A Review on Recent Approaches," Methods of Video Quality Assessment

amitesh singam[1] and Amitesh Kumar Singam[2]

[1]Affiliation not available
[2]Halmstad University

July 25, 2023

A Review on Recent Approaches,"Methods of Video Quality Assessment"

Amitesh Kumar Singam[a]

[a]Blekinge institute of Technology,

Abstract

Generally the evaluation of video quality is classified into two methods. They are subjective and objective analysis of video quality and Subjective analysis is conducted based on human perception since it is concerned with how video is perceived by a viewer or subject and expresses subjects opinion on a particular video sequence in comparison with its original video sequence. The subject has to vote for the video sequences under certain test environment conditions for example the ITU-Recommendations. Human perception is considered as the true judgment of video quality and precise measurement of perceptual quality but it is quite expensive and tedious in terms of time such as preparation, running and human resources. Objective quality assessment is therefore essential. Objective Video Quality Metric should be designed based on HVS (Human Visualizing System) characteristics. Some aspects of HVS like contrast, orientation sensitivity, spatial and temporal masking effects, frequency selectivity and color perception are incorporated in the design of objective quality metrics. Even though it is computationally very expensive and complex to design a quality metric with above aspects. It is useful for a wide range of applications if it correlates well with human perception. The impairments visibility which is related to video processing system is subjected to spatial and temporal properties of video content, since subjective analysis is quite expensive and time conservative method, objective metric has been developed considering HVS. In our research work we have done experiment on both subjective and objective analysis.

Keywords: VQM, NRVQM, RRVQM, FRVQM, MOS, H.264

1. Introduction

The true judges of video quality assessment are humans at end users of any video services field and the scientific evaluation based on test conditions of reconstructed or distorted video sequences rated by humans is called subjective quality assessment. However, in past subjective evaluation was time-consuming, expensive and inconvenient. It has to be done by specific test conditions based on special recommendations in order to produce reproducible and standard results. These reasons give rise to the need of some intelligent ways of automatically predicting the perceived quality that can be performed swiftly and economically. but these day due to huge demand and tough competition between User experience and quality of service, Subjective quality assessment has become too easy and convenient and can be was conducted under laboratory viewing environment specified by ITU-R BT.500-12 Standards[1].

2. Subjective Video Quality Assessment

Single Stimulus Continuous Quality Evaluation(SSCQ) process was selected out of Single Stimulus and Stimulus Comparison Quality Evaluation. This method considers hidden reference video during subjective analysis which leads in subjects grading values given for hidden reference that can be only used to consider the seriousness of the subject and is not included in final results. In this method subjects will be able to observe video sequence once and grade it in the given 10 seconds time based on his/her perception and All Test sequences have been played in pseudo random order. Each video sequence was displayed for 10 seconds with grading scale of 0-100 [2]

3. Objective Video Quality Assessment

The properties of encoded videos are acquired from bit stream data of H.264/AVC which has been generated as a trace file while encoding process. Rather than using completely decoded frame, our interest lies in reversing the entropy encoding of bit stream. By analyzing three successive Nal, Slice and Macro Block Layer the following features were extracted.

3.1. No Reference Video Quality Assessment

In general, while transmitting multimedia information within a network which is also refereed as bit steam data, Its a low computational complexity between Extracted features of bit-stream data at Nal layer and the corresponding quality scores towards finding relation between both of them. This metric is deployed towards validating relation between input features and target scores towards classical linear regression model which is based on Pearson correlation and moreover we understood that characteristics of Pearson coefficient is completely based on goodness of fitness towards linearity nature. We considered maximum selected number of features out of bit-stream data at network layer and the complexity of selected features was based on individual variance which is calculated by principle component analysis as we selected it. After our investigation, we confirmed it as no reference quality assessment where reference video is partially available [3].

