Correlation between Objective and Subjective Methods Used for Video Quality Evaluation

Miroslav UHRINA[1], *Jan HLUBIK*[1], *Martin VACULIK*[1]

[1]Department of Telecommunications and Multimedia, Faculty of Electrical Engineering, University of Zilina, Univerzitna 8215/1 010 26, Zilina, Slovak republic

miroslav.uhrina@fel.uniza.sk, jan.hlubik@fel.uniza.sk, martin.vaculik@fel.uniza.sk

Abstract. *The article deals with the correlation between the objective and the subjective methods. In the first part the short characteristic of the MPEG compression standards is written. The second part focuses on the objective metrics and the third part on the subjective methods, which were used for testing. In the fourth part regression and correlation analysis are written. In the fifth part the measurements and the experimental results using the objective metrics and in the fifth part the measurements and the experimental results using the subjective methods are described. The last part deals with the correlation between the objective and the subjective methods, mapping and approximation.*

Keywords

Correlation, mapping, MPEG, objective assessment, subjective assessment.

1. Introduction

The demand of the multimedia services has rapidly increased in recent years that mean the broadcasting, transmission and receiving the video, audio and other data in one stream, called the multimedia stream. Because of this progress, it is important to measure the video quality as a part of the multimedia technology. The compression and the transmission link imperfection are two main factors that influence the video quality. Many new compression standards are being developed and most of them are based on the MPEG technology. There are many objective and subjective methods used for the video quality measuring and evaluating. For the practice it is very important that the objective methods correlate well with subjective methods.

2. MPEG Compression Standards

MPEG, which stands for Moving Picture Experts Group, is the name of a family of standards used for coding audio-visual information (e.g. movies, video, music) in a digital compressed format [1].

2.1. MPEG-2

MPEG-2 is one of the most used compression standards. It was approved in 1994. MPEG-2 is built on the MPEG-1 and his video coding scheme is a refinement of MPEG-1 standard. The advantage of the MPEG-2 standard is that it is suitable for coding both progressive and interlaced video. Many functionalities such as scalability were introduced. MPEG-2 also defines the Profiles and Levels. The Profile describes a degree of functionality whereas the Level describes resolution and bitrates. But not all Levels are supported at all Profiles. The most important application of the MPEG-2 standard is digital television broadcasting (DVB-T, DVB-S, DVB-C) but it also specifies the format of movies and other programs that are distributed on DVDs and similar disks [2], [3], [5].

2.2. MPEG-4 AVS

MPEG-4 Part 2 (Visual) called also MPEG-4 AVS is the combination of standard coding and object coding. It was approved in 1998 and improves on the popular MPEG-2 standard both in terms of compression efficiency and flexibility. It achieves this in two main ways, by making use of more advanced compression algorithms and by providing an extensive set of "tools" for coding digital media. Some of the key features that distinguish MPEG-4 AVS from previous coding standards include efficient compression of progressive and interlaced video sequences, coding of video objects (irregular-shaped regions of a video scene), support for effective transmission over networks, coding of still "texture" (image data), coding of animated visual objects such as 2D and 3D polygonal meshes, animated faces and animated human bodies, coding for specialist applications such as "studio" quality video [5].

2.3. MPEG-4 H.264/AVC

The latest and today most used compression standard designed for a wide range of applications, ranging from mobile video to HDTV, is MPEG Part 10, called also MPEG-4 H.264/AVC. Some of the feature enhancements in MPEG-4 H.264/AVC standard over the earlier codecs are:

- DCT algorithm works at 4×4 pixels instead of 8×8, but supports also 8×8,
- DCT is layered using Hadamard transforms,
- color sampling supported at 4:2:2 and 4:4:4,
- up to 12 bits per pixel are possible,
- motion compensation blocks are of variable sizes,
- arithmetic variable-length coding,
- built-in de-blocking filter and hinting mechanism,
- rate-distortion optimizer,
- weighted bi-directional prediction,
- redundant pictures,
- flexible macroblock ordering,
- direct mode for B-frames,
- multiple reference frames,
- sub-pixel motion compensation.

MPEG-4 H.264/AVC defines the Profiles and Levels, too, but its organization is much simpler than in MPEG-4 Part 2. There are only three Profiles currently defined (Baseline, Main, Extended) [2], [4], [6], [7].

3. Objective Methods

Objective assessment consists of the use of computational methods called "metrics" which produce values that score the video quality. They measure the physical characteristics of a video signal such as the signal amplitude, timing, signal to noise ratio. They are repeatable. Well-known and most used objective metrics are Peak Signal-to-Noise Ratio (PSNR), Video Quality Metric (VQM) and Structural Similarity Index (SSIM).

3.1. PSNR and MSE

The PSNR in decibels is defined as:

$$\mathrm{PSNR} = 10 \log \frac{m^2}{\mathrm{MSE}} [\mathrm{dB}], \quad (1)$$

where m is the maximum value that pixel can take (e. g. 255 for 8-bit image) and MSE (Mean Squared Error) is the mean of the squared differences between the gray-level values of pixels in two pictures or sequences I and $\widetilde{I}$:

$$\mathrm{MSE} = \frac{1}{TXY} \sum_t \sum_x \sum_y \left[I(t,x,y) - \widetilde{I}(t,x,y) \right]^2, \quad (2)$$

for pictures of size X $\times$ Y and T frames.

Technically, MSE measures image difference, whereas PSNR measures image fidelity. The biggest advantage of the PSNR metric is easy and fast computing [2].

3.2. SSIM

The SSIM metric measures three components - the luminance similarity, the contrast similarity and the structural similarity and combines them into one final value, which determines the quality of the test sequence (Fig. 1). This method differs from the methods described before, from which all are error based, by using the structural distortion measurement instead of the error one. It is due to the human vision system that is highly specialized in extracting structural information from the viewing field and it is not specialized in extracting the errors. Owing to this factor, SSIM metric achieves good correlation to subjective impression [8]. The results are in interval [0,1], where 0 is for the worst and 1 for the best quality.

