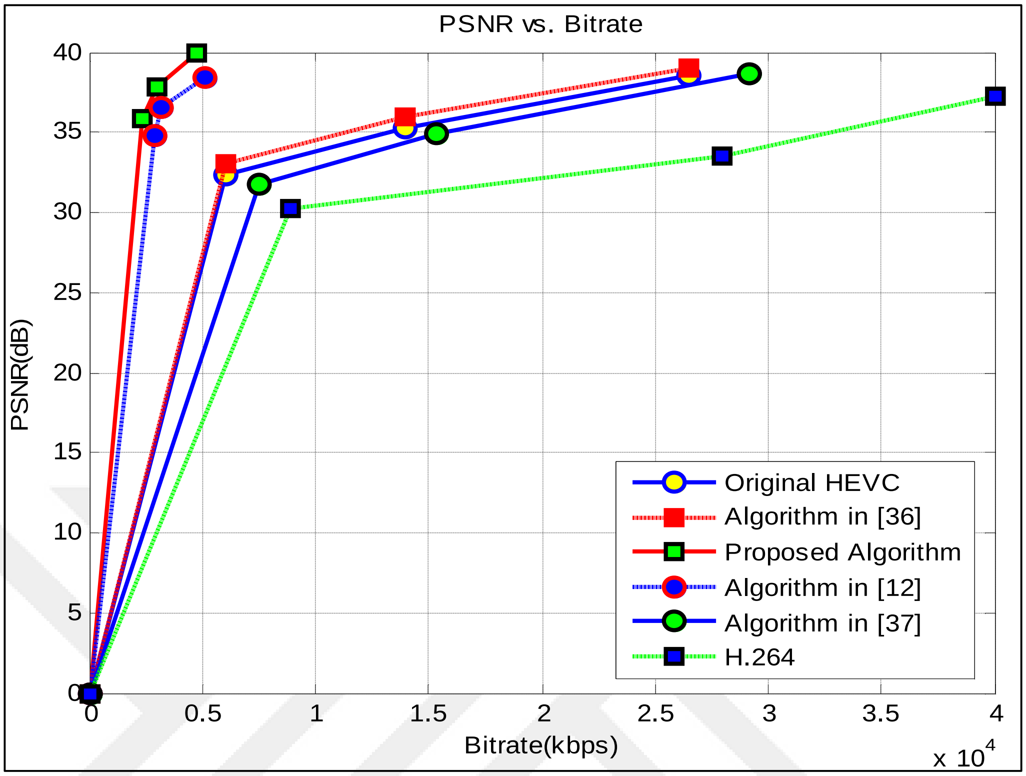


Figure 4.28 Bitrate (kbps) vs. PSNR (dB)

Figure 4.28 shows a comparison between our algorithm and the algorithms in [12], [36], [37], original HEVC (H.256) and H.264. Where, there is a PSNR's increasing in average using our proposed algorithm compared with the others by about (3.5%, 5.4%, 8.2%, 7.3%, 8.7%) respectively, and a Bitrate's reduction in average by about (23%, 74%, 77.8%, 74%, 83.8%) respectively. The PSNR's increment and the Bitrate's decrement represent a good indication to the codec's efficiency. Table 4.5 shows a comparison between our proposed algorithm and the algorithms in [12], [36], and [37]. Note that we mentioned all algorithms that decrease the encoding time compared to the original HEVC except H.264 as it has a higher encoding's time (slower in speed) than the original one. The above mentioned algorithms obtained (27.75%, 35.19%, 16.53%) reduction in time and (1.64%, 1.34%, 1.12%) loss in the BD-Rate respectively, while a (57.8%) reduction in time have been obtained using our algorithm with only a 0.88% loss in the BD-rate in average. As a result, our proposed algorithm is faster in coding than the other algorithms on average.

Table 4.5 Coding Efficiency & Time Reduction compared to original HEVC

Class	Video Name	Algorithm in [12]		Algorithm in [36]		Algorithm in [37]		Proposed	
		BD Rate (%)	ΔT (%)	BD Rate (%)	ΔT (%)	BD Rate (%)	ΔT (%)	BD Rate (%)	ΔT (%)
A	PeopleOnStreet_2560x1600	1.3	25	1.58	35.99	0.73	15.7	0.81	61.6
B	ParkScene_1920x1080	1.24	36	0.39	41.73	0.4	16.9	1.35	56.0
C	BasketballDrill_832x480	1.16	31	1.84	37.49	1. 3	16.7	0.90	53.2
D	BlowingBubbles_416x240	0.82	19	1.61	25.55	2.03	16.8	0.44	60.3
Total Average		**1.64**	**27.75**	**1.34**	**35.19**	**1.12**	**16.53**	**0.88**	**57.8**

4.5.3 Third Method

A faster technique for Intra Prediction partitioning is presented based on the GlobalEdge (GE) and LocalEdge (LE) complexities (vertical, horizontal, 45° diagonal, and 135° diagonal directions). Where, each CU is categorized under three types (split, nonsplit, and undefined). Moreover, Shannon entropy is only computed for the undefined type, to reduce the computational overhead, bitrate and encoding time. The system flowchart is shown in Figure 4.29.

· ***Global and Local Edge Complexity***

Splitting a CU or not and determining the CU's size depend on the complexity of GE/LE edges. In general, the CU's complexity can be determined as follows:

$$\mathbf{Cox} = \sum_{\mathbf{i}=0}^{\mathbf{M}}\sum_{\mathbf{j}=0}^{\mathbf{M}} |\mathbf{Lu}(\mathbf{i},\mathbf{j}) - \mathbf{Lu_m}| \quad (4.16)$$

Where $\boldsymbol{Lu_m}$ defines the CU's mean luminance and $\boldsymbol{Lu(i,j)}$ defines the pixel's luminance at **(i, j)**. Figure 4.30 shows a complexity's difference of the two halves that is used for determining the direction and GE's complexities in four directions as shown in equations (4.27-4.30). The least value of the four determined directions specifies the type of the CU's GE complexity.