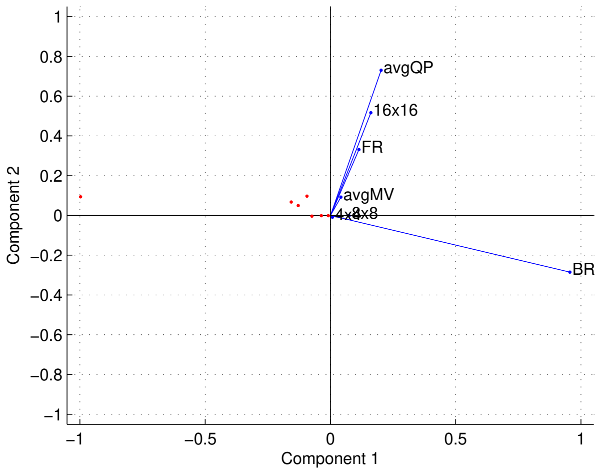

Figure 1: Typical 2 dimensional of chrominance and luminance component

1. Bitrate
2. Frame rate
3. Percentage of inter macroblocks of size16x16.
4. Percentage of inter macroblocks of size 4x4
5. Percentage of inter macroblocks of size 8x8
6. Average quantization parameter.
7. Average Motion Vector

Even hypothetically it's Impossible to combine Speech with Visual characteristics together. Based on ITU Recommendations, we initially assumed and considered the typical two-dimensional plane. In technical terms it should be Chrominance and Luminance plane. In order to quantify impairments of spatial and temporal domain, firstly it should be based on technical assumptions, i.e, we should do mathematical operations based on spatial information within chrominance plane and temporal information within luminance plane. Secondly colour domain exists between two planes.

3.2. Reduced Reference Video Quality Assessment

It is a reduced complexity between individual features extracted of H.264 bit stream data at slice layer without involvement of corresponding quality scores, for instance its an error concealment between spatial of chrominance component and temporal of luminance component, the role of error concealment is to replace duplicate data within frames of video sequences towards original and error concealment is quantified based on poor coding error not because of packet loss or delay'and it can be only traced at slice level after extraction of bit stream data moreover feature extraction of h.264 bit stream data at slice level is based on spatial and temporal complexity. This process of evaluating visual quality is referred as reduced reference video quality assessment.

3.3. Full Reference Video Quality Assessment

Its an high computational complexity between individual features with corresponding quality scores based on statistics of Transform coefficients,specifically this type of full reference video quality assessment is evaluated only after extraction h.264 bit stream data at macro block level

- Avg QP- Average Quantization Parameter.
- Avg bitrate [kbps]-Average Bits per second
- Inter[%]-Percentage of Inter Coded Macro Blocks
- Intra[%]-Percentage of Intra Coded Macro Blocks
- Skip[%]-Percentage of Skip Coded Macro Blocks
- P16x16[%]-Percentage of Inter Coded Macro Blocks with 16x16 subdivision
- P8x8[%]-Percentage of Inter Coded Macro Blocks with 8x8 subdivision
- P4x4[%]-Percentage of Inter Coded Macro Blocks with 4x4 subdivision
- MV_X- Average of Horizontal absolute Motion Displacement
- MV_Y- Average of Vertical absolute Motion Displacement
- MVD_X- Average of the Motion Displacement difference in horizontal direction

$$MVD_X = |MV_x(i_l.j) - MV_x(i_r, j)| \tag{1}$$

Where (i_l, j) and (i_r, j) positions at left and right edge image or frame

- MVD_Y- Average of the Motion Displacement difference in vertical direction

$$MVD_Y = |MV_y(i_l.j) - MV_y(i_r, j)| \tag{2}$$

Where (i_l, j) and (i_r, j) positions at left and right edge image or frame

- Zero MVs[%]-Percentage of Zero absolute Motion vectors
- Zero MVDs[%]-Percentage of Zero Motion vector Difference
- Motion Intensity

$$MI1 = \sum_{i=0}^{N} \sqrt{MV_{X^2_i} - MV_{Y^2_i}} \tag{3}$$

where N is the total number of macro blocks in each frame. MV_{X_i} and MV_{Y_i} are the absolute motion vector of the ith macro block in Horizontal(X) and Vertical(Y) directions respectively.

- Motion Intensity II

$$MI2 = \sqrt{MV_{X^2} - MV_{Y^2}} \tag{4}$$

where MV_X and MV_Y are the average of absolute motion vectors in each frame in X and Y directions respectively.

- IinPframes[%]-Percentage of Intra coded macro blocks in P frames.

4. Test Methodology for Objective Methods

We conducted experiments on three types of databases, 288 video sequence in test methodology I, 195 video sequences in subjective experiments and 120 video sequences in our research work as mentioned in test methodology III.

