Comparison of Video Quality Assessment Methods

Agata Jung

This thesis is presented as part of Degree of
Master of Science in Electrical Engineering

Blekinge Institute of Technology
August 2017

Blekinge Institute of Technology
Department of Applied Signal Processing
Supervisors: Doc. Benny Lövström, Dr. Muhammad Shahid
Examiner: Dr. Sven Johansson

ABSTRACT

Context: The newest standard in video coding High Efficiency Video Coding (HEVC) should have an appropriate coder to fully use its potential. There are a lot of video quality assessment methods. These methods are necessary to establish the quality of the video.

Objectives: This thesis is a comparison of video quality assessment methods. Objective is to find out which objective method is the most similar to the subjective method. Videos used in tests are encoded in the H.265/HEVC standard.

Methods: For testing MSE, PSNR, SSIM methods there is special software created in MATLAB. For VQM method downloaded software was used for testing.

Results and conclusions: For videos watched on mobile device: PSNR is the most similar to subjective metric. However for videos watched on television screen: VQM is the most similar to subjective metric.

Keywords: Video Quality Assessment, Video Quality Prediction, Video Compression, Video Quality Metrics

TABLE OF CONTEST

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Nowadays people are using more and more media communication and videos are one of the most significant part of media. People are watching many videos for example on the Internet and streaming media. The quantity of devices on which we can watch videos are also increasing. There is a huge improvement in technologies. Because of that there is always a need to improve the quality of media.

Nowadays the usage of digital images and videos has been growing each year. In 2014 the global mobile data traffic grew 69% comparing to a previous year and date traffic was 30 times more that the global Internet in 2000. The mobile video traffic achieved 50% of total mobile data traffic in 2012. In 2014 a smartphone usage grew for 45% and tablets increased to 74 million items. Each table is generating about 2,5 and each laptop 3,2 times more traffic than a smartphone. The mobile data traffic will grow to 57% from 2014 to 2019. The mobile network speed will increase from 1,7 Mbps in 2014 to 4,0 Mbps in 2019. The forecast is that more than the half of the devices will be a "smart: devices by the end of 2019 and tablets will generate double the traffic generated in 2014 by entire mobile network. The video will be 75% of the total mobile data traffic by 2019 [9].

The most important property of the video is quality. In this case quality means the highest perceptual quality with the smallest size of the video file. The video decoders are being improved because of a huge improvement in technologies. Technologies are developing very fast. In the figure 1 we can see how video coding standards were developed during previous years.

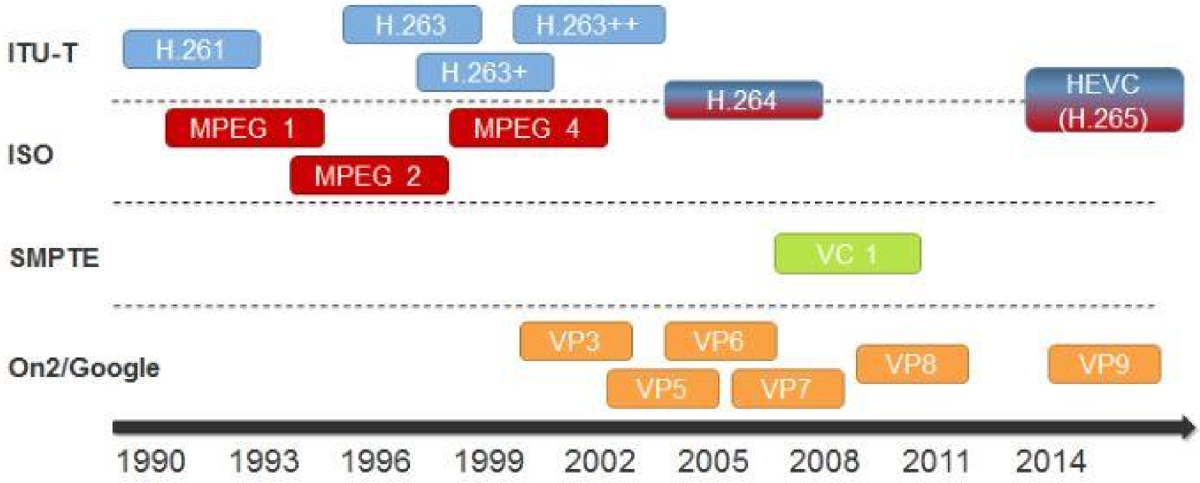

Figure 1. History of video coding standards [1]

As it is shown the video coding standards are developing really fast and the quality of videos is being improved as well. There are also many video quality assessment methods which are needed for testing video encoders. It is necessary to improve these methods to obtain the best quality of the video.

1.1 THESIS STRUCTURE

First chapter is an introduction of the thesis, motivation and objective of the thesis. Then, second chapter is an overview of related work. Next chapter is about problem which this thesis is focusing on. Fourth chapter is an overview of video compression – mostly about video coding standards. In the fifth chapter video quality metrics are described. Next chapter shows tests of video files for comparison of video quality metrics. The seventh chapter is analysis of the results. The last chapters are conclusion, future work and reference list.

1.2 MOTIVATION AND OBJECTIVE OF THE THESIS

The motivation is focusing on comparison of the video quality metrics. There is a need to provide the best quality of videos. This thesis is about comparing video quality assessment methods - subjective and objective tests with videos encoded using the newest standard H.265/HEVC. Comparison is needed to ensure the best quality of the video.

The objectives of the thesis are: analysis of video quality assessment methods, testing these methods and show which objective method is the most similar to the subjective method.

2 RELATED WORK

Paper [2] is about objective video quality metrics. Compared metrics are SSIM, MSE, VQM and MOS (Mean Opinion Score). Only two videos were tested with different bitrates (from 16 to 2500kbit/s). Subjective test was also conducted. Authors indicated that tests results not only depend on the methods used for testing but also on content. Objective quality metrics show a compatibility with subjective quality metrics.

Paper [10] is about a classification and review of no-reference image and video quality assessment methods. It is an overview of pixel-based, bitstream-based and hybrid methods. This article was useful to conduct comparisons for video quality assessment methods.

However paper [16] is about quality assessment for videos encoded in HEVC standard. Five objective metrics has been tested: PSNR, SSIM, UIQI, VFI and VSNR. There were also subjective tests carried on humans. This paper allows to compare results from objective tests to results from subjective tests. In this paper is concluded that SSIM is not good to estimate quality for video with random packet loss and low contrast.

3 PROBLEM DESCRIPTION

This thesis focuses on comparison of video quality assessment methods. There are several problems related to appropriate analysis. To evaluate video prediction completely it should be prepared video files which have a variety of bitrates and content. Below the list of problems which occurred during this thesis is presented:

- choosing appropriate video quality assessment methods to compare,
- finding suitable video files,
- creating video quality prediction software,
- appropriate analysis of results.

4 VIDEO COMPRESSION

Video compression is a technology where the size of a video file is minimized but the quality is kept high, preferably with no noticeable distortion. Compression allows more efficient storage and transmission of the data. Reduction in the file size can be accomplished with little effect on the visual quality or it could be no effect on visual quality. When the size of the file is reduced by raising the compression level for a compression technique, the quality of the video file can be a little bit affected. Removing and reducing unwanted data is necessarily for sending a digital video file over network more effectively [3].

The video compression is needed because uncompressed video file produces a huge amount of data. For example a resolution of *720x576* pixels, with a refresh rate of 25 fps and 8-bit color depth needs bandwidth: *720 x 576 x 25 x 8 + 2 x (360 x 576 x 25 x 8) = 1.66 Mb/s*. For High Definition Television 1920 x 1080 x 60 x 8 + 2 x (960 x 1080 x 60 x 8) = 1.99 Gb/s. In that case such data amount would cause extreme high computational requirements for the computer systems [4].

There are two types of compression: lossy and lossless. Lossy compression does not allow to recover the whole amount of the original data. It is used for data which includes a lot of redundancies and which is insensitive to losses. It is used in images, video or sound files. Compering to lossless compression, lossy compression provides higher compression ratios. Lossless compression provides a whole recovery of the original data. It is used in files where loss causes a lot of damage such as text and executable files [4].

4.1 OVERVIEW OF THE VIDEO CODING STANDARDS

There are different compression technologies and different standards are available. Standardization is necessarily to guarantee compatibility. They are a lot of video compression formats for example: H.120, H.261, MPEG-1, MPEG-2 Part 2, H.263, Motion JPEG, MPEG-4, H.264 and HEVC. HEVC is the latest and most efficient video compression standard [3].

These standards are developed by two organizations ISO/IEC and ITU-T. The ITU Telecommunication Standardization Sector (ITU-T) is one of the three sectors of the International Telecommunication Union (ITU). International Telegraph Union was formed in

1865. The International Telegraph and Telephone Consultative Committee was created in 1956, and in 1993 it was renamed ITU-T. It is responsible for coordinating standards for telecommunication. This organization developed following standards H.120, H.261, H.262, H.263, H.264, H.265. The second organization that is also developing standards for video coding is ISO/IEC. It is the International Standardization Organization (founded in 1947) and International Electrotechnical Commission (founded in 1906). ISO had joint committees with the IEC for developing standards and terminology in the field of electrical, electronic and related technologies. It had developed following standards MJPEG, Motion JPEG 2000, MPEG-1, MPEG-2 Part 2, MPEG-4 Part 2/ASP and Part 10/AVC, MPEG-H Part 2/HEVC. Some of the standards were developed by two organizations (ITU-T and ISO/IEC) [5].

