AIN SHAMS UNIVERSITY
FACULTY OF ENGINEERING
Electronics and Communications Engineering Department

Development of MPEG-4 Video Codec for WCDMA Mobile System

A Thesis submitted in
partial fulfillment of the requirements for the Degree of
Master of Science in Electrical Engineering.

Submitted by

Sherif Mohamed Gouda Hassanein

B.Sc. in Electronic Engineering
(Electrical Communications Engineering, Menoufia University, 1999)

Supervised by

Prof. Dr. Abd El-halim Abd El-naby Zekry
Department of Electronics and Communications Engineering
Faculty of Engineering, Ain Shams University

Prof.Dr. Abd El-fattah Ahmed El-sohly
Chairman and CEO of
Banha Company for Electronic Industries

Dr. Mohammed Nabil Mostafa
Department of computer science Engineering
Faculty of Engineering, Ain Shams University

Faculty of Engineering, Ain Shams University
EGYPT, CAIRO
2009

AIN SHAMS UNIVERSITY
FACULTY OF ENGINEERING
Electronics and Communications Engineering Department

JUDGMENT COMMITTEE

Name	:	Sherif Mohamed Gouda Hassanein
Thesis	:	Development of MPEG-4 Video Codec for WCDMA Mobile System
Degree	:	Master of Science in Electrical Engineering

NAME, TITLE, AND AFFILIATION	SIGNATURE
Prof.Dr. El-sayed Mostaa Saad Professor of Electronics and Communications Engineering Faculty of Engineering, Helwan University	-------------- (Examiner)
Prof.Dr. Adel Ezzat El-henawy Professor of Electronics and Communications Engineering Faculty of Engineering, Ain Shams University	-------------- (Examiner)
Prof.Dr. Abd El-haleem Abd El-naby Zekry Professor of Electronics and Communications Engineering Faculty of Engineering, Ain Shams University	-------------- (Supervisor)
Prof.Dr. Abd El-fattah Ahmed El-sehly Chairman of the Board of Directors of Banha Company for Electronic Industries	-------------- (Supervisor)

Date : 1/11/2009

Faculty of Engineering, Ain Shams University
EGYPT, CAIRO
2009

AIN SHAMS UNIVERSITY
FACULTY OF ENGINEERING
Electronics and Communications Engineering Department

STATEMENT

This Dissertation is submitted to Ain Shams University in partial fulfillment of the Degree of Master of Science in Electrical Engineering (Electronics and Communications Engineering).

The work include in this thesis was received by author at the Department of Electronics and Communications Engineering, Faculty of Engineering, Ain Shams University, Cairo, Egypt.

No part of this thesis was submitted for a degree or qualification at any other university or institution.

Name : Sherif Mohamed Gouda Hassanein

Signature : ……………………………

Date : ……/……../………

Faculty of Engineering, Ain Shams University
EGYPT, CAIRO
2009

AIN SHAMS UNIVERSITY
FACULTY OF ENGINEERING
Electronics and Communications Engineering Department

CURRICULUM VITAE

Name	: Sherif Mohamed Gouda Hassanein
Date of birth	: 11/11/1976
Place of birth	: Egypt, Qalioubeya.
Nationality	: Egyptian
First university degree	: B.Sc. in Electronic Engineering (Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, 1999)
Certification date	: May, 1999.

Name	: Sherif Mohamed Gouda Hassanein
Signature	: ………..………………………
Date	: ……/……../………

Faculty of Engineering, Ain Shams University
EGYPT, CAIRO
2009

Acknowledgements

I would like to express my great gratitude to my major advisor **Prof. Dr. Abd El-haleem Zekry** for his great support, guidance and encouragement. His support stands behind all my achievements in my thesis. I realy appreciate his care to constructively criticize my work.

Also, I would like to thank **Dr. Abd El-fattah El-Sohly** for his continuous support and care.

I would like to express my sincere gratitude to **Dr. Mohamed Nabil** for his great help. I realy enjoyed working with him.

Special Appreciation must be dedicated to **Dr. Mohamed Shedeed** and **Dr. Fawzy Elmansy** for their support during the preparation of this thesis.

Finally, I would like to thank my family, specially my parents, and my wife for their great support

Faculty of Engineering, Ain Shams University
EGYPT, CAIRO
2009

Chapter 1

Introduction

1-1 What's Video?