$$Xh = \sum_{i=1}^{M}\sum_{j=1}^{\frac{M}{2}}|Lu(i,j) - Lu_m| - \sum_{i=1}^{M}\sum_{j=\frac{M}{2}+1}^{M}|Lu(i,j) - Lu_m| \quad (4.17)$$

$$Xv = \sum_{i=1}^{M/2}\sum_{j=1}^{M}|Lu(i,j) - Lu_m| - \sum_{i=\frac{M}{2}+1}^{M}\sum_{j=1}^{M}|Lu(i,j) - Lu_m| \quad (4.18)$$

$$X45 = \sum_{i=1}^{M}\sum_{j=1}^{i}|Lu(i,j) - Lu_m| - \sum_{i=1}^{M}\sum_{j=1}^{M}|Lu(i,j) - Lu_m| \quad (4.19)$$

$$X135 = \sum_{i=1}^{M}\sum_{j=1}^{M-i-1}|Lu(i,j) - Lu_m| - \sum_{i=1}^{M}\sum_{j=M-i-1}^{M}|Lu(i,j) - Lu_m| \quad (4.20)$$

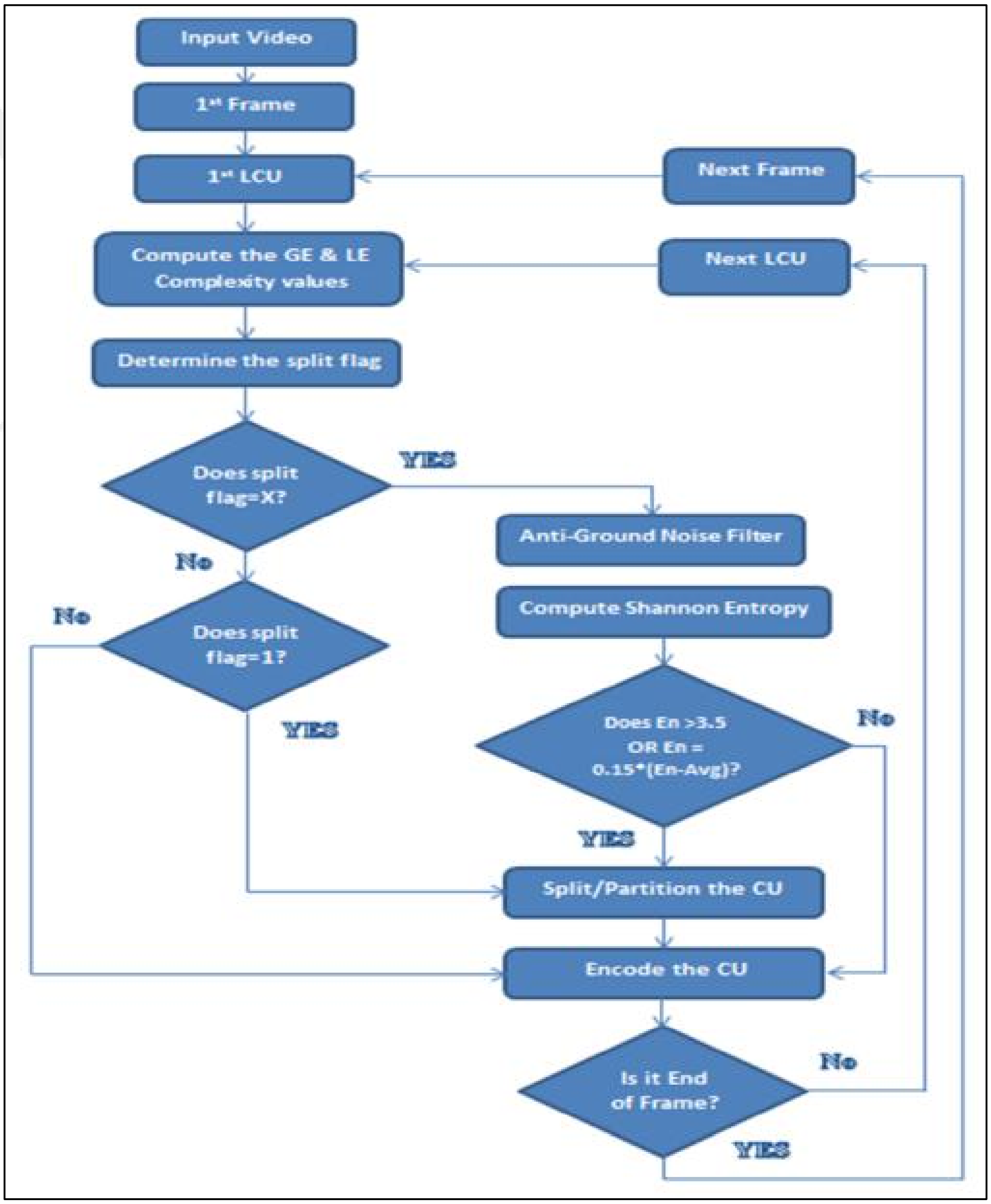


Figure 4.29 Flowchart of the proposed algorithm.

The GE complexities of the sub-blocks are also used in deciding the CU's splitting as shown in equations (4.21, 4.22); in which they are helpful in refining the CU's partitioning. *Where*: (Xv, Xh, X135, & X45) are the CU's GE complexities in the previous mentioned directions.

