# Objective Video Quality Metrics for HDTV Services: A Survey

Reuben A. Farrugia [1], Chris Galea [2], Saviour Zammit [3], Adrian Muscat [4]

*Department of Communications and Computer Engineering*
*University of Malta*
*Msida, Malta*
[1] reuben.farrugia@um.edu.mt
[2] chris.p.galea@gmail.com
[3] saviour.zammit@um.edu.mt
[4] adrian.muscat@um.edu.mt

***Abstract*—The exponential growth of video traffic is expected to reach 62% of the global Internet traffic by the end of 2015 [1]. This presents as a significant challenge for the television service providers who need to employ networking technologies to monitor specific Quality of Service (QoS) parameters such as packet loss rate, jitter and delay, to ensure an acceptable level of quality. However, recent research has demonstrated that the quality experienced by the end-user does not correlate to the QoS parameters employed by most service providers [2].**

**This paper investigates the correlation between the QoS parameters and the quality perceived by the end. user. These results indicate that although the QoS parameters may sometimes achieve high correlation with respect to the quality perceived by the viewer, they still have large variances. This suggests that the QoS parameters are not enough to quantify the subjective quality with a high level of confidence. This work further compares a number of existing objective video quality metrics. The results presented in this paper show that the Full-Reference Motion based Video Integrity Evaluation (MOVIE) metric and the Spatio-Temporal Reduced Reference Entropic Differences (STRRED) metric achieve excellent correlation with the subjective scores. This research also demonstrates that the STRRED metric and its derivatives have several advantages over the MOVIE metric since less information needs to be transmitted and it is less computationally intensive.**

***Index Terms*—High definition television, IP networks, image and video quality metrics, quality of experience, quality of service, wireless networks**

## I. Introduction

Digital videos are continuously increasing in popularity thanks to the upsurge of applications such as video on demand (VoD), video teleconferencing, video sharing services such as YouTube, digital television, home videos, digital cinema, video streaming over the Internet and other Multimedia services [3]. In addition, advances in technology allow more efficient capture, transmission, storage and sharing of videos and enable new forms of multimedia such as Ultra high definition television (UHDTV) [2], 3D Television (3DTV) services and mobile TV [4]. However, video quality can be seriously affected during its course from a transmitter to the receiver as a result of impairments accumulated during processing, compression, storage, packet losses/bit errors, variations in data rate (jitter) and even distortions introduced by the display device itself [5].

Even though audio-visual service providers specify that they can guarantee specific Quality of Experience (QoE) criteria, recent studies have demonstrated that the quality experienced by the end-user does not correlate well to Quality of Service (QoS) parameters (namely packet loss rate, data rate, delay and jitter) commonly used today to quantify quality [2]. The best way to measure the quality perceived by a user is using subjective evaluation, where humans view the content and are asked for their opinion of the quality of the image or video. This is mainly due to the fact that humans can almost instantaneously assess the quality of the image or video that they are viewing using prior experience acquired from watching thousands or even millions of different images on a daily basis [3], [6]. However, subjective evaluation is time-consuming and cannot be integrated into automatic systems which adjust themselves on the fly according to the output quality. These issues have thus led to increased research for the development of reliable objective metrics focusing on prediction of the perceived video quality automatically. Subjective evaluations are however still the benchmark with which Image Quality Assessment (IQA) and Video Quality Assessment (VQA) algorithms are judged [3].

This paper evaluates the correlation between the QoS parameters commonly used today and the corresponding quality perceived by the end-user. The results presented in this work indicate that these parameters are correlated, however large variations are present which indicate that the QoS parameters on their own are not sufficient to quantify quality with a high level of confidence. This work further compares several objective video quality metrics which are found in literature and identifies that the Motion based Video Integrity Evaluation (MOVIE) metric and the Spatio-Temporal Reduced Reference Entropic Differences (STRRED) metric achieve excellent correlation with respect to the mean opinion scores (MOS). Moreover, the STRRED is less computational intensive and needs significantly less information to be transmitted as side information, making it more attractive for future implementa-

978-1-4673-2232-4/13/$31.00  

tions.