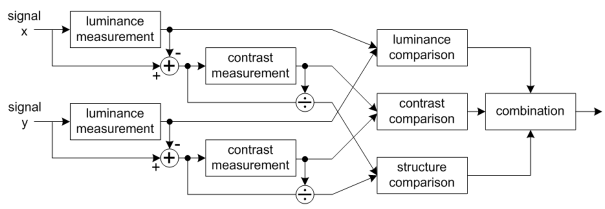

Fig. 1: The block diagram of SSIM metric.

3.3. VQM

The VQM metric computes the visibility of artefacts expressed in the DCT domain. Figure 2 shows the block diagram of this metric, which can be divided into 9 steps. The input of the metric is a pair of colour image sequences - the reference one and the test one.

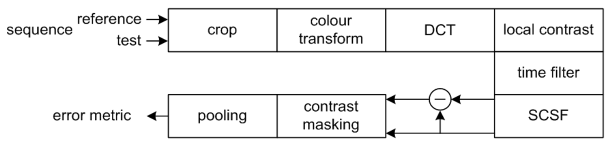

Fig. 2: The block diagram of VQM metric.

Both sequences are cropped, then converted from the input colour space to the YOZ colour space, then transformed to blocked DCT and afterwards converted to units of local contrast. In the next step the input sequences are subjected to temporal filtering, which implements the temporal part of the contrast sensitivity function.

The DCT coefficients, expressed in a local contrast form, are then converted to just-noticeable-differences (jnds) by dividing by their respective spatial thresholds. This implements the spatial part of the contrast sensitivity function. In the next step, after the conversion to jnds, the two sequences are subtracted to produce a difference sequence. In the following step the contrast masking operation to the difference sequence is performed.

Finally the masked differences are weighted and pooled over all dimensions to yield summary measures of visual error [9]. The output value of the VQM metric indicates the amount of distortion of the sequence - for no impairment the value equal to zero and for rising level of impairment the output value rises, too.

4. Subjective Methods

The subjective assessment is based on the use of human observers (people) who watch the sequences and score the video quality. It is the most reliable way to determine the video quality and should not be replaced with an objective assessment. The disadvantage of this method is that it is time consuming and human resources are needed. Owing to this fact, the objective methods are mostly preferred and used.

The well-known subjective methods are Double Stimulus Impairment Scale (DSIS) also known as Degradation Category Rating (DCR), Double Stimulus Continuous Quality Scale (DSCQS), Single Stimulus Continuous Quality Evaluation (SSCQE), Absolute Category Rating (ACR) also known as Single Stimulus (SS), Simultaneous Double Stimulus for Continuous Evaluation (SDSCE).

To achieve reliable results, at least 15 observers should be used. They should be non-experts in the sense that they are not directly concerned with television picture quality as a part of their normal work and they are not experienced assessors. Before the test, session assessors should be familiar with:

- the method of the assessment,
- the types of impairments,
- the grading scale,
- the sequence,
- the timing (time duration of the reference and the test sequence, the time duration of voting).

The presentation structure of the test session is shown in the Fig. 3. The source signal provides the reference sequence directly and input for the system under test. It should be of optimum quality (without failing). To obtain stable results the absence of defects in the reference sequence is crucial.

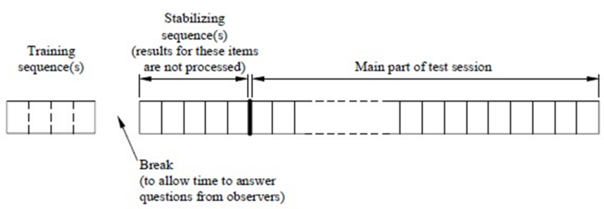

Fig. 3: The presentation structure of the test session.

The training sequences demonstrate the range and the type of the impairments that are assessed. They should be used with illustrating pictures different from those used in the test but of comparable sensitivity.

The whole session should last up to 30 minutes. At the beginning of the first session some sequences (from three to five) should be shown to stabilize the observers opinion. The data obtained from these presentations must not be taken into account in the results of the test. A random order should be used for the presentations but the test condition order should be arranged so that any effects on the grading of tiredness or adaptation are balanced out from session to session. Some of the presentations can be repeated from session to session to check coherence.

Finally after the test session the calculation of the mean score is done:

$$\overline{u_{jkr}} = \frac{1}{N}\sum_{i=1}^{N} u_{ijkr}, \quad (3)$$

where: u_{ijkrs}: score of observer i for test condition j, sequence k, repetition r; N is number of observers [2], [8], [11].

In our experiments DSIS, DSCQS and ACR methods were used.

4.1. The Double-Stimulus Impairment Scale Method (DSIS)

In this method two sequences with the same content are shown to the assessor - the unimpaired (the reference) one and the same impaired (the test one). The reference sequence is shown before the test one (Fig. 4) and the viewer knows which one is the reference and which one the test.

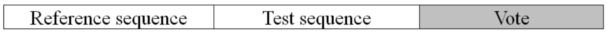

Fig. 4: The presentation structure of the test material in DSIS method.

After watching both sequences the assessor is asked to rate the second one, keeping in mind the first one. The five-grade impairment scale is used [2], [8], [11], [12]:

- imperceptible,
- perceptible, but not annoying,
- slightly annoying,
- 2 annoying,
- 1 very annoying.

4.2. The Double-Stimulus Continuous Quality-Scale Method (DSCQS)

By this method also two sequences with the same content are shown to the assessor - the unimpaired (the reference) sequence and the same sequence impaired (the test one) but the viewer is not informed which one is the reference and which one is the test (Fig. 5). The position of the reference sequence is changed in pseudo-random fashion.

