

Evaluation of Video Quality Assessment Algorithms

Sophie M Greene

November 28, 2016

Abstract

Since human subjects are the end users of videos, they are considered the most reliable in evaluating video quality. Hence, the most accurate video quality assessment method is the subjective Mean Opinion Square MOS, which is acquired using human subjects in a controlled viewing environment, this process is time consuming and expensive, thus the need for computational models that can predict perceived video quality accurately and automatically. These computational models are commonly known as objective Video Quality Assessment VQA algorithms and they have many applications commercially and otherwise. This report describes a set of experiments conducted to evaluate the different existing objective VQA algorithms using a subjective study of the same videos. The video sequences were acquired from LIVE Video Quality Database [2, 1]; which consist of 10 original videos as well as 15 distorted copy of each. The distortion methods vary from compression to simulated transmission over different communication channels. The Peak Signal to Noise Ratio PSNR, Structural Similarity Index SSIM and the Visible Signal to Noise Ratio VSNR objective algorithms are applied to the videos and the outcome of each is plotted against the corresponding subjective scores. Finally, the Spearman Rank Order Correlation Coefficient SROCC and Pearson Linear Correlation coefficient LCC are computed to evaluate the relationship between the aforementioned objective algorithms and the subjective scores. The results of the evaluation are compared to study in [2, 1].

1 Introduction

During acquisition, processing, compression, storage, transmission and reproduction, digital videos are subject to a wide variety of distortions which could cause a degradation of the videos' visual quality. In applications where human beings are the ultimate viewers of videos, subjective evaluation is the most reliable method for quantifying visual video quality. However, subjective evaluation is costly and time consuming thus the need for automated video quality assessment VQA algorithms to provide quantitative measures capable of predicting the perceived video quality. Objective VQA metrics can play a variety of roles in video processing applications; one of which is dynamically monitoring and adjusting video quality. For example, a TV channel broadcasting live over the Internet can examine the quality of video being transmitted and allocate streaming resources accordingly. VQA metrics can also be used to optimise algorithms and parameter settings of video processing applications. Objective VQA metrics can be classified according to many criteria. In this context, the criterion we are interested in is the availability of an original, distortion-free video, with which the distorted video is compared. A VQA metric is referred to as full-reference FR whenever a complete reference video is available. Practically, the reference video is not always available, which inspires video quality assessment algorithms that try to quantify the quality of a video without a reference video being available, this approach is referred to as no-reference NR, VQA algorithm. Reduced Reference RF is the third class of VQA algorithms in which the reference video is only partially available in the form of a subset of features from the original video used to judge the quality of a distorted one. For more on VQA algorithms classification refer to [3]. For any objective VQA algorithm to be truly successful, it needs to be validated. Our experiments aim to validate and inspect the effectiveness of some of the most commonly used FR VQA such as PSNR, SSIM, MS-SSIM and VSNR. Section ?? describes the the videos used in the experiment and the subjective video quality assessment measure, the Difference Mean Opinion Score DMOS and the procedure used to calculate this measure . Section ?? describe the objective VQA algorithms being validated. in section ref

2 Subjective Video Quality Assessment

Since human beings are the ultimate receivers in most video processing applications, the most reliable way of assessing the quality of an image is by subjective evaluation. Indeed, the mean opinion score (MOS), a subjective quality measure requiring the services of a number of human observers, has been long regarded as the best method of image quality measurement.

2.1 Reference Videos

The reference videos are ten uncompressed videos available from [2]in uncompressed planar YUV 4:2:0 format, the each of the videos has a 768x432 pexels resolution.

