Handbook of Research on Mobile Multimedia

Second Edition

Ismail Khalil Ibrahim
Johannes Kepler University Linz, Austria

Volume I

Information Science REFERENCE

INFORMATION SCIENCE REFERENCE
Hershey · New York

Director of Editorial Content:	Kristin Klinger
Director of Production:	Jennifer Neidig
Managing Editor:	Jamie Snavely
Assistant Managing Editor:	Carole Coulson
Typesetters:	Jeffrey Ash and Michael Brehm
Cover Design:	Lisa Tosheff
Printed at:	Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanbookstore.com

Library of Congress Cataloging-in-Publication Data

Handbook of research on mobile multimedia / Ismail Khalil Ibrahim, editor. -- 2nd ed.
p. cm.
Includes bibliographical references and index.
Summary: "The book is intended to clarify the hype, which surrounds the concept of mobile multimedia through introducing the idea in a clear and understandable way, with a strong focus on mobile solutions and applications"--Provided by publisher.
ISBN 978-1-60566-046-2 (hardcover) -- ISBN 978-1-60566-047-9 (ebook)
1. Mobile communication systems. 2. Wireless communication systems. 3. Multimedia systems. 4. Multimedia communications. 5. Mobile computing. I. Ibrahim, Ismail Khalil.
TK6570.M6H27 2008
384.3'3--dc22
2008013114

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter X
Comparison of Video Coding Standards Used in Mobile Applications

Goran Gvozden
RTL Televizija, Croatia

Mislav Grgic
University of Zagreb, Croatia

Sonja Grgic
University of Zagreb, Croatia

Miran Gosta
Croatian Telecommunication Agency, Croatia

ABSTRACT

This chapter gives description and comparison of video coding standards currently used in bandwidth limited mobile communications emphasizing at the same time the importance of coding efficiency and robustness, particularly for video applications. Due to its exceptional efficiency and performance a number of mobile service operators recognized and embraced relatively new H.264/AVC compression method. Utilization of this efficient compression method in the bandwidth limited and distortion prone mobile environment enables and provides transport of high quality video on low data rates. In order to demonstrate these abilities a comparison of H.264/AVC relative to MPEG-4 SP and H.263 method is presented. Comparison is performed using objective video quality assessment methods accompanied with description of different issues related to video quality measurement and its implication on coding process. Finally, comparison showed and confirmed great efficiency and performance possibilities, which will make H.264/AVC the ubiquitous coding technique of multimedia world in time to come.

INTRODUCTION

Although already designed and enhanced for high data rates and flexible communication capabilities, today's modern mobile telecommunication systems are experiencing growth of mobile user requirements for delivering of high quality multimedia information. Mobile users are used to being able to access different kind of rich multimedia content such as pictures, moving clips, movie trailers, animations or real time video streaming on their mobile terminals. Provision of this high quality multimedia information puts more demanding requirements on planning and designing of mobile communications systems for improving throughput, transfer delay, and data error rate.

Important issues arise from this situation. The first one is cost of services that forms and affects the acceptance and attractiveness of certain mobile applications, the second is the capacity of modern mobile communication channels, and the third is storage capability of mobile terminals. Access to rich multimedia content through the mobile network is very expensive since it requires high data rate point-to-point connection for each mobile user. The importance of mobile terminal storage capacity lies in the possibility to allow the time of information transmission to be decoupled from its time of use, thereby enabling the best possible economy of use of the available spectrum. Furthermore, large number of users and limited bandwidth capacity can cause system congestion that can lead to poor quality of service requiring network operators to carefully balance network traffic (Holma et al., 2007). Concerning the real time video distribution as an alternative to point-to-point oriented, limited and expensive mobile networks such as 3G (Kumar, 2007), new point-to-multipoint mobile systems were developed to enable and allow reception of video for a number of mobile users simultaneously without consuming limited channel resources (Kornfeld, 2004). Additionally, Faria (2006) describes how representatives of these new mobile networks employ techniques and spectrum of digital terrestrial broadcasting technology. There are several standards around the world designed for broadcasting to mobile terminals such as DVB-H (ETSI EN 302 304, 2001) and T-DMB/DAB (ETSI EN 102 427, 2005; ETSI EN 102 428, 2005). Probably a combination of technologies will be used together in order to provide interactivity and high quality of delivered rich multimedia content on mobile terminals as depicted in Jordan et al. (2006).

Nevertheless, regardless of used mobile system for efficient utilization of available transmission channel capacity and improvement of the quality of service, the appropriate coding method needs to be selected for mobile systems. Since video signal is composed of successive frames alternating in time domain, there is a high correlation present between neighboring image elements (pixels) inside and between successive frames. This means that the video has a large amount of spatial and temporal redundancy, which can be removed in order to reduce the irrelevant information and achieve the required transmission bit rate. Furthermore, while human visual system is less sensitive to higher spatial frequencies, these higher frequencies can be eliminated without any effects on subjective picture quality degradation (Zovko-Cihlar et al., 1998; Bauer et al., 1998).

To select the most efficient and robust video coding method for some application, a complete performance analysis of influence of compression on transferred video quality has to be made (Joch et al., 2002). This chapter analyses the effects of video compression methods on picture quality in low bit rate communication. Three video coding methods are tested: H.263 (ITU-T Rec. H.263, 2000), MPEG-4 SP (ISO/IEC 14496-2, 2000) and H.264/AVC (ITU-T Rec. H.264, 2005). The focus of this chapter has been on H.264/AVC coding technique, which is relatively new technique in mobile applications in comparison with H.263 and MPEG-4 SP coding techniques, already established in mobile environment. Two test sequences with different spatial and frequency characteristics suitable for low bit rate transmission and display on mobile terminals were used for testing on different bit rates.

Evaluation of video quality was performed using objective video quality assessment methods, PSNR (peak signal to noise ratio), VQM [video quality metrics –(Xia, 2000)], and SSIM [structural similarity –(Wang et al., 2002)]. Besides that, description of usefulness of these methods in coding performance optimization is given. The next section describes the H.264/AVC advanced video coding method characterized by high compression ratios and coding efficiency that makes it a number one candidate for systems with limited bandwidth capacity.

