International Journal of Trend in Scientific Research and Development, Volume 1(1), ISSN: 2456-6470
www.ijtsrd.com

Video Display for Subjective Video Quality Assessment

Daisy Varghese
Lecturer in Electronics, Department of Electronics and Communication Engineering,
Government Polytechnic College, Perumbavoor, Ernakulam, Kerala, India

ABSTRACT

We present the findings of a recent large-scale subjective study of video quality on a set of videos that have been distorted by a variety of application-relevant processes. Methods for Assessing the Visual Quality of Digital Videos as Perceived by Human Observers are Increasingly Important, Due to the Large Number of Applications that Target Humans as Video End Users. Because of the several approaches to video quality assessment (VQA) that are being developed, there is a need for a diverse independent public database of distorted videos and subjective scores that is free to use. The resulting Laboratory for Image and Video Engineering (LIVE) Video Quality Database includes 150 distorted videos (derived from 10 uncompressed reference videos of natural scenes). That were made using four different types of commonly seen distortions. 38 human subjects evaluated each video, and the difference mean opinion scores (DMOS) were recorded. On the New Database, we also evaluated the performance of many cutting-edge, publicly available full reference VQA algorithms. There is also a statistical evaluation of the relative performance of these algorithms. We shall explore Video Display for Subjective Video Quality Assessment in this study.

Keywords: *Video, Display, Subjective, Video, Quality, Assessment, Image, Video Engineering, Live, Natural Scenes, Recorded.*

INTRODUCTION

Video quality as seen by humans is referred to as subjective video quality. It is concerned with how a viewer (also known as a "observer" or "subject") perceives video and expresses their perspective on a certain video sequence. It is related to the Quality of Experience field. Because objective quality evaluation algorithms, such as PSNR, have been demonstrated to correlate poorly with subjective judgements, subjective video quality must be measured. Subjective assessments can also be utilised to design new algorithms as ground truth. [1]

Subjective video quality tests are psychophysical investigations in which a group of people score a set of stimuli. These tests are highly costly in terms of time (planning and execution) and human resources, so they must be properly developed.

Typically, in subjective video quality testing, SRCs ("Sources", i.e., original video sequences) are treated with various conditions (HRCs for "Hypothetical Reference Circuits") to yield PVSs ("Processed Video Sequences").

Video: Subjective and Objective Quality Assessment:

Understanding how different elements of your distribution chain might affect your video in different ways is a critical factor in guaranteeing sustained quality in the service and speedy fault discovery where problems are detected.

Abdul Rehman of SSIMWAVE discusses at the Kitchener-Waterloo Video Technology Meetup on subjective quality assessment, in which humans judge the quality of the video, and objective quality assessment, in which computers analyse large amounts of video, frequently terabytes, to determine the quality.

Abdul describes how many things can go wrong in the chain, beginning with a video that shows examples of several problems that might occur in the chain. These include lost or delayed data, erroneous content, and service configuration checks. Display devices presently come in a variety of shapes, sizes, and resolutions, which can cause display issues, as can the player and viewing conditions. These are only about half of the various alternatives for the sort of person - a golden eye or a pure consumer.

One approach is subjective video quality assessment, which employs people who are considerably better at spotting creative quality than machines. As a result, many false positives are avoided, where video may be evaluated as terrible yet there is intent in the use. Furthermore, it represents direct feedback from your target audience. Abdul goes over the various parts of what you need to control for when using subjective video quality assessment to maximise its use and allow results from multiple sessions and experiments to be directly compared.

This will be contrasted to objective video quality assessment, which will use a computer to sift through the films. This can be quite effective for various applications, which means it can shine in terms of throughput and measurement number. Furthermore, it can greatly simplify regression testing. Depending on the application, the disadvantages can include expense, false positives, and possibly speed. You can then choose amongst algorithms such as MS-SSIM, VMAF, and others. Abdul concludes by going over the advantages and warning signs. [2]

Subjective Quality Assessment

Humans as end users are the legitimate evaluators of visual quality, and their judgements can be gained through subjective experiments. Subjective studies use a panel of non-experts, often known as test subjects, to evaluate the perceptual quality of a specific test material, such as a video sequence. Even though crowdsourcing-based quality assessment has promising correlation values with laboratory-based testing, subjective trials are normally conducted in a controlled laboratory environment.