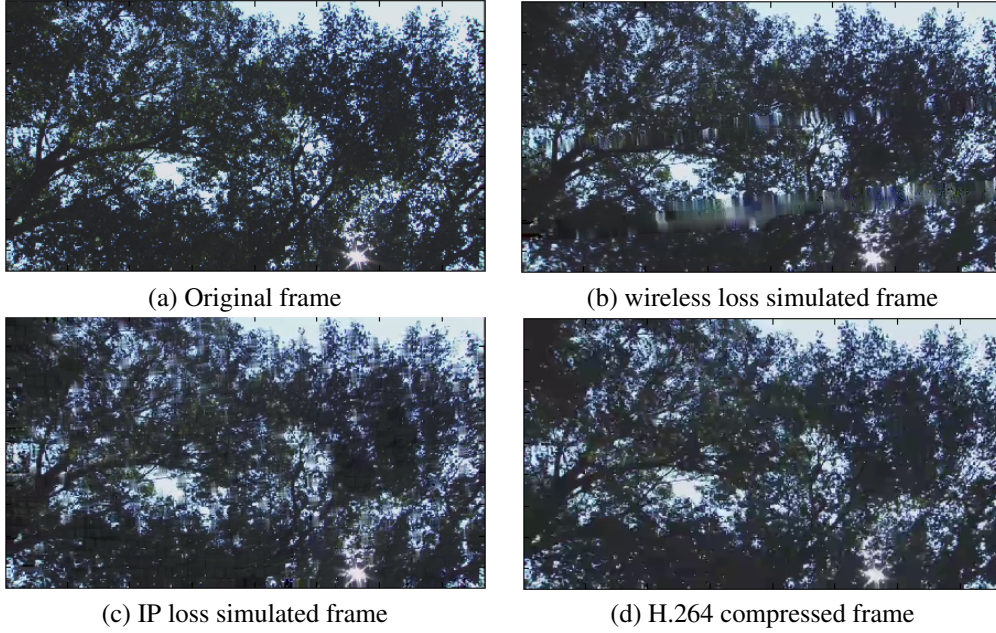


Figure 1: Examples of distorted frames.

2.2 Distorted Videos

We used 15 distorted videos for each reference video. the videos were created using four different distortion processes - MPEG-2 compression (4 test videos per reference), H.264 compression (4 test videos per reference), lossy transmission of H.264 compressed bitstreams through simulated IP networks (3 test videos per reference) and lossy transmission of H.264 compressed bitstreams through simulated wireless networks (4 test videos per reference) [2, 1]. Figure 1 illustrates sample frames from the reference and distorted videos.

2.3 Difference Opinion Mean Square DMOS

we used the DMOS scores provided by [2] which were obtained by the following procedure

1. in two half-hour sessions; show a playlist consisting of the test and reference videos to 38 participants, asking them to rate each video after it is displayed. The rating categories are “Excellent”, “Good”, “Fair”, “Poor” and “Bad”.
2. Compute the difference score for each of the test videos by subtracting each score from each corresponding reference video.
3. The difference scores were then normalised by subtracting the mean of the scores assigned by the participant in the session from each difference score and dividing the result by the standard deviation of scores assigned by the participant in the corresponding session.
4. The results were statistically manipulated to check for outliers and scores from 9 participants were rejected
5. The remaining normalised scores were re-scaled in the range [1:100]
6. The DMOS of each video was computed as the mean of the re-scaled normalised scores

3 Objective Video Quality Assessment

The goal of objective image quality assessment research is to design computational models that can predict perceived image quality accurately and automatically. We use the term predict here, since the numerical measures of quality that an algorithm provides are useless unless they correlate well with human subjectivity. In other words, the algorithm should predict the quality of an image that an average human observer will report. Objective VQA metrics can be classified according to many criteria. The criterion we are interested in, in this context, is the availability of an original, distortion-free video, with which the distorted video is compared. Reduced Reference RF is the third class of VQA algorithms in which the reference video is only partially available in the form of a subset of features from the original video used to judge the quality of a distorted one. A VQA metric is referred to as full-reference FR whenever a complete reference video is available. Practically, the reference video is not always available, which inspires video quality assessment algorithms that try to quantify the quality of a video without a reference video being available, this approach is referred to as no-reference NR, VQA algorithm. For more on VQA algorithms classification refer to [3].