A video image is a projection of a 3-D scene onto a 2-D plane (Figure 1.1). A 3-D scene consisting of a number of objects each with depth, texture and illumination is projected onto a plane to form a 2-D representation of the scene. The 2-D representation contains varying texture and illumination but no depth information. A still image is a 'snapshot' of the 2-D representation at a particular instant in time whereas a video sequence represents the scene over a period of time.

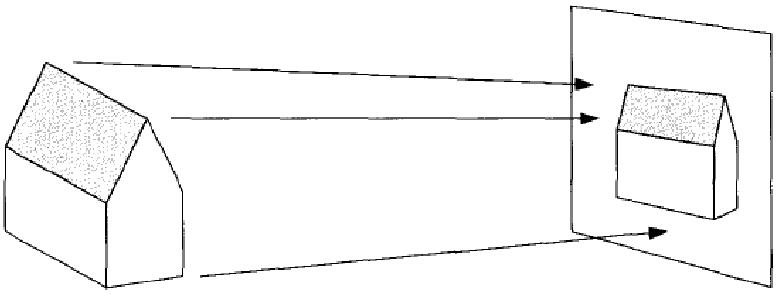

Figure 1-1 Projection of 3-D scene onto a video image

A real visual scene is continuous both spatially and temporally. In order to represent and process a visual scene digitally, it is necessary to sample the real scene spatially (on a rectangular grid in the video image plane) and temporally (as a series of still images or frames sampled at regular intervals in time) as shown in Figure 1-2. Digital video is the representation of a spatio-temporally sampled video scene in digital form. Each spatio-temporal sample (pixel) is represented digitally as one or more numbers that describe the brightness (luminance) and color (chrominance) of the sample.

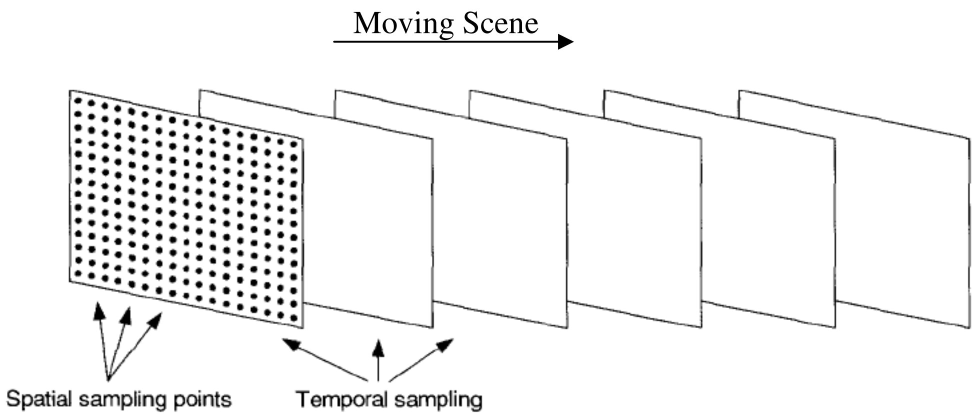

Figure 1-2 Spatial and Temporal Sampling

Video compression is reducing the quantity of data used to represent digital video images, and is a combination of spatial image compression and temporal motion compensation. It is a tradeoff between disk space, video quality, and the cost of hardware required to decompress the video in a reasonable time. Video compression is needed because raw or uncompressed digital video typically requires a large bitrate (approximately 216 Mbits for 1 second of uncompressed TV-quality video) and compression is necessary for practical storage and transmission of digital video [1].

Compression uses two systems, a compressor (encoder) and a decompressor (decoder). The encoder converts the source data into a compressed form (occupying a reduced number of bits) prior to transmission or storage and the decoder converts the compressed form back into a representation of the original video data. The encoder/decoder pair is often described as a CODEC (enCOder/ DECoder) (figure 1-3) [2].

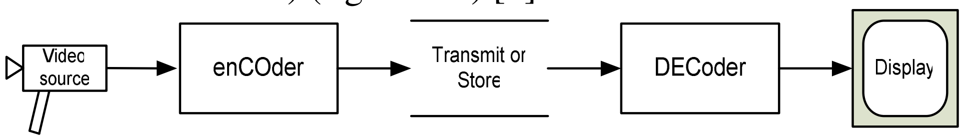

Figure 1.3 Encoder/Decoder

1.2 Video Compression

Video compression has two important benefits. First, it makes it possible to use digital video in transmission and storage environments that would not support uncompressed video. Second, video compression enables more efficient use of transmission and storage resources. If a high bit rate transmission channel is available, then it is a more attractive proposition to send high-resolution compressed video or multiple compressed video channels than to send a single, low-resolution, uncompressed stream. Even with constant advances in storage and transmission capacity, compression is likely to be an essential component of multimedia services for many years to come.