4.1.1 H.120

H.120 was the first video codec standard developed in 1984. It is not used nowadays. The first version had scalar quantization, switch for quincunx sampling and variable-length coding. Second version was made in 1988. Comparing to the first version motion compensation and background prediction was added. Video bitrates in this standard are between 1544 kbit/s and 2048 kbit/s [6].

4.1.2 H.261

H.261 is the basis of the modern video compression. First version was developed in 1990. Scalar quantization, zig-zag scan, loop filter, variable-length coding these are key aspects for this standard. Color motion compensation with resolutions QCIF (176 x 144) and CIF (352 x 288) are used. Second version from 1993 had backward-compatible high-resolution graphics trick mode and video bitrates in this are between 64-2048 kbps [6].

4.1.3 MPEG-1

In 1993 MPEG-1 was developed. MPEG-1 is standard for lossy compression of video and a development of H.261. It has better quality comparing to H.261. Another features added to this standard are bi-directional motion prediction, half-pixel motion, slice-structured coding and quantization weighting matrices [6].

4.1.4 H.262/MPEG-2 PART 2

Developed in 1994 and is in wide use for DVD and high-definition DTV. It supports interlaced-scan pictures, increased DC quantization precision and it is not useful below 2-3 Mbps. The range is between 2-5 Mbps [6].

4.1.5 H.263

H.263 was developed in 1995. It has better quality comparing to earlier standards at all bitrates. Second and third version had a lot of new features. For example half-pel motion compensation, median motion vector prediction, increased motion vector range with picture extrapolation, improved compression efficiency and macroblock and block-level reference picture selection [6].

4.1.6 MOTION JPEG

Motion JPEG is made of a series of separate JPEG images. Every image in a video sequence has the same quality which is determined by the compression level. That level is chosen for specific network camera or video encoder. When the file size and image quality is decreasing, the compression level is increasing [3].

4.1.7 MPEG-4

This standard supports applications with low-bandwidth as well as applications that require high quality of the image. Another specification is that frame rate has no limitation and bandwidth is unlimited virtually [3].

4.1.8 H.264

H.264 is also known as MPEG-4 Part 10/AVC.H.264. In this standard we can achieve reduced video file for about 80% comparing to Motion JPEG and for 50% comparing to MPEG-4 maintaining the same image quality. In the same bitrate there is higher video quality comparing to previous standards [3].

4.1.9 HEVC/ H.265/ MPEG-H PART 2

H.264/MPEG-4 AVC was providing technology for digital video in almost every field that was on covered by H.262/MPEG-2 and it displaced previous standard over the existing application areas. It is widespread for many applications: broadcast HD TV signals over satellite, cable or terrestrial transmission systems; systems of video purchase and editing; cameras; mobile network and Internet video; Blu-ray Discs; security and real-time conferencing applications. HEVC was developed to solve all applications of H.264/MPEG-4 AVC and to focus on increasing video resolution and increasing use of a parallel processing architectures. High Efficiency Video Coding (HEVC) is the newest video coding standard of the ITU-T Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group. The most important feature of this standard is that compression is improved comparing to previous standards - 50% bitrate reduction for equal perceptual video quality. This standard includes text specifications as well as reference software source code, example of encoding and decoding HEVC video file [7].

In this thesis work all tested videos are encoded in this standard because it is the newest standard, it provides the best quality and the use of this coding standard is increasing.

4.1.10 HEVC ENCODER

For encode the video in this standard first picture should be divided into macroblocks. Next step is using intra-frame compression for reducing the spatial redundancy. Then using inter-frame compression the temporal redundancy is reduced. In the next step transformation and quantization for reduce data compression is used. In the last step using entropy coding the final redundancy and motion vectors transmission are reduced [8].

HEVC encoder generates a valid sequence of bits and it could be divided into few steps. Figure 2. shows the block diagram of the HEVC encoder. HEVC as well as previous standards since H.261 follows the classical block-based video coding approach. The coding algorithm is a joint of inter-picture prediction to use spatial and temporal statistical dependencies. For future use of spatial statistical dependencies code of the prediction of residual signals is transformed [7].

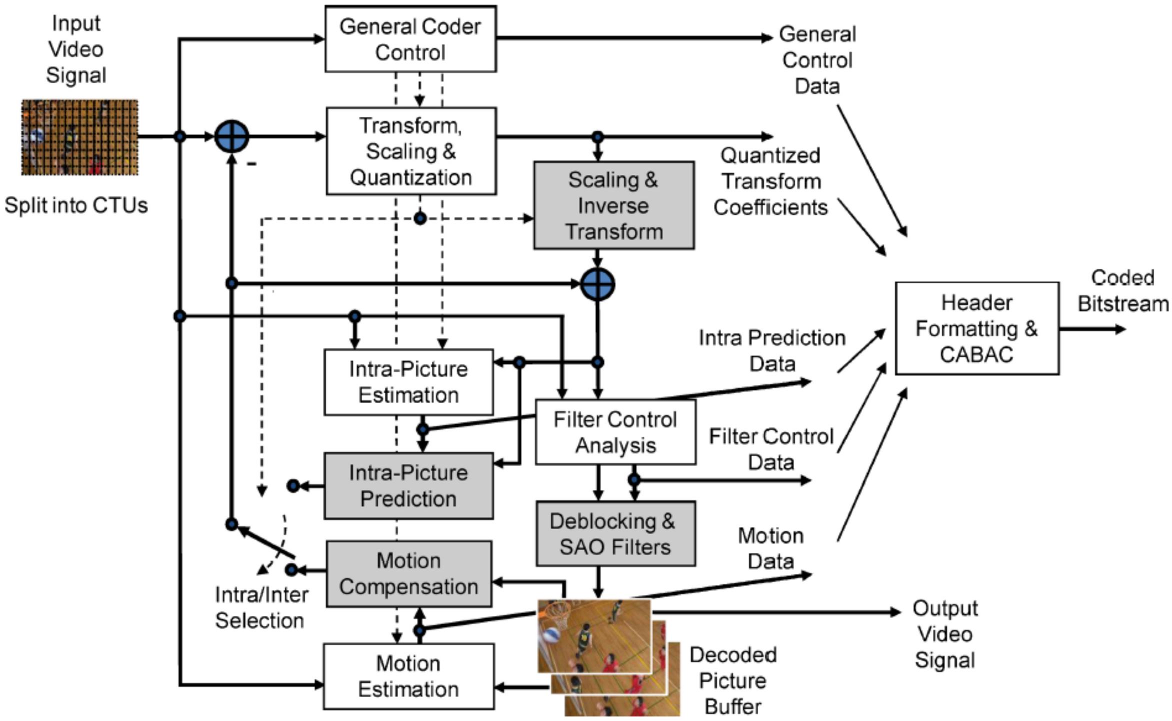

Figure 2 HEVC video encoder [7]

HEVC uses tristimulus YCbCr color space with 4:2:0 sampling for representing colors in a video. It divides a color representation into Y, Cb and Cr elements. The Y is responsible for brightness and it is also called luma. Cb and Cr are called chroma and are responsible for in which way color goes from gray to blue and red. In sampling structure, each chroma element has one fourth of the luma element. It is a typical used structure because the human visual system is more sensitive for luma than for chroma elements. Luma CB and two chroma CBs are creating a coding unit (CU). Each CU can be divided into Transform Units (TUs). Transform sizes are 32×32, 16×16, 8×8 and 4×4. Larger TUs are applicable for encoding stationary signals, smaller TUs are applicable for encoding impulsive signals. This transforms are based on Discrete Cosine and Discrete Sine Transforms (DCT and DST) [7].

In HEVC standard a picture is divided into Coding Tree Units (CTUs) which includes chroma and luma CTBs. The figure 3 shows how the picture is transforming into a code tree units. A picture of LxL samples is covered by luma CTBs and chroma CTBs cover L/2×L/2 samples of the two chroma elements. L value can be 16, 32 or 64. HEVC supports deferent size of CTBs which are used in different encoders for computational and memory requirements. For the high resolution of the video support of bigger CTBs are very beneficial in HEVC standard. The CTU is a processing unit that specifies decoding process. Luma CTB and two chroma CTBs with syntax create a CTU. CTBs can be used as CBs of can be divided into few CBs which is performed using tree structures. The CTU allows to divide the CBs into specific size which depends on the signal characteristics. The prediction can be intra-picture or inter-picture. In intra prediction the size of the block (PB) is equal to CB size for all block except the smallest one. The figure 4 shows possible options for dividing inter-picture-predicted CBs [7].

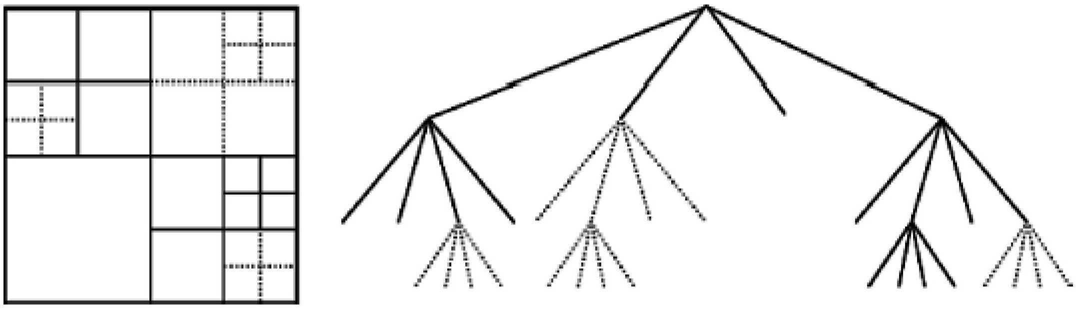

Figure 3 Fragmentation of a CTB into CBs [7]

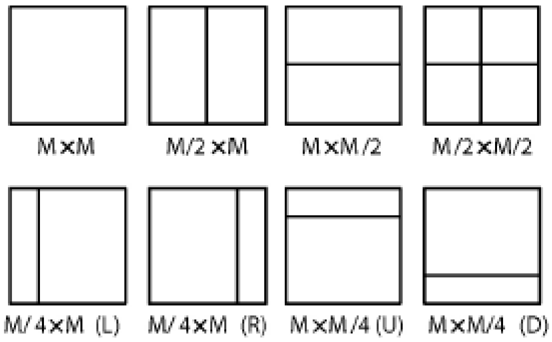

Figure 4 Modes of splitting a CB into PBs [7]

Intra prediction is responsible for recovering the information from blocks. In this standard there are 33 directional modes for intra prediction. The figure 5 is representing the possible prediction directions.