VQA	SROCC	LCC
PSNR	0.4268	0.4617
VSNR	0.6731	0.6889
SSIM	0.5209	0.5384
MSSIM	0.7349	0.6812

Table 1: SROCC & LCC

3.1 Peak Signal to Noise Ratio PSNR

Equation 2 defines the peak signal to noise ratio.

$$MSE = \frac{1}{N} \sum_{i=1}^N |x_i - y_i|^2 \quad (1)$$

$$PSNR = 10 \log_{10} \frac{L^2}{MSE} \quad (2)$$

3.2 Structural Similarity Index SSIM

Natural image signals are highly structured. Samples taken from image signals have strong dependencies amongst themselves, and these dependencies carry important information about the structures of the objects in the visual scene. The principal idea underlying the structural similarity approach is that the HVS is highly adapted to extract structural information from visual scenes, and therefore, a measurement of structural similarity (or distortion) should provide a good approximation to perceptual image quality.

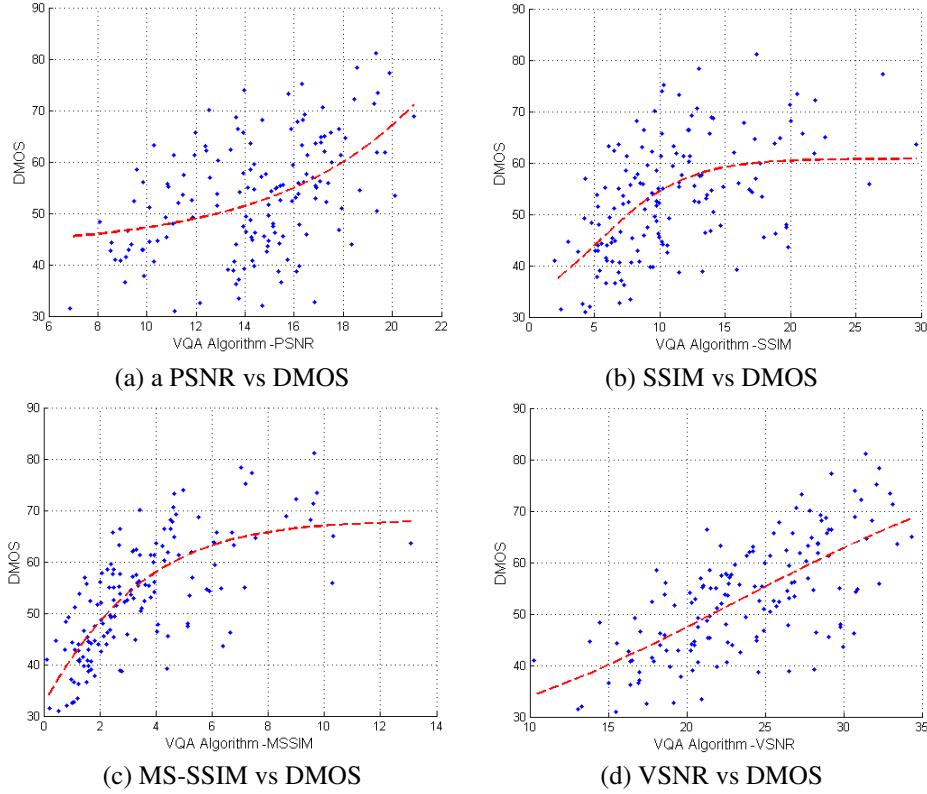


Figure 2: Evaluation of different objective VQA. a plot showing the relationship. small values of DMOS indicates a good quality of the video assessed while

References

- [1] Kalpana Seshadrinathan, Rajiv Soundararajan, Alan C. Bovik, and Lawrence K. Cormack. Study of subjective and objective quality assessment of video. *IEEE Transactions on Image Processing*, 19(6):1427–1441, 2010.

- [2] Kalpana Seshadrinathan, Rajiv Soundararajan, Alan C. Bovik, and Lawrence K. Cormack. A subjective study to evaluate video quality assessment algorithms. In *SPIE Proceedings Human Vision and Electronic Imaging*, 2010.
- [3] Zhou Wang and Alan C. Bovik. Modern image quality assessment. *Synthesis Lectures on Image, Video, and Multi-media Processing*, 2(1):1–156, 2006.