This paper is organized as follows: Section II provides an overview of the different objective video quality metrics found in literature while section III describes the testing methodology employed to compare the different video quality metrics considered. Section IV contains the major results which are discussed in some detail. The final comments and conclusion are drawn in the final section.

## II. Background

There are several objective video quality metrics that can be used to quantify the quality perceived by the end-user. These metrics can be categorized into three distinct groups [6]:

- *Full-reference (FR) quality metrics* which are computed using the original image/video as the reference
- *Reduced-reference (RR) quality metrics* where some information is transmitted as side information to improve the performance of the metric
- *No-reference (NR) quality metrics* which exploit the human vision system (HVS) properties and require only the received image/video.

The objective video quality metrics considered in this work are summarized in the following subsections.

### A. Full-Reference Quality Metric

Computation of FR metrics requires that both the received video and also the original (unimpaired) video be available. The Mean Square Error (MSE) and the related Peak Signal-to-Noise (PSNR) are two of the most common FR metrics used due to their simplicity and straightforward mathematical definition, and are computed using

$$\text{MSE} = \frac{1}{W \times H \times N} \sum_{t=1}^{T} \sum_{w=1}^{W} \sum_{h=1}^{H} (O_{t,w,h} - R_{t,w,h})^2 \quad (1)$$

$$\text{PSNR} = 10 \log_{10} \left( \frac{2^n - 1}{\text{MSE}} \right) \quad (2)$$

where $O_{i,w,h}$ represents the pixel at coordinates $(w, h)$ of the $i^{\text{th}}$ of the original video, $R_{i,w,h}$ is the corresponding pixel in the reference frame, $T$ represents the number of frames, $n$ is the bit depth and $H$ and $W$ represent the number of rows and columns in the frame respectively. Both these metrics have been criticised for not being reliable predictors of the image/video quality, and therefore more advanced schemes which take into consideration the human visual system (HVS) characteristics have been presented in literature [6].

The Structure Similarity Index (SSIM) [6] is based on the assumption that the HVS is highly adapted to extract structural information from a person's field of vision and as a result performs three types of comparisons between the reference and distorted image, namely luminance, contrast and structure to determine the loss of structural information. These components are then combined to obtain the overall similarity measure. The Multi-Scale SSIM (MS-SSIM) [7] extends the capabilities of SSIM by including a low-pass filter, down-sampled and indexed at multiple scales so that the overall SSIM evaluation is computed by combining measurements at various scales.

The authors in [8] have adopted advanced statistical models of natural images to derive the optimal perceptual weights. Information content weighting was used to improve the performance of PSNR and SSIM, where the best performance was registered when using Information content weighting with SSIM (IW-SSIM).

The Visual Signal-to-Noise (VSNR) metric [9] employs a two-stage approach to predict the quality of experience. The first stage adopts wavelet models to determine the presence of distortions by exploiting the HVS. The second stage is employed whenever the distortions are higher than a predefined threshold to model the visual quality of the perceived contrast and global precedence.

The authors in [10] have presented the Feature Similarity Index (FSIM) which employs phase congruency as the primary feature together with the image gradient magnitude. This metric is based on the fact that the human vision system primarily realizes an image using its low-level features. The FSIM quality metric can produce a quality score by considering both grayscale images and color images, with the latter metric FSIM$_c$ being computed using the YIQ color space model.

The Motion based Video Integrity Evaluation (MOVIE) [3] was specifically designed to assess the quality of videos. The MOVIE metric considers both spatial and temporal distortions and adopts Gabor receptive field models to dissemble video related information to multi-scale space-time primitives.

The Visual Information Fidelity (VIF) [11] adopts three models to determine the perceived quality of a distorted image. These are the Natural Scene Statistics (NSS), distortion and HVS models. Like some other image quality metrics, VIF was extended to work on videos using temporal derivatives and is denoted by V-VIF.

### B. Reduced Reference Quality Metrics

Reduced-reference metrics predict quality perception using only a limited amount of information from the reference video. The algorithm proposed in [12] predicts the quality score based on a natural image statistic model in the wavelet domain, using as a measure of image distortion the Kullback-Leibler distance between the marginal probability distributions of wavelet coefficients of the reference and impaired images.