Fig. 5: The presentation structure of the test material in DSCQS method.

The participant sees sequence A then sequence B twice and then is asked to rate both sequences. The grading scale is divided into five equal intervals [2], [8], [11], [12]:

- excellent (80–100),
- good (60–79),
- fair (40–59),
- poor (20–39),
- bad (10–19).

4.3. The Absolute Category Rating Method (ACR)

In this method only the impaired (the test) sequence is shown to the assessor (Fig. 6) so the viewer does not know of which quality is the reference sequence.

Fig. 6: The presentation structure of the test material in ACR method.

The assessor is asked to rate the quality of the test sequence based on the level of the quality he has in his opinion for it after watching it. The five-level grading scale is used [2], [9], [11], [12], [13]:

- excellent,
- good,
- fair,
- 2 poor
- 1 bad.

5. Statistical Analysis

Statistical analysis consists of:

- regression analysis,
- correlation analysis.

5.1. Regression Analysis

Regression analysis is one of the most commonly used statistical. Its main objective is to explore the relationship between a dependent variable and one or more independent variables (which are also called predictor or explanatory variables). The relationship is expressed with regression function. Regression should be divided into:

- linear: relationships can be readily described by straight lines,
- nonlinear regression: relationships cannot be readily described by straight lines.

Simple linear regression is the simplest and most used regression in praxis. It is appropriate when the following conditions are satisfied:

- the dependent variable Y has a linear relationship to the independent variable X,

- for each value of X, the probability distribution of Y has the same standard deviation σ,
- for any given value of X:
  - the Y values are independent, as indicated by a random pattern on the residual plot,
  - the Y values are roughly normally distributed.

Linear regression finds the straight line using least squares estimation procedure. If supposed Y is a dependent variable, and X is an independent variable, then the population regression line is expressed as:

$$Y = \beta_0 + \beta_1 X, \tag{4}$$

where β_0 is a constant, β_1 is the regression coefficient, X is the value of the independent variable, and Y is the value of the dependent variable.

Given a random sample of observations, the population regression line is estimated by:

$$\widehat{y} = b_0 + b_1 x, \tag{5}$$

where b_0 is a constant, b_1 is the regression coefficient, x is the value of the independent variable, and $\widehat{y}$ is the predicted value of the dependent variable.

The difference between the observed value of the dependent variable (y) and the predicted value ($\widehat{y}$) is called the residual (e). Each data point has one residual. Residual = Observed value - Predicted value:

$$e_i = y_i - \widehat{y}_i. \tag{6}$$

The least squares estimation procedure uses the criterion that the solution must give the smallest possible sum of squared deviations of residuals, i.e.:

$$\sum_{i=1}^{n} \left(y_i - \widehat{y}_i\right)^2 = \mathbf{min}. \tag{7}$$

The estimators for b_0 and b_1 are obtained by using calculus to find the values that minimize e_i. The derivatives of e_i with respect to b_0 and b_1 in turn are set equal to zero. This gives two equations in two unknowns called the normal equations. Solving the normal equations simultaneously for b_0 and b_1 gives the estimates of b_1 and b_0 as:

$$b_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \left(n\sum_{i=1}^{n} x_i\right)^2}, \tag{8}$$

$$b_0 = \frac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i \cdot \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \tag{9}$$

where b_0 is the constant in the regression equation called the intercept - the value of y where x equals to zero, b_1 is the regression coefficient called the slope - shows the lines steepness, y_i is the Y value of observation i, x_i is the X value of observation i, $\overline{x}$ is the mean of X, $\overline{y}$ is the mean of Y.

Simple non-linear correlation dependence is such dependence, in which between the two monitored quantitative characters is a non-linear relationship, which can be characterized by non-linear function. Nonlinear regression can be fairly easily converted by the transformation to a linear model. Some transformations are shown in Tab. 1. Letter F means type of function: 1-linear, 2-power, 3-logarithmic, 4-linear fractional (hyperbola), 5-quadratic (parabola).

Tab. 1: Nonlinear functions.

F	Regresion function	Transformed function	v_i	β_0	β_1	u_i
1	$y_i = \beta_0 e^{\beta_1 x_i}$	$\ln y_i = \ln \beta_0 + \beta_1 x_i$	$\ln y_i$	$\ln \beta_0$	β_1	x_i
2	$y_i = \beta_0 x_i^{\beta_1}$	$\ln y_i = \ln \beta_0 + \beta_1 \ln x_i$	$\ln y_i$	$\ln \beta_0$	β_1	$\ln x_i$
3	$y_i = \beta_0 + \beta_1 \ln x_i$	-	y_i	β_0	β_1	$\ln x_i$
4	$y_i = \beta_0 + \frac{\beta_1}{x_i}$	-	y_i	β_0	β_1	$\frac{1}{x_i}$
5	$y_i = \beta_0 + \beta_1 x_i^2$	-	y_i	β_0	β_1	x_i^2

The least squares estimation procedure can be used also for estimating parameters of nonlinear functions. First, the regression function should be transformed to linear function $v_i = \beta_0 + \beta_1 u_i$ using Tab. 1. The estimation of parameters β_0 and β_1 is done using least squares estimation procedure where variable Y is replaced by V and variable X is replaced by U. The final result of the estimated model is obtained from reverse transformation of modified function covered in the second column of the Tab. 1 to the original function in the first column of the table [14], [15].

5.2. Correlation Analysis

The correlation analysis is used for:

- determination of degree of dependence between the variables (correlation coefficient),
- providing the tightness assessment of regression function to empirical values (index of determination, residual standard deviation).