Prior conducting a subjective experiment, meticulous planning and various elements such as evaluation method, test material selection, viewing settings, grading scale, and presentation timing must be considered. Different recommendations have been standardised to provide guidelines and allow comparisons between subjective experiments. ITU-R BT.500 and ITU-T P.910, for example, provide specific rules for conducting several sorts of subjective tests for quality assessment. These strategies are either single stimulus or double stimulus-based. In single stimulus approaches, individuals are shown variations of test films with no reference for comparison, such as Absolute Category Rating (ACR). [3]

In other cases, a hidden reference can be included, but the evaluation is based solely on the participants' no-reference ranking, for example, Absolute Category Rating including Hidden Reference (ACR-HR). In double stimulus approaches, a pair of films, one with the reference video and one with a degraded video, is shown once or twice, and the subject is asked to rate the quality or difference in quality between the two video streams. Degradation Category Rating (DCR) or Double Stimulus Impairment Scale (DSIS), for example. There is also a multi stimulus method, such as Subjective Assessment of Multimedia Video Quality (SAMVIQ), ITU-T Rec., where the subject rates the quality of various test videos, including reference and hidden reference. The subjects are also permitted to watch the films multiple times.

Individual scores given by test subjects are the findings of a subjective experiment. Depending on how the experiment is designed, these scores are used to compute Mean Opinion Score (MOS) or Differential Mean Opinion Score (DMOS). The major distinction between MOS and DMOS is that MOS is the outcome when the subject rates a stimulus in isolation, but DMOS is the difference in quality between two copies of the same stimulus, for example, MOS is used for ACR, ACR-HR, SAMVIQ, and DSIS, and DMOS is used for ACR-HR and DSIS. [4]

Review of Literature:

Jumisko et al. demonstrate that content recognition influences the rating of perceived video quality. They discovered that video clips recognised by participants received lower ratings than unrecognised clips, whereas fascinating contents received higher ratings than those deemed dull (whether recognised or not). It is suggested that assessors with prior knowledge of the genre are more exacting in terms of quality acceptance. In line with previous research, they discovered that the audio component (when accessible in the experiment) compensated to a good extent for deficiencies in the visual aspect of the film, whereas deficits in speech were found to be quite distracting. This is explained by the fact that, for that specific experimental design, audio provides the crucial information and the visual component only supports it (e.g., music videos and news with a narrating voice in the background). [5]

According to Gulliver et al. They discovered that the sequence's content had a greater impact on a user's amount of information transfer than frame rate or display device type. When information transfer is removed from the equation, participants in the same study found frame-rate and device type to be important for perceived video quality, demonstrating that they could distinguish between their subjective enjoyment of a video clip and the level of quality that they perceived the video clip to have. This implies that there is a link between clip contents and user-perceived video quality, but the components of the equation that lead to a final score must be carefully analysed. [6]

Objectives:

- Subjective methods are regarded as the most trustworthy ways for assessing picture quality.
- Subjective quality assessment entails the observer explicitly assigning a quality score to a given image.
- Personal judgement or standards that are less systematic than those employed in objective examinations are utilised to score an assessment tool.

Research Methodology:

This study's overall design was exploratory. According to related study, assessors are not indifferent to the content they evaluate. This has potential implications for real-world applications, because subjective perception of quality of some "neutral" content chosen for its suitability for introducing specific impairments to video streams (i.e., standardised VQA databases) may differ from subjective perception of quality of content that the assessor is familiar with or interested in and pays more attention to (i.e., content that the assessor regularly consumes, such as TV shows and films). [7] We included a number of videos that should be fairly familiar to assessors, rather than focusing solely on content that was supposed to induce different emotions or other mental responses, because familiar video content was shown to sometimes receive a lower quality rating at the same impairment level as unrecognised video content. First, we identified numerous kinds of TV programmes that are most frequently viewed by average Serbian citizens (since participants were all Serbian nationals) based on information from national television services as well as the findings of a study done by a media research firm. [8]

Result and Discussion:

Objective Quality Assessment:

Due to the time-consuming nature of conducting subjective experiments, considerable work has been expended in developing objective quality measurements, often known as objective quality methodologies. The goal of such objective quality methodologies is to predict MOS with high accuracy automatically. Psychophysical and engineering approaches are two types of objective quality methodologies.