The Video Quality Metric (VQM) [13] is a metric designed for videos at the National Telecommunications and Information Administration (NTIA). The algorithm requires only around 4% of the reference data and is based on measuring colour impairments and evaluating losses in spatial gradients of luminance components and features using the product of luminance contrast and motion.

The Reduced Reference Entropic Differencing (RRED) metric proposed in [14] was primarily designed to be used on images and measures the information changes between reference and distorted images by finding the differences in the entropies of wavelet coefficients of these images. The algorithm is also flexible in terms of the side-information transmitted. An

978-1-4673-2232-4/13/$31.00 ©2013 IEEE

extension to the RRED metric was proposed in [15] such that spatial and temporal entropic differences (SRRED and TRRED respectively) are both used. The STRRED is reported to be able to produce encouraging results suggesting that these two indices are considering complementary distortions and thus their combination is quite beneficial.

### C. No-Reference Quality Metrics

The Blind Image Quality Index (BIQI) [16] is based on Natural Scene Statistics and only requires a training stage before it can be used; that is, unlike several other no-reference metrics, no knowledge of the distortion type affecting the image/video to be evaluated is required. The algorithm is primarily intended for use on images and is flexible in the sense that it can be adapted to cater for several other distortions other than those considered in the original implementation (namely JPEG, JPEG2000, white noise, fast fading Rayleigh channel and Gaussian blur). The algorithm works by first determining the most probable type of distortion in an image using a multiclass Support Vector Machine (SVM) classifier with a radial-basis function (RBF) kernel and then quantifies the distortion using Natural Scene Statistics, to result in an image quality expressed as a probability-weighted summation.

The BLind Image Integrity Notator using Discrete Cosine Transform (DCT) Statistics (BLIINDS) [17] metric also does not require knowledge on distortion types and uses natural scene statistics of DCT coefficients to predict image quality. Since the features extracted by the algorithm are independent from the distortion type, these are quite successful across various types of distortions and thus BLIINDS can be used in several applications. Two versions of BLIINDS have been developed, namely BLIINDS-I and BLIINDS-II: BLIINDS-I uses no statistical modelling and a different set of DCT statistics than the considerably improved second model, BLIINDS-II.

The Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) index [4] also uses NSS to qualify the distortion affecting an image and quantify the amount of distortion present. In fact this index first identifies the distortion type and then performs quality assessment based on the distortion type deduced.

A recent no-reference metric proposed by the authors in [18] is the Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE), which also uses NSS to assess image quality in the spatial domain. This metric does not require computation of distortion-specific features but the features extracted can be used for distortion identification.

## III. Testing Methodology

The performance of the different QoS and objective video quality metrics mentioned in this paper were evaluated using a total of five databases, four of which were part of the Video Quality experts Group (VQEG) HDTV Phase I project [19] and one obtained from the Laboratory for Image and Video Engineering [20], [21].

The VQEG databases considered in this work contain subjective scores of every video contained in each database. The subjective scores were acquired using the five-scale single-stimulus Absolute Category Rating with Hidden Reference (ACR-HR) method, where the ACR scores given to the distorted videos were subtracted from ACR scores of the corresponding reference videos to obtain the DMOS values. All four databases contain 9 different source sequences, each processed by 15 distortions for a total of 135 impaired video sequences. The video sequences were compressed using either MPEG-2 or H.264/AVC, with impairments consisting of both coding-only and coding-with-transmission errors, where both wireless and IP networks were considered.

The LIVE video database consists of 10 different sources and 15 distortions, for a total of 160 videos, including the reference videos. The subjective scores were obtained using a single-stimulus procedure, where a continuous scale was adopted. The test sequences were created using four different types of distortion processes, namely H.264/AVC compression, MPEG-2 compression and simulated transmission of H.264/AVC coded bitstream through wireless and IP networks.