VIDEO CODING IN MOBILE COMMUNICATIONS

While broadcast networks are able to provide the high quality video streaming to a large number of users, unicast networks are experiencing congestion issues when large number of users are accessing video streaming services simultaneously. Constant growth of more demanding mobile services requires redefinition and standardization of new, advanced, and complex coding techniques. Accordingly, the basic principles of advanced H.264/AVC video compression methods will be described as well as already mentioned H.263 and MPEG-4 SP video coding methods designed and used for low bit rate communication in mobile environment. The progress of video compression techniques combined with advanced mobile network systems has made possible to reach the point when the mobile user will be able to access the high quality multimedia content, where the real time streaming video applications are the most demanding concerning the capacity, coding algorithm, complexity, and power consumption.

Transferring video over capacity constrained and distortion submissive mobile channels presents very demanding and difficult tasks. A number of different parameters have to be taken into consideration when choosing the convenient coding method. A compromise has to be made among bit rate, video quality, transfer delay, implementation complexity, storage capacity, and power consumption. The video signal consists of two parts of information, variable and unexpected as well as invariable and expected information. The first information is called entropy of the video signal that defines the minimum transmission bit rate for reception without quality degradation, while the second information represents redundancy. Hereafter, we will describe techniques used to compress the video signal in order to achieve required bit rate.

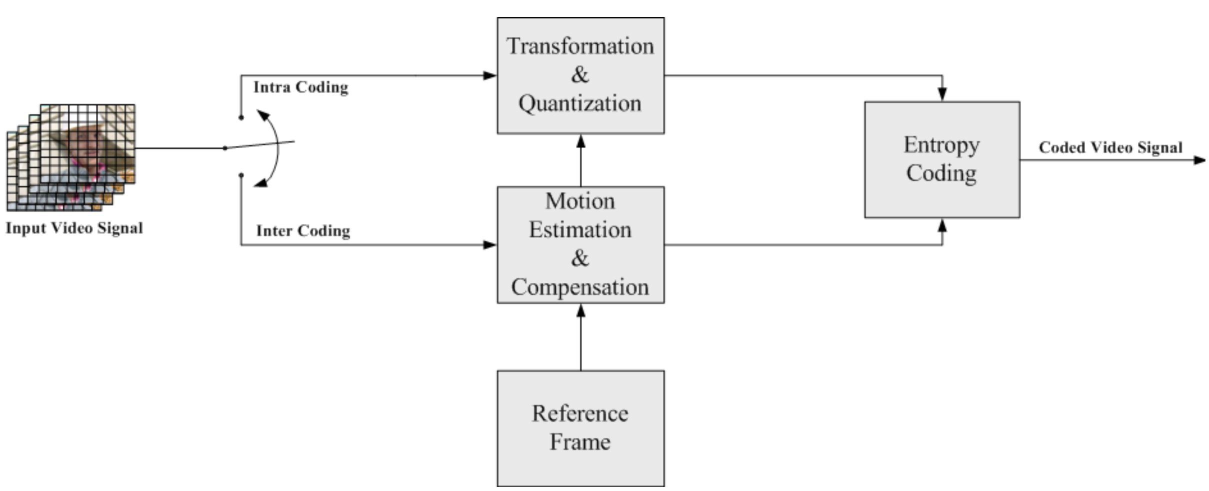
Video Compression Basics

All of today's video coding techniques are based on conventional block-based hybrid video coding concepts. This concept implies combination of transform coding for exploitation of spatial redundancy and inter-frame prediction that utilize motion compensation process for exploitation of temporal redundancy. Prior to transformation and prediction process each video frame is divided into blocks of picture elements. The maximum and common size of these blocks is 16x16 picture elements but it can be smaller depending on used compression standard. Blocks of picture elements can be processed in different coding mode depending on the frame or slice coding type. According to applied coding, three types of pictures are defined in all standards. There are intra-coded frames called I-frames where macro blocks are coded separately without referring to successive frames. Frames that are referring to successive frames during coding process are defined as inter predicted frames, further divided on P-frames (using previously coded frames as reference) and B-frames (using both, previous and future coded pictures as a reference). Further bit rate reduction is done using entropy coding methods that map the input elements to a series of code words where the length of codeword depends on statistical probability of input elements. Block scheme of video coding process is shown in Figure 1.

Hereafter, we will explain transformation and prediction process as well as entropy coding. To transform the blocks of image elements to a domain more suitable for compression a discrete cosine transformation (DCT) is used. Essentially, DCT transformation is performing a frequency analysis of image content that results in transform coefficients. DCT is followed by quantization process that assigns higher values to lower spatial frequency components according to characteristics of human visual system. Since the human visual system is more sensitive to lower spatial frequencies, quantized coefficients representing picture area with higher spatial frequencies can be represented with smaller number of bits per picture element or even discarded. Thus, compression ratio depends on amount of discarded spatial redundancy presented in high frequency area of input image. The inverse procedure is done during decompression process. Transformation and quantization of image blocks is done for all frame types: I-frame, P-frame, and B-frame.

Another process for bit rate reduction is inter-frame coding, which removes temporal redundancy between successive frames. Temporal redundancy is reduced by forming a predicted frame and subtracting it from current frame. The predicted frame is assumed to be a motion translated version of one or more previous frames (P frame) or future frames (B frame). The motion translation vectors are calculated using different techniques that should determine the location of macroblock from current frame in the previous frame based on highest correlation between macroblocks (Atkinson, 2004). After the motion estimation a motion compensation process is carried out by subtracting the motion translated, predicted frame from the current frame. Subtraction result is a residual frame that is to be transformed, entropy coded, and sent to the decoder for image reconstruction. Furthermore, successfulness of this process depends mostly on the motion search area and block size used for motion estimation process. Using smaller block sizes implies more accurate estimation and compensation process. However, this increases the complexity and produces a large number of transmitted motion vectors that can cause extra overhead that outweighs the advantage gained with less residual information. Therefore a compromise has to be made to adapt the block size to the picture characteristics in such a way that large blocks are used in large homogenous

Figure 1. Block scheme of the video coding process



areas and small blocks are used in complex and heterogeneous areas (Richardson, 2003).