The goal of psychophysical metrics is to simulate the human visual system (HVS) by utilising characteristics such as contrast and orientation sensitivity, frequency selectivity, spatial and temporal pattern, masking, and colour perception. These metrics can be utilised for a wide range of video degradations, however the computation is generally time-consuming and is rarely employed in the streaming context. [9]

Engineering metrics are typically based on the extraction and analysis of certain elements or artefacts in a video, but they do not necessarily disregard the properties of the HVS because they frequently account psychophysical impacts as well. However, rather than fundamental vision modelling, the conceptual basis for their design is to analyse video content and distortion. A collection of video features or quality-related criteria is pooled together to create an objective quality technique that can be mapped to predict MOS.

The objective approaches are further classified into full reference (FR), reduced reference (RR), and no-reference (NR) methods based on the amount of information available from the original video as a reference in the quality assessment: [10]

FR Methods: With this approach, the entire original video is available as a reference. Accordingly, FR methods are based on comparing a distorted video with the original video.

RR Methods: In this case, it is not required to give access to the original video but only to provide representative features of the characteristics of the original video. The comparison of the reduced information from the original video with the corresponding information from the distorted video provides the input for RR methods.

NR Methods: This class of objective quality methods does not require access to the original video but instead searches for artefacts in a video's pixel domain, uses information embedded in the bitstream of the related video format, or performs quality assessment as a hybrid of pixel-based and bitstream- based approaches. [11]

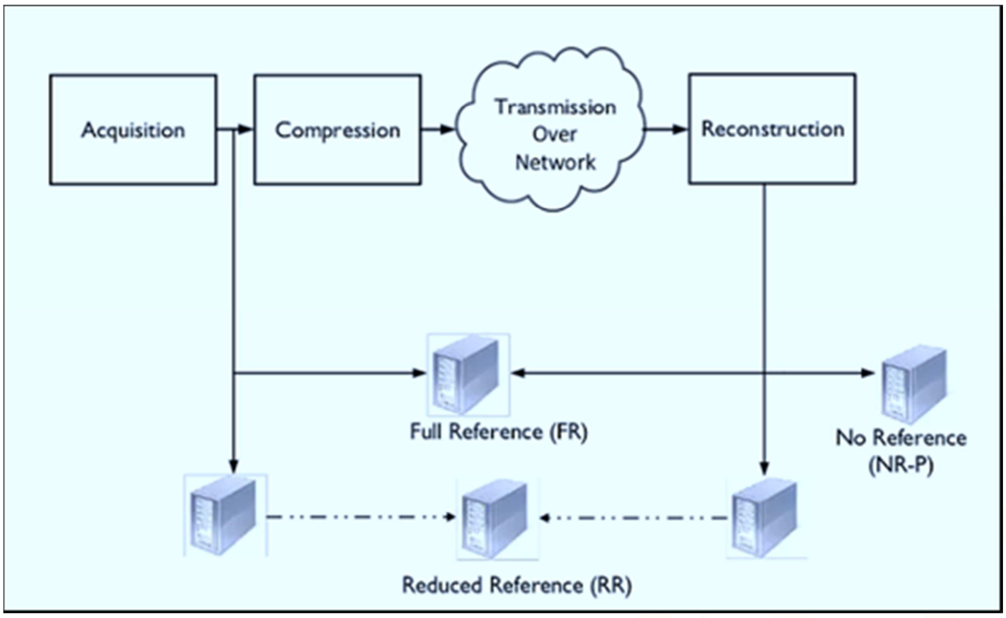

Figure 1: Objective Assessment, FR, RR, NR

Subjective Examination In general, we use the Absolute Category Rating (ACR) technique, as recommended by ITU-T recommendation P.910. Although SSCQE is intended to track immediate video quality with time, it is not used in our experiment for the following reasons. First, in practise, human subjects frequently choose to rate video quality on a per-segment basis, ignoring immediate quality fluctuations between frames within a scene. Second, the identical coding setup and settings are applied throughout the whole duration of each scene in our database, which is likewise relatively constant in terms of content and complexity. As a result, a single quality score is adequate to describe its overall quality. [12]