The VQEG databases include the QoS measurements, and therefore can be used to evaluate the performance of several QoS parameters with respect to the quality perceived by the end user. The QoS parameters considered in this work include packet loss rate, data rate and jitter. However, the VQEG databases were not used to evaluate the objective video quality metrics since several distorted videos have significant spatial and temporal differences with respect to the original video sequences. Compensation for these artefacts to ensure correct values obtained from the metrics used would take a considerable amount of time and was thus not performed. The LIVE database was used instead, since the videos available are already calibrated.

The performance of the quality metrics considered in this work was evaluated using the Spearman Rank Order Correlation Coefficient (SROCC) and the Pearson Linear Correlation Coefficient (LCC). The correlation was computed after the objective model data was non-linearly regressed with a four parameter monotonic logistic function following the recommendations in [19] , which is computed using

$$\hat{Q}_j = \beta_2 + \frac{\beta_1 - \beta_2}{1 + \exp\left(-\frac{Q_j - \beta_3}{|\beta_4|}\right)} \tag{3}$$

where $Q_j$ is the predicted quality of video $j$ by the metric, $\hat{Q}_j$ is the vector of fitted objective scores and the $\beta$ values are derived using least squares optimization. The SROCC and LCC are then computed between the vector $\hat{Q}_j$ and the subjective DMOS values. The optimal parameters $\beta$ and their initial values are computed following the recommendation in [19]. The objective scores were also linearly rescaled prior to using the above procedure to aid numerical convergence.

## IV. Results

The correlation of the different QoS metrics considered in this work are summarized in tables I - IV. The PSNR objective metric was included in these tables so that it can be used as a baseline to which the performance of the other objective video quality metrics could be compared. It can be seen that the QoS parameters, namely packet loss rate, data rate and jitter are correlated to the quality perceived by the end user. However, this correlation has a large variance and fluctuates between 0.4 and 0.9 for different databases. Moreover, it can be seen that the PSNR is generally more correlated to the subjective scores than the other QoS metrics and is generally more stable, especially for H.264/AVC encoded video.

TABLE I: Correlation of QoS parameters with subjective scores for the VQEGHD1 database

| QoS Metric | LCC | SROCC |
|---|---|---|
| Jitter | 0.7609 | 0.7546 |
| Packet Loss | 0.7330 | 0.6972 |
| Data Rate | 0.4007 | 0.3320 |
| **PSNR** | **0.8678** | **0.8627** |

TABLE II: Correlation of QoS parameters with subjective scores for the VQEGHD2 database

| Codec | QoS Metric | LCC | SROCC |
|---|---|---|---|
| H.264/AVC | Packet Loss | 0.5016 | 0.4670 |
| | Data Rate | 0.9512 | 0.9087 |
| | **PSNR** | **0.6415** | **0.6176** |
| MPEG-2 | Packet Loss | NA | NA |
| | Data Rate | 0.9159 | 0.9118 |
| | **PSNR** | **0.2344** | **0.2834** |

TABLE III: Correlation of QoS parameters with subjective scores for the VQEGHD3 database

| Codec | QoS Metric | LCC | SROCC |
|---|---|---|---|
| H.264/AVC | Packet Loss (CBR) | 0.7528 | 0.6762 |
| | Data Rate (CBR) | 0.4352 | 0.3984 |
| | Packet Loss (VBR) | NA | NA |
| | Data Rate (VBR) | 0.9317 | 0.9193 |
| | **PSNR (CBR)** | **0.9540** | **0.9203** |
| | **PSNR (VBR)** | **0.6893** | **0.6804** |
| MPEG-2 | Packet Loss (CBR) | 0.9325 | 0.8674 |
| | Data Rate (CBR) | NA | NA |
| | **PSNR (CBR)** | **0.9730** | **0.9195** |

TABLE IV: Correlation of QoS parameters with subjective scores for the VQEGHD5 database

| Codec | QoS Metric | LCC | SROCC |
|---|---|---|---|
| H.264/AVC | Packet Loss | 0.8935 | 0.8926 |
| | Data Rate | 0.5912 | 0.6620 |
| | **PSNR** | **0.7318** | **0.6494** |
| MPEG-2 | Packet Loss | NA | NA |
| | Data Rate | 0.4336 | 0.5050 |
| | **PSNR** | **0.8222** | **0.8212** |

Tables V - VI show the correlation of the objective video quality parameters with respect to the subjective scores. The MSE and the related PSNR were found to be quite poor when considering the LIVE database, with correlation below 0.6 for both SROCC and LCC. This affirms that the classical MSE and PSNR metrics are quite unreliable in predicting human subjectivity. The best full-reference metrics considered in this research are the MOVIE and its components, namely Spatial MOVIE and Temporal MOVIE which performs approximately 1.5 times better than PSNR. This is mainly attributed to the fact that MOVIE adopts important human vision perception related models in space and time and a model for visual motion processing. These results also indicate that using the temporal information is important for Video Quality Assessment.