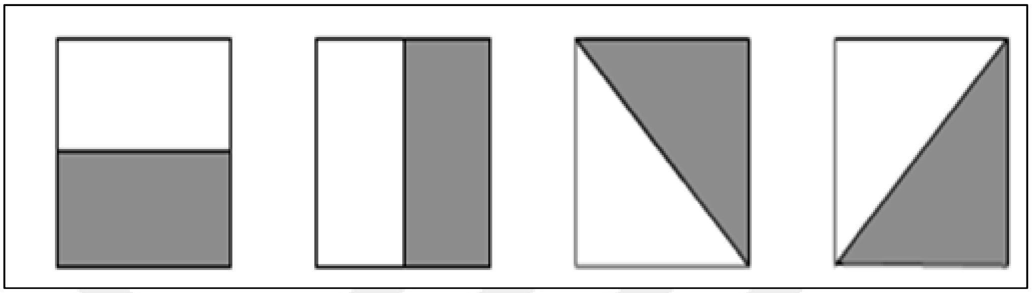


Figure 4.30 GE for the difference of 2 halves /CU

$$\mathbf{H(k)} = \sum_{\mathbf{i}=\mathbf{Z}}^{\mathbf{N}}\sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{M/2}} \left(|\mathbf{Lu(i,j)} - \mathbf{Lu_m}| - \left|\mathbf{Lu}\left(\mathbf{i,j} + \frac{\mathbf{M}}{\mathbf{2}}\right) - \mathbf{Lu_m}\right|\right) \quad (4.21)$$

$$\mathbf{V(k)} = \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{M/2}}\sum_{\mathbf{j}=\mathbf{Z}}^{\mathbf{N}} \left(|\mathbf{Lu(i,j)} - \mathbf{Lu_m}| - \left|\mathbf{Lu}\left(\mathbf{i j} + \frac{M}{\mathbf{2}}\right) - \mathbf{lu_m}\right|\right) \quad (4.22)$$

Where: **V(k)** and **H(k)** are the GE complexities of the sub-blocks. ($\mathbf{Z} = \frac{\mathbf{(k-1)}}{\mathbf{4}} * \boldsymbol{M} + \mathbf{1}$), ($\mathbf{N}=\mathbf{k} * \frac{\mathbf{M}}{\mathbf{4}}$), **M** defines the number of selected pixels; **k** defines the sub-block's position.

The LE complexity is also useful for deciding the CU partitioning, as it is difficult to cover the edge of a small region in local scale using only GE's complexity. As shown in Figure 4.31, four edge filters are applied to the CU's pixels that define the LE's directions. The LE's complexity of the filtered CU is computed using the following equation.

$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}$$

Figure 4.31 LE filters for each pixel/CU

$$\mathbf{Y_d} = \sum_{i=1}^{M}\sum_{j=1}^{M} |\mathbf{F_d}(\mathbf{Lu}(\mathbf{i},\mathbf{j})) - \mathbf{F_d}(\mathbf{Lu_m})| \quad (4.23)$$

Where: $\mathbf{Y_d}$ defines the LE's complexity in the previous mentioned directions and ($\mathbf{d}$) represents the direction as (h, v, 45^o, 135^o) within the local region. $\mathbf{F_d}(\mathbf{Lu}(\mathbf{i},\mathbf{j}))$ defines the luminance of pixel after applying the local filter on the specified direction at **(i,j)**. $\mathbf{F_d}(\mathbf{L\,u_m})$ defines the mean value.

· ***Split Thresholds***

As mentioned before, we made three categories for the CU (nonsplit, split, undefined). The 1st and 2nd ones are defined using the GE/LE complexity; in which the determination of their thresholds are made based on the similarities between our algorithm and the original one. So, Tx= 4850 and Ty=1000 represent the thresholds of LE and GE respectively at layer 0 (depth 0), as they have the least percentage of wrong-hit as shown in Figure 4.32. The threshold of the LE at the next sub-layer (layer1/depth1) will be 1/2 the first Local threshold and for the GE will be 1/4 the first Global threshold.

So these thresholds are useful for determining the value of the split flag. For the undefined category, the CU's are defined to be non-split or not using a Shannon entropy threshold that is computed using the same computation shown in the fast splitting process of the 1st method. Many standard videos have been tested to find out the relationship between the partitioned CUs using the original HEVC, and their entropy values. As a conclusion, each partitioned CU has an entropy value larger than 3.5. The pseudo code shown in Algorithm 2 presents the way of determining the split flag. *Where*: Tx & Ty are the global & local thresholds. (Qsh, Qsv, Qs45, & Qs135) are the maximum GE complexities of its four sub-blocks in the mentioned directions. (Yv, Yh, Y135, & Y45) are the CU's LE complexities in the mentioned directions. (Zsv, Zsh, Zs135, & Zs45) are the maximum LE complexities of its four sub-blocks in the mentioned directions.

Algorithm 2 Proposed Algorithm

Input:

Xh, Xv, $X45$, $X135$, Yh, Yv, $Y45$, $Y135$, Qsh, Qsv, $Qs45$, $Qs135$, Zsh, Zsv, $Zs45$, $Zs135$, $H(k)$, $V(k)$, Tx & Ty.

Output: Split flag

if $((Xh \le Tx, Yh \le Ty, Qsh \le Tx/4, H(k) \le Tx/4$ & $Zsh \le Ty/4)$ **then**

  $flag \leftarrow 0;$

end

else

if $(Xv \le Tx, Yv \le Ty, Qsv \le Tx/4, V(k) \le Tx/4,$ *and* $Zsv \le Ty/4)$ **then**

  $flag \leftarrow 0;$

end

else if $(X45 \le Tx, Y45 \le Ty, Qs45 \le Tx/4$ *and* $Zs45 \le Ty/4)$ **then**

  $flag \leftarrow 0;$

end

else if $(X135 \le Tx, Y135 \le Ty, Qs135 \le Tx/4$ & $Zs135 \le Ty/4)$ **then**

  $flag \leftarrow 0;$

end

else if $min(Xh,Xv,X45,X135) > Tx$ **OR** $min(Yh,Yv,Y45,Y135) > Ty$

then

  $flag \leftarrow 1;$

end

else [$flag \leftarrow X$]

  - *perform Quantization with a step size=8*
  - *Compute the probability of pixel occurrence per CU*

  if $En > 3.5$

   $flag \leftarrow 1;$

  else

   $flag \leftarrow 0;$

  end

end

Figure 4.32 Threshold Values versus wrong HitRate.

A comparison between our proposed HEVC and H.264 encoder is shown in Figure 4.33. Where, a 2.42 (kbits/s) and 0.945(kbits/s) bitrates have been saved in comparison with the H.264 codec and original HEVC respectively.

Figure 4.33 Comparison between bitrate and QP.

