A No-Reference Video Quality Metric Using a Natural Video Statistical Model

Christian Galea
Department of Communications and Computer Engineering, University of Malta, Msida, MSD2080, Malta
E-mail: christian.galea.09@um.edu.mt

Reuben A. Farrugia
Department of Communications and Computer Engineering, University of Malta, Msida, MSD2080, Malta
E-mail: reuben.farrugia@um.edu.mt

***Abstract*— The demand for high quality multimedia content is increasing rapidly, which has resulted in service providers employing Quality of Service (QoS) strategies to monitor the quality of delivered content. However, the QoS parameters commonly used do not correlate well with the actual quality perceived by the end-users. Numerous objective video quality assessment (VQA) metrics have been proposed to address this problem. However, most of these metrics rely on the availability of additional information from the original undistorted video to perform adequately, which will increase the bandwidth required. This paper presents a No-Reference (NR) VQA algorithm, which extracts a Natural Video Statistical Model using both spatial and temporal features to model the quality experienced by the end-users without needing additional information from the transmitter. These features are based on the observation that the statistics of natural scenes are regular on pristine content but are significantly altered in the presence of distortion. The proposed method achieves a Spearman Rank Order Correlation Coefficient (SROCC) of 0.8161 with subjective data, which is statistically identical and sometimes superior to existing state-of-the-art full and reduced reference VQA metrics.**

Keywords — no-reference video quality assessment, quality of experience, natural statistics, visual perception

I. Introduction

The continuous advances being made in technology have allowed electronic devices to become smaller yet more powerful, arguably turning them into portable computers. This has in turn led to the provision of a number of Multimedia services such as video on demand (VoD), video sharing, videoconferencing, digital television and video streaming over the Internet [1], which has increased the popularity of videos. However, video quality is significantly affected by the amount of packets lost during transmission [2]–[8]. Service providers use Quality of Service (QoS) approaches to monitor the quality experienced by the end-users. Nevertheless, recent research has demonstrated that the quality perceived by humans does not correlate to the QoS parameters employed by service providers [2].

Recent research is focussing on the development of algorithms designed to model human perception of quality. Existing Video Quality Assessment (VQA) metrics can be categorised into three groups: Full-Reference (FR) metrics directly compare the original and received video content, Reduced-Reference (RR) metrics only use partial information from the reference while No-Reference (NR) metrics operate solely on the received content [9]. Most of the methods found in literature rely on the former two approaches that increase the bandwidth requirements while the performance of NR metrics is generally poor due to the limited amount of information available at the receiver.

The most commonly used FR metrics are the Peak Signal-to-Noise Ratio (PSNR) [5] and the Structural SIMilarity index (SSIM) [6]. One of the best extensions to SSIM is Multi-scale SSIM (MS-SSIM) [10] since it is computed over multiple scales to cater for different viewing conditions. The Feature SIMilarity (FSIM) and related $FSIM_C$ [11] indices use low-level features to understand an image and hence employ measures for gradient magnitude and phase congruency. The reduced-reference metrics of Li and Wang [12] and the RR Entropic Differences (RRED) metric [13] are based on a Gaussian Scale Mixture (GSM) model of wavelet coefficients. These metrics utilise Natural Scene Statistics (NSS) modeling, which is based on the observation that images acquired from the natural environment have statistical properties which are perturbed in the presence of distortions [13], [14].

The No-Reference metrics found in literature generally exploit NSS to model the quality perceived by the end user. The Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) [3] metric uses steerable pyramid decomposition to extract features based on sub-band statistics. The BLind Image Integrity Notator using DCT Statistics (BLIINDS) [15] also uses statistical modelling, via the Multivariate Gaussian (MVG) distribution. The Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [14] is based in the spatial domain and considers the luminance coefficients whose statistical characteristics change in the presence of distortions. This is quantified through a Generalised Gaussian Distribution (GGD) applied on the coefficients of each pixel. The Naturalness Image Quality Evaluator (NIQE) [16] metric uses the same features of BRISQUE but employs a different feature pooling process, which virtually does not require any training.

For VQA, the above mentioned image quality assessment (IQA) metrics can be merely applied on each frame and then the frame-level scores are averaged for the final quality rating [1]. However, recent research has demonstrated that the exploitation of temporal information is necessary if the performance of VQA metrics is to be made robust [1], [4], [7], [9], [17], [18]. State-of-the-art metrics designed specifically for VQA that consider temporal and/or motion information include the full-reference MOtion-based Video Integrity Evaluation (MOVIE) algorithm [1], which is primarily based on Gabor coefficients obtained via linear decomposition. The full reference Video MS-SSIM (ViMSSIM) metric [18]

