COMPARISON OF VIDEO QUALITY METRICS ON MULTIMEDIA VIDEOS

Mei Hwan Loke, Ee Ping Ong, Weisi Lin, Zhongkang Lu, and Susu Yao

Institute For Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613
Email: {mhloke, epong, wslin, zklu, ssyao}@i2r.a-star.edu.sg

ABSTRACT
The state-of-the-art objective video quality metrics reported in the literature have so far been focused on TV signals with large frame sizes, full TV frame rates, and high compressed bit rates. The aim of this paper is to present the performance of three widely-known objective video quality metrics but tested on multimedia videos with small frame sizes, various frame rates, and low bit rates. The video quality metrics used were the NTIA video quality metrics (the “general” and the “videoconferencing” models), a modified Watson’s DVQ metric, and the VSSIM metric. The test videos consist of both H.264 and MPEG-4 compressed videos with CIF and QCIF frame sizes, at various bit rates (24kbps-384kbps) and frame rates (7.5fps-30fps). Here, results and conclusions derived from the comparison of the three metrics will be provided.


1. INTRODUCTION
The growing digital video industry brings the need for a standardized benchmark to measure the objective quality of digital videos. A reliable video quality metric (VQM) would be useful in optimizing video coding algorithms and examining hypothetical reference circuit (HRC). Presently, there are metrics that give acceptable performance for digitally compressed high to medium bit rate video sequences [1,2], but they either are not illustrated or fail for measuring the visual quality of low bit rate videos. This is a realistic problem as there is increasingly more applications that operate at low bit rates (e.g. 3G mobile videophone applications). A good VQM should reflect the human visual system (HVS), which has limited sensitivity in lower spatial and temporal frequencies [1,3]. The difficulty of creating a good VQM lies in the nonlinear behavior of HVS, and deciding what parameters to be extracted for measurement. Due to the lack of video quality metric that has been explicitly illustrated to work well on low bit rates, lower frame rates, and small frame size videos, the Video Quality Expert Group (VQEG) is in the progress of investigating and consolidating contributions for video quality metrics designed for such multimedia videos [4]. For this paper, the VQMs described are full-reference quality metrics that assumes that the best quality of the picture is known and is used as reference to the processed video.

In this paper, we analyze the performance of 3 widely-known objective video quality metrics on multimedia videos. The video quality metrics that will be discussed are the National Telecommunication and Information Administration (NTIA) Video Quality Metric (VQM) [5], a modified Watson’s Digital Video Quality (DVQ) metric [6], and the Video Structural Similarity (VSSIM) metric [7]. The paper is organized as follows: Section 2 gives a brief overview of the metrics, presenting their underlying logic and algorithms. The details of the experimental materials, results, and analysis are presented in Section 3. Finally, Section 4 gives a conclusion of the comparison study being made.

2. OVERVIEW OF OBJECTIVE VIDEO QUALITY METRICS
2.1. National Telecommunication and Information Administration (NTIA) Video Quality Metric
The NTIA Video Quality Metric (VQM) [5] has 5 video quality models that extract different parameters to be compared. For this paper, only the “General” and the “Videoconferencing” video quality models are examined in detail. The parameters of the general model are optimized using a wide range of video quality and bit rates, while the videoconferencing model is optimized for low bit rate videos. The VQM extracts parameters from both the processed and the original video sequences and compares the features. Parameters which this VQM used were: 1) Features to measure spatial impairments; 2) Features to measure distortions in
chrominance signals; 3) Features to measure localized contrast information; and 4) Features to measure distortions in motion flow.

The VQM algorithm consisted of 2 parts: feature extraction and comparison of parameters. Feature extraction involves applying a perceptual filter, dividing the video sequence into spatial-temporal regions and extracting the needed parameters. The second part uses a comparison function to compare the extracted parameters, followed by spatial and temporal collapsing.