Figure 4.34 Comparison of Bitrate (kbps) vs. PSNR (dB)

Another comparison is shown in Figure 4.34 that occurs between our algorithm and others on the image (RaceHorse). The results presents an average PSNR increasing using our proposed algorithm compared to the original codec, the algorithm in [19] and [20] by about (8.28%, 9.1%, 7.01%) respectively, in addition to the Bitrate's reduction by about (22.7%, 23%, 4.42%); which represents a good indication for the system coding efficiency.

Table 4.6 shows a comparison with [4], [5], [16], [8] & our algorithm. Where, the time reduction has reached to (18.46%, 38.57%, 34.57%, 59.99%) in the above mentioned algorithms with an increasing in the BD-Rate by about (0.91%, 1.5%, 0.87%, 1.91%) respectively, while we obtained (66.95%) reduction in time and a 0.3131% increase in the BD-rate on average.

Table 4.6 Coding Efficiency & Complexity Reduction

Class Type	**Sequence Name**	**Algorithm in [38]**		**Algorithm in [5]**		**Algorithm in [12]**		**Algorithm in [3]**		**Proposed Algorithm**	
		BD-Rate (%)	**ΔT (%)**	**BD-Rate (%)**	**ΔT (%)**	**BD-Rate (%)**	**ΔT (%)**	**BD-Rate (%)**	**ΔT (%)**	**BD-Rate (%)**	**ΔT (%)**
A	PeopleOnStreet	2.3	20.91	1.41	30	0.27	35.99	2.44	63	0.43	58.72
	SteamLocomotive	0.49	12.44	--	--	0.27	27.72	1.37	48.34	0.32	62.04
B	Kimono	0.81	31.73	0.76	45	0.18	40.48	2.87	71.07	0.14	67.83
	ParkScene	0.03	8.74	1.41	47	0.39	41.73	2.33	62.08	0.26	66.7
C	BasketballDrill	--	--	1.71	46	1.84	37.49	1.37	58.20	0.54	65.47
	BQMall	--	--	3.5	50	1.11	40.53	2.18	59.15	0.41	71.2

D	RaceHorses	--	--	0.84	21	1.31	27.04	2.15	64.74	0.18	69.81
	BlowingBubbles	--	--	0.88	31	1.61	25.55	0.52	53.35	0.31	73.8
Avg.		**0.91**	**18.46**	**1.5**	**38.57**	**0.87**	**34.57**	**1.91**	**59.99**	**0.313**	**66.95**

There is an average similarity reaches to 89.77% with the original algorithm regarding Table 4.7.

Table 4.7 Average Similarity

Video Sequence Name	Similarity (%)
ParkScene_1920x1080_24	87.2%
RaceHorses_832x480_30	89.45%
BasketballPass_416x240_50	90.3%
Mobile_Cif_352x288_24	92.13%
Total Average	89.77%

4.6 GUI Interfaces

When starting a session, the server should be initialized at first as shown in Figure 4.35, and then it will wait for any input calls from the clients. Where, there are different ports and threads for each call. The Multimedia Panel in each interface displays the Local Camera.

Figure 4.35 Socket Listening and Connecting at Server

Moreover, the Administrator can share a blackboard screen with the Teleconference members' which is under the Balckboard panel, in order to write some lectures as shown in the GUI interface of server in Figure 4.36.

Fig. 4.36 Teleconference-Server GUI Interface

The benefit of each tool within the Blackboard toolbar is demonstrated in Figure 4.37.

Figure 4.37 Blackboard Toolbar

The administrator has the ability to control the call session of each joined member after selecting him/her from the list shown in the online members as shown in Figure 4.38.

Figure 4.38 Online Members Control Options

Note that the multimedia operation will not begin if the administrator didn't enable the multimedia session for such a user, Figure 4.39 shows that the administrator gave permission for one of the clients to start an audio session only.

Figure 4.39 Audio Activation Message

In our system, the Administrator will have the ability to broadcast an open session in real time with other people who are not members in our system to watch a live teleconference session using a displayed HTTP link under the chat panel, as shown in the GUI client interface of Figure 4.40. The broadcasting can also be accomplished using a second camera and microphone so that first devices remain with the main session.

Moreover, there is a screen sharing for capturing the Client Windows or activities on the PC as shown in the button next to the camera button at the Multimedia Panel in the Figure below.

Figure 4.40 Teleconference-Client GUI Interface

At the client side, the user should start the connection by clicking on the connect button from the teleconference menu in the main bar and input the Server Port Number/IP address in the connect dialog box with the registered name in the teleconference system. Then, the system will start to setup the socket connection with the destination after clicking on the “connect” button as shown in Figure 4.41.

Figure 4.41 Connect Dialog Box

The Authentication process for the login user will begin as shown in Figure 4.42.

Figure 4.42 Authentication Message

The server will receive an Authorization request from the connected client to either accept or reject his joining to the teleconference session as shown in Figure 4.43.

Figure 4.43 Authorization Message

But if the user name is not registered in the Teleconference system at all, a failure message will appear to the client user and his program will be closed immediately as shown in Figure 4.44.

Figure 4.44 Logon Failure Message

After a successful Authentication the user can start the Teleconference session as a multimedia-call (video and audio) or video or voice calls only with the ability to send a text message. The Captured Screen or Camera is displayed by the SDL YUV Display as shown in Figure 4.45.

Figure 4.45 SDL Display of Remote Camera

The client teleconference session will end after clicking on the “Stop” button to end the call and “Exit” button to exit the program. The administrator can also add new users by writing their names in the “New Registration dialog” from the Help menu as shown in Figure 4.46.

Figure 4.46 New Registration Dialog Box

Or an authenticated user can also register a new user, but then an authorization process will also be held in order to either accept or reject the new registered user by the administrator as shown in Figure 4.47.

Figure 4.47 Authentication Process for New Registration

Moreover, in the broadcasting session, there will be No need for a registration and having the client program for non-members, where they can watch the live teleconference by typing the announced URL on the Player Dialog of our Teleconference program as shown in Figure 4.48 or any other media player.

Figure 4.48 GUI Player Dialog Box

There is a decoding buffer for the variable sizes of image. Figure 4.49 shows how the HEVC video stream occupies the buffer in accordance with the different quantization level. Where, the buffering delay represents the time interval between the first byte of the coded picture to the last one. This delay occurs between two key events which are:

1) BTC: begins to receive the coded picture.
2) BTD: begins to decode the coded picture.