Video compression is achieved by removing redundancy (components that are not necessary for faithful reproduction of the data). Video signals contain three types of redundancy:

1- **Statistical redundancy** which is present because certain data patterns are more likely than others. This is mainly due to the high spatial (intraframe) and temporal (interframe) correlations between neighboring pels.

2- **Psychovisual redundancy** which is due to the fact that the Human Visual System (HVS) is less sensitive to certain visual information than to other visual information.

3- **Coding redundancy** which is due to the usage of more and/or longer code symbols than necessary [2].

1.3 Compression Methods

A form of data compression is **lossless**. This means that when the data is decompressed, the result is a bit-for-bit perfect match with the original. While lossless compression of video is possible, it is rarely used.

Another form of video compression is **lossy**. It operates on the premise that much of the data present before compression is not necessary for achieving good perceptual quality, so the reconstructed data is not identical to the original data. Such method is therefore irreversible, and it usually achieves higher compression than lossless methods.

1.4 Video Encoder

The video encoder (figure 1-4) consists primarily from three main elements.

1- **The transformer** which transforms the input raw data into a representation that is designed to reduce statistical redundancy and make the data more amenable to compression in later stages. The transformation is a one-to-one mapping and is, therefore, reversible.

2-**The quantizer** reduces the accuracy of the mapper's output in an attempt to reduce psychovisual redundancy. This is a many-to-one mapping and is, therefore, irreversible.

3- **The symbol encoder** (or codeword assigner) assigns a codeword, a string of binary bits, to each symbol at the output of the quantizer. The code must be designed to reduce coding redundancy. This operation is reversible [3].

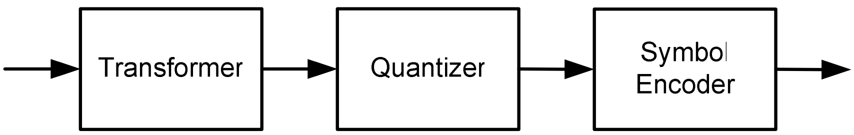

Figure 1-4 Elements of a video encoder

1.5 Video Coding Applications

Video coding has too many applications that became very important in our living style. Some of them are:

- Video-on-demand.

- Broadcast television.
- Distance learning and training.
- Interactive gaming.
- Remote shopping.
- Online media services, such as news reports.
- Videotelephony. (figure 1-5)
- Videoconferencing. (figure 1-6)
- Telemedicine for remote consultation and diagnosis.
- Telesurveillance.(figure 1-7)
- Remote consultation or scene-of-crime work.
- Collaborative working and telepresence. [3]

Figure 1-5 A Video call

Figure 1-6 A Video conference

Figure 1-7 Telesurveillance

1.6 Video Coding Standards

Video coding standardization started in 1980 by **International telegraph and Telephone Consultative Committee (CCITT),** which is currently known as **International Telecommunications Union – Telecommunication Standardization Sector (ITU-T)**. This first action was followed by **CCIR (now ITU-R), ISO** and **IEC**.

1.6.1 H.120

H.120 was published by the CCITT in 1984 and targeted for videoconferencing applications at 1.544 Mb/s and 2.048 Mb/s.

H.120 streams ran at 1544 kbit/s for National Television System Committee (NTSC) and 2048 kbit/s for Phase Alternating Line (PAL). H.120 quality was not good because of its poor temporal quality.

1.6.2 H.261

H.261 was the first practical video codec published in 1990 by ITU-T for transmission over Integrated Services Digital Network (ISDN) lines on which data rates are multiples of 64 kbit/s. The coding algorithm was designed to be able to operate at video bit rates between 40 kbit/s and 2 Mbit/s. The standard supports two video frame sizes: Common Intermediate Format (CIF) (352x288 luma with 176x144 chroma) and Quarter CIF (QCIF) (176x144 with 88x72 chroma) using a 4:2:0 sampling scheme.

H.261 introduced the macroblock concept. Each macroblock consists of a 16x16 array of luma samples and two corresponding 8x8 arrays of chroma samples, using 4:1:1 sampling and a YCbCr color space. The adopted techniques of H.261 *{16 × 16 macroblocks for MC, 8 × 8 blocks for Discrete Cosine Transform (DCT), SKIP/INTER/INTRA mode switching, zigzag scanning, Run Length Encoding (RLE), scalar quantization, Variable Length Encoding (VLC) entropy coding, deblocking filtering}* have become key elements in most video coding standards.