2.2. Modified Watson’s Digital Video Quality (DVQ) Metric

This Modified Watson’s Digital Video Quality (DVQ) Metric [6] is based on Watson’s Digital Video Quality (DVQ) model [8,9] which uses the Discrete Cosine Transform (DCT). The DVQ metric computes the visibility of artifacts expressed in the DCT domain. The metric makes use of DCT coefficients to make it closer to human perception. The algorithm for this VQM is as follows: Both the processed and reference video sequences are converted to the YOZ color space, and undergo DCT transformation. The DCT coefficients are converted to units of local contrast, which is defined as the ratio of the AC amplitude to the temporally low-pass filter DC amplitude. The local contrasts are subjected to spatial contrast sensitivity functions for the static and dynamic frames, and the DCT coefficients are converted to just noticeable differences. The video sequences are subtracted to produce a difference sequence, and this is subjected to a contrast masking in a maximum operation and a weighted pooling mean distortion.

2.3. Video Structural Similarity (VSSIM) Metric

The Video Structural Similarity (VSSIM) metric [7] measures the luminance, contrast, and structure of signals between two video sequences, assuming that these 3 components are independent of each other. For discrete signals, luminance is estimated as the mean intensity. The contrast is estimated as the standard deviation after removal of luminance from the signal. Contrast is removed from video sequence by normalizing the signal with its standard deviation, which leaves the structural component. The 3 components are subjected to their respective comparison functions before being pooled into an overall similarity measure equation that gives a value of the similarity between the processed and reference video sequences. VSSIM measures the distorted video in 3 levels. The first level is the local region level, where random areas of 8x8 windows size are selected and the VSSIM measurement is applied to each of the Y, Cb, Cr color planes before combining into a local quality measure. The second level is the frame level, where the local quality values are merged into the frame-level index using weighing. The third level is the sequence level that averages all the frames level values to produce the overall quality of the video sequence.

3. RESULTS AND DISCUSSIONS

3.1. Test Conditions

A total of 180 QCIF and CIF video sequences were generated from 12 reference sequences (“Coast Guard”, “Container”, “Foreman”, “Japan League”, “News” and “Tempete”). They were subjected to H.264 and MPEG-4 video compression, with different bit rates (24kbps to 384kbps) and frame rates (7.5Hz to 30Hz). Each of the video sequences consists of 250 frames. The VQMs were run on each of the video sequence under test.

3.2. Subjective Test Method

The subjective video quality tests were carried out similar to the tests conducted for the evaluation of JVT video sequences [10]. This used the Double-Stimulus Impairment Scale variant II (DSIS-II) subjective test method and is performed by 20 subjects. The decoded sequences with frame rate lower than 30 fps are displayed with repeated frames on the 30 Hz display device when performing the subjective test.

3.3. Performance

Performance is measured by comparing the metric output Q with the subjective rating of subjective tests between the original and distorted sequences. To facilitate monotonic prediction and a common analysis space of comparison, Q is fitted via a 4-parameter cubic polynomial to the corresponding subjective rating [2] as:

\[ q = a_0 + a_1Q + a_2Q^2 + a_3Q^3 \]

Two performance measures are used for comparison: Pearson correlation coefficient \( r_p \), and Spearman rank-order correlation coefficient \( r_s \). The best possible values that can be achieved is \( r_p = 1, r_s = 1 \).

Table 1 shows the Pearson and Spearman correlation results of the 3 VQMs. The upper bound (UB) and lower bound (LB) of the Pearson correlation were calculated with a confidence interval of 95%. Table 2 shows the Pearson and Spearman correlation results of the 3 VQMs, with reference to compression types.
Figure 1 shows the scatterplot of subjective ratings versus the NTIA general model video quality ratings, Figure 2 shows the scatterplot of subjective ratings versus the NTIA videoconferencing model video quality ratings, Figure 3 shows the scatterplot of subjective ratings versus the modified Watson’s DVQ, and Figure 4 shows the scatterplot of subjective ratings versus the VSSIM video quality ratings. In these figures, the middle solid line portrays the logistic fit using the 4-parameter cubic polynomial [2], while the upper dotted curve and the lower dotted curve portray the upper bound and lower bound respectively obtained with a confidence interval of 95%.