The bitrate and the size of buffer manage the maximum buffering delay that occurs when a decoded frame fills the buffer totally as shown in the following equation. As the buffer limit is fixed to the size of the decoded image, then the controlling of maximum delay is only depends on the bitrate value.

$$\text{Buffering-Delay}_{(\max)} = \text{Buffer-Size(Bytes)}*8/\text{Bitrate(bps)} \qquad (4.24)$$

Figure 4.49 Reduction in the Bitrate by the proposed Algorithm

The congestion of each link has been determined in order to manage the congestion and to notify the user about switching from the High-Quality to the Medium-Quality video streaming regarding the following equation:

$$Cn = ST/AT \tag{4.25}$$

Where Cn represents the degree of Congestion, St and TA represent service-time and local-time of the local packet respectively. Figure 4.50 shows the Throughput gain to about (20.3%, 14%, and 10.8%) on average compared to the Skype, original algorithm and BigO respectively using the proposed congestion control.

Figure 4.50 Throughput versus Time.

4.7 Conclusions

In our project, we developed our Teleconference system using our proposed video encoding (H.265) technology. We have obtained a better understanding of real-time systems and socket-communications by implementing the theory knowledge into practice in addition to the knowledge of hardware/software interactions. By implementing the new codec, it guaranteed the smoothing streaming of voice and video media to various individuals in a real time communication. The proposed algorithm in method (1) improved the video encoding time by 57.2% with an acceptable loss of BD rate of 0.33%, saved the bitrates to about 21% and achieved a similarity percentage with the original HEVC of 85.3%, as demonstrated in the results. The proposed algorithm in method (2) improved the video encoding time by 57.8%, increased the data compression ratio with an acceptable loss of BD rate by 0.88%, and achieved a similarity percentage with the original HEVC of 85.3%, as demonstrated in the results. The proposed algorithm in Method (3) reduced the encoder complexity suitable for real-time processing and Teleconferencing. It improved the video encoding time by 66.95% with an acceptable loss in the BD-rate by 0.313% on average and achieved a similarity percentage with the original HEVC of 89.77%, as demonstrated in the results. Also, the overall system obtained an average gain in throughput of 20% compared to Skype. We proposed a solution of pure software multipoint teleconference system, which implemented on the visual C++ developer. The Server acts as the manager of a conference setup. One of the key functions of our server model is the management of the conference, such as the user registration, member authentication, and authorization. The server has many other functions besides the session's management, such as the Blackboard Sharing, Screen

Sharing, Text Chats, Multimedia Broadcasting for members and Broadcasting for non-members using the HTTP streaming link of the live teleconference for watching a live lecture (for example) or saving it as a video file. Moreover, a list of the login members with their login/logoff time will be displayed for each member with the available audio/video capture devices. The system has a simple and extensible platform, and can work efficiently in the Intranet/Internet Environment.

4.8 Future Plans

As a future plan, we will consider more factors in the design of the system such as Echo cancellation, Noise removal. Another option that might be considered in the future is the hardware implementation such as controlling the camera of each user remotely by the administrator. In addition to that, many features can be added to system software design such as presentation sharing, file transmission. Finally, the software program of the encoder/decoder can be burned on a chip to get a faster performance than a software implementation.

REFERRENCES

[1] Roger L. Freeman. (2013). Fundamentals of Telecommunications. Second: Edition. John Wiley & Sons, Inc., England.

[2] Bruce Hartpence. (2013). Packet Guide to Voice over IP. First: Edition. O'Reilly Media, United States of America.

[3] Wes Simpson. (2008). Second: Edition. Video over IP-IPTV, Internet Video, H.264, P2P, Web TV, and Streaming: A Complete Guide to Understanding the Technology. Focal Press, England.

[4] Seunghyun Cho and Munchurl Kim. (2013). Fast CU splitting and pruning for sub-optimal CU partitioning in HEVC intra-coding, *IEEE TCSVT*. **2**. 1-10.

[5] T. L. da Silva, L. V. Agostin, and A. L. da Silva Cruz. (2012). Fast HEVC intra prediction mode decision based on EDGE direction information, in *Proc. 20th Conf. (EUSIPCO)*. **1**. 1214–1218.

[6] Shahid, Z., Puech, W. (2014). Visual protection of HEVC Video by Selective Encryptions of CABAC Bins. *IEEE Trans. Multimedia*. **16**, 24–36.

[7] Iain E. G. Richardson. (2003). H.264 and MPEG-4 Video Compression-Video Coding for Next-generation Multimedia. First: Edition. John Wiley & Sons, Inc., England.

[8] Scott Firestone, Thiya Ramalingam, and Steve Fry. (2007). Voice and Video Conferencing Fundamentals. Cisco Press. Indianapolis, USA.

[9] Eskinder Anteneh Ayele and S.B. Dhok. (2012). Review of Proposed High Efficiency Video Coding (HEVC) Standard, *International Journal of Computer Applications.* **59.** 1-9.

[10] Tamer Shanableh, Eduardo Peixoto, and Ebroul Izquierdo. (2013). MPEG-2 to HEVC Video Transcoding with Content-Based Modeling, *IEEE TCSVT*. **23**. 1191-1196.

[11] Jens-Rainer Ohm, Gary J. Sullivan, Heiko Schwarz, Thiow Keng Tan and Thomas Wiegand. (2012). Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC), *IEEE TCSVT*. **22**. 1669- 1684.

[12] Liquan Shen, Zhi Liu, Xinpeng Zhang, Wenqiang Zhao, and Zhaoyang Zhang. (2013). An Effective CU Size Decision Method for HEVC Encoders, *IEEE Transaction On Multimedia.* **15.** 465 - 470.

[13] Frank Bossen, Benjamin Bross, Karsten Sühring, and David Flynn. (2012). HEVC Complexity and Implementation Analysis, *IEEE TCSVT*. **22**. 1685-1696.