Third, there is a time delay in SSCQE between the recorded instantaneous quality and the video content, which varies between participants and is also a function of slider "stiffness." This is an unresolved issue with the SSCQE methodology in general, although it is avoided when only a single score is obtained.

We believe that ACR is significantly simpler than SSCQE and gives more valid and realistic quality ratings under the subjective study conditions. The subjective test included thirty naive subjects, all of whom were university undergraduate and graduate students. [13]

To calculate the fatigue factor, the first few video sequences were repeated at the end of the test. We discovered that there was no bias or substantial difference in the scores achieved for the same set of video sequences at the start and end of the test. Subjects rated the test video sequences under the viewing conditions shown in Table 1.

The subjects were given instructions in both written and oral form. The test was preceded by a training session in which the subject was shown samples of distorted video sequences predicted in the test. [14]

Table1: Display Devices and Viewing Conditions Employed in the Subjective Study [15]

Display Device	Diag. Screen Size (in)	Resolution	Brightness (cd/m^2)	Viewing Distance (in)
iPhone 5S	4”	1136x640	556	10
iPhone Air	9.7”	2048x1536	421	16
Lenovo laptop	15.6“	1920 X 1080	280	20
Sony TV	55“	1920 X 1080	350	90
Sony TV (TV-Expert)	55”	1920x1080	350	40

We propose SSIM plus, a unique VQA measure that takes into consideration human visual system features, video content, and viewing conditions. The VQA measure provides simple predictions on what an average customer thinks about the quality of video material being delivered on a scale of 0-100, as well as categorises the quality as bad, poor, fair, good, or exceptional. The underlying algorithm employs an advanced perceptual model, allowing the VQA measure to adapt to any display device and viewing conditions. The first step in assessing video QOE is to execute a multi-scale transform on reference and distorted video frames, which decomposes a video frame into numerous scales, each associated with a particular frequency range. Following that, the quality maps for each scale are produced using a structure comparison of consecutive reference and distorted scales. Following that, the quality of all scales is established by spatial pooling of quality maps depending on local information content and distortion. A weighted combination of the scale-wise quality values is used to calculate the perceptual quality of the warped frame. [16] The weights are calculated using a process that takes into account the display device's features as well as the viewing conditions. The perceptual quality of video content is determined by the sampling density of the signal, the viewing conditions, the display device, and the observer's visual system's perceptual capability. In practise, when these parameters change, the subjective perception of a specific film changes. The human visual system's contrast perception capability is heavily dependent on the spatial or spatio-temporal frequency of a visual input, which is modelled using a function known as the contrast sensitivity function (CSF). [17]

To measure the contrast sensitivity of the human visual system, one can use one or a combination of the following devices and viewing parameters:
- Average or range of user viewing distances;
- Screen and viewing window sizes;
- Screen resolution;
- Video scaling;
- Screen contrast;
- Replay temporal resolution;
- Viewing environment illumination condition;
- Viewing angle;
- Resolution of the viewing window;
- Post-filtering and picture scaling procedures;
- Device model;
- Gamma correction adjustment for the screen;
- Video scan type (interlaced or progressive).