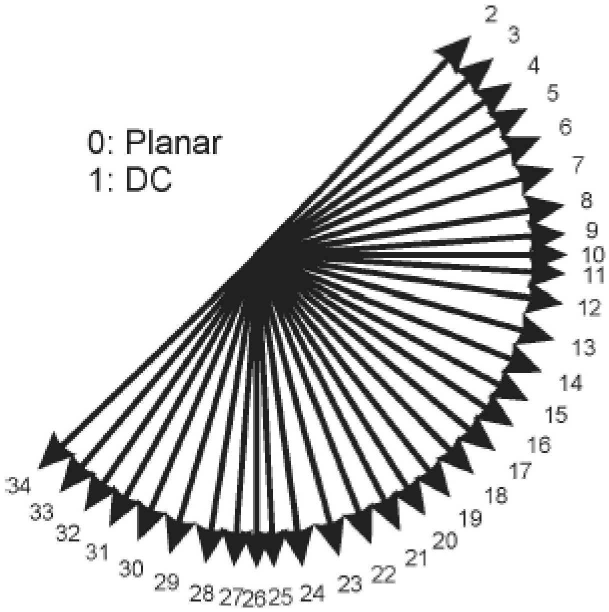

Figure 5 Modes and directions for intra-picture prediction [7]

In HEVC for entropy coding is using CABAC (Context adaptive binary arithmetic coding). However it is similar to previous standards, it has some improvements. For example in parallel-processing architectures the throughput speed, compression and memory reduce is improved [7].

5 VIDEO QUALITY ASSESSMENT

Generally video quality assessment can be divided into two main categories: objective video quality metrics and subjective video quality metric. Subjective metric is based on the tests conducted on a group of people who are judging the quality of the video by watching it. There are a few important factors to perform a subjective experiment for example careful planning, assessment method, selection of test material, viewing conditions or timing of the material. Subjective method is hard to conduct in real time, that is why objective metrics are created. Objective metrics are calculated by the computer. Some metrics are more and some are less similar to how the human perceives quality of the video.

Objective video quality metrics can be divided into three categories: full reference (FR), reduced reference (RR), and no-reference (NR). With FR method original image or video is available as a reference when distorted image or video is compared with the original one. In RR method we provide features about texture or other characteristic of the original image or video. The input in this method is the comparison between reduced information from original file and information from distorted file. NR method does not require access to original image or video but it use information in bitstream or search for information in pixel domain [10]. The figure 6 is presenting an overview of no-reference image and video quality assessment methods.

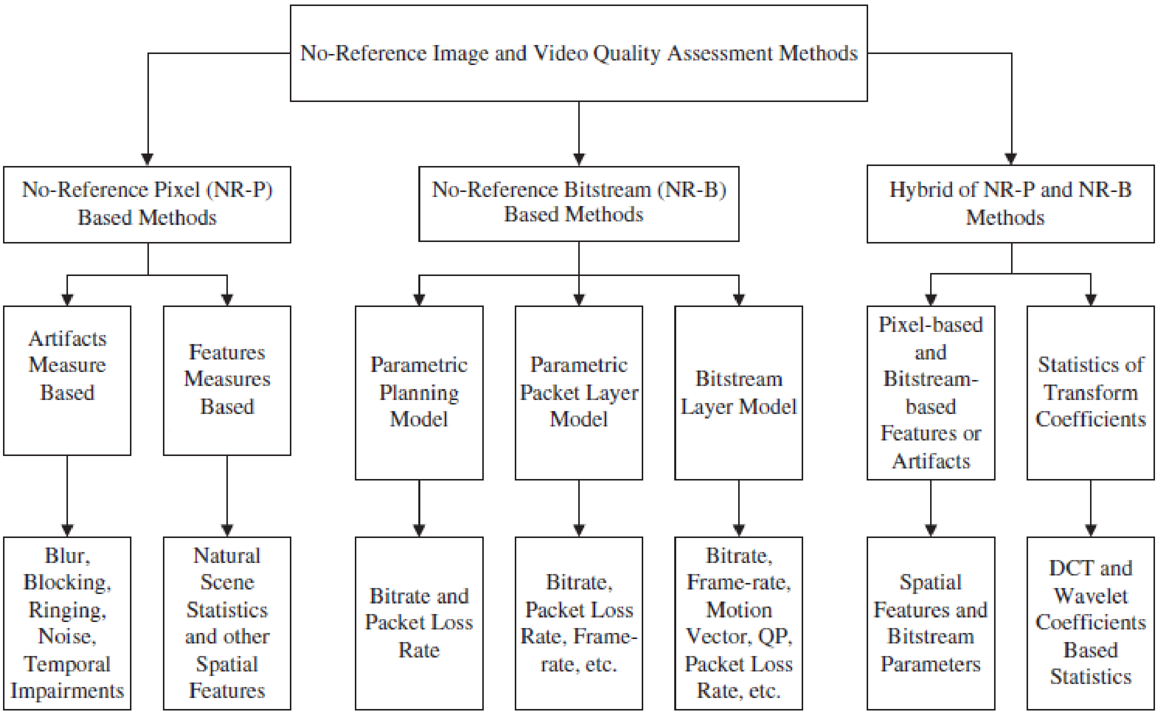

Figure 6 An overview of No-Reference image and video quality assessment methods [10]

In No-reference methods are divided into three categories: No-reference Pixel Based Methods, Bitstream Based Methods and Hybrid of Pixel and Bitstream Based Methods.

No-reference-pixel-based method has one relevant method to classify approaches. We should test it in reference of involvement of the artifacts that indicate degradation of the visual quality. Quality assessment is proportional to a number of these artifacts. This method requires more computational power than No-Reference-Bitstream-based methods.

No-Reference-Bitstream-based methods are simpler to compute than No-Reference-Pixel-based methods. It also can be computed without a full decoder. These methods are designed for specific coding technique and bitstream format. These methods are based on encoding information derived from bitstream, packet header or both. It is a suitable approach for network video applications.

Hybrid method is a combination of No-Reference-Pixel-Based method and No-Reference-Bitstream-Based method. This method gain computational simplicity of Pixel-Based method and robustness from Bitstream-Based method.

No-Reference-Bitstream-based methods are divided into three categories: parametric planning model, parametric packet-layer model and bitstream layer model. These categories are based on the amount of information that is used for processing. The figure 7 shows the visualizing classes of objective quality assessment methods [11].

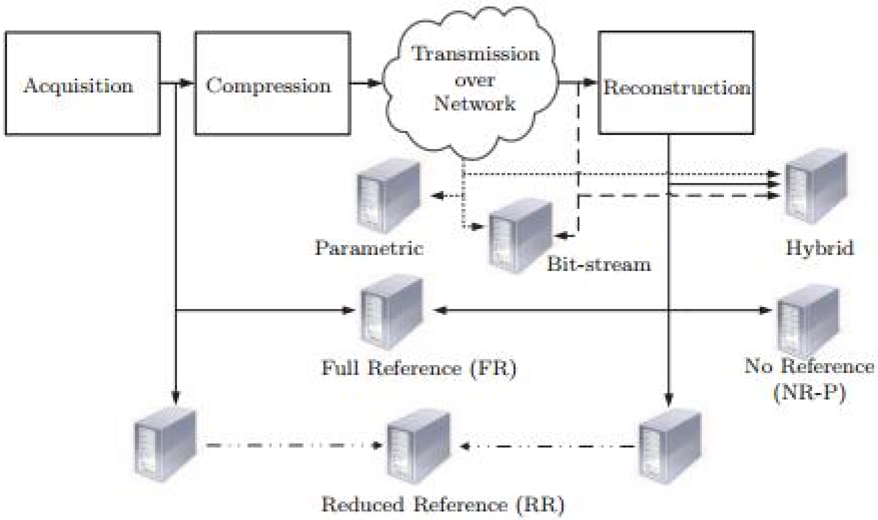

Figure 7 Visualizing classes of objective quality assessment methods [11]

Parametric models are divided into planning and packet-layer model. Planning model don't have access to bitstream but it utilize bitrate, codec type and packet loss rate. This model has lower complexity than packet layer model. This type has access to bitstream and can pull out such parameters as bitrate on sequence, frame level, frame rate and type or packet loss rate. Packet layer model is also known as QoS-based method [11].

Using Bitstream layer model we can have access to majority of the data used for video quality estimation. With this model it is possible to do any kind of analysis except the usage of the pixel data. The input information consists of the parameters from the packet header and payload. Parametric model uses parameters such as quantization parameter (QP), DCT coefficients of the coded video or pixel information. This model gives better performance but it is more complex [11].