The reduced reference STRRED is remarkable since it achieves correlation comparable to that of the MOVIE index. However, the major advantage of the STRRED metric is that it needs to transmit significantly less information with respect to the other full reference approaches. In addition, the STRRED metric and its derivatives are also adjustable in terms of the amount of side-information required by the decoder, and is reported to achieve significantly high correlation while significantly reducing the number of scalars per frame to be transmitted. An implementation where only a single coefficient per frame is transmitted, which despite the significantly reduced amount of information required still achieved a respectable performance that is superior to several of the full-reference schemes evaluated. The STRRED is computationally simpler with respect to the MOVIE metric and therefore might be viable to be employed in real-time applications. The results of SRRED, TRRED and STRRED also reaffirm that both spatial information and in particular temporal information offer complementary information which is important for the accurate prediction of video quality.

The performance of all no-reference quality metric schemes considered in this work was relatively poor. These metrics were found to perform well when tested using images but fail when considering videos. This poor performance is mainly attributed to the fact that they were designed for image quality assessment, and therefore temporal information is not considered. Also, the distortion types considered are specific to image-related impairments only. These results affirm that no-reference metrics are still an open research problem and much work is needed for no-reference quality assessment, especially for video sequences.

## V. Conclusion

This work has demonstrated that objective video quality metrics are typically more reliable than QoS metrics since the correlation of the latter was found to be quite unstable achieving correlations as high as 0.95 in some instances and as low as 0.33 in other cases. These results could also be explained by the distributions of the DMOS scores (Fig. 1) considered in the VQEG databases upon which the QoS metrics evaluation was performed. It can be seen that the Live database includes more videos which are considered to provide medium quality, while the VQEG database has more videos containing excellent quality. The high correlation of PSNR with subjective data on three of the four VQEG databases could also be explained by this phenomenon, since when

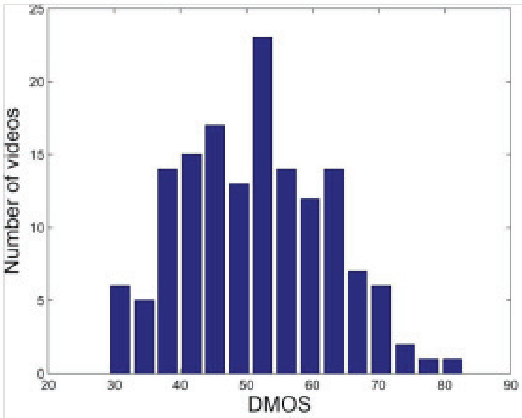
978-1-4673-2232-4/13/$31.00 ©2013 IEEE

TABLE V: SROCC Correlation of objective video quality parameters with subjective scores for the LIVE database