[14] Chong Luo, Wei Wang, Jian Tang, Jun Sun, and Jiang Li. (2007). A Multiparty Videoconferencing System over an Application-Level Multicast Protocol, *IEEE Transaction on Multimedia*. **9.** 1621-1632.

[15] Yi-Fei Xu and Hai-Bo Hu (2012). Development of a Network Adaptive H.264/AVC Medical Video Transmission System, *International Journal of Future Computer and Communication.* **2.** 292- 295.

[16] Scott Firestone, Thiya Ramalingam, and Steve Fry. (2007) Voice and Video Conferencing Fundamentals. Cisco Press. Indianapolis, USA.

[17] Masoud Nouripayam, and Nima Sheikhipoor. (2014). HEVC Intra-frame prediction implementation using MATLAB. Bleking Institute of Technology, Department of Signal Processing, Karlskrona. 36 p.
[18] Günther Liebl, Thomas Schierl, Thomas Wiegand, and Thomas Stockhammer. (2006). Advanced wireless multiusers video streaming using the salable video coding extensions of h.264/mpeg4-avc, *The IEEE International Conference on Multimedia & Expo (ICME)*. **6**. 625-628.
[19] Hao Zhang and Zhan Ma. (2014). Fast Intra Mode Decision for High Efficiency Video Coding (HEVC), *IEEE TCSVT*. **24**. 660-668.
[20] Mengmeng Zhang, Jianfeng Qu, and Huihui Bai. (2013). Entropy-Based Fast LargestCoding Unit Partition Algorithm in High-Efficiency Video Coding. *Entropy*. **15**. 2277-2287.
[21] Min Xing, Siyuan Xiang, and Lin Cai. (2014). A Real-Time Adaptive Algorithm for Video Streamings over Multiple Wireless Access Networks, *IEEE Journal On Selected Areas In Communications.* **32.** 795-805.
[22] Omer Boyaci, Andrea G. Forte, Henning Schulzrinne. (2009). Performance of Video-Chat Applications under Congestion, IEEE *International Symposium on Multimedia*. **7**. 213-218.
[23] Vivienne Sze and Madhukar Budagavi. (2012). High Throughput CABAC Entropy Coding in HEVC. **22**. 1778-1791.
[24] Gary J. Sullivan, Jens-Rainer Ohm and Woo-Jin Han and Thomas Wiegand. (2012). Overview of the High Efficiency Video Coding (HEVC) Standard, *IEEE TCSVT*. **22**. 1649- 1668.
[25] Do-Kyung Lee, Miso Park and Je-Chang Jeong. (2014). Fast Intra-Coding by using RD cost candidate Elimination for High Efficiency Video Coding, *Proceedings of the World Congress on Engineering and Computer Science*. **1**. 978-988.
[26] Nitin . R. Talhar, K. S. Thakare. (2011). Real-time and Object-based Video Streaming Technique with Application to Communication System, *ISCCC*. **1**. 25-29.
[27] Jie Jiang, Kefeng Fan, Baolong Guo and Wei Mo. (2012). Block-Based Parallel IntraPrediction Scheme for HEVC, *Journal of Multimedia*. **7**. 289-294.
[28] Mark D. Pesce. (2003). Programming Microsoft®DirectShow® for Digital Video and Television . First: Edition. Microsoft Press.
[29] Jian Xiong, Hongliang Li, Qingbo Wu, and Fanman Meng. (2014). A Fast HEVC interCU Selection Method based on Pyramid Motions Divergence, *IEEE Trans.Multimedia*. **16**. 559-564.
[30] J. Chakareski and P. Frossard. (2008). Distributed collaboration for enhanced sender-driven video streaming , *IEEE Trans.Multimedia*. **10**. 858–870.
[31] Il-Koo Kim, Junghye Min, Lee T., Woo-Jin Han, and Jeong Hoon Park. (2012). Block partitioning structure in the HEVC stand, *IEEE TCSVT*. **22**. 1697-1706.
[32] Chua, T., and Pheanis, D.C. (2006). Bandwidth Conserving Real-Time VoIP Teleconference System, In Proceeding of the Third Intern Conf. On Info. Technology IEEE. **1**, 10-15.
[33] Yang Jie, Shen Sun. (2006). Implementation of Multipoint Video Conferences in Software, *IEEE conference*. **1**, 5-12.

[34] Yi-Fei Xu and Hai-Bo Hu. (2013). Development of Network Adaptive H.264/AVC Medical Video Transmission Systems, *International Journal of Future Computer and Communication*. **2**. 4-8.
[35] Feng Pan, Xiao Lin, Rahardja S., Lim, K.P., Li, Z.G., Dajun Wu, and Si Wu. (2005). Fast Mode Decision Algorithm for Intraprediction in H.264/AVC Video Coding, *IEEE TCSVT*. **15**. 813-822.
[36] Wen Shi, Xiantao Jiang, Tian Song and Takashi Shimamoto. (2015). Edge Detector based fast level decision algorithm for intra-prediction of HEVC, *Journal of signal processing.* **19**. 67-73.
[37] Zhang, Moh. (2013). Fast intra prediction mode decision algorithm for HEVC. *TELKOMNIKA*. **11.** 5703–5710.
[38] Anis BenHajyoussef, Tahar Ezzedine, and Ammar Bouallegue. (2014). Fast Gradient Based Intra-mode decision for High Efficiency Videos Coding, *International Journal of Emerging Trends & Technology in Computer Science*. **3**. 223-228.
[39] Chen-Kuo Chiang, Wei-Hau Pan, Chiuan Hwang, Shin-Shan Zhuang and Shang-Hong Lai. (2011). Fast H.264 encoding based on statistical learning, *IEEE TCSVT.* **21**. 1304-1315.
[40] Yan Ye, and Andrivon P. (2014). The Scalable Extension of HEVC for Ultrahigh Definition Video Delivery, *IEEE Multimedia*. **21**. 58–64.
[41] Siwei Ma, Shiqi Wang, and Wen Gao. (2014).Low Complexity Adaptive View Synthesis Optimization in HEVC Based 3D Video Coding, *IEEE Trans.Multimedia*. **16**. 266-271.