Index of determination I_{yx}^2 is the proportion of the variation in the dependent variable explained by the regression model, and is a measure of the goodness of fit of the model. It can range from 0 to 1, and is calculated as follows:

$$I_{yx}^2 = \frac{\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}, \quad (10)$$

where y_i is the Y value of observation i, $\widehat{y}_i$ is the predicted value of the dependent variable of observation i, $\overline{y}$ is the mean of Y.

If:

- an I_{yx}^2 of 0 means that the dependent variable cannot be predicted from the independent variable,
- an I_{yx}^2 of 1 means the dependent variable can be predicted without error from the independent variable,
- an I_{yx}^2 between 0 and 1 indicates the extent to which the dependent variable is predictable.

Residual standard deviation is the standard deviation of the residuals. It is calculated as follows:

$$adjs_{rez}^2 = \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{n-p}, \quad (11)$$

where y_i is the Y value of observation i, $\widehat{y}_i$ is the predicted value of the dependent variable of observation i, n is the size of statistics set, p is the number of parameters.

Correlation coefficient describes the direction and the magnitude of the relationship between two variables. It is calculated as follows:

$$r_{yx} = \frac{k_{xy}}{\sigma_x \sigma_y}, \quad (12)$$

where σ_x a σ_y are standard deviations of variables x a y, k_{xy} is covariance calculated as:

$$k_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}). \quad (13)$$

The value of a correlation coefficient ranges between −1 and 1. The greater the absolute value of a correlation coefficient, the stronger the linear relationship. The strongest linear relationship is indicated by a correlation coefficient of −1 or 1. The weakest linear relationship is indicated by a correlation coefficient equal to 0. A positive correlation means that if one variable gets bigger, the other variable tends to get bigger. A negative correlation means that if one variable gets bigger, the other variable tends to get smaller [14], [15].

6. Video Quality Assessment Using Objective Methods

In these experiments two test sequences were used - one with slow motion (called the "Train" sequence - Fig. 8) and one with dynamic scene (the "Football" sequence - Fig. 7). Both sequences were in the resolution of 720 × 576 px with 25 fps (frames per second). The length of these sequences was 220 frames, i.e. 8,8 seconds. The measurement procedure consists of four steps:

- First, both sequences were downloaded from [16] in the uncompressed format (*.yuv) and used as the reference sequences.
- Afterwards, they were encoded to different MPEG compression standards (MPEG-2, MPEG-4 AVS, MPEG-4 H.264/AVC) using the program FFmpeg [17] or x264 [18]. The parameters of the encoded sequences were set to Main Profile, Level 3. The target bitrates were in the range from 2 Mbps to 10 Mbps, changed in 1 Mbps step.
- Then, the sequences were decoded using the same program FFmpeg or x264 back to the format (*.yuv).
- At the end, the quality between these sequences and the reference (uncompressed) sequence was compared and evaluated. This was done using the MSU Measuring Tool version 2.7.3 [19] and PSNR, SSIM and VQM objective metrics were used. There are other tools that can be the video quality confirmed with (for instance Elecard Video Studio or Video Quality Studio), but the MSU Measuring tool is the best one and mostly used in this field of research.

The whole process of video quality assessment using objective methods is shown in Fig. 9.

Fig. 7: The "Football" sequence.

The Fig. 10, Fig. 11 and Fig. 12 show the results of the performed tests of the "Football" sequences using

Fig. 8: The "Train" sequence.

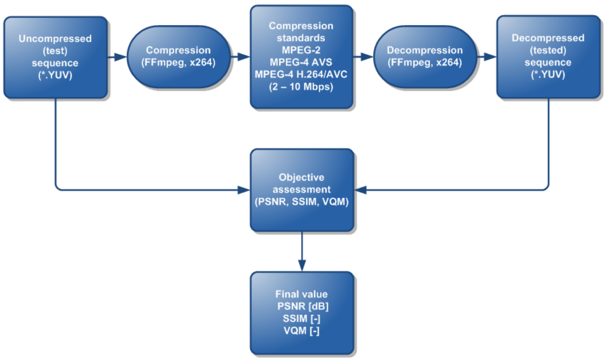

Fig. 9: The process of video quality assessment using objective methods.

PSNR, SSIM and VQM metrics and the Fig. 13, Fig. 14 and Fig. 15 show the results of the performed tests of the "Train" sequences using PSNR, SSIM and VQM metrics.

As it can be seen from the graphs, in both test sequences the MPEG-4 H.264/MPEG-4 compression standard can be considered as the best one. The quality of the MPEG-2 compression standard is similar to MPEG-4 AVS compression standard. In the dynamic "Football" sequence is the difference between the best one standard (MPEG-4 H.264/MPEG-4) and the rest ones a bit bigger than in the scene with the slow motion.

7. Video Quality Assessment Using Subjective Methods

In these experiments the same test sequences and the same measurement procedure as in previous (objective) measurements was used - just the target bitrates ranged from 3 Mbps to 15 Mbps, changed in 2 Mbps step - it was made due to the fact that the quality of the sequences in an 1 Mbps interval would be hardly recognizable by observers.

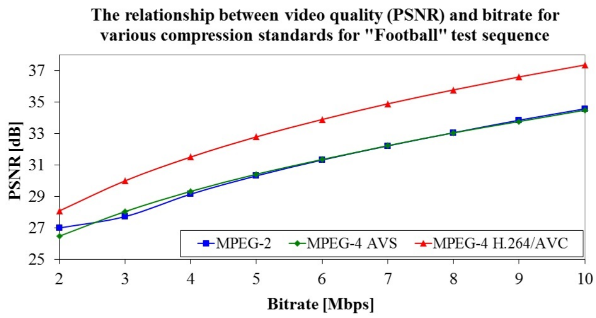

Fig. 10: The relationship between the video quality measured with SSIM metric and bitrate for MPEG-2, MPEG-4 AVS, MPEG-4 H.264/AVC compression standards for "Football" test sequence.