These parameters are used to determine the human visual system's sensitivity to the individual scales of the incoming video signals. The sensitivity values are then normalised to determine the scales' weight/importance. [18]

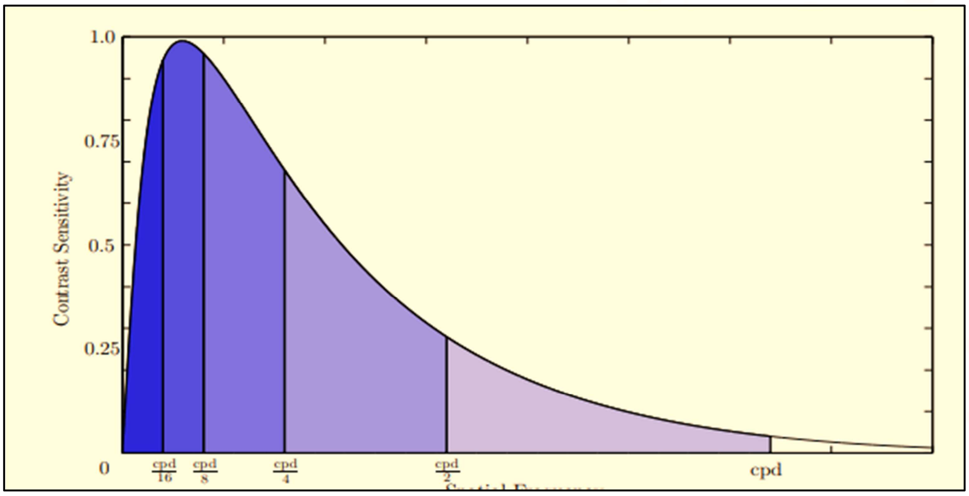

Figure 2: Sensitivity of the Human Visual System to the Individual Scales of the Input Video Signals.

The viewing window/screen size, device screen resolution, replay temporal resolution, viewing distance, device screen contrast, viewing angle, and viewing window resolution, or a subset of these parameters, are converted into a viewing resolution factor in the unit of pixels per degree of visual angle. The CSF of the human visual system is likewise computed using these characteristics. The frequency coverage range of each scale in the multi-resolution transform is then determined using the viewing resolution factor. In the multi-resolution transform, the frequency spanning ranges of all scales divide the entire CSF into many regions, each corresponding to one scale. The area under the CSF function inside the frequency coverage range of each scale is then calculated to yield a weighting factor for that scale. Because the viewing resolution factor and CSF computation are device and viewing condition dependent, the frequency covering ranges and, as a result, the weighting factor of each scale are also device and viewing condition dependent, which is an important factor that distinguishes the proposed method from existing approaches. These devices and viewing condition-dependent parameters are utilised to assess the importance of each scale in the overall image or video signal quality evaluation. Figure 2 depicts an example of the details of a multi-scale weights computation technique based on device and viewing condition-dependent factors. The cycles per degree of visual angle (cpd) in Figure 2 are controlled by the viewing resolution factor. Starting with the finest scale, the frequency coverage ranges of the scales in the multi-resolution transform are between cpd/2 and cpd, cpd/4 and cpd/2, cpd/8 and cpd/4, and so on. Dynamically computed integrals of the CSF curve under the respective frequency coverage range are utilised to define the weighting factor and consequently the visual relevance of the related scale. Following that, the scale-wise scores are pooled to obtain frame-level and sequence-level QOE scores. [19]

Provides insight into investigations conducted to establish objective methods of video quality measurement and subjective VQA experiments. The H.264/AVC standard was utilised to encode the test videos used in this research. [20] This standard is based on the Differential Pulse Code Modulation/Discrete Cosine Transform (DPCM/DCT) hybrid codec model, which includes motion estimation and compensation (DPCM), followed by transform (DCT) and entropy coding processes. Throughout these steps, several decisions regarding video frame (sub)partitioning, the use of an appropriate reference frame, and the level of quantization are made. [21] These decisions are expressed as precise coding parameter values that are utilised to (de)code the visual stream. Certain coding-based criteria have been discovered to be beneficial in assessing the perceived quality of encoded video. Some of the papers in this dissertation give details on this type of quality estimation. [22]

Figure 3: Original Image (b) processed with blurring in (a) and with Salt & Pepper Noise in (c)

Conclusion:

The most important conclusion we took from the experiment results is that the content of the video sequence has a strong influence on activating different cognitive, emotional, and conative components within assessors. These components, in turn, have been demonstrated to play an important role in VQA tasks (both in this and previous studies), whether assessors are aware of it or not. As a result, we offer a number of guidelines that should be considered while performing subjective video quality assessment tests. A larger-scale study may also show other characteristics relevant to diverse populations (gender, culture, and demographics) that influence subjective assessment of video quality and should thus be considered when VQA tasks are in issue.