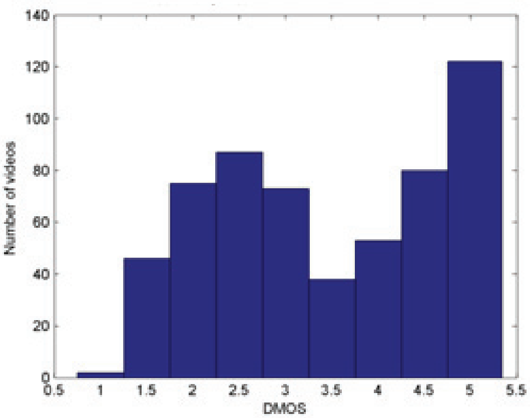
| | Metric | Wireless (H.264) | IP (H.264) | H.264 | MPEG-2 | All Data |
|---|---|---|---|---|---|---|
| FR | MSE | 0.6574 | 0.4167 | 0.4585 | 0.3862 | 0.5398 |
| | PSNR | 0.6574 | 0.4167 | 0.4585 | 0.3862 | 0.5398 |
| | SSIM | 0.6553 | 0.6182 | 0.7129 | 0.6652 | 0.6947 |
| | MS-SSIM | 0.7328 | 0.6792 | 0.7311 | 0.6701 | 0.7479 |
| | IW-SSIM | 0.7326 | 0.6770 | 0.6756 | 0.7116 | 0.7417 |
| | IW-MSE | 0.7131 | 0.6418 | 0.6152 | 0.5694 | 0.6860 |
| | IW-PSNR | 0.7174 | 0.5853 | 0.5794 | 0.5688 | 0.6536 |
| | FSIM | 0.7413 | 0.7090 | 0.6944 | 0.6941 | 0.7318 |
| | FSIMc | 0.7291 | 0.6970 | 0.6717 | 0.6846 | 0.7175 |
| | VIF | 0.5317 | 0.5506 | 0.6349 | 0.6331 | 0.5541 |
| | V-VIF | 0.5507 | 0.4736 | 0.6807 | 0.6116 | 0.5710 |
| | VSNR | 0.6951 | 0.6390 | 0.6405 | 0.5874 | 0.6726 |
| | MOVIE | 0.8109 | 0.7157 | 0.7664 | 0.7733 | 0.7890 |
| | Spatial MOVIE | 0.7927 | 0.7046 | 0.7066 | 0.6911 | 0.7270 |
| | Temporal MOVIE | 0.8114 | 0.7192 | 0.7797 | 0.8170 | 0.8055 |
| RR | RR IQA [12] | 0.0790 | -0.1622 | 0.6043 | 0.5502 | 0.2105 |
| | VQM | 0.7214 | 0.6383 | 0.6520 | 0.7810 | 0.7026 |
| | RRED | 0.6143 | 0.5720 | 0.5784 | 0.6731 | 0.6133 |
| | SRRED | 0.7925 | 0.7624 | 0.7542 | 0.7249 | 0.7592 |
| | TRRED | 0.7765 | 0.7513 | 0.8189 | 0.5879 | 0.7802 |
| | STRRED | 0.7857 | 0.7722 | 0.8193 | 0.7193 | 0.8007 |
| | STRRED (1 coeff) | 0.7208 | 0.5075 | 0.7197 | 0.7247 | 0.7319 |
| NR | BRISQUE | 0.0098 | 0.1168 | 0.1837 | 0.2785 | 0.0785 |
| | BLIINDS-II | -0.3618 | -0.3916 | 0.4794 | 0.4400 | -0.1781 |
| | BIQI | 0.1231 | -0.0830 | 0.2831 | 0.3793 | 0.0614 |
| | DIIVINE | -0.0328 | 0.1697 | 0.2677 | 0.2761 | 0.1010 |

TABLE VI: LCC Correlation of objective video quality parameters with subjective scores for the LIVE database