With these methods we have a variety of video quality estimation. The bitstream-based methods have access to bitrate, frame rate, QP or motion vectors. These methods are less applicable because the quality estimation depends on the limited access to the bitstream. With the bitstream layer model the complexity can be flexible due to accuracy. Packet layer model do not need complex processing and decrypting data so it is more applicable in intermediate nodes [11].

The hybrid method is a combination of the No-reference-Pixel and No-reference-Bitstream methods. This method has the simplicity from the bitstream approach and accuracy of the quality estimation from the pixel-based approach. In this method we can avoid the difficulties [11].

5.1 VIDEO QUALITY METRICS

Different metrics are used to estimate a quality of the video. The most popular are the objective, full reference metrics peak signal-to-noise ratio (PSNR) and the mean-squared-error (MSE). These metrics are based on pixel-by-pixel comparison. There are also metrics such as SSIM, VSSIM (video version of SSIM), NTIAs VQM, PVQM or PEVQ. These metrics use reference information or original frame to estimate relation between this and decoded frame. The original frames are not available in streaming video, video telephony, MBMS or DVB-H. There are a lot of reference free quality metrics which are implemented at the decoder. These metrics focus on parameters like blur, blockiness or motion. They need processing and they are less applicable. However this thesis is focusing on reference metrics. All of these metrics are described below [12].

5.1.1 MSE – MEAN SQUARED ERROR

It is the average of the squared differences between the luminance values of corresponding pixels in two frames. It allows to evaluate the degree of image reconstruction by a decoder. Values are from 0 (no difference) to 65025 (maximum difference at 8 bit color depth). This factor should be as small as possible [12]. It is defined by:

$$MSE = \frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{j=1}^{M} ([f(i,j) - f'(i,j)])^2 \qquad (1)$$

5.1.2 PSNR – PEAK SIGNAL-TO-NOISE-RATIO

One of the most commonly used measures describes the ratio of peak to noise. This is a mean square error referenced to the maximum possible difference. Accepted values for 8 bit color: from 0 (max difference - red) to infinitive dB - (no difference - black) however for practical situations PSRN value is not bigger than 50 dB.. The bigger the PSNR, the better. It is computing using MSE factor. $\mathrm{MAX_I}$ is maximum pixel value of the image [12].

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

MSE and PSNR parameters are often used for ease of identification, but they do not reflect the perceptions of the recipient. In the figure below there are presented pictures with different distortions with MSE and PSNR values. As it is shown, for the best quality MSE should be the smallest and PSNR should be the biggest. According to metrics results picture with the best quality is picture *"c"* with JPEG Artifacts but the picture with the worse quality is picture *"d"* with brightness shift.

Sample picture
MSE = 0
PSNR = Inf

(a)

(b)

Gaussian Blur
MSE = 117, PSNR = 27.43 dB

(c)

JPEG Artifacts
MSE = 26, PSNR = 33.96 dB

(d)

Brightness shift
MSE = 3293, PSNR = 12.95 dB

(e)

Impulsive noise (pixels on and off),
MSE = 986, PSNR = 18.19 dB

(f)

Gaussian Noise,
MSE = 632, PSNR = 20.12 dB

(g)

Poissonian Noise,
MSE = 93, PSNR = 28.43 dB

Figure 8 Comparison of MSE and PSNR values [13]

5.1.3 SSIM – STRUCTURAL SIMILARITY INDEX MEASURE

It is a factor that shows the similarity of the two images. It uses structural distortions to evaluate the perceptual distortion base on the difference of the luminance values. This metric compares not the pixel values but the image elements perceived by the human. It better describes the image quality differences than for example PSNR. It includes three types of distortion: luminance, contrast and texture. The final SSIM index takes together all of these distortions. Acceptable values from -1 (maximum difference) to 1 (no difference). Higher value means better quality [12].

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \tag{3}$$

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1'} \tag{4}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2'} \tag{5}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + + C_3'} \tag{6}$$

$$C_3 = \frac{C_2}{2} \tag{7}$$

and when $\alpha = \beta = \gamma = 1$ then

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_x + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{8}$$

$$SSIM_{ij} = W_Y \cdot SSIM_{ij}^{Y} + W_{Cb} \cdot SSIM_{ij}^{Cb} + W_{Cr} \cdot SSIM_{ij}^{Cr} \tag{9}$$

where:

α, β, γ - importance coefficients

μ_x - mean value of luminance sample for signal x

μ_y - mean value of luminance sample for signal y

σ_x - deviation value of the luminance sample for image x

σ_y - deviation value of the luminance sample for image y

σ_{xy} - covariance of the luminance samples of two images x and y

C_1, C_2 – stabilizing constants

W_Y – the weight of Y

W_{Cb} - the weight of Cb

W_{Cr} - the weight of Cr

In the figure below there are presented pictures with different SSIM values. As we can see the bigger SSIM value, the better quality. The original (not distorted) image has SSIM value 1 and has the best quality. In other images the quality is decreasing as well as the metric value.

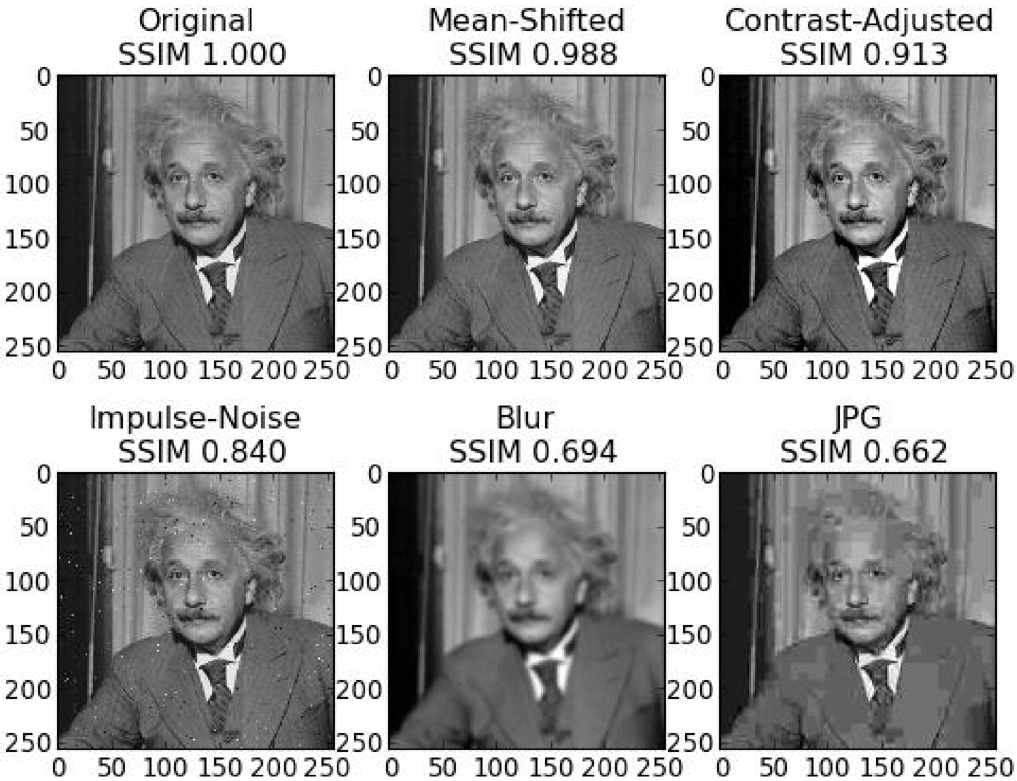

Figure 9 Presentation of different SSIM values [14]

5.1.4 VSSIM - VIDEO STRUCTURAL SIMILARITY

It is metric for quality evaluation for video and it is calculated as [11]:

$$Q_i = \frac{\sum_{j=1}^{R_S} w_{ij} SSIM_{ij}}{\sum_{j=1}^{R_S} w_{ij}} \tag{10}$$

Q_i - the quality index of the i-th frame

w_{ij} - weight value

R_s - quantity of sampling windows per video frame

VSSIM for the entire video of length N:

$$VSSIM = \frac{\sum_{i=1}^{N} W_i Q_i}{\sum_{i=1}^{N} W_i} \tag{11}$$

W_i - weight for the i-th frame based on global motion and w_{ij} .

5.1.5 PVQM – PERCEPTUAL VIDEO QUALITY METRIC

It estimates the perceptual quality and is a linear combination of distortion indicators: edginess, temporal decorrelation and color error [12].

5.1.6 VQM - VIDEO QUALITY MODEL

A reduced reference method that includes combination of objective parameters for estimate the perceptual effects of an extensive range of distortions such as blurring, block distortion, unnatural motion, noise or error blocks. The perceptual distortion is estimated using comparison functions for spatial and temporal distortions. Features use comparison function - Euclidean distance between two original and two processed streams, ratio comparison or log comparison function [12].

This metric is for evaluating image quality - the degree of distortion seen by human. VQM values are roughly correlated with subjective viewer ratings. The algorithm performs operations on the DCT cosine transform coefficients (local contrast calculations and comparison with the contrast perception function). Values: 0 means no difference (best quality), the higher the value, the greater the difference (worst quality) [12].

5.1.7 PEVQ – PRECEPTUAL EVALUATION OF VIDEO QUALITY

This metric estimates measures from the differences in luminance and chrominance between frames. Motion information is used in estimating the final measure. PEVQ was developed for low bit resolutions and rates (352×288 and 176×144) [11].