APPENDIX

(Teleconference Basics)

1. Definition:

The term Teleconference/videoconferencing is a confusing one. Some commercial companies (AT & T in the States) are now advertising "videoconferencing: as a new technology. The fact is that videoconferencing is a function which can be hosted on a variety of technologies and has been for some years. It is not a technology in itself. In America, the term is fast becoming defined as any use of television to join people in some live interaction. However, the term is actually applied to a wide range of situations from live video lecturing to large audiences, to a point-to-point, individual-to-individual desktop PC chats. One possible categorisation is into large scale and small scale. The majority of large scale set-ups are currently satellite-based in the form of "interactive television" i.e., one-way video, two-way audio. This allows for broadcast from a central point to many different locations regardless of distance. Small scale refers to compressed video for meetings between relatively few points for small meetings. A technology used for this function is ISDN. ISDN promises to make two-way video equally as cost effective, with potential for greater interactivity. Traditional video conferencing requires expensive, fixed delivery and reception installations and high transmission costs over full band width analogue video channels or high capacity digital channels. Such high grade services allow full two-way audio and video communication between several locations at a price; a more common configuration is that of Interactive TV (Full service out, audio only in). High costs and lack of flexibility has limited the past educational uses in the past to research projects. Recent developments in video compression and codec technology are increasing the use of relatively low bandwidth ISDN using a variety of display formats.

2. Technological Issues:

The technologies used to deliver video conferences currently have a dramatic effect on the quality of the communication achievable. This report concentrates on the use of ISDN with compressed video where as earlier examples, often referred to as Interactive Television use transportation media such as microwave links or satellite links and provide high cost, full motion video links. In order to understand the issues in video conferencing, a basic understanding of transmission technology is required. The issues are those of:

- Bandwidth
- video compression
- delivery method

2.1. Bandwidth:

(Or baud rate) refers to the amount of information (bits) which can be transmitted along a carrier every second. The bandwidth required depends on the application. Thus textual data can be transmitted slowly, or using a narrow bandwidth because it is not required in real time and the printed information contains little information relative to sound and video. Sounds such as speech, contain more information than the printed word and have to be carried at the same speed as normal speech for conversation to be possible; thus a wider bandwidth is required. If moving pictures have also to be sent and real time transmission is required then we have a lot of information to send very quickly, thus a high bandwidth is required. Consider the amount of information that must be transmitted if a video image is to be sent in real time. A screen has 625 lines and there are 625 points on each line. A point is stored in 24 bits. To transmit in real time, 25 pictures must be sent every second. This adds up to 234,375,000 bits of information to be transmitted every second. Telephone lines are required to carry voice information at roughly 2400 bps which they manage. A single ISDN telephone line can carry information at 64 Kbps. So consider how many telephone lines would be required to carry the amount of information for a video image! It would not be possible. Thus alternative to telephone lines, such as satellite or fibre optics which have extremely high bandwidth would have to be used. This would reduce accessibility of the technology and increase costs. An alternative is to compress the video image thus less information is transmitted along a lower bandwidth medium. We will look at both solutions.

2.2. *Video compression*:

An analogue, full motion, signal takes up a great deal of bandwidth. It can be converted to digital signals (and vice versa) and compressed using codecs. One way to reduce the bandwidth required is to compress the video image; that is digitise the signal and then remove as much extraneous data as possible. A video signal changes 25 times per second, not the entire picture changes in each frame and so with a compressed image only the changes are sent. Thus the more changes, the less compression can occur. Compressed video can be squeezed into as few as two telephone lines, or up to 24. The greater the compression means the greater the loss of clarity, continuity of motion and colour information. There are already several levels of compression being adopted:

- Video on desktop computers: 64 kbps. Allows video integrated into the screen. However this rate is not good enough for full video conferences, it would suffice for one to one video phone situations. Many people in education do not feel it is adequate.
- Group video conferencing: between 128 kbps and 2 Mbps. 384 kbps is providing a good quality reception for conferencing and is used in many educational environments.
- Digital Broadcasting: Uses 2- 6 Mbps rates. The quality is greatly increased over the previous compression levels but costs are also higher.
- HDTV: 25-45 Mbps is adopted by High Definition Television. This is a relatively new technology and it is not universally accepted. Some vendors provide equipment that can operate at a variety of compression levels. Remember the more visuals and movements to be transmitted the greater the transmission requirements and hence the higher the cost of both transmission and site equipment. There are recommended ways of coping with video compression.

2.3. *Delivering Video Conferences*:

- By standard telephone lines: The advantage of this method is one of accessibility. However, this only allows 64 kbps transmission rates. This is relatively untested for education. The picture and sound quality will be poor and the picture jerky.

- By ISDN: ISDN can in theory be carried through any telecommunications delivery medium: fibre optics, or telephone. This method is rapidly being taken up by educational establishments. There are several advantages for the education sector:
    - Two way voice, data and graphics can be carried simultaneously over the telephone network.
    - May become the standard telephone system and therefore reach all homes, offices and educational establishments.
    - It is relatively inexpensive.
- By satellite broadcast: A common use of satellites is for one way television transmission with audio only feedback. The advantage is that its cost is independent of distance, whereas cable transmission costs increase with distance. This method is used where very large distances and many sites are involved or there are natural barriers to the laying of any cable technology, such as mountain ranges or oceans. The disadvantages include lack of visuals from receive site, unavailability of transmission time, or particular time may be expensive and the cost of the equipment installation.
- By VSAT (very small aperture terminal): This method uses narrow band transmission (256-384 kbps). They can be used as receive only, or for data transmission to a central point, with a more powerful “hub”. For education, it is possible to use VSAT in a “mesh” system with each site capable of both transmit and receive, to any other site on the system, thus allowing any site to originate teaching materials. Full two ways motion video with sound is possible. Transmission costs are likely to be low but ground stations are in the range of 50,000 dollars.
- By co-axial cable: It is a useful medium if delivering from a single point (TV Station) to many different sites. It is a common form of television broadcasting in North America and becoming more common in the UK. Co-axial cable can allow up to 40 TV channels to be carried without compression. In Canada, at least one of these must be dedicated to educational programming (The Knowledge Network). It can also be used for two way video communication

provided there is suitable equipment at both ends. The cost of laying cable is high so would only be of benefit for educational video conferencing if the cable was already laid. This is the case in the USA, Canada and some parts of the UK.