| | Metric | Wireless (H.264) | IP (H.264) | H.264 | MPEG-2 | All Data |
|---|---|---|---|---|---|---|
| FR | MSE | 0.6703 | 0.4609 | 0.5724 | 0.3754 | 0.5576 |
| | PSNR | 0.6698 | 0.4689 | 0.5725 | 0.4045 | 0.5621 |
| | SSIM | 0.6718 | 0.6919 | 0.7423 | 0.6862 | 0.7065 |
| | MS-SSIM | 0.7362 | 0.7331 | 0.7369 | 0.6942 | 0.7551 |
| | IW-SSIM | 0.7326 | 0.7492 | 0.6650 | 0.7346 | 0.7493 |
| | IW-MSE | 0.7180 | 0.6661 | 0.6620 | 0.5997 | 0.6941 |
| | IW-PSNR | 0.7127 | 0.6299 | 0.6246 | 0.5753 | 0.6785 |
| | FSIM | 0.7389 | 0.7496 | 0.7002 | 0.7211 | 0.7402 |
| | FSIMc | 0.7314 | 0.7264 | 0.6839 | 0.7014 | 0.7254 |
| | VIF | 0.5863 | 0.5993 | 0.6513 | 0.6609 | 0.5683 |
| | V-VIF | 0.5488 | 0.5102 | 0.6911 | 0.6145 | 0.5756 |
| | VSNR | 0.6975 | 0.7372 | 0.6501 | 0.5880 | 0.6885 |
| | MOVIE | 0.8386 | 0.7622 | 0.7902 | 0.7595 | 0.8116 |
| | Spatial MOVIE | 0.7883 | 0.7378 | 0.7252 | 0.6587 | 0.7451 |
| | Temporal MOVIE | 0.8371 | 0.7383 | 0.7920 | 0.8252 | 0.8217 |
| RR | RR IQA [12] | 0.0302 | 0.0011 | 0.5217 | 0.5516 | 0.2419 |
| | VQM | 0.7325 | 0.6480 | 0.6459 | 0.7860 | 0.7236 |
| | RRED | 0.6013 | 0.6478 | 0.6070 | 0.6981 | 0.6342 |
| | SRRED | 0.8067 | 0.7977 | 0.7540 | 0.7415 | 0.7732 |
| | TRRED | 0.7795 | 0.7713 | 0.8324 | 0.6203 | 0.7854 |
| | STRRED | 0.8041 | 0.7919 | 0.8237 | 0.7474 | 0.8054 |
| | STRRED (1 coeff) | 0.7340 | 0.5549 | 0.7138 | 0.7467 | 0.7325 |
| NR | BRISQUE | 0.0013 | 0.1465 | 0.1172 | 0.2500 | 0.1972 |
| | BLIINDS-II | -0.4843 | -0.5093 | 0.3261 | 0.5146 | -0.1513 |
| | BIQI | -0.1448 | -0.0516 | 0.2389 | 0.3245 | 0.0362 |
| | DIIVINE | -0.0241 | 0.1956 | 0.2004 | 0.2456 | 0.1056 |




(a) LIVE Database




(b) VQEG Database

Fig. 1: Histograms of the DMOS subjective scores for the Fig. 1a Live Database and Fig. 1b VQEG Database.

tested on the LIVE Video database the PSNR and related MSE achieved correlations below 0.6 in terms of both SROCC and LCC. They were overall also worse than all the algorithms considered except the no-reference metrics. Future work will consider the evaluation of more databases and other video content including 3DTV.

From the 25 algorithms considered in this project, the MOVIE and Temporal MOVIE full-reference indices were found to be the best-performing metrics with overall correlations of 0.7890 and 0.8055 in terms of SROCC respectively and correlations of 0.8116 and 0.8217 in terms of LCC respectively. The reduced-reference metric STRRED however achieved performance close to these two full-reference metrics with SROCC and LCC correlations of 0.8007 and 0.8054 respectively. It was also superior to MOVIE and Temporal MOVIE in the case of the ’IP’ and ’H.264’ distortion categories. However, since STRRED is a reduced-reference metric, less information needs to be transmitted to a receiver for the latter to be able to judge the perceived image quality; the optimal performance resulted in data that is approximately 576 times smaller than if a full-reference metric was implemented for the video sequences in the LIVE video database. STRRED was in fact designed to be adjustable in terms of the side-information sent and even when using just a single coefficient, performance was superior to that of several full-reference metrics. This metric is thus very promising considering the high bandwidth requirements of full-reference metrics and the poor performance of no-reference algorithms, the latter which require no information to be transmitted. However, it should be noted that the no-reference metrics evaluated in this project were designed and tested mainly for image quality assessment. The importance of designing algorithms specifically for video quality assessment was in fact also discussed, since temporal information and visual motion processing prediction were found to be factors which can significantly improve the metrics’ performance.

## VI. Acknowledgements

The authors would like to thank the Malta Communications Authority (MCA) for financing this project.

## References

[1] “Cisco Visual Networking Index: Forecast and Methodology, 2010-2015,” June 2011.