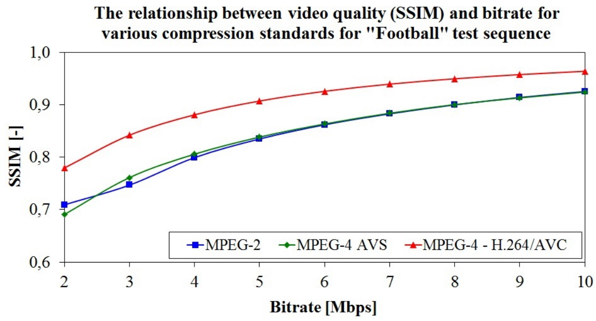

Fig. 11: The relationship between the video quality measured with SSIM metric and bitrate for MPEG-2, MPEG-4 AVS, MPEG-4 H.264/AVC compression standards for "Football" test sequence.

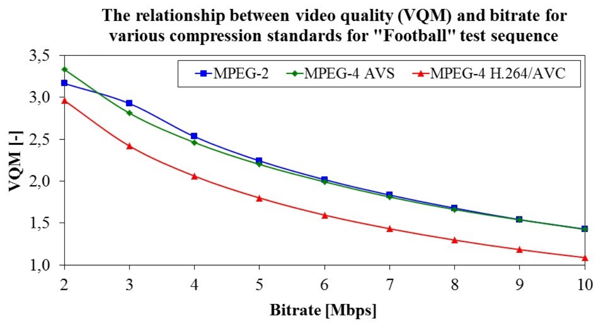

Fig. 12: The relationship between the video quality measured with VQM metric and bitrate for MPEG-2, MPEG-4 AVS, MPEG-4 H.264/AVC compression standards for "Football" test sequence.

The whole process of video quality assessment using subjective methods is shown in Fig. 16.

For evaluating the quality of the sequences the DSIS, DSCQS and ACR subjective methods were used. Different groups of assessors were used by these tests. The statistics of these observers is illustrated in Tab. 2.

The Fig. 17, Fig. 18 and Fig. 19 show the results of the performed tests of the "Football" sequences using DSIS, DSCQS and ACR methods and the Fig. 20,

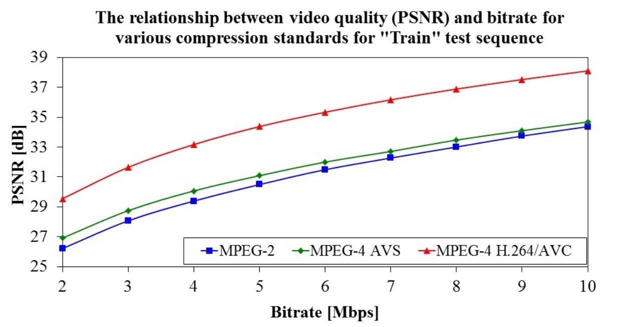

Fig. 13: The relationship between the video quality measured with PSNR metric and bitrate for MPEG-2, MPEG-4 AVS, MPEG-4 H.264/AVC compression standards for "Train" test sequence.

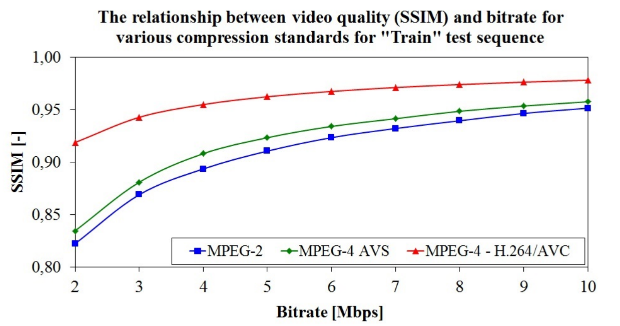

Fig. 14: The relationship between the video quality measured with SSIM metric and bitrate for MPEG-2, MPEG-4 AVS, MPEG-4 H.264/AVC compression standards for "Train" test sequence.

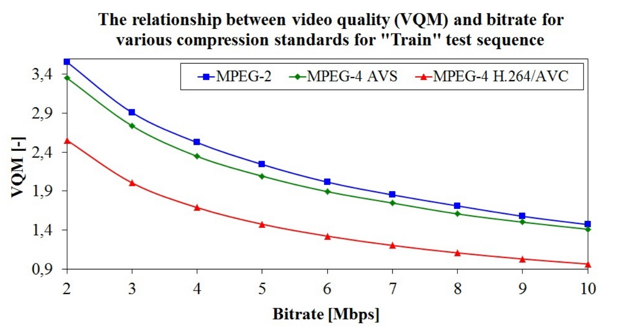

Fig. 15: The relationship between the video quality measured with VQM metric and bitrate for MPEG-2, MPEG-4 AVS, MPEG-4 H.264/AVC compression standards for "Train" test sequence.

Tab. 2: Statistics of the observers used by the measurements of the video quality using the subjective methods.

Method:		DSIS	DSCQS	ACR
Number of observers:		19	17	19
From that:	men:	16	15	10
	women:	3	2	9
Mean age of observers:		27,42105	21,35294	31,63158

Fig. 21 and Fig. 22 show the results of the performed tests of the "Train" sequences using DSIS, DSCQS and ACR methods.

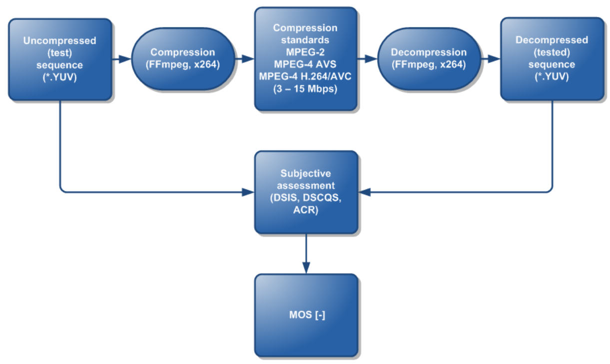

Fig. 16: The process of video quality assessment using subjective methods.