References:

[1] ITU-T Tutorial: Objective perceptual assessment of video quality: Full reference television, 2004.

[2] ITU-T Rec. P.910: Subjective video quality assessment methods for multimedia applications, 2008.

[3] Methodology for the Subjective Assessment of the Quality for Television Pictures. Standard

ITU-R BT.500, revision 13. ITU-R, January 2012.

[4] Subjective video quality assessment methods for multimedia applications. Standard ITU-T P.910, revision 3. ITU-T, April 2008.

[5] Jumisko, V. P. Ilvonen, and K. A. Väänänen-Vainio-Mattila, "Effect of TV content in subjective assessment of video quality on mobile devices," in IS & T Electronic Imaging-Multimedia on Mobile Devices, vol. 5684 of Proceedings of SPIE, pp. 243–254, January 2005.

[6] S. R. Gulliver, T. Serif, and G. Ghinea, "Pervasive and standalone computing: the perceptual effects of variable multimedia quality," International Journal of Human Computer Studies, vol. 60, no. 5-6, pp. 640–665, 2004.

[7] Methodology for the subjective assessment of video quality in multimedia applications. Standard ITU-R BT.1788, revision 1. ITU-R, January 2007.

[8] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," IEEE Transactions on Image Processing, vol. 19, no. 6, pp. 1427–1441, 2010.

[9] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "A subjective study to evaluate video quality assessment algorithms," in Human Vision and Electronic Imaging XV, Proceedings of SPIE, January 2010.

[10] H. R. Wu, K. R. Rao, Digital Video Image Quality and Perceptual Coding., Signal Processing and Communications. (CRC, Boca Raton, 2005).

[11] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, M. Carli, New full-reference quality metrics based on HVS, CD-ROM Proceedings of the Second International Workshop on Video Processing and Quality Metrics, Scottsdale, USA, 2006.

[12] Nikolay Ponomarenko, Flavia Silvestri, Karen Egiazarian, Marco Carli, Jaakko Astola, Vladimir Lukin, On between-coefficient contrast masking of DCT basis functions, CD-ROM Proceedings of the Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM-07, Scottsdale, Arizona, USA, January 2007.

[13] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," tech. rep., International Telecommunication Union, Geneva, Switzerland (Apr. 2008).

[14] ITU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," (Mar. 2002).

[15] Baroncini, V., Ohm, J. R., and J., S. G., "Report on preliminary subjective testing of hevc compression capability," in [JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11], (Feb. 2012).

[16] Sullivan, G., Ohm, J., Han, W., and Wiegand, T., "Overview of the high efficiency video coding (HEVC) standard," IEEE Trans. Circuits Syst. Video Techn. 22(12), 1649–1668 (2012).

[17] Wang, Z., Simoncelli, E. P., and Bovik, A. C., "Multi-scale structural similarity for image quality assessment," in [Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers], 1398–1402 (Nov. 2003).

[18] Pinson, M. H. and Wolf, S., "A new standardized method for objectively measuring video quality," IEEE Trans. Broadcasting 50(3), 312–322 (2004).

[19] Seshadrinathan, K. and Bovik, A. C., "Motion tuned spatio-temporal quality assessment of natural videos," IEEE Trans. Image Processing 19, 335–350 (Feb. 2010).

[20] Zeng, K., Rehman, A., Wang, J., and Wang, Z., "From H.264 to HEVC: coding gain predicted by objective video quality assessment models," in [7th Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics], 42–46 (Jan. 2013).

[21] N. Westerlund, M. Dahl, and I. Claesson. Real-time implementation of an adaptive gain equalizer for speech enhancement purposes. WSEAS., 2003.

[22] M. Dahl, I. Claesson, B. Sällberg, and H. Akesson. A mixed analog -digital hybrid for speech enhancement purposes. ISCAS., 2005.