Third process included in video compression schemes is entropy coding. According to the information theory, entropy coding represents the process of assigning variable length codes to transformed and quantized coefficients where the code word length is a function of the probability distribution of coded syntax elements. Hence, the code word L_c is proportional to negative logarithm of the syntax element probability and is calculated as:

$$L_c = -\mathrm{Log}_b P \qquad (1)$$

where P is probability value of syntax element and b is number of symbols used for generating code words. There are various entropy coding techniques used in video compression schemes. For a detailed inquiry interested reader is referred to Viterbi (1991) and Mackay (2003)

Video Compression Standards for Low Bit Rate

Above-mentioned and described methods present the core functionalities of the video compression techniques standardized for utilization in mobile communications. Accordingly, we will give a brief overview of these techniques selected by mobile network operators around the world. The emphasis will be put on a relatively new and advanced H.264/AVC standard.

ITU-T Recommendation H.263

H.263 standard (Cote et al., 1998), first time published in 1995, was designed for low bit rate primarily in video telephony and videoconferencing applications over circuit and packet-based networks. The standard evolved over time, as new improved algorithms were included in coding process. Video coding algorithm is based on hybrid of transform coding and inter frame prediction. There are five standardized video formats covered with coding algorithm: sub-QCIF (SQCIF), QCIF, CIF, 4CIF and 16CIF (Table 1). The motion compensation is performed on 16x16 macro blocks and 8x8 sub blocks of frame elements. Besides that, prediction process is improved using the more precise interpolation techniques for motion estimation. Furthermore, use of multiple reference pictures is enabled as well as deblocking filter for blockiness effect reduction. Overall compression efficiency was improved by intra prediction mode that used neighboring, previously processed transform coefficients to predict and code current transform coefficients. Moreover, entropy scheme also experienced changes within the meaning of using multidimensional run-level-last variable length codes with tables optimized for lower bit rate. Due to above mentioned functionalities, flexible video formats, error resilience features, and video scalability suitable for utilization in mobile environment, current mobile networks selected this standard as an essential factor for delivery of quality multimedia content to mobile user.

ISO/IEC Standard 14 496-2: MPEG-4 Visual

MPEG-4 Visual standard, first time published in 1998, covers a large number of functionalities that include efficient coding of natural and synthetic video information. MPEG-4 Visual covers a wide range of different profiles and levels developed for various applications. For the purpose of this chapter we will focus on Simple and Advanced Simple profiles suitable for bandwidth constrained mobile

Table 1. Input video formats

Format	16CIF	4CIF	CIF	QCIF	SQCIF
Resolution	1408 × 1152	704 × 576	352 × 288	176 × 144	128 × 96

environment. MPEG-4 efficiency and flexibility is built on top of the H.263 functional elements. Differences can be found in quantization methods and the more precise prediction enabled by use of ¼ pixel interpolation process. Variable block sizes are allowed as well as bidirectional coded frames (B frames). The main goal of MPEG-4 coding for low bit rate was to achieve high coding efficiency while providing scalable, reliable, low latency, and robust transmission at moderate complexity and low power consumption.

ITU-T Recommendation H.264 and ISO/ IEC Standard 14 496-10: H.264/AVC

H.264/AVC standard was developed and standardized jointly by the ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Experts Group (MPEG). It shows great improvements in compression efficiency and error robustness in comparison with H.263, MPEG-2, and MPEG-4 Visual for a number of applications ranging from mobile services and videoconferencing to broadcasting and digital storage media. Essentially, it uses the same fundamental functional elements as previous video coding standards (MPEG-2, MPEG-4 Part 2, H.261, and H.263), such as block-based spatial prediction for reduction of spatial redundancy, motion estimation, and compensation for reduction of temporal redundancy and residual entropy coding. However, the main difference lies in particular functional elements presented in Wiegand et al. (2003), such as adaptive variable block size, accuracy of the prediction process, in-loop deblocking filter, multiple reference frames, and entropy coding based on context modeling. Another functionality that affects great coding efficiency is quantization parameter. Increment of one in quantization step matches the increase in the quantization step of approximately 12% that actually represents bit rate reduction in same percentage. Furthermore, for robustness to bit error and operational flexibility, variable slice sizes, arbitrary slice ordering, and flexible macro blocks ordering is responsible.

The standard has seven profiles and fifteen levels. The profile can be described as a subset of different coding algorithms and the level presents bit rate constraints on parameter values (such as frame size and bit rate) and thus restricts computational complexity within certain profiles. This section will focus on baseline profile designed for mobile communications, video telephony, and videoconferencing. Thereafter, functional elements for baseline profile will be explained and described.

Variable block size is very usable during motion compensation process. There are seven different block sizes defined for inter-frame prediction and three block sizes defined for intra prediction modes. Macro block division is allowed from 16x16 to 4x4 block size, which increases motion search precision.

Motion estimation implies the search of the area (macro block) in the previous frame to find the best match with the current macro block in order to calculate the translation. In some cases, when motion vector doesn't point to integer pixels, to find the best match interpolation technique has to be applied in order to generate non-integer pixel positions. More detailed information concerning the motion compensation can be found in Wedi (2003).

Baseline profile has possibility to use multiple reference pictures, which enables more efficient and accurate prediction process. These pictures are stored in encoder's buffer and organized in lists of short and long term pictures (Wiegand et al., 1999; Wiegand et al., 2001).

Each coded video frame is decoded internally in encoder and processed with deblocking filter (List et al., 2003). In-the-loop deblocking filtering function compares samples near vertical and horizontal block boundaries and smoothes the block edges if necessary. It has great significance because it reduces ringing and blocking distortion while preserving image structure, object characteristics and subjective quality.

Redundant picture represents part or entire coded picture, which is discarded in normal operation. This functionality is very useful in case of missing or damaged primary pictures normally used for reconstruction in decoder. Therefore redundant picture can replace damaged or missing primary picture.

Using arbitrary slice order (ASO) means that slices inside the video frame do not have to follow precise decoding order. If first macro block in a slice of decoded frame has a smaller address comparing to first macro block in a previously decoded slice of the same frame, ASO is in use.

Flexible macro block ordering (FMO) enables grouping of macro blocks to different slices inside the frame (Wenger, 2003).

For entropy coding of the bit stream syntax (macro block-type, motion vectors) H.264/AVC baseline profile uses advanced functionality called context-based adaptive variable length coding (CAVLC). CAVLC is using different code word tables, which depend on context of previously coded syntax elements. It is a lossless method but it prolongs time of coding and decoding.