According to the graphs, in all sequences MPEG-4 H.264/MPEG-4 compression standard can be considered as the best one. In the "Train" sequence with slow motion the difference between the MPEG-4 H.264/AVC standard and the rest ones is bigger than in the dynamic "Football" sequence. This can be due to the fact that the assessors perceived the difference in the video quality between of the compression standards better in the scene with slow motion. The graphs also show that the assessors did not rate the video quality with extremities, which confirm the fact that people mostly do not like to give the extreme values.

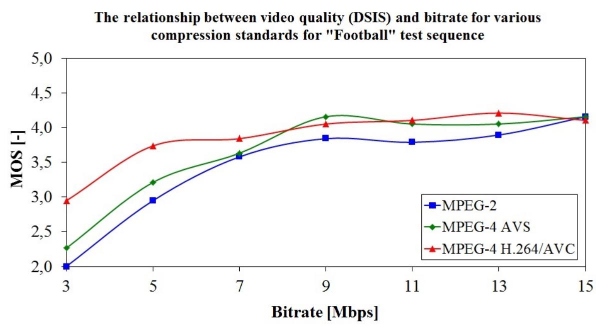

Fig. 17: The relationship between the video quality measured with DSIS metric and bitrate for MPEG-2, MPEG-4 AVS, MPEG-4 H.264/AVC compression standards for "Football" test sequence.

8. Correlation Between Objective And Subjective Methods

For both test sequences the correlation coefficients between particular objective and subjective methods were calculated. It was done using the formula (12) for test sequences coded to 3, 5, 7, 9 and 11 Mbps. These correlation coefficients for both test sequences are reported below in the Tab. 3 and Tab. 4.

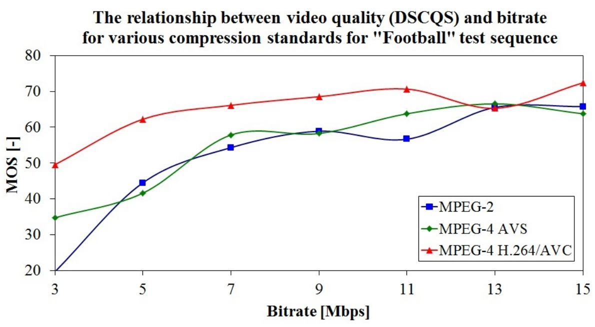

Fig. 18: The relationship between the video quality measured with DSCQS metric and bitrate for MPEG-2, MPEG-4 AVS, MPEG-4 H.264/AVC compression standards for "Football" test sequence.

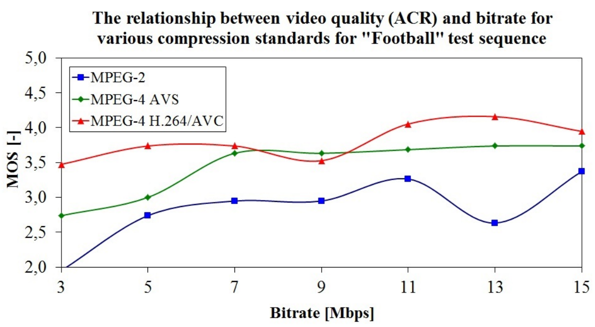

Fig. 19: The relationship between the video quality measured with ACR metric and bitrate for MPEG-2, MPEG-4 AVS, MPEG-4 H.264/AVC compression standards for "Football" test sequence.

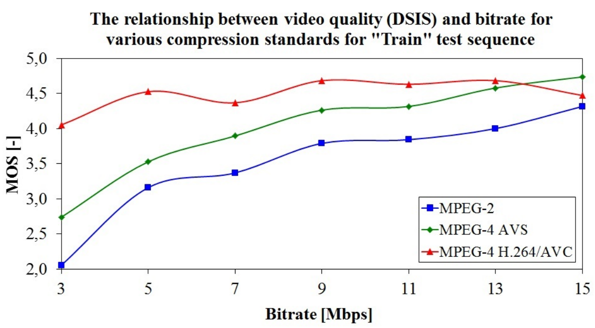

Fig. 20: The relationship between the video quality measured with DSIS metric and bitrate for MPEG-2, MPEG-4 AVS, MPEG-4 H.264/AVC compression standards for "Train" test sequence.

The results show that the best correlation is between the objective metric SSIM and the subjective method DSIS. It is due to the fact that in these methods the process of video quality evaluation is very similar. All these methods belong to full reference methods; it means that they compare the reference sequence with the test sequence.

Then the mapping of values of the objective SSIM method and subjective DSIS method for both sequences was done (Fig. 23 and Fig. 24). Due to mathe-

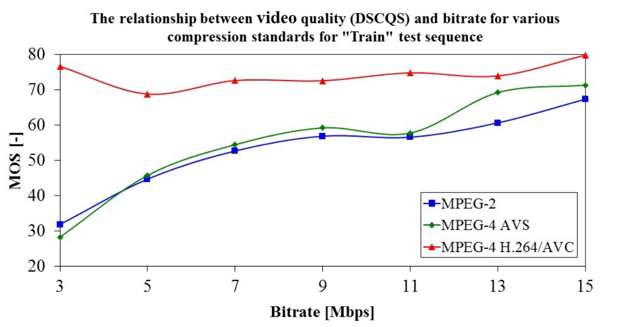

Fig. 21: The relationship between the video quality measured with DSCQS metric and bitrate for MPEG-2, MPEG-4 AVS, MPEG-4 H.264/AVC compression standards for "Train" test sequence.