H.261 is still used as a backward-compatibility mode in some video conferencing systems and for some types of internet video.

1.6.3 MPEG-1

MPEG-1 was an early standard for lossy compression of video and audio. It was designed to compress VHS-quality raw digital video and CD audio down to 1.5 Mbit/s (26:1 and 6:1 compression ratios respectively) without excessive quality loss, making Video CDs, digital cable/satellite TV and Digital Audio Broadcasting (DAB) possible.

The MPEG-1 standard is published as ISO/IEC-11172. The standard consists of the following five Parts:

1. Systems (storage and synchronization of video, audio, and other data together)
2. Video (compressed video content)
3. Audio (compressed audio content)
4. Conformance testing (testing the correctness of implementations of the standard)
5. Reference software (example software showing how to encode and decode according to the standard)

The MPEG-1 file extensions includes (mpg, mpeg, mp1, mp2, mp3, m1v, m1a, m2a, mpa, mpv). It supports resolutions up to 4095×4095 (12 bits) and bitrates up to 100 Mbits/s.

MPEG-1 videos use Source Input Format (SIF) resolution: 352x240, 352x288, or 320x240. These low resolutions, combined with a bitrate less than 1.5 Mbit/s, make up what is known as a **constrained parameters bitstream (CPB)**, later renamed the "Low Level" (LL) profile in MPEG-2. This is the minimum video specifications any decoder should be able to handle, to be considered MPEG-1 compliant.

MPEG-1 has several frame types such as Intra-frames, Predicted-frames, Bidirectional-frames and a unique frame type not found in later video standards which is DC-frames. These frames are intra frames that have been DC encoded only (AC coefficients are removed) and hence are very low quality. DC-frames are only used for fast previews of video, for instance when seeking through a video at high speed.

1.6.4 MPEG-2

MPEG-2 is a standard for the generic coding of moving pictures and associated audio information. It describes a combination of lossy video compression and lossy audio compression (audio data compression) methods which permit storage and transmission of movies using currently available storage media and transmission bandwidth.

1.6.5 H.263

H.263 was developed by ITU-T Video Coding Experts Group (VCEG) in 1996 as a video codec standard designed as a low-bitrate compressed format for videoconferencing.

H.263 was the principle codec for many applications like:

1. Flash Video Content (You Tube, Google Video, etc.)
2. Erlier versions of Real Player before version 8.

The codec was first designed to be utilized in H.324 based systems (Public Switched Telephone Network-PSTN and other circuit-switched network videoconferencing and videotelephony)

1.6.6 MPEG-4

MPEG-4 was introduced in 1998 by the ISO/IEC Moving Picture Experts Group (MPEG) under the formal standard ISO/IEC 14496. Uses of MPEG-4 include streaming media, CD distribution, voice (telephone, videophone) and broadcast television applications.

MPEG-4 has new features such as (extended) VRML support for 3D rendering, object-oriented composite files (including audio, video and VRML objects).

AAC (Advanced Audio Codec) was standardized in MPEG-2 (as Part 7) before MPEG-4 was issued.

Initially, MPEG-4 was aimed primarily at low bit-rate video communications. MPEG-4 provides the following functionalities:

- Improved coding efficiency
- Ability to encode mixed media data (video, audio, speech)
- Error resilience to enable robust transmission
- Ability to interact with the audio-visual scene generated at the receiver

1.6.7 H.264

H.264 is a standard for video compression, and is equivalent to MPEG-4 Part 10, or MPEG-4 AVC (for Advanced Video Coding). As of 2008, it is the latest block-oriented motion-compensation-based codec standard developed by the ITU-T Video Coding Experts Group (VCEG) together with the ISO/IEC Moving Picture Experts Group (MPEG), and it was the product of a partnership effort known as the Joint Video Team (JVT). The ITU-T H.264 standard and the ISO/IEC MPEG-4 Part 10 standard (formally, ISO/IEC 14496-10) are jointly maintained so that they have identical technical content. The final drafting work on the first version of the standard was completed in May 2003.

The intent of the H.264/AVC project was to create a standard capable of providing good video quality at substantially lower bit rates than previous standards (e.g. half or less the bit rate of MPEG-2, H.263, or MPEG-4 Part 2), without increasing the complexity of design so much that it would be impractical or excessively expensive to implement. An additional goal was to provide enough flexibility to allow the standard to be applied to a wide variety of applications on a wide variety of networks and systems, including low and high bit rates, low and high resolution