4.1. Test Methodology I

We deployed an traditional linear regression formulation for training and testing of data and below equation formulates it, selection criteria for training of our proposed model was based on randomly permuting of input features(X) with respective target scores(Y) and later testing of our proposed model was done accordingly.

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\epsilon}_i \qquad (5)$$

Table 1: Proposed model statistical results

VQM	PSNR	PEVQ	SSIM
PCC	**0.99**	**0.93**	**0.62**
MSE	**4.74**	**0.12**	**0.00**
STD	**1.43**	**0.33**	**0.04**
MAE	**1.77**	**0.27**	**0.05**

4.2. Test Methodology II

It has been traced out that error concealment is completely based on coding and moreover our approach towards reduced reference video quality is evaluated based on spatial, temporal and colour domain. Practically correlating input features without involvement of Target scores based on distortion is only possible through rate distortion algorithm which is not yet developed for bit-stream data. Therefore we conclude that it need advanced understanding towards coding or compression techniques of respective encoded videos.

4.3. Test Methodology III

According to Video Quality Experts Group phase II test, performance of objective quality prediction model is evaluated by prediction accuracy. These attribute is evaluated by following performance metric **Metric**: Linear correlation metrics related to prediction accuracy of proposed prediction model. Mathematical model of linear correlation coefficient metrics[4] is

$$r_p = \frac{\sum_{i=1}^{N} (\bar{Y}_i - \bar{Y}) * (Y_i - Y)}{\sum_{i=1}^{N} \sqrt{((\bar{Y}_i - \bar{Y}))^2 * (Y_i - Y)^2}}. \qquad (6)$$

where N denotes total number of values. $\bar{Y}_i, Y_i$ represents estimated and target values of video sequence. $Y, \bar{Y}$ represents mean of target and estimated values respectively. The below plot illustrates that out of all 17 features P8x8 and P4x4 are low correlated for all quality metrics while other are acceptable. This low correlated features where eliminated out of 17 features for training and testing of our model. Our Model is trained and tested with 120x15 input features out of 120 live video sequences which were generated in our thesis [5].

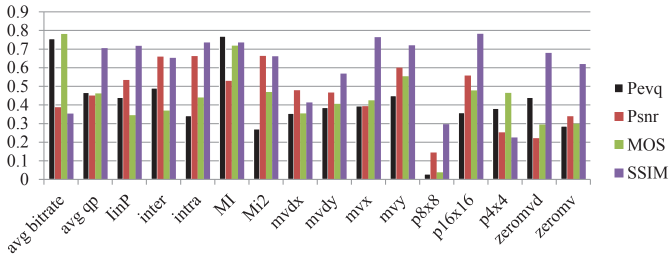

Figure 2: Linear correlation between Individual Feature and Target Values

5. Conclusion

After through investigation of our research work, we believed that objective metric should solely used for estimation of video coding not for quality assessment and the scope of subjective quality assessment is essential but even subjective analysis is limited to few concepts, because selectively subjective scores should be considered as true values judged by humans. Finally, we concluded that even after achieving consistency within subjective scores, hypothetically we must assume that our test configuration as sampling distribution not normal distribution because subjective scores are considered as independent variables moreover subjective experiments are based on human visualization characteristics.

6. Acknowledgments

I am very grateful to Prof. Benny Lovstrom and Mohammad Shahid for all his support. I also would like to thank our reviewers for their valuable time and effort.

References

[1] ITU-R BT.500-12,radio communication sector of itu, 2009. http://www.itu.int/.

[2] Muhammad Shahid, Amitesh Kumar Singam, Andreas Rossholm, and Benny Lövström. Subjective quality assessment of h.264/avc encoded low resolution videos. In *2012 5th International Congress on Image and Signal Processing*, pages 63–67, 2012.

[3] Amitesh Kumar Singam and Venkat Raj Reddy Pashike. Low complex blind video quality predictor based on support vector machines. 2012.

[4] Andreas Rossholm and Benny Lovstroem. A new low complex reference free video quality predictor. In *2008 IEEE 10th Workshop on Multimedia Signal Processing*, pages 765–768, 2008.

[5] H.264/avc jm reference software version 16.0, tech.rep, joint video team, 2011.