6 TESTS

6.1 RESULTS FROM THE FIRST PART

In the first part four metrics are compared MSE, PSNR, SSIM, Subjective. The comparison is carried out based on comparing bitrate to metric values and which objective metrics have most similar values to subjective method. Metric values for MSE, PSNR, SSIM are from MATLAB tests but results from Subjective are from [15]. Subjective tests were conducted on 54 people who watched videos on mobile devices.

To respectively perform the video quality prediction different video files should be tested. Sequences with different amount of motion, animations, color, frame rate, bitrates were tested. The original video file is be compared to the distorted video file. Videos are with spatial distortions, unexpected camera shakes and at the beginning a short delay followed by few stalls. For testing database from Laboratory for Image & Video Engineering from University of Texas at Austin was used [15]. In the table below there are parameters of videos which were used in tests.

Table 1 Video parameters

Video ID	Size [Pixel]	Frame rate [Frames/second]	Bitrate [kb/s]
1	1280x720	25	1648
2	1280x720	29	2523
3	1280x720	29	1313
4	484x360	29	407
5	640x360	25	432
6	1280x720	29	2999
7	640x360	24	303
8	1280x720	23	2553
9	1280x720	25	1341
10	1280x720	23	2438

All of these videos have different content. In the table below an overview of the videos used in tests is presented.

Table 2 Description of tested videos

ID	Photo	Content
1		People, animals, moving camera, different light.
2		Advertisement video, moving, shaking camera, communicates on the screen.
3		Advertisement video, people, zooming camera.
4		Newscast, still camera.
5		Tv show, still camera, one scene.
6		Basketball game, zooming camera.

7		Conference, two scenes.
8		Bike ride in forest.
9		Music video, many different scenes.
10		Surfing, fast moving action.

Software in MATLAB which is calculating SSIM, PSNR and MSE was created. MATLAB has build in functions for calculating these factors but it is only suitable for images. Because of that each metric value for each frame of the video was calculated. Then the average value of each video was calculated. Test results are showed in a table below. Yellow color indicates values with the best quality and blue color values with the worst quality.

Table 3 Video quality metrics values

Video ID	SSIM value	PSNR value	MSE value	Subjective DMOS
1	0,986	36,64	19,24	1,47
2	0,965	38,89	75,01	0,48
3	0,995	43,11	4,72	1,85
4	0,958	31,73	58,69	2,34
5	0,909	28,38	95,58	1,85
6	0,931	27,33	138	1,04
7	0,972	34,06	29,57	1,43
8	0,961	33,3	53,45	1,67
9	0,987	38,02	12,86	0,84
10	0,951	32,15	55,65	1,51

Results from Subjective tests are in difference mean opinion scores (DMOS) scale. DMOS is the difference between original video and distorted video. The scale is presented in a table below [20].

Table 4 DMOS scale

Value	Description
3,1-4,0	Most Users Dissatisfied
2,1-3,0	Many Users Dissatisfied
1,1-2,0	Some Users Satisfied
0,7-1,0	Most Users Satisfied
0,0-0,6	Very Satisfied

6.2 RESULTS FROM THE SECOND PART

In this part four metrics were compared. Values from SSIM, PSNR, VQM tests are compared to values from the Subjective method. Test results for SSIM, PSRN and Subjective method are from master thesis „Quality Assessment for HEVC Encoded Videos: Study of Transmission and Encoding Errors” from Sohaib Ahmed Siddiqui and Yousuf Hameed Ansari. In a table below there is an overview of video files used in tests [16].

Table 5 An overview of videos

ID	Photo	Content
1		People diving in a pool.
2		Woman drinking a coke.
3		Man playing on a harmonica
4		Airplane landing.
5		Students are walking.

Tested videos are also from Laboratory For Image And Video Engineering - The University Of Texas At Austin database. There are videos with no packet loss (p0 in name), with 1% packet loss (p1 in name) and 2% packet loss (p2 in name). Packet loss occurs when some data did not reach destination and is missing during the transmission. There are also two levels of QP (quantization parameter), 28 and 41. Quantization is compressing range of values to smaller range. Tests of a Subjective method were conducted in the study room at the Blekinge Institute of Technology. The environment was friendly and quiet with white light and background. 18 test subjects were selected for the experiments (13 males, 5 females at age between 20 to 40 years). For these tests a specific tool was developed by students from Blekinge Institute of Technology - perceptual video quality measure tool. The hardware for these tests was also prepared carefully to ensure watching videos without any technical problems. Videos were showed in random order. Then the mean value of all the subjects per video was calculated. Scale for all metrics in this part is presented below [16]:

Table 6 Scale for the subjective method

MOS	Quality
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Tests results are showed below [16]:

Table 7 Results for video ID1

NAME	PSNR	SSIM	Subjective
dv28p0.mp4	1,72	4,53	3,51
dv28p1.mp4	1,71	4,52	2,74
dv28p2.mp4	1,53	4,31	2,3
dv41p0.mp4	1,56	4,19	3,06
dv41p1.mp4	1,55	4,18	2,28
dv41p2.mp4	1,56	4,18	2,12

Table 8 Results for video ID2

NAME	PSNR	SSIM	Subjective
fc28p0.mp4	2,25	4,96	3,08
fc28p1.mp4	2,24	4,96	2,76
fc28p2.mp4	1,55	4,84	2,57
fc41p0.mp4	2,13	4,94	2,68
fc41p1.mp4	1,86	4,90	2,22
fc41p2.mp4	1,81	4,87	1,94

Table 9 Results for video ID3

NAME	PSNR	SSIM	Subjective
hc28p0.mp4	1,72	4,62	3,27
hc28p1.mp4	0,98	2,30	2,58
hc28p2.mp4	0,90	2,21	2,55
hc41p0.mp4	1,50	4,02	2,74
hc41p1.mp4	1,33	3,59	2,67
hc41p2.mp4	1,06	2,63	2,51

Table 10 Results for video ID4

NAME	PSNR	SSIM	Subjective
la28p0.mp4	2,03	4,89	3,06
la28p1.mp4	2,01	4,88	2,86
la28p2.mp4	1,53	4,49	2,83
la41p0.mp4	1,87	4,80	2,92
la41p1.mp4	1,75	4,71	2,43
la41p2.mp4	1,62	4,62	2,31

Table 11 Results for video ID5

NAME	PSNR	SSIM	Subjective
ss28p0.mp4	1,68	4,59	3,18
ss28p1.mp4	1,68	4,59	2,83
ss28p2.mp4	1,23	3,87	2,68
ss41p0.mp4	1,53	4,25	2,94
ss41p1.mp4	1,26	3,83	2,74
ss41p2.mp4	1	3,35	1,97

6.2.1 VQM METRIC TESTS

In this thesis, for the comparison of previous results - VQM metric was tested. Software for VQM was found on the Institute for telecommunication sciences the research laboratory of the National Telecommunications and Information Administration website. Command Line Video Quality Metric (CVQM) software tool was chosen for tests. CVQM is a Windows command line program that allows testing video files. CVQM performs video processing on two files - original video sequence and processed video sequence. Each video file has to go through two main steps. First - the calibration is run and then the requested model. For tests general model and frtime (full reference temporal registration and valid region estimation) calibration were chosen. Reasons for choosing these options are that other calibrations need median filtering because running tests can give different results and general model is the most commonly used model [19].

CVQM software does not support large video files and tests can only be conducted on the first 15 seconds of video. However all videos used in these tests are just 15 seconds long. Videos chosen for testing were the same as in thesis „Quality Assessment for HEVC Encoded Videos: Study of Transmission and Encoding Errors" from Sohaib Ahmed Siddiqui and Yousuf Hameed Ansari. In the table below is presenting test results.

Table 12 Values for VQM

NAME	VQM
dv28p0.mp4	0,687
dv28p1.mp4	0,652
dv28p2.mp4	0,624
dv41p0.mp4	0,633
dv41p1.mp4	0,618
dv41p2.mp4	0,61
fc28p0.mp4	0,187
fc28p1.mp4	0,18
fc28p2.mp4	0,149
fc41p0.mp4	0,173
fc41p1.mp4	0,153
fc41p2.mp4	0,131
hc28p0.mp4	0,785
hc28p1.mp4	0,769
hc28p2.mp4	0,766
hc41p0.mp4	0,779
hc41p1.mp4	0,77
hc41p2.mp4	0,741
la28p0.mp4	0,251
la28p1.mp4	0,247
la28p2.mp4	0,235
la41p0.mp4	0,24
la41p1.mp4	0,229
la41p2.mp4	0,223
ss28p0.mp4	0,888
ss28p1.mp4	0,884
ss28p2.mp4	0,86
ss41p0.mp4	0,87
ss41p1.mp4	0,856
ss41p2.mp4	0,85

7 ANALYSIS OF THE RESULTS

Analysis and an overview of the results from the first and second part are presented in this chapter.

7.1 ANALYSIS OF THE FIRST PART

According to metric values it is not clearly shown which video has the worst quality. PSNR and MSE values show that video with the worst quality is ID6, however SSIM indicates video ID 5. It is because MSE and PSNR are calculated based on the same criteria. All objective metrics indicate video ID 3 as video with the best quality.

These metrics values could be compared to bitrate. On the picture below there are presented four graphs representing metrics values for each video file and bitrate values also for each video file.