VIDEO QUALITY ASSESSMENT

As we described, in today's mobile environment different coding techniques and models are used depending on the application area. Selection of convenient coding technique suitable for certain application and video system is a very demanding task dependant on a quality of received video signal after the process of acquisition, compression, transmission, and reconstruction. Consequently, an efficient method of video quality measurement is required to identify and determine the level of reconstructed video quality degradation. In this respect a process of correction and improvement of video signal can be easier. Hereafter we will explain different methods for video quality evaluation and the difficulties encountered during such process.

A digital video content is representation of a natural video scene composed of different objects with their own characteristic shape, texture, dimension, color, and luminance, which represent different spatial frequencies. These spatial (objects' shape and dimension) and temporal (moving objects, camera movement) characteristics of the picture affect the compression process and determine the level of encoding process complexity. To determine the level of compression used in coding process and its effects on video quality, different methods for video quality measurement are used. Furthermore, quality evaluation and comparison of encoded video content is very difficult and demanding process principally because the video quality depends on the subjective experience of a viewer, which is the result of a complex interaction between eye and brain. Methods for picture quality measurement can be roughly separated on subjective and objective quality assessment methods.

Subjective quality measurement is time consuming and expensive, affected by a number of parameters (active or passive viewing, viewing environment). However, it is more accurate than objective method, which is faster and computationally easier. In Recommendation ITU-R BT. 500-11, 2002, ITU-T P.910 (1999), methodology for subjective quality assessment is given accompanied with the description of viewing conditions, source signals, criteria for test sequences selection, and results presentations.

As opposed to subjective quality assessment methods, objective methods are based on the faster and easier mathematical models, which make them more attractive for implementation in video systems. The purpose of objective methods is development of quantitative measures that automatically predict video quality. Currently, contributing efforts are being made to develop and standardize objective assessment method that would align with subjective quality assessment methods. The International Telecommunication Union is the organization leading the process of developing recommendations for the objective measurement of video quality (ITU-T Tutorial, 2004).

A wide range of different objective assessment methods is used today in the video systems. Most of these methods determine video quality calculating the difference between original and reconstructed video after compression. Thus, the amount of difference represents the degradation level of reconstructed video. In Wang et al. (2003), a thorough analysis and description of different objective methods is given. Wang gives and explains difficulties of designing and developing an objective video quality metric that would correlate well with subjectively perceived video quality. For better understanding a brief introduction to the relevant physiological and psychophysical components of the HVS is given. Furthermore, weaknesses of error sensitivity approach for determining the video quality are particularly emphasized. Although some methods tried incorporating spatial-temporal characteristics of the human visual system (HVS), they are still insufficiently aligned with it. In the next section we will describe methods used for comparison of coding methods for mobile communications based on error sensitivity philosophy as well as on spatial-temporal characteristics of human visual system.

Objective Quality Assessment Methods

Comparison of H.263, MPEG-4 SP and H.264/AVC was performed using PSNR, VQM, and SSIM objective quality measurement methods. All three quality assessment methods belong to full reference method, which means that the original sequence was used as a reference for comparison with compressed sequences.

Peak signal to noise ratio (PSNR) is the most widely used objective video quality assessment method. As we mentioned, video signal is composed of successive rectangular frames. Calculation of video distortion is based on measuring the amount of error between frame elements of reconstructed and original frames. It is expressed as a logarithmic ratio of maximum amplitude of picture elements and mean square error (MSE). PSNR is defined as:

$$PSNR[dB] = 10\log_{10}\frac{\left(2^n - 1\right)^2}{MSE} \tag{2}$$

where n is number of bits per picture element of original video sequence. MSE is defined as:

$$MSE = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left[a_{ij} - a'_{ij}\right]^2 \tag{3}$$

where $M \times N$ is frame size of video signal processed by the compression system. a_{ij} represents original picture element while a'_{ij} represents corresponding picture element in the reconstructed frame. MSE is squared difference between picture element in original frame and picture element in reconstructed frame.

Due to simple and fast calculation PSNR is a very popular method. It is useful for quality assessments of same test sequences coded with different coding methods on different bit rates. This enables selection of appropriate coding methods for use in certain video system. The higher PSNR values represent higher quality of reconstructed video.

The second method used is video quality metric (VQM) (Xiao, 2000), which is a variant of a digital video quality method (DVQ) (Watson, 2001). It brings some advantages comparing to PSNR because it simulates the spatial-temporal characteristics of human visual system. Original and processed picture are transformed using DCT according to relation (3):

$$\begin{aligned} DCTa_i'(u,v) &= DCT\left(a_i(x,y)\right) \\ DCTa_i'(u,v) &= DCT\left(a_i'(x,y)\right) \end{aligned} \tag{4}$$

where $DCTa_i$ and $DCTa_i$’ are transformed coefficients of original and reconstructed frame respectively. Afterwards, local contrast for original and reconstructed block is calculated from transformation coefficients using DC component of each block respectively:

$$LC_i\left(u,v\right)=\frac{DCTa_i\left(x,y\right)}{DC_i}\left(\frac{DC_i}{1024}\right)^{0.65}$$
$$DC_i\left(u,v\right)=DCTa_i\left(0,0\right) \tag{5}$$

where $LC_i(u,v)$ and $LC_i'(u,v)$ represent local contrast of original and reconstructed frame respectively, number 1024 represents mean DCT value, while the 0.65 represents parameters that best suits psychophysics data. $LC_i(u,v)$ and $LC_i'(u,v)$ coefficients are processed regardless of static or dynamic frames, with human spatial contrast sensitivity function (SCSF) in order to be converted to just noticeable difference values of original ($JND_i(u,v)$) and reconstructed frame ($JND_i'(u,v)$). Subsequently subtraction of JND coefficients of original and reconstructed frame is performed to produce $Diff_i(t)$ values. Contrast masking is incorporated into maximum operation, which is then weighted with the pooled distortion according to:

$$Dist_{Mean}=1000\,mean\left(mean\left(Diff_i\left(t\right)\right)\right)$$
$$Dist_{Max}=1000\max\left(\max\left(Diff_i\left(t\right)\right)\right)$$
$$VQM=Dist_{Mean}+0.005\,Dist_{Max} \tag{6}$$

The quality of a video signal measured with VQM method decreases with higher VQM values. Zero VQM value represents video signal without degradation.