[2] Touradj Ebrahimi, “Quality of multimedia experience: past, present and future,” in *Proceedings of the 17th ACM international conference on Multimedia*, New York, NY, USA, 2009, MM ’09, pp. 3–4, ACM.

[3] Kalpana Seshadrinathan, Rajiv Soundararajan, Alan Conrad Bovik, and Lawrence K. Cormack, “Study of subjective and objective quality assessment of video,” *Trans. Img. Proc.*, vol. 19, no. 6, pp. 1427–1441, June 2010.

[4] Anush Krishna Moorthy and Alan Conrad Bovik, “Blind image quality assessment: From natural scene statistics to perceptual quality,” *Trans. Img. Proc.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[5] R.A. Farrugia and C.J. Debono, “A support vector machine approach for detection and localization of transmission errors within standard h.263++ decoders,” *Multimedia, IEEE Transactions on*, vol. 11, no. 7, pp. 1323 –1330, nov. 2009.

[6] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 13, no. 4, pp. 600–612, 2004.

[7] Z. Wang, E.P. Simoncelli, and A.C. Bovik, “Multiscale structural similarity for image quality assessment,” in *Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on*, nov. 2003, vol. 2, pp. 1398 – 1402 Vol.2.

[8] Zhou Wang and Qiang Li, “Information content weighting for perceptual image quality assessment,” *Trans. Img. Proc.*, vol. 20, no. 5, pp. 1185–1198, May 2011.

[9] D. M. Chandler and S. S. Hemami, “Vsnr: A wavelet-based visual signal-to-noise ratio for natural images,” *Trans. Img. Proc.*, vol. 16, no. 9, pp. 2284–2298, Sept. 2007.

[10] Lin Zhang, Lei Zhang 0006, Xuanqin Mou, and David Zhang, “Fsim: A feature similarity index for image quality assessment.,” *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[11] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *Image Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[12] Zhou Wang and Eero P. Simoncelli, “Reduced-reference image quality assessment using a wavelet-domain natural image statistic model,” in *in Proc. of SPIE Human Vision and Electronic Imaging*, 2005, vol. 5666, pp. 149–159.

[13] M. H. Pinson and S. Wolf, “A new standardized method for objectively measuring video quality,” *IEEE Transactions on broadcasting*, vol. 50, no. 3, pp. 312–322, Sept. 2004.

[14] Rajiv Soundararajan and Alan C. Bovik, “Rred indices: Reduced reference entropic differencing for image quality assessment,” *Trans. Img. Proc.*, vol. 21, no. 2, pp. 517–526, Feb. 2012.

[15] Rajiv Soundararajan and Alan C. Bovik, "Spatio-temporal reduced reference entropic differencing (rred) index," *Trans. Circuits and Systems for Vudei Technology*, in review.

[16] A.K. Moorthy and A.C. Bovik, "A two-step framework for constructing blind image quality indices," *Signal Processing Letters, IEEE*, vol. 17, no. 5, pp. 513 –516, may 2010.

[17] M.A. Saad, A.C. Bovik, and C. Charrier, "A dct statistics-based blind image quality index," *Signal Processing Letters, IEEE*, vol. 17, no. 6, pp. 583 –586, june 2010.

[18] A Mittal, Moorthy A.K., and Alan C. Bovik, "No-reference image quality assessment in the spatial domain," *Trans. Image Processing*, to appear.

[19] A. Takashi, Schmidmer C., C. Lee, F. Speranza, J. Okamoto, K. Brunnstrm, L. Janowski, M. Barkowsky, M. Pinson, Q.H. Staelens, N. Thu, R. Green, R. Bitto, R. Renaud, S. Borer, T. Kawano, Baroncini V., and Dhondt Y., "Report on the validation of video quality models for high definition video content," Tech. Rep., 2010.

[20] H.R. Sheikh, K. Seshadrinathan, A.K. Moorthy, Z. Wang, A.C. Bovik, and L.K. Cormack, "Image and Video Quality Assessment at LIVE," 2012.

[21] Kalpana Seshadrinathan A, Rajiv Soundararajan B, Alan C. Bovik B, and Lawrence K. Cormack B, "A subjective study to evaluate video quality assessment algorithms," 2010.