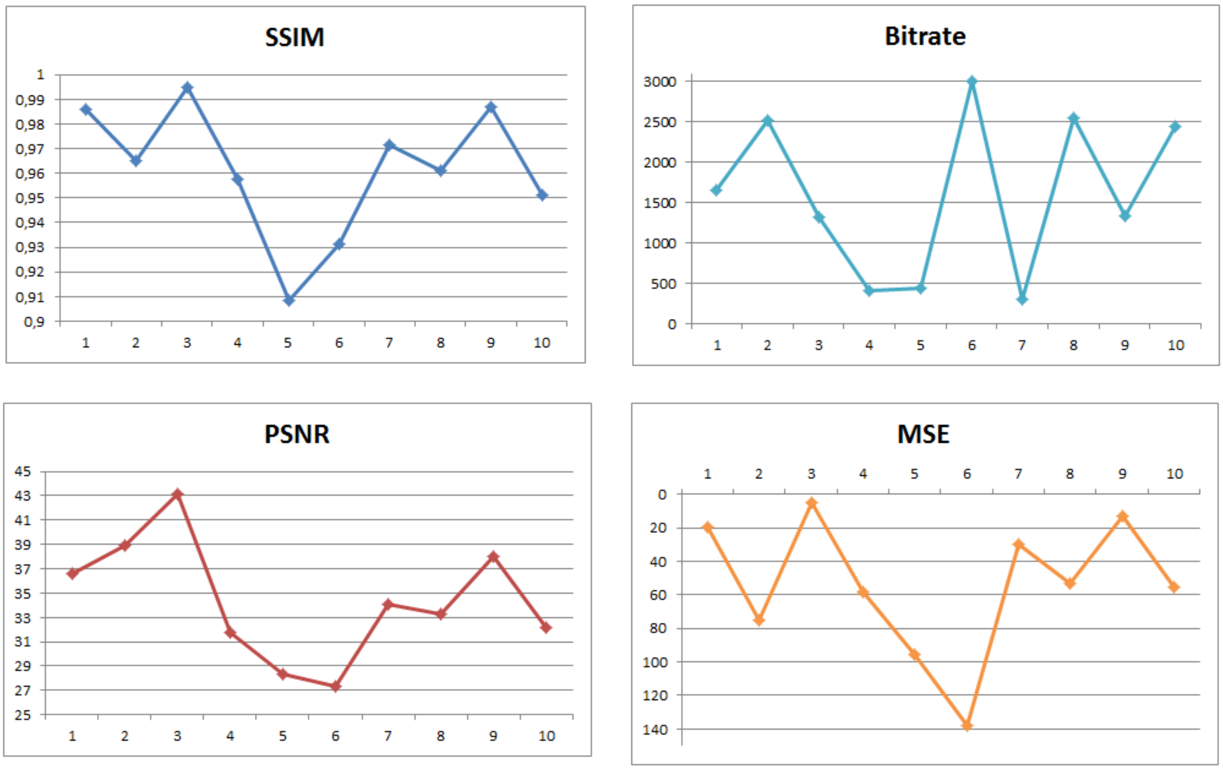

Figure 10 Values for metrics and bitrate

The scale for MSE graph is decreasing because it is now more simple to compare the shapes of the graphs. In MSE metric for the higher value the quality is decreasing. Another situation is with other metrics and that is why the scale has to decrease. Three graphs for metric values (MSE, SSIM, PSNR) should compare the shape with the graph of bitrate. It is visible that metrics are describing the quality of the video differently but there are some common points in which all metrics show the same quality. For example video ID 8 and ID 10 has approximately the same quality according to all graphs. The points are almost at the same value. The same situation is with video ID 3 and ID 9. All metrics do not have the best quality for the highest bitrate, however many points show that for the higher bitrate the better quality. The graph for subjective method is shown below.

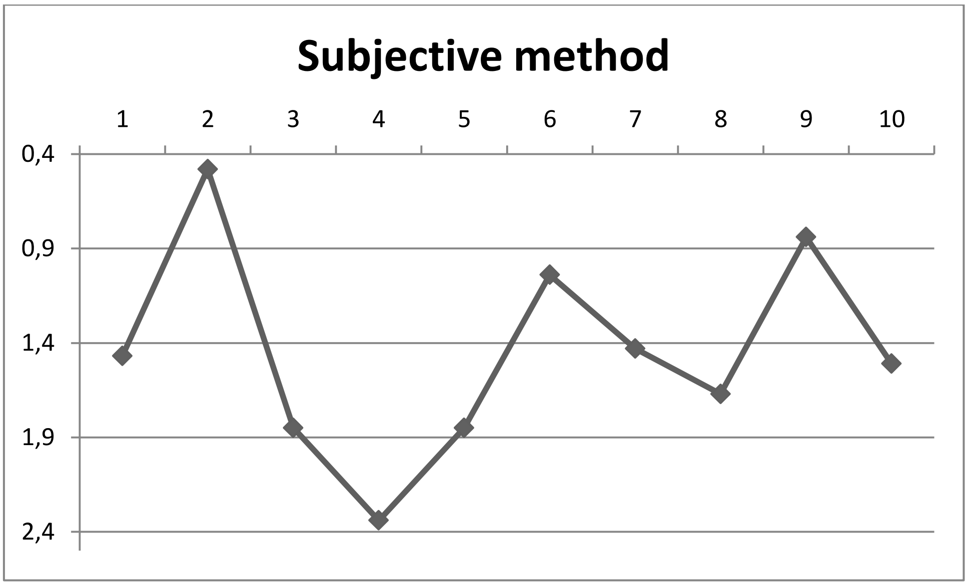

Figure 11 Graph for Subjective method

Comparing objective metrics MSE, PSNR, SSIM to subjective metric it is visible that the shape of the graph is similar to other metrics. Subjective results show more compatibility to bitrate than other metrics. There are visible differences between subjective and objective results. Comparing the shape of the graph, PSNR is the most similar metrics to subjective. Next it is SSIM and at least similar is MSE.

7.2 ANALYSIS THE SECOND PART

According to tests results, the quantization parameter (QP) should be the smallest for the best quality of the video. The same situations is with packet loss – it also should be the smallest for the best quality. According the subjective tests packet loss is more disturbing than quantization.

The results from testing VQM are similar to each other. Generally the bigger VQM value, the worse quality. VQM is in scale from 0 to infinity (0 is best quality), however other metrics are in scale 1-5 (5 is best quality). Because of that it is hard to compare it with results of other metrics. In this thesis different comparison was proposed. Below there are graphs showing values of metrics (SSIM, VQM, PSNR, Subjective) for each video.