The third method is based on a SSIM Index. Wang et al. (2002, 2004) proposes this video quality metric, which simulates human visual system and its natural ability to extract structural information from the viewing field. Based on this ability a quality assessment algorithm is implemented in the following way. SSIM objective video quality assessment method compares the information about luminance, contrast, and structural similarity between original and processed picture. SSIM is defined as:

$$SSIM\,(x,y)=l(x,y)\,c(x,y)\,s(x,y) \tag{7}$$

where $l(x,y)$ is luminance comparison, $c(x,y)$ is contrast comparison and $s(x,y)$ is structure comparison. They are defined as:

$$l\left(x,y\right)=\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1} \tag{8}$$

$$c\left(x,y\right)=\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2} \tag{9}$$

$$s\left(x,y\right)=\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \tag{10}$$

x and y represent two aligned nonnegative picture signals for comparison, μ_x and μ_y represent mean intensity of x and y picture, σ_x and σ_x are standard deviation of both pictures, and σ_{xy} is covariance of both pictures. C_1, C_2, C_3 are constants, which are defined as:

$$C_1=\left(K_1L\right)^2,\quad C_2=\left(K_2L\right)^2,\quad C_3=\frac{C_2}{2} \tag{11}$$

where L is dynamic range of pixel values (255 for 8 bit grayscale pictures) and K_1, $K_2 << 1$. Since human visual system is sensitive to structural distortion inside picture, quality assessment with SSIM parameter can be considered as a deservingly approximation of subjective quality assessment methods.

The maximum SSIM value of 1 indicates that the two compared frames are identical while lower values indicate degradation and structural dissimilarity.

Importance of Video Quality Measurement

Above-mentioned objective quality methods can be very applicable in a wide range of video applications. Results of these video quality measurements are source of important information about certain coding methods, video communication, and storage systems. Accordingly, information, identification of weaknesses and disadvantages in different video systems can be realized and thus used to control and improve overall system. Consequently, objective quality assessment algorithms can be incorporated in coding process for the purpose of performance optimization. This is very useful in error prone wireless networks utilized for low bit rate transmission. Since quality of a reconstructed video depends on a wide range of different parameters, the main goal in optimization process is shaping the coding parameters to find the minimal amount of video information that is to be transmitted over communication channel in a way that the original signal is reconstructed with acceptable distortion (Sullivan et al., 1998). Schuster (1996) gives detailed overviews of rate distortion optimized video compression that include selection process among coding parameters strengthened with mathematical fundamentals.

TESTING SETUP AND RESULTS

Performance evaluation was performed for three coding techniques, H.263, MPEG-4 SP, and H.264/AVC BP. In order to make comparison of these techniques we have encoded available test sequences on different bit rates suitable for mobile communication (3GPP Technical specification TR 26.902, 2007).

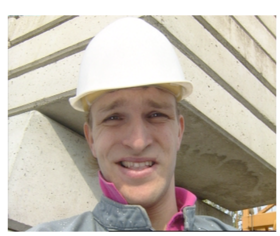
Test Setup

Sequences were coded using free software coder available at www.erightsoft.com. Encoded video sequences were composed of I-frames and P-frames only. Every twentieth frame was intra coded while other frames were inter-coded. During coding, rate-distortion optimization, faster encode (single-pass), and frame reordering was used. Coding was performed using baseline profile for all three coding techniques with differences in level setup. Thereby, MPEG-4 SP used level 3, H.263 used level 20 and H.264/AVC used level 1.3.

For testing purposes, two different test sequences with duration of twelve seconds were used. Both sequences were in YUV 4:2:0 formats, which are converted to audio video interleave format convenient for processing with used software. Sequences were available in two different formats suitable for display on mobile terminals. Hence, the first format is Common Intermediate Format (CIF) with 352x288 frame size (progressive), while the second format is Quarter CIF (QCIF) with 176x144 frame size (progressive). First sequence—named Foreman—describes a man standing and talking in front of shaking camera. In the end of Foreman sequence camera turns to the building with constant movement. The second test sequence—Mobile—depicts a toy train passing in front of moving detailed and colorful background. Figure 2 depicts both sequences.

In order to perform comparison, identical bit rates were used for all three coding techniques. Used bit rates for QCIF sequences are follow-

Figure 2. Original test sequences used for performance evaluation a) Foreman and b) Mobile



a)

b)

ing: 64 kbps, 128 kbps, 192 kbps, 288 kbps, and 384 kbps, whereas CIF sequences were coded on 64 kbps, 128 kbps, 288 kbps, 567 kbps, and 768 kbps.

Test Results and Analysis

All coded test sequences were processed with the application that outputs results for PSNR, VQM, and SSIM objective measurement values. Table 2 and Table 3 depict results for QCIF Foreman and Mobile sequences respectively while Table 4 and Table 5 depict results for CIF Foreman and Mobile sequences respectively. PSNR, SSIM, and VQM are given for luminance component (Y). It can be seen that the best results for QCIF sequences are accomplished with H.264/AVC coding technique followed by MPEG-4 SP and H.263 coding techniques, taking into account all three assessment metrics. Figure 3 depicts frames extracted from QCIF sequences Foreman and Mobile, coded at 192 kbps. It can be seen that images have different perceptual quality depending on used coding techniques listed in the same order as that gained with objective results. Obtained results for CIF format showed different behavior on lower bit rates comparing to assessment results obtained for QCIF format. On very low bit rates (64 kbps) coding efficiency is changed in favor of MPEG-4 SP coding technique, followed by H.263 and H.264/AVC BP.

With higher bit rates (after 128 kbps) H.264/AVC again shows better coding efficiency in comparison with MPEG-4. Quality assessment results were also shown in a form of rate distortion curves (for PSNR values only). Rate distortion curves are useful for presentation of coding efficiency among different coding techniques where one curve for each encoder is being evaluated. Curves are generated using and plotting average PSNR values obtained on different bit rates on vertical axis while the corresponding bit rates are plotted on horizontal axis. Therefore, looking at rate distortion curves depicted in Figure 4 and Figure 5, it can be clearly seen that H.264/VAC has great coding efficiency in respect to MPEG-4 SP and H.263.