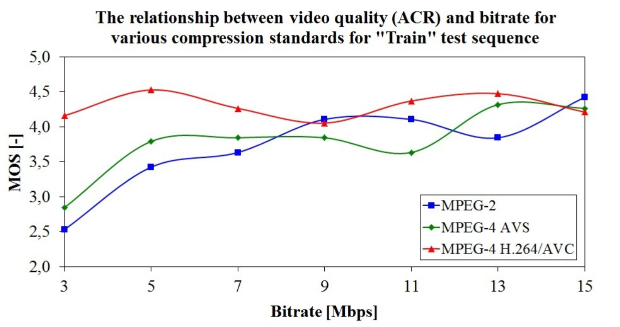

Fig. 22: The relationship between the video quality measured with ACR metric and bitrate for MPEG-2, MPEG-4 AVS, MPEG-4 H.264/AVC compression standards for "Train" test sequence.

Tab. 3: Correlation coefficients between the objective and subjective methods for the "Football" sequence.

MPEG-2			
	ACR	DSCQS	DSIS
PSNR	0,94316	0,91943	0,95370
SSIM	0,97218	0,97008	0,98693
VQM	-0,96546	-0,95652	-0,97906
MPEG-4 AVS			
	ACR	DSCQS	DSIS
PSNR	0,93820	0,96692	0,96383
SSIM	0,95442	0,96646	0,98742
VQM	-0,95177	-0,96905	-0,98234
MPEG-4 H.264/AVC			
	ACR	DSCQS	DSIS
PSNR	0,64271	0,95928	0,93910
SSIM	0,61383	0,99503	0,98387
VQM	-0,62563	-0,98432	-0,96915

matical description of these values, the approximation had to be done. Four different regression functions were tested. The fidelity of these functions was determined by the index of determination and residual standard deviation using formula (10) and (11), and recorded in Tab. 5.

As could be seen from the table, for both test sequences the exponential regression function had the

Tab. 4: Correlation coefficients between the objective and subjective methods for the "Train" sequence.

MPEG-2			
	ACR	DSCQS	DSIS
PSNR	0,96999	0,96595	0,96036
SSIM	0,98963	0,99205	0,98832
VQM	-0,98587	-0,98684	-0,98178
MPEG-4 AVS			
	ACR	DSCQS	DSIS
PSNR	0,71390	0,95253	0,98508
SSIM	0,83609	0,96646	0,99662
VQM	-0,79090	-0,98023	-0,99680
MPEG-4 H.264/AVC			
	ACR	DSCQS	DSIS
PSNR	0,00286	-0,14760	0,86850
SSIM	0,07088	-0,29386	0,89497
VQM	-0,04309	0,24617	-0,88833

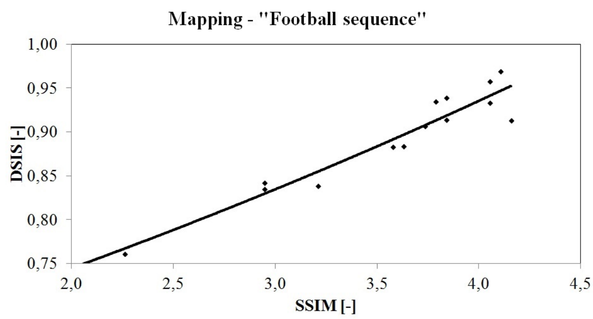

Fig. 23: Mapping of objective SSIM method into subjective DSIS method for the "Football" sequence.

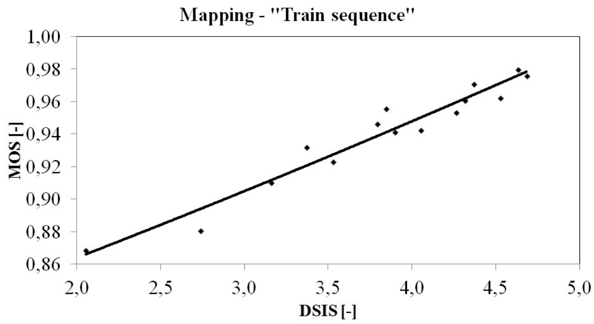

Fig. 24: Mapping of objective SSIM method into subjective DSIS method for the "Football" sequence.

best fidelity. The equations expressing these regression functions are:

- for the "Football" sequence: $y = 0,5924e^{0,1142x}$,
- for the "Train" sequence: $y = 0,7873e^{0,0464x}$.

Tab. 5: Indexes of determination and residual standard deviations of regression functions.

Test sequence	**Regression function**	**Index of determination**	**Residual standard deviation**
Football	Linear	0,9360	0,01763
	Logarithmic	0,9219	0,01947
	Exponential	0,9449	0,01759
	Power	0,9369	0,01859
Train	Linear	0,9443	0,00803
	Logarithmic	0,9299	0,00901
	Exponential	0,9444	0,00810
	Power	0,9339	0,00872

From these equations the critical (acceptable) boundary of degradation of the video signal for objective method SSIM according to subjective DSIS method could be determined

- for the "Football" sequence the SSIM value should be minimum $0,83446$,
- for the "Train" sequence the SSIM value should be minimum $0,90489$,

according to DSIS values 3 which means that observer assessed the quality of the sequence as good.