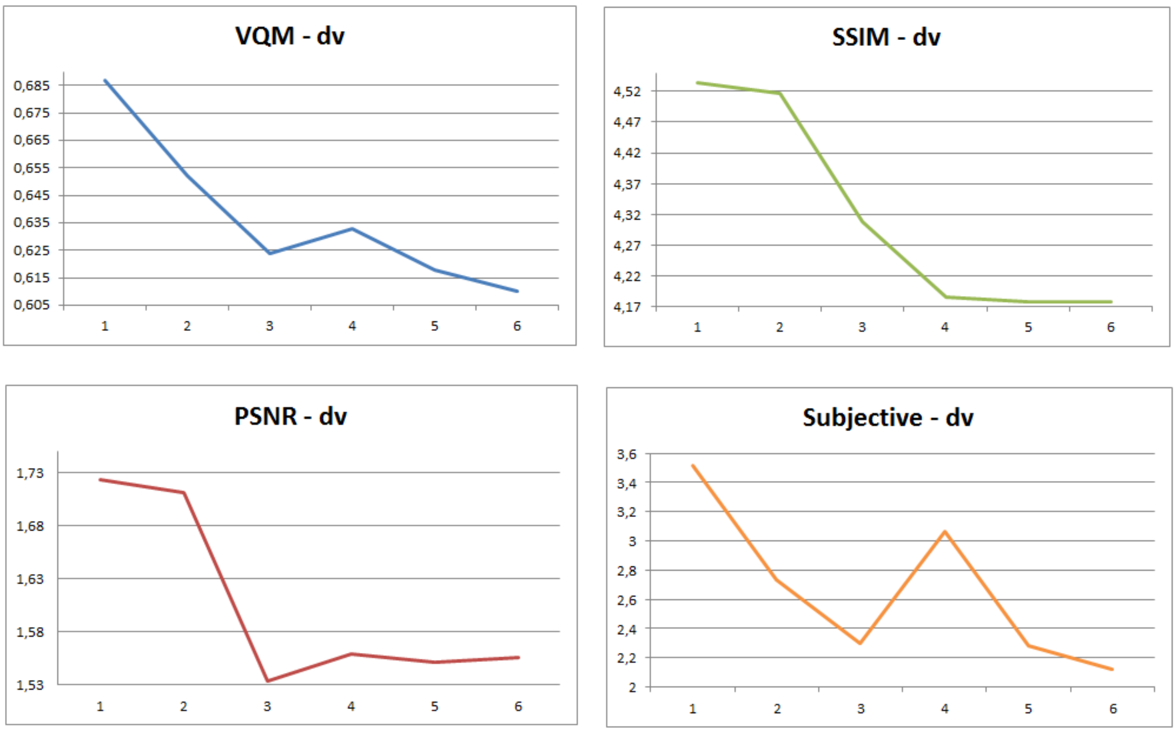

Figure 12 Graphs for dv video

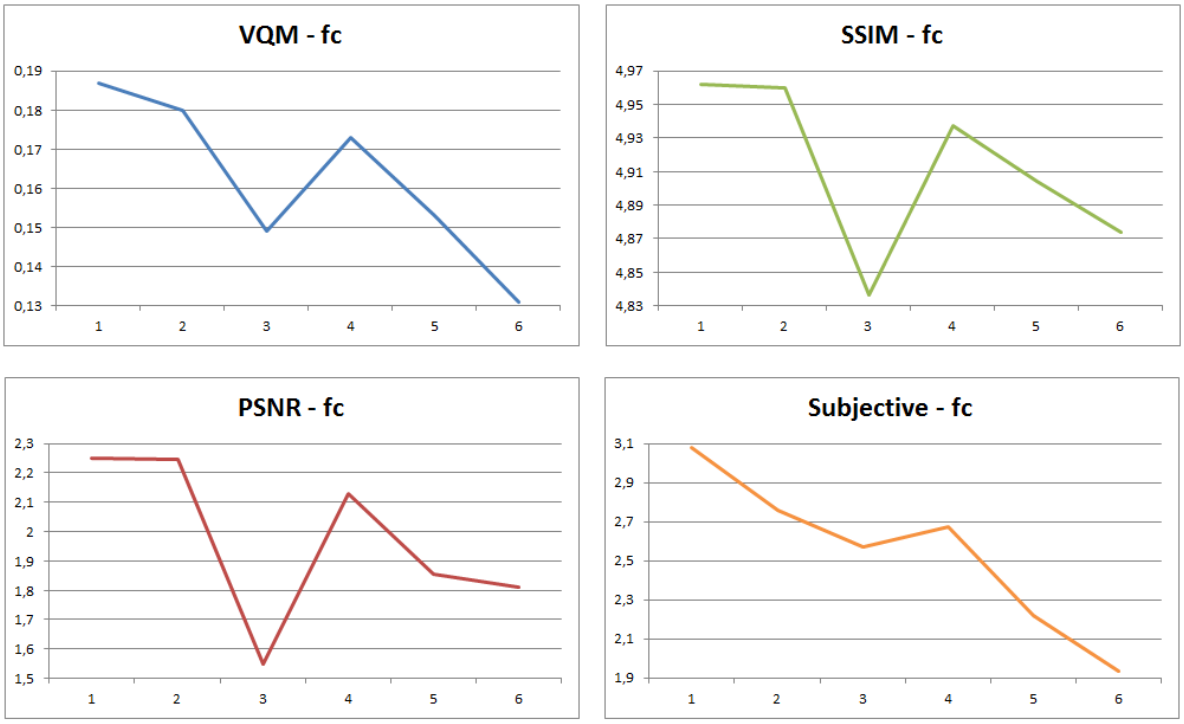

Figure 13 Graphs for fc video

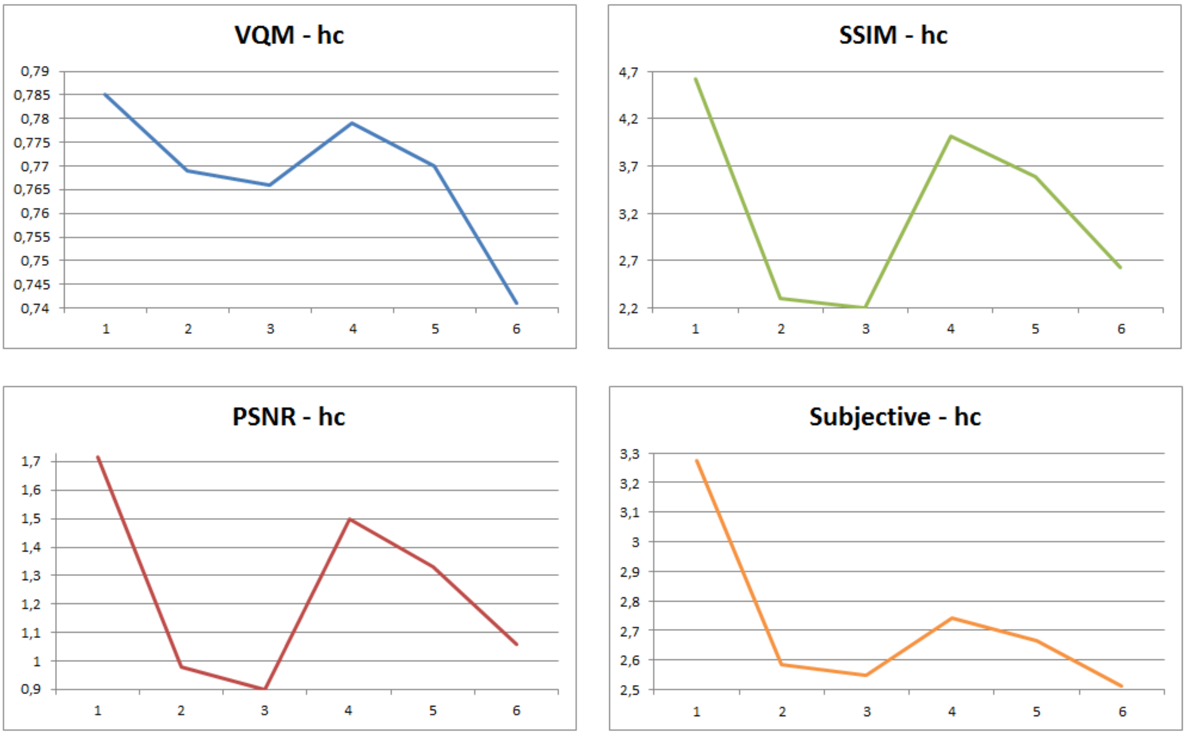

Figure 14 Graphs for hc video

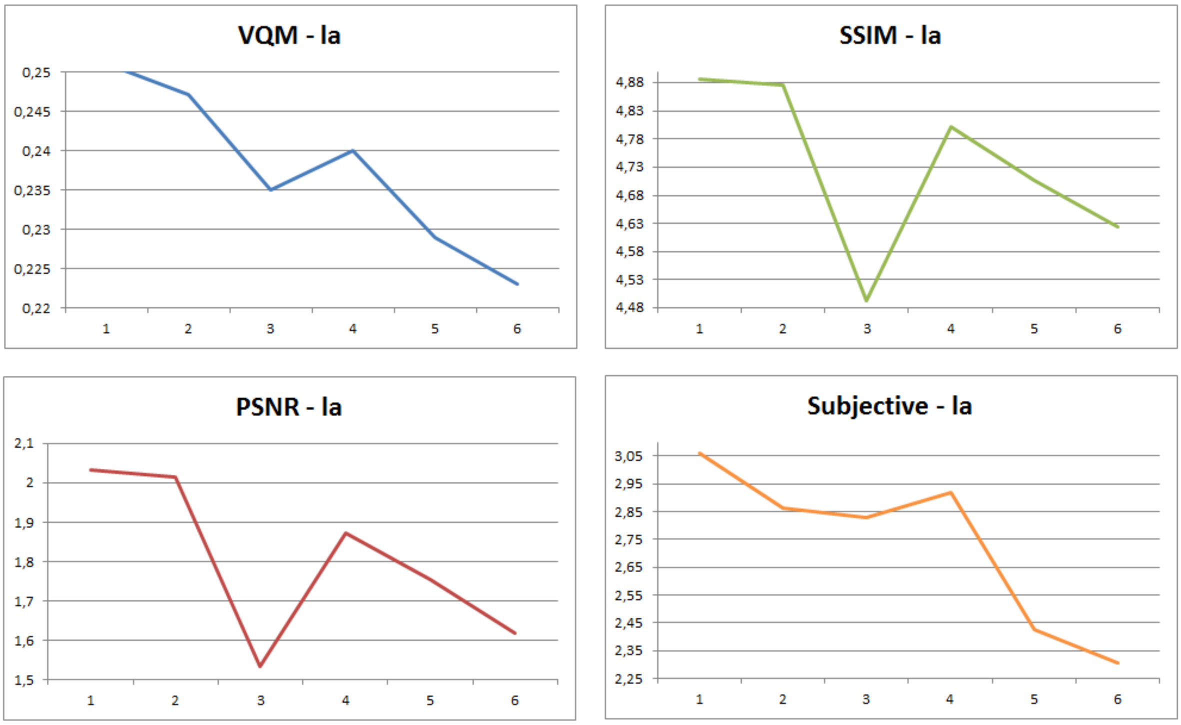

Figure 15 Graphs for la video

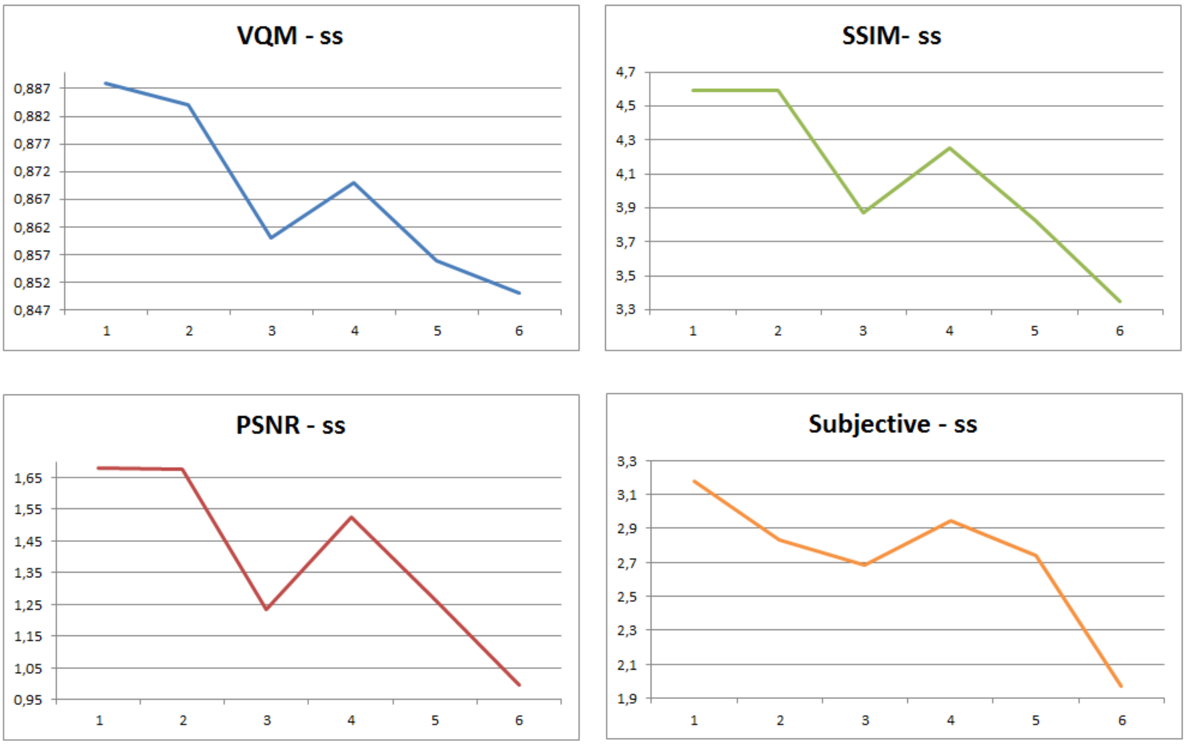

Figure 16 Graphs for ss video

Comparison of shapes of the graphs for each video is described below:

- for **dv** videos – metrics are not similar as in other videos; metric which is the most similar to subjective method is VQM, then PSNR is at least similar is SSIM,
- for **fc** videos – all shapes of graphs are similar to each other, however VQM is the most similar to subjective, then SSIM and PSNR,
- for **hc** videos – SSIM graph is very similar to PSNR graph, but metrics which graph is the most similar to subjective is VQM,
- for **la** videos – the situation is the same as for hc videos,
- for **ss** videos – all shapes of graphs are similar to each other, but VQM is the most similar to subjective method, then SSIM and PSNR.

According to the comparison described above, VQM is the most similar method to subjective method. SSIM and PSNR are equally similar to the subjective method.

8 CONCLUSIONS

There are many video quality assessment methods. We can divide them into subjective and objective methods. Subjective are conducted by a human perception and objective are conducted by a computer software which is calculating the video quality. All of these methods have theirs advantages and disadvantages.

Subjective method is time-consuming because there are several things need to be prepared: computer setup, specific software for showing videos, group of carefully selected people, software for collecting the results, non-disturbing environment. The results of the same videos could be different depending on group of subjects, computer setup and the environment. Regardless it all, the Subjective method reflect the most the human view of the quality of the video.

There are many objective methods. This thesis is mostly focused on four of them: MSE, PSNR, SSIM and VQM. MSE is easy and fast to calculate. PSNR is a little bit more complex and it take a little bit more time to calculate it. SSIM is the most complex metric, it takes much more time to calculate it. VQM cannot be compared to previous metrics in terms of duration of calculating because VQM was calculated using different software.

MSE, PSNR and SSIM may give results not compatible with bitrate. However most of the results are compatible but some of them will not be compatible with bitrate.

According to the first part of the tests where subjective method was conducted on videos watched on mobile devices, PSNR is the most similar method to subjective. Then SSIM and MSE is the least similar metric.

According to the second part of the tests and comparing PSNR, SSIM and VQM to subjective method, VQM is the most similar method. It reflects the most the human perception of the quality of the video. SSIM and PSNR gives very similar results to subjective method but they are still worse than VQM. In this part in subjective tests videos were watched on the television screen.

Depending on how much time we have for tests we have to select suitable method for testing video quality but in a realtime scenario only objective methods are possible.

9 FUTURE WORK

Future work for this thesis could be testing other video quality assessment methods. For example there is a new method developed by Netflix. Netflix is a multinational entertainment company formed in 1997 in USA. It offers streaming media and video on demand. From 2013 Netflix provides film, television production and online distribution [17].

Netflix has recently released their documentation and source code for their own video quality measure. Their method Video Multimethod Assessment Fusion (VMAF) is trying to reflect viewer's perception of the quality of the streamed video. Netflix is trying to provide the best perceptual quality of the streaming media. When a video is compressed improperly there are impairments known as compression artifacts like blocking, ringing or mosquito noise. Netflix is often comparing codec vendors to find the best quality of their streaming videos. They compare H.264/AVC, HEVC and VP9 standards and they are planning to experiment with next generation codecs. They are also experimenting on encoding optimization within existing standards [18].

The Netflix solution for encoding videos is distributed cloud-based media pipeline. With growth of video streaming service it was necessarily to build an encoding pipeline. This system is designed to guarantee high quality while it's more scalable. To achieve the best efficiency and minimize the impact of bad source deliveries or software bugs they apply quality monitoring at various points in pipeline. This solution is helpful for detect issues of video quality [18].

Netflix gather their own dataset for video quality assessment. There are publicly available databases but they are not very useful for other purposes because of the lack of diversity relevant to streaming services. Netflix has a wide range of collection of films and TV shows which are for example kids content, animation, fast-moving films, documentaries. These content deal with different distortions and different characteristics. They are using Transmission Control Protocol (TCP) and there are two types of artifacts which impact viewer's quality of experience (QoE): compression and scaling artifacts. Compression artifacts are caused by lossy compression and scaling artifacts are caused by downsampling and unsampling videos for lower bitrates [18].

Using PSNR, SSIM, FastSSIM, PSNR-HVS metrics the results are not satisfying for streaming content. Netflix invented a machine-learning based model for designing a metric that reflects more closely the human perception of video quality. The Video Multimethod Assessment Fusion (VMAF) predicts subjective video quality by combining many metrics. Each metric has its now advantages and disadvantages. Joining these metrics using machine-learning algorithm Support Vector Machine (SVM) which assigns weights to metrics gives as a result the final metric. This final metric has all advantages of many metrics and is more accurate to streaming media. The model is trained and tested [18].

Metrics used for creating SVM are Visual Information Fidelity (VIF), Detail Loss Metric (DLM) and Motion. VIF is basing on premise that quality is quality is supplementary the estimation of information accuracy loss. DLM is measuring loss of details which affect visibility and redundant distortions. Both VIF and DLM are image quality metrics. Motion is a measure for videos that measures differences between consecutive frames [18].

For future work this video quality assessment method from Netflix could be tested as well as the other method mentioned in this thesis like PVQM or PEVQ. Also no-reference video quality metrics could be tested. All of these methods could be compared to subjective method.

10 REFERENCE LIST

1. https://www.linkedin.com/pulse/brief-history-video-codecs-yoav-nativ, Page visited 22.03.2015
2. M. Vranješ, S. Rimac-Drlje, D. Žagar “Objective Video Quality Metrics”, University of Osijek, Faculty of Electrical Engineering
3. http://www.axis.com/global/en/learning/web-articles/technical-guide-to-network-video/video-compression, Page visited 02.04.2015
4. http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV0506/s0561282.pdf, Page visited 17.06.2015
5. http://web.stanford.edu/class/ee398b/handouts/lectures/03-StandardsH261H263.pdf, Page visited 5.11.2015
6. Gary J. Sullivan “Overview of International Video Coding Standards (preceding H.264/AVC)”, presentation ITU-T VICA Workshop 2015
7. G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard", IEEE Trans. Circuits and Systems for Video Technology, Vol. 22, No. 12, pp. 1649-1668, Dec. 2012.
8. https://sonnati.wordpress.com/2014/06/20/h265-part-i-technical-overview/, Page visited 02.02.2016
9. “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018,” Cisco. [Online]. Available: http://cisco.com/c/en/us/solutions/collateral/service-provider/visualnetworking-index-vni/white_paper_c11-520862.html. Page visited 28.07.2016
10. M. Shahid, A. Rossholm, B. Lövström, and H.-J. Zepernick, “No-reference image and video quality assessment: a classification and review of recent approaches,” EURASIP J. Image Video Process., vol. 2014, Aug. 2014.
11. Rossholm A. “ON ENHANCEMENT AND QUALITY ASSESSMENT OF AUDIO AND VIDEO IN COMMUNICATION SYSTEMS”, Doctoral Dissertation Series No. 2014:16, Blekinge Institute of Technology

12. Panasiuk J., Kowalczuk Ł. “ON BITSTREAM FEATURES BASED PERCEPTUAL QUALITY ESTIMATION OF HEVC CODED VIDEOS”, Master thesis, Blekinge Institute of Technology 2015
13. http://emanuelecolucci.com/2011/04/image-and-video-quality-assessment-part-one-mse-psnr/, Page visited 3.10.2016
14. http://mubeta06.github.io/python/sp/ssim.html, Page visited 7.12.2016
15. D. Ghadiyaram, A.C. Bovik, H. Yeganeh, R. Kordasiewicz, and M. Gallant, "LIVE Mobile Stall Video Database," Online: http://live.ece.utexas.edu/research/LIVEStallStudy/index.html, 16 Nov. 2016.
16. Sohaib Ahmed Siddiqui, Yousuf Hameed Ansari “Quality Assessment for HEVC Encoded Videos: Study of Transmission and Encoding Errors”, Master thesis Blekinge Institute of Technology 2016
17. https://en.wikipedia.org/wiki/Netflix, Page visited 11.01.2017
18. http://techblog.netflix.com/2016/06/toward-practical-perceptual-video.html, Page visited 13.01.2017
19. https://www.its.bldrdoc.gov/resources/video-quality-research/software.aspx, Page visited 16.05.2017
20. http://videoclarity.com/wpunderstandingjnddmospsnr/, Page visited 23.06.2017