Furthermore, there is another method for presenting the video quality assessments results expressed in form of a bit rate saving. Determination

Figure 3. Comparison of picture quality for test sequences coded at 192 kbps with H.263, MPEG-4 SP and H.264/AVC BP

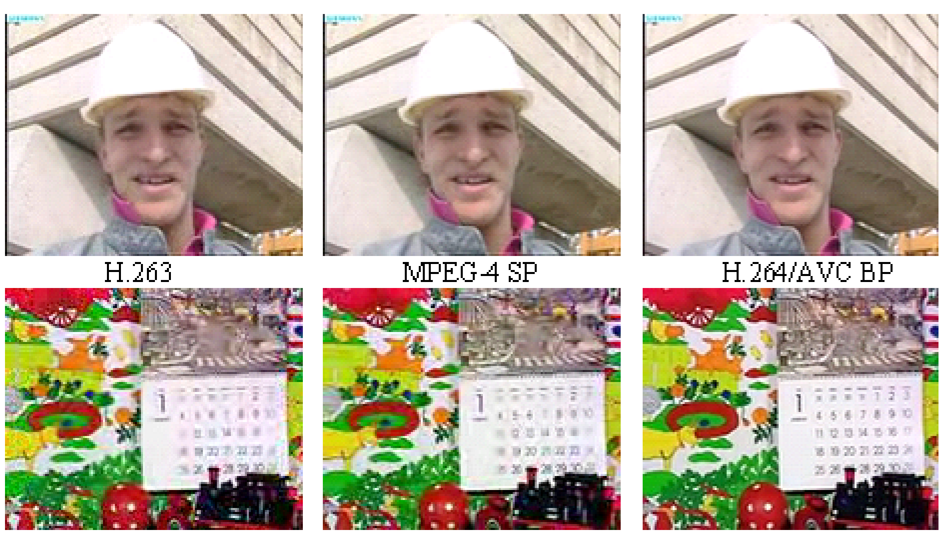


Table 2. The objective quality assessment results for QCIF sequences Foreman coded with H.263 with MPEG-4 SP and H.264/AVC on different bit rates

Bit rate [kbps]	Foreman QCIF 25 Hz								
	H.263			MPEG-4 SP			H.264/AVC BP		
	PSNR	VQM	SSIM	PSNR	VQM	SSIM	PSNR	VQM	SSIM
64	30,66	2,00	0,88	31,67	1,78	0,90	33,11	1,58	0,93
128	33,28	1,53	0,93	34,41	1,35	0,94	36,69	1,07	0,97
192	35,01	1,28	0,95	36,08	1,13	0,96	38,72	0,85	0,98
288	36,79	1,07	0,96	37,76	0,95	0,97	40,68	0,68	0,99
384	38,03	0,93	0,97	38,91	0,84	0,98	42,01	0,58	0,99

Table 3. The objective quality assessment results for CIF sequences Foreman and Mobile coded with H.263 with MPEG-4 SP and H.264/AVC on different bit rates

Bit rate [kbps]	Mobile QCIF 25 Hz								
	H.263			MPEG-4 SP			H.264/AVC BP		
	PSNR	VQM	SSIM	PSNR	VQM	SSIM	PSNR	VQM	SSIM
64	23,53	5,27	0,80	24,38	4,73	0,83	26,65	3,67	0,90
128	24,81	4,58	0,85	25,90	3,99	0,87	29,57	2,62	0,95
192	25,95	4,03	0,88	27,11	3,49	0,90	31,55	2,08	0,97
288	27,40	3,41	0,91	28,59	2,97	0,93	33,60	1,64	0,98
384	28,69	2,95	0,93	29,86	2,57	0,95	35,15	1,37	0,99

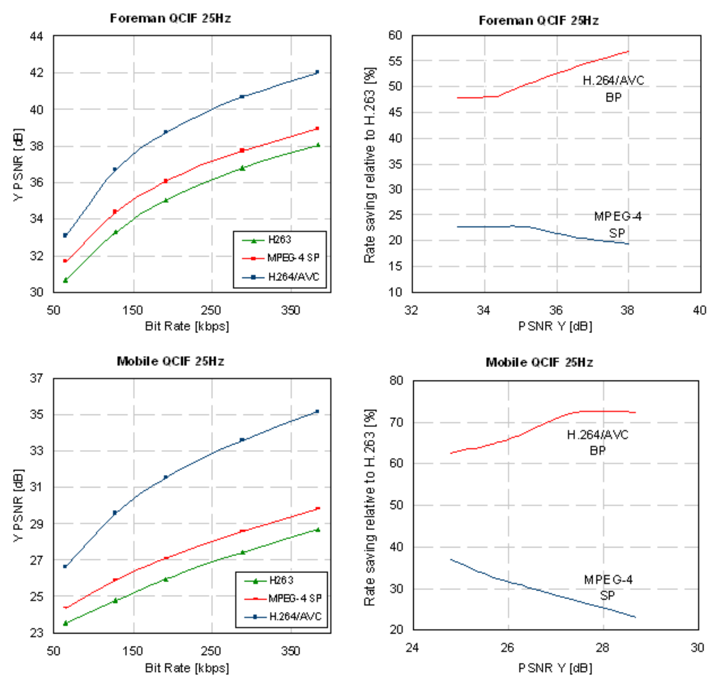
Table 4. The objective quality assessment results for CIF sequences Foreman coded with H.263 with MPEG-4 SP and H.264/AVC on different bit rates

Bit rate [kbps]	Foreman CIF 25 Hz								
	H.263			MPEG-4 SP			H.264/AVC BP		
	PSNR	VQM	SSIM	PSNR	VQM	SSIM	PSNR	VQM	SSIM
64	29,19	2,35	0,80	29,48	2,29	0,81	28,87	2,56	0,80
128	30,42	2,06	0,84	31,92	1,83	0,88	31,92	1,83	0,88
288	32,83	1,62	0,89	33,55	1,49	0,90	35,45	1,24	0,94
576	34,93	1,31	0,92	35,62	1,22	0,93	37,83	0,95	0,96
768	36,30	1,14	0,94	36,95	1,06	0,95	39,26	0,81	0,97

Table 5. The objective quality assessment results for CIF sequences Mobile coded with H.263 with MPEG-4 SP and H.264/AVC on different bit rates

Bit rate [kbps]	Mobile CIF 25 Hz								
	H.263			MPEG-4 SP			H.264/AVC BP		
	PSNR	VQM	SSIM	PSNR	VQM	SSIM	PSNR	VQM	SSIM
64	22,34	5,16	0,77	22,75	5,01	0,77	21,94	6,30	0,70
128	23,34	5,05	0,78	23,81	4,83	0,79	24,73	4,52	0,84
288	24,44	4,63	0,81	25,04	4,31	0,83	28,08	3,09	0,92
576	25,83	3,95	0,86	26,64	3,58	0,88	30,89	2,26	0,96
768	26,77	3,57	0,88	27,63	3,23	0,90	32,13	1,96	0,97