It can be seen that the NTIA videoconference model gives the best performance, followed by the NTIA general model, then the modified DVQ, and finally the VSSIM method. Between the two NTIA models, the videoconferencing model performed better than the general model, because it was trained for low bit-rate videoconferencing sequences, thus removing the error blocks and color distortions measurements which the general model use. The NTIA videoconference model also does not take the measurement of chrominance components into its computations whilst the other metrics do.

<table>
<thead>
<tr>
<th></th>
<th>$r_p$</th>
<th>$r_p$ UB</th>
<th>$r_p$ LB</th>
<th>$r_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTIA VQM General</td>
<td>0.707</td>
<td>0.773</td>
<td>0.625</td>
<td>0.656</td>
</tr>
<tr>
<td>NTIA VQM VideoConf.</td>
<td>0.716</td>
<td>0.781</td>
<td>0.637</td>
<td>0.661</td>
</tr>
<tr>
<td>Modified DVQ</td>
<td>0.646</td>
<td>0.724</td>
<td>0.552</td>
<td>0.551</td>
</tr>
<tr>
<td>VSSIM</td>
<td>0.513</td>
<td>0.614</td>
<td>0.397</td>
<td>0.484</td>
</tr>
</tbody>
</table>

*Table 1: Performance of the VQMs*

<table>
<thead>
<tr>
<th></th>
<th>H.264</th>
<th>MPEG-4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r_p$</td>
<td>$r_s$</td>
</tr>
<tr>
<td>NTIA VQM General</td>
<td>0.724</td>
<td>0.690</td>
</tr>
<tr>
<td>NTIA VQM VideoConf.</td>
<td>0.747</td>
<td>0.722</td>
</tr>
<tr>
<td>Modified DVQ</td>
<td>0.676</td>
<td>0.629</td>
</tr>
<tr>
<td>VSSIM</td>
<td>0.593</td>
<td>0.599</td>
</tr>
</tbody>
</table>

*Table 2: Performance of the VQMs, with reference to compression types*

Referring to Table 2, another characteristic that was seen for most of the video metrics was that there was a better correlation among the MPEG-4 video sequences compared to the H.264 video sequences. This is except the NTIA videoconferencing model that showed a better correlation for the H.264 video sequences than the MPEG-4 video sequences.
While the modified DVQ showed a less linear relationship, it was noted in [6] that the training database used was small compared to the databases used for the other algorithms. The VSSIM is a computationally-simple method which is consistent with many observations of HVS behavior, but does not address some, such as vertical distortions being more noticeable than horizontal distortions.

The best metric provides a correlation of about 0.716 to human visual subjective scores, which thus indicates the needs for further research in objective video quality metrics for multimedia videos. The test sequences used were all low-bit rate sequences. Most video metrics do not perform as well in the low bit-rate range as they do in the mid-high bit-rate range because different artifacts dominate in the lower bit-rate ranges. For example, noise would be less noticeable compared to blocking artifacts in the low bit-rate video. A characteristic of low bit-rate videos is that it results in serious blocking artifacts (e.g. MPEG-4) because block motion detection can straddle objects, and adjacent blocks may have very different estimated motion vectors. In the case of H.264 compressed videos, the low bit rate videos look blurred across block boundaries, mainly because of the use of in-loop filter. Overall, it can be seen that there is still a need to improve the video quality metrics to measure the quality rating of low bit-rate video sequences.

4. CONCLUSION

This paper has presented a description of three widely-known objective video quality metrics, followed by a comparison of their results and conclusions that can be drawn from the comparison.

This study suggests that current state-of-the-art objective video quality metrics that so far have been designed for quality measurement of TV signals with large frame sizes, full TV frame rates, and high compressed bit rates may not work well on multimedia videos with low bit rates, various frame rates, and small frame sizes. Thus, there is a need for further research in this area, and it is encouraging that a major step has been taken by the Video Quality Expert Group (VQEG) which is in the progress of investigating and consolidating contributions for video quality metrics designed for such multimedia videos [4].

5. REFERENCES