9. Conclusion

This article deals with the correlation between the objective and subjective methods. First some theory of the MPEG compression standards, objective and subjective methods were mentioned. These methods were used for evaluating the video quality of selected MPEG compression standards. The results in both measurements show that the MPEG-4 H.264/AVC compression standard can be considered as the best one and that the quality of the MPEG-2 compression standard is similar to MPEG-4 AVS compression standard. From the measurements also results that in the dynamic "Football" sequence the difference between the best one standard (MPEG-4 H.264/MPEG-4) and the rest ones is a bit bigger than in the scene with the slow motion. After these measurements, the correlation coefficients between particular objective and subjective methods were calculated. This was performed for both test sequences. The results show that the best correlation is between the objective metric SSIM and the subjective method DSIS. Then the mapping of values of the objective SSIM method and subjective DSIS method for both sequences was done. Afterwards the approximation of these values was performed. From approximation the equations expressing the regression functions were divided and then from these equations the critical (acceptable) boundary of degradation of the video signal for objective method SSIM according to subjective DSIS method was determined.

Acknowledgment

This contribution/publication is the result of the project implementation: Centre of excellence for systems and services of intelligent transport, ITMS 26220120050 supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] MPEG. *MPEG.ORG* [online]. 2013. Available at: http://www.mpeg.org.

[2] WINKLER, Stefan. *Digital Video Quality: Vision Models and Metrics*. Chichester: John Wiley and Sons Ltd., 2005. ISBN 978-0-470-02404-1.

[3] WATKINSON, John. *The MPEG Handbook: MPEG-1, MPEG-2, MPEG-4*. 2nd edition. Oxford: Elsevier Inc., 2004. ISBN 978-0240805788.

[4] WOOTON, Cliff. *A Practical Guide to Video and Audio Compression*. Oxford: Elsevier Inc., 2005. ISBN 0-240-80630-1.

[5] RICHARDSON, Iain E. G. *H.264 and MPEG-4 Video Compression*. Chichester: John Wiley and Sons Ltd, 2003. ISBN 0-470-84837-5.

[6] RICHARDSON, E.G Iain. *The H.264 Advanced Video Compression Standard*. 2nd edition. Chichester: John Wiley and Sons Ltd., 2003. ISBN 978-0-470-51692-8.

[7] MPEG-4 Part 10 AVC (H.264) Video Encoding. In: *Scientific Atlanta* [online]. 2005. Available at: http://www.scientificatlanta.com/products/customers/white-papers/7007887b.pdf.

[8] WU, H. R. and K. R. RAO. *Digital Video Image Quality and Perceptual Coding (Signal Processing and Communications)*. Boca Raton: CRC Press, 2006. ISBN 978-0824727772.

[9] LOKE, H. M., P. E. ONG, W. LIN, Z. LU and S. YAO. Comparison of Video Quality Metrics on Multimedia Videos. In: *IEEE International Conference on Image Processing, 2006*. Atlanta: IEEE, 2006, pp. 457–460. ISBN 1-4244-0480-0. DOI: 10.1109/ICIP.2006.312492.

[10] ITU-T J.249. *Perceptual video quality measurement techniques for digital cable television in the presence of a reduced reference*. Geneva: International Telecommunication Union ITU, 2010.

[11] ITU-R BT.500-12. *Methodology for the subjective assessment of the quality of television pictures*. Geneva: International Telecommunication Union ITU, 2009.

[12] Subjective Video Quality Assessment. *IRISA* [online]. 2003. Available at: http://www.irisa.fr/armor/lesmembres/Mohamed/Thesis/node2.html.

[13] ITU-T P.910. *Subjective video quality assessment methods for multimedia applications*. Geneva: International Telecommunication Union ITU, 2008.

[14] RAWLINGS, John O., Sastry G. PANTULA and David A. DICKEY. *Applied Regression Analysis: A Research Tool*. 2nd edition. Berlin: Springer Verlag, 1998. ISBN 0-387-98454-2.

[15] Linear Correlation Coefficient. *Stat Trek: Teach yourself statistics* [online]. 2013. Available at: http://stattrek.com/statistics/correlation.aspx?tutorial=ap.

[16] Video Quality Test Sequences. *Xiph.org* [online]. 2006. Available at: http://media.xiph.org/vqeg/TestSeqences/Reference/.

[17] Encoder FFmpeg. *FFmpeg* [online]. 2013. Available at: http://ffmpeg.org/download.html.

[18] Encoder x264. *VideoLAN Organisation* [online]. 2013. Available at: http://www.videolan.org/developers/x264.html.

[19] MSU Quality Measurement Tool: Download Page. *Everything about the data compression* [online]. 2011. Available at: http://compression.ru/video/quality_measure/vqmt_download_en.html#start.

About Authors

Miroslav UHRINA was born in 1984 in Zilina, Slovak Republic. He received his M.Sc. and Ph.D. degrees in Telecommunications at the Department of Telecommunications and Multimedia, Faculty of Electrical Engineering at the University of Zilina in 2008 and 2012, respectively. Nowadays, he is an assistant lecturer at the Department of Telecommunications and Multimedia, at the University of Zilina His research interests include audio and video compression, TV broadcasting (IPTV, DVB-T, DVB-H) and IP networks.

Jan HLUBIK was born in 1982 in Humenne, Slovak Republic. He received his M.Sc. and Ph.D. degrees at the University of Zilina, Faculty of Electrical Engineering, Slovak Republic, in 2000 and

2008, respectively. Nowadays, he is an assistant lecturer at the Department of Telecommunications and Multimedia, University of Zilina. Areas of his research include video and audio compression, video streaming, 3D modeling and animation, augmented reality and multimedia technologies. From 2008 his research interests cover system, services and terminals of DVB-H technology.

Martin VACULIK was born in 1951. He received his M.Sc. and Ph.D. in Telecommunications at the University of Zilina, Slovak Republic in 1976 and 1987 respectively. In 2001 he was habilitated as associate professor of the Faculty of Electrical Engineering at the University of Zilina in the field of Telecommunications. Currently he works as a head of Department of Telecommunications and Multimedia, University of Zilina. His interests cover switching and access networks, communication network architecture, audio and video applications.