Figure 4. Luminance rate distortion curves and average bit rate savings for QCIF tested sequences

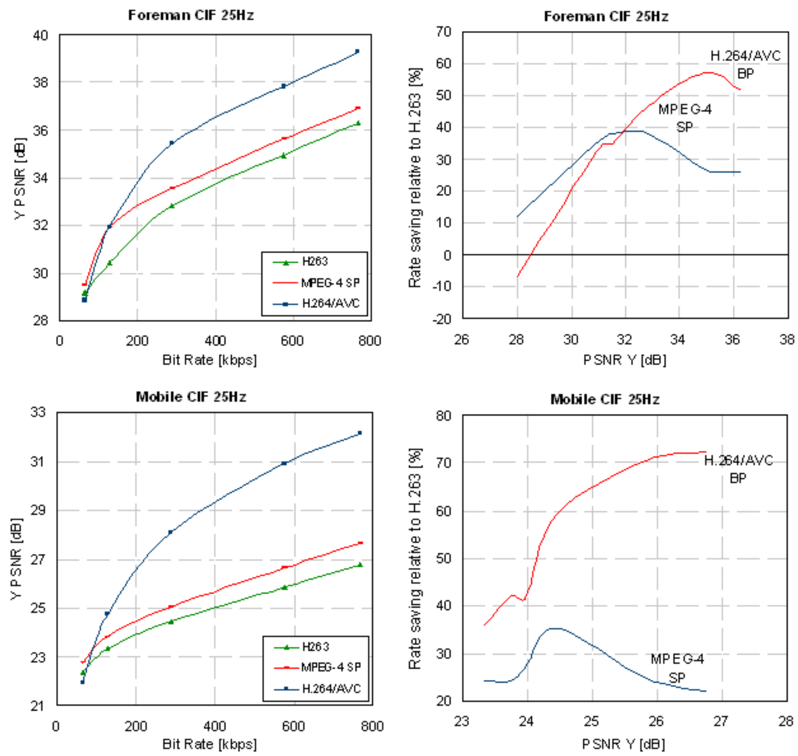


of bit rate saving of one coding technique relative to another can be achieved with interpolation of two points on two rate distortion curves, which is followed by finding the points of same distortion and calculating the difference in bit rates between these points. A curve with the poorest results is used as a reference during calculation of bit rate savings. Bit rate saving is defined as:

$$S_{bit} = \frac{A(PSNR) - B(PSNR)}{A(PSNR)} \quad 100 \quad [\%] \qquad (12)$$

where A represents the bit rate of the inferior coder necessary to achieve certain PSNR value while B represents the bit rate of better coder necessary to achieve the same PSNR value as inferior coder. Besides this interpolation process,

Figure 5. Luminance rate distortion curves and average bit rate savings for CIF tested sequences



an easier method for bit rate calculation is given in Bjontegaard (2001). This method proposes calculation of either the change in bit rate or change in PSNR in order to represent the results for average bit rate saving between different curves.

Based on described methods we made a comparison of coding techniques concerning the amount of bit rate savings. Curves of average bit rate saving are depicted in Figure 4 and Figure 5. Comparing the bit rate saving curves it is even more obvious that H.264/AVC BP outperforms MPEG-4 SP in all range of tested bit except the very low bit rates. Difference between two techniques is increasing if using higher bit rates. H.264/AVC achieves the same results as MPEG-4 SP at three to four times lower bit rates. Comparison of coding techniques in Wiegand et al. (2003) and Gvozden et al. (2007) resulted in similar results.

CONCLUSION

In this chapter we described important issues that have implications on mobile communication systems. Delivering high quality multimedia content to mobile users over distortion prone and capacity limited wireless channels depends greatly on high efficiency and robustness of video coding techniques especially in unicast networks. Description of core functionalities of coding techniques for low data rates communication showed the core functionalities needed for achieving the high video quality at low bit rates. The results of performed comparison demonstrated significant advantage and enhanced compression efficiency of H.264/AVC relative to MPEG-4 SP and H.263 coding technique. PSNR objective video quality assessment evidently showed extraordinary performance and H.264/AVC superiority. VQM and SSIM methods considered as a deservingly approximation of subjective quality assessment methods showed and confirmed great H.624/AVC coding efficiency.

Test results show that H.264/AVC coding technique achieves same video quality as MPEG-4 SP and H.263 at three to four times lower bit rates. Coding tools like variable block size, motion compensation with small block size, quarter sample accurate motion compensation, multiple reference pictures, flexible macroblock ordering, deblocking filter and CAVLC represent functional elements responsible for demonstrated results. With these characteristics and performance, H.264/AVC will easily substitute current coding methods and become dominant coding technique used for video applications in mobile systems.

REFERENCES

3rd Generation Partnership Project. (2007). *Video Codec Performance (Release 7)*. Technical Specification Group Services and System Aspects, 3GPP TR 26.902 V7.0.0 (2007-06).

Bauer, S., Zovko-Cihlar, B., & Grgic, M. (1996). Objective and subjective evaluations of picture quality in digital video systems. In *International Conference on Multimedia Technology and Digital Telecommunication Services*, ICOMT'96, Budapest, Hungary, (pp. 145-150).

Bjontegaard, G. (2001). Calculation of average PSNR differences between RD-curves. In *VCEG-M33*, ITU-T Q6/16, Austin, TX, USA.

Cote, G., Erol, B., Gallant, M., & Kossentini, F. (1998). H.263+: Video coding at low bit rates. *IEEE Transactions on Circuits and Systems for Video Technology*, *8*(7), 849-866.

ETSI EN 102 427 Digital Audio Broadcasting (DAB). (2005). *Digital audio broadcasting (DAB); data broadcasting - MPEG-2 TS streaming* (Version 1.1.1). European Broadcasting Union, European Telecommunications Standards Institute.

ETSI EN 102 428 Digital Audio Broadcasting (DAB). (2005). *Digital audio broadcasting (DAB); DMB video service; user application specification* (Version 1.1.1). European Broadcasting Union, European Telecommunications Standards Institute.

ETSI EN 302 304 Digital Video Broadcasting (DVB). (2004). *Transmission system for handheld terminals (DVB-H)* (Version 1.1.1). European Broadcasting Union, European Telecommunications Standards Institute.