- By fibre-optic cable: This has a far greater capacity than co-axial cable. • It is currently being installed by major telecommunication companies. It will form a backbone for services that will often transfer to copper wire for local delivery. There is therefore a growing range of ways to deliver video conferencing. The most appropriate choice of system will depend partly on the physical configuration of sites to be connected, the applications which are required, the amount of traffic to be carried, and the distances between sites.

3. Range of equipment:

One site to one site connections is typically called point to point connections. This reflects the number of sites connected rather than the number of people present at a site. Multipoint connections are also possible. However this is achieved by using a multipoint bridge and switching between sites can be carried out manually or can be voice activated (if you make a noise, you are seen). Multipoint bridges can be rented from British Telecom for the duration of a conference. They are prone to technical problems. 3.1.

3.1. Full screen TV image:

The Roll about system is used by corporate meetings when full screen, life-like intimacy is required. The codec, TV monitor, camera and microphone are all integrated into a single unit. Other audio-visual equipment can be used in conjunction with this equipment:

- Slide to video devices for displaying slides
- Micro-video systems for displaying microscopic images
- 3-D imaging devices for displaying real images
- PhotoCD presenters
- Video players
- Video Lecterns (visualizer) for documents and prepared presentations.

The choice of devices is dependent on the nature of the content of communication between the participants. This form of equipment is more reliable than its Desktop version. It is also more expensive. A recent setup at Heriot Watt University required investment of £25,000 at both ends of the connection. This type of equipment can be linked with stations supplied by other vendors. We have used this system with Norway, Holland, Australia and California on a regular basis. Although some functionality such as remote control of cameras is lost, the audio and visual connection is manageable. This equipment can be used if the focus of the conference is on the speakers and the materials being discussed are not in a computerized format. The range of audio-visual equipment means that little reworking of material is necessary.

3.2. ***Desk Top Video Conferencing (DTVC)***:

DTVC is the technology which is bringing video conferencing back into focus. However, the frame rate and the tiny picture window make standalone DTVC an uncompelling application. But when it is used in conjunction with other collaborative work software, such as whiteboards, shared screen and shared control, there is adequate functionality to entice users. This type of video conference is most useful when the documents and information to be exchanged are stored on the computer and of importance rather than the presenter. Information can be shared and discussed quickly over the network. There are a number of problems associated with this equipment: The best DTVC offers 15 fps (frames per second) or 16 fps. This is like watching a pixelated home video. Typical rates are even lower. This may never prove adequate if full screen live interaction is required. Crashes and bandwidth problems are common. Many systems do not support sound. This is handled via the telephone! This means that the sound and video image are not kept synchronised. The odd thing about DTVC is that it is not the video aspect that sells the system. This is simply incidental to file transfer and other co-operative working activities. This equipment is not currently 100% reliable. The H261 standard is not well implemented. The equipment is cheaper than others, requiring an investment of around £5,000 per station. But both stations must be identical in terms of hardware and software, including versions. The Physical Environment Often people think they can set up a video conferencing system anywhere. This would be analogous to holding a seminar in the coffee lounge or your office while the person you are working with is trying to get on with their work! To maximise the chance of successful interaction the quality of the input must be maximised. The content of the conference should be the central issue, but if the student is uncomfortable, sound is poor or inadequate lighting is reducing the quality of images, then the learning process will be interfered with. The conference equipment should be based in a special room if possible. It can then be available for use whenever it is required rather than setting it up every time it is requested. There is justification for support personnel to maintain and run the equipment and leave the lecturers free to concentrate on the learning process. There are some guidelines for room set up and organising a video conference. Summary anyone considering video conferencing as a solution to an educational need should understand the nature of the technology. The technology is still evolving. ISDN and desktop systems are still problematic and fall-back scenarios should always be in place in case of system failure. This supports the need for specialised technicians who understand the technology and keep it up and running. This would be an added burden on lecturers if they were to maintain the equipment and get the best out of it. The quality of the signal is going to be reduced due to compression and therefore someone who knows how to maximise input in terms of sound and vision would be an indispensible member of staff. To reduce the potential for problems, the equipment at both ends of the connection should be identical.

Summary:

Anyone considering video conferencing as a solution to an educational need should understand the nature of the technology. The technology is still evolving. ISDN and desktop systems are still problematic and fall-back scenarios should always be in place in case of system failure. This supports the need for specialised technicians who understand the technology and keep it up and running. This would be an added burden on lecturers if they were to maintain the equipment and get the best out of it. The quality of the signal is going to be reduced due to compression and therefore someone who knows how to maximise input in terms of sound and vision would be an indispensible member of staff. To reduce the potential for problems, the equipment at both ends of the connection should be identical.

VITA

PERSONAL INFORMATION

Name and Surname: *SHAIMA SAFA ALDIN BAHA ALDIN*

Natioality: *IRAQ*

Birth place and date: *IRAQ /31-12-1982*

Marial status: *Married*

Email: *eng_shaima1183@yahoo.com*

EDUCATION	**Graduate school**	**Year**
Master	Al-Nahrain University	2006
Bachelor	Al-Nahrain University	2003
High School	Al-Ma'aly Highschool	2000

Work experience	**Place**	**Enrollment**
2006-Present	Al-Nahrain University	Instructor

PUBLICATIONS

- **(Article) Design and implementation of a teleconferencing system using improved HEVC coding *at (Journal of Real Time Image Processing | February 2016)***
- **(Article) A new teleconference system with a fast technique in HEVC coding *at (Multimedia Tools and Applications | December 2015)***
- **(Article) A Real-Time H.265 codec at (International Journal of Science and Research (IJSR) | May 2014)**
- **(Book) A New Internet Service Provider Billing System at (LAP LAMBERT Academic Publishing (January 26, 2013))**
- **(Article) A New Internet Service Provider Billing System at (*IJCSET |November 2011*)**

FOREIGN LANGUAGE ***English***

HOBBIES ***Drawing and Reading***