Faria, G., Henriksson, J.A., Stare, E., & Talmola, P. (2006). DVB-H: Digital broadcast services to handheld devices. *Proceedings of the IEEE*, *94*(1), 194-209.

Gvozden, G., Gosta, M., & Grgic, S. (2007). Comparison of H.264/AVC and MPEG-4 ASP coding techniques designed for mobile applications using objective quality assessment methods. In *Proceedings of the 49th International Symposium ELMAR-2007 focused on Mobile Multimedia*, Zadar, Croatia (pp. 51-54).

Holma, H., & Toskala, A. (2007). *WCDMA for UMTS: HSPA Evolution and LTE*. Chichester, England: John Wiley & Sons.

ISO/IEC JTC 1. (2000, February). *Coding of audio-visual objects - Part 2: Visual* (MPEG Visual Version 2). ISO/IEC 14496-2.

ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG4-AVC). (2005, July). *Advanced video coding for generic audiovisual services* (Version 4).

ITU-R Recommendation BT. 500-11. (2002). *Methodology for the subjective assessment of the quality of television pictures*.

ITU-T Recommendation P.910. (1999). *Subjective video quality assessment methods for multimedia applications*. Geneva, October 1999.

ITU-T Recommendation H.263. (2000). *Video coding for low bit rate communication* (Version 1, 1995; version 2, 1998; version 3, 2000).

ITU-T Tutorial. (2004). *Objective perceptual assessment of video quality: Full reference television*.

Joch, A., Kossentini, F., Schwarz, H., Wiegand, T., & Sullivan, G.J. (2002). Performance comparison of video coding standards using Lagrangian coder control. In *IEEE 2002 International Conference on Image Processing*, Rochester, NY, USA (Vol. 2, pp. 501-504).

Jordan, N., & Schatz, R. (2006). Broadcast television services suited for mobile handheld devices. In *IEEE International Conference on Digital Telecommunications* (pp. 55-59).

Kornfeld, M. (2004). DVB-H - The emerging standard for mobile data communication. *IEEE Symposium on Consumer Electronics*. Reading, UK (pp. 193-198).

Kumar, A. (1) (2007). *Mobile TV: DVB-H, DMB, 3G systems and rich media applications*. Burlington, US: Focal Press

Lappalainen, V., Hallapuro, A., & Hämäläinen, T.D. (2003). Complexity of optimized H.26L video decoder implementation. *IEEE Transactions on Circuits and Systems for Video Technology*, *13*(7), 717-725.

List, P., Joch A., Lainema, J., Bjøntegaard, G., & Karczewicz, M. (2003). Adaptive deblocking filter. *IEEE Transactions on Circuits and Systems for Video Technology*, *13*(7), 614-619.

MacKay, D.J. (2003). *Information theory, inference and learning algorithms*. Cambridge, UK: Cambridge University Press.

Richardson, I.E.G. (2003). *H.264 and MPEG-4 video compression*. Chichester/Southern Gate, UK: John Wiley & Sons/The Atrium.

Schuster, M.G., & Katsaggelos, A. (1996). *Rate-distortion based video compression: Optimal video frame compression and object boundary encoding*. Boston, US: Kluwer Academic Publishers.

Sullivan, G.J., & Wiegand, T. (1998). Rate-distortion optimization for video compression. *IEEE Signal Processing Magazine*, *15*(6), 74-90.

Viterbi, A.J. (1991). Wireless digital communications: A view based on three lessons learned. *IEEE Communication Magazine*, *29*(9), 33-36.

Wang, Z., Bovik, A.C., Sheikh, H.R., & Simoncelli, E.P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600-612.

Wang, Z., Sheikh, H.R., & Bovik, A.C. (2003). Objective video quality assessment. In B. Furht (Ed.), *The handbook of video databases: Design and applications* (pp.1041-1078). Boca Raton, FL, US: CRC Press.

Wang, Z., Bovik, A.C., & Ligang, L. (2002). Why is image quality assessment so difficult? In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 4, pp. 3313-3316).

Watson, A.B., Hu, J., & McGowan, J.F. III. (2001). DVQ: A digital video quality metric based on human vision. *Journal of Electronic Imaging*, *10*(1), 22-29.

Wedi, T. (2003). Motion compensation in H.264/AVC. *IEEE Transactions on Circuits and Systems for Video Technology*, *13*(7), 560-576.

Wenger, S. (2003). H.264/AVC over IP. *IEEE Transactions on Circuits and Systems for Video Technology*, *13*(7), 645-656.

Wiegand, T., & Girod, B. (2001). *Multi-frame motion-compensated prediction for video transmission*. Boston, US: Kluwer Academic Publishers.

Wiegand, T., Schwarz, H., Joch, A., Kossentini, F., & Sullivan, G.J. (2003). Rate-constrained coder control and comparison of video coding standards. *IEEE Transactions on Circuits and Systems for Video Technology*, *13*(7), 688-703.

Wiegand, T., Sullivan, G.J., Bjontegaard, G., & Luthra, A. (2003). Overview of the H.264/AVC video coding standard. *IEEE Transactions on Circuits Systems for Video Technology*, *13*(7), 560-576.

Wiegand, T., Zhang, X., & Girod, B. (1999). Long-term memory motion-compensated prediction. *IEEE Transactions on Circuits and Systems for Video Technology*, *9*(1), 70-84.

Zovko-Cihlar, B., Grgic, S., & Grgic, M. (1998). Standard procedures for video coding in multimedia communications. In *Second International Conference on Telecommunications*, BIHTEL'98, Sarajevo, Bosnia, and Herzegovina (pp. 43-52).

KEY TERMS

Coding Algorithm: A technique composed of different instructions adapted to certain video characteristics and conditions and set in advance in order to improve video coding process.

Mobile Communication: Communication between mobile devices via wireless networks.

Multimedia: Integration of different self-contained media formats such as text, audio, still images, and video.

Test Sequences: Video sequences available in a variety of different formats used in video quality assessment process.

Video Compression: Process of bit rate reduction of the original video sequence to achieve a form more convenient for transmission over capacity-constrained video communication systems.

Video Format: Video structure composed of successive alternating frames characterized by certain vertical and horizontal resolution.

Video Quality Assessment Method: Method used for picture quality assessment after the process of coding and decoding.