978-1-4799-8569-2/15/$31.00 ©2015 IEEE

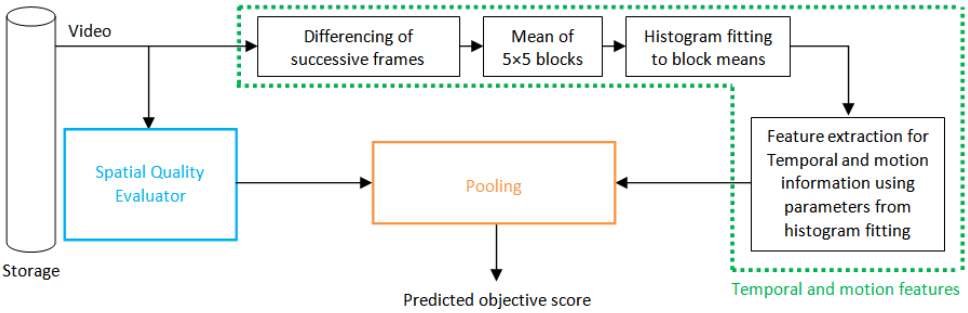


Fig. 1. Framework of the proposed NR-VQA system.

is an extension of MS-SSIM and employs a moving average window to pool MS-SSIM scores after they have been obtained from each frame. Temporal degradations are obtained by applying MS-SSIM on frame differences. The RR Spatio-Temporal RRED (STRRED) metric [17] is based on the RRED metric and is one of the leading metrics currently available. Lastly, the NR Video-BLIINDS metric [19] is an extension of BLIINDS-II [15] but also quantifies motion through motion coherency and egomotion. Spatial quality is specifically assessed with the NIQE metric. Video-BLIINDS, to the best of the authors' knowledge, is virtually the only NR-VQA metric designed to operate robustly on multiple distortions and the first to consider frame difference statistics.

Unlike existing methods found in literature, the work proposed in this paper combines both spatial- and frequency-domain features to obtain frame-level scores. In addition, the pooling strategy adopted by current metrics only captures the overall quality of the video by assigning equal importance to each frame. The pooling strategy proposed in this work augments this with the fact that the HVS is more sensitive to large distortions [18]. The temporal and motion features extracted are then combined with the spatial scores to obtain the final quality score of the video. Simulation results demonstrate that the proposed method achieves a Spearman Rank Order Correlation Coefficient (SROCC) of 0.82, which makes it statistically identical and sometimes superior to existing state-of-the-art full and reduced reference VQA metrics despite using less videos for training than other NR metrics.

The rest of this paper is organised as follows: the proposed metric is described in Section II and evaluated in Section III. Concluding remarks and proposals for future work are then given in Section IV.

II. Proposed Objective Video Quality Metric

The framework of the proposed system is shown in Fig. 1, demonstrating that videos are processed by two main blocks. The *Spatial Quality Evaluator* block extracts spatial and frequency domain features and models the quality of each frame using a Support Vector Regressor (SVR). The *Temporal and motion features* block extracts temporal and motion information based on the statistics of frame differences across video sequences. The resultant features are then passed to the *Pooling* block where the spatial scores are also pooled across an entire video in two ways to yield two spatial features per video. All the features are combined in another SVR to obtain the final video quality score. Details on each of these blocks are provided in the following subsections.

A. Spatial Quality Evaluator

A number of spatial features that provide some of the highest performance in the literature were extensively evaluated and the set of features that provided the best performance were chosen. A total of 55 features were found to provide the best performance on the TID 2008 database [20], which is one of the largest image databases and contains a wide spread of distortions. The first 36 features considered are those extracted by BRISQUE, which are computed over two different scales [14] and complemented with the spatial score provided by NIQE [16]. The authors of [21] noted that the combination of spatial- and frequency-based features could yield substantial improvements. However, very few metrics combine these two feature types. Hence, the DIIVINE [3] metric was considered, since it is based on frequency domain features and was found to achieve good correlation in terms of image quality assessment [3]. However, the most discriminative features for DIIVINE were found to be features 13-24 [3] corresponding to the shape parameter obtained after fitting a GGD to each of the 12 sub-band coefficients. As a result, only these features from DIIVINE are used. The GGD is defined using:

$$f(x|\sigma,\gamma,\mu) = \alpha e^{-(\beta|x-\mu|)^{\gamma}} \tag{1}$$

where μ is the mean and γ is the shape parameter while α and β are normalising and scale parameters as follows:

$$\alpha = \frac{\beta\gamma}{2\Gamma(1/\gamma)} \tag{2}$$

$$\beta = \frac{1}{\sigma}\sqrt{\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}} \tag{3}$$

where σ is the standard deviation and Γ is the ordinary gamma function.

In this evaluation, the Phase Congruency (PC) and Gradient Magnitude (GM) features adopted by $FSIM_C$ [11] were found to be among the most discriminative features. In $FSIM_C$, the image quality was evaluated by computing PC and GM on both the reference and distorted images. The proposed method only computes PC and GM on the distorted image to eliminate the requirement of the reference image.

Another source of distortions commonly present in compressed videos is blocking artefacts. However, none of the features discussed so far explicitly consider this artefact. As a result, the *blockiness* measure proposed by Chen and Bloom

[22] was implemented due to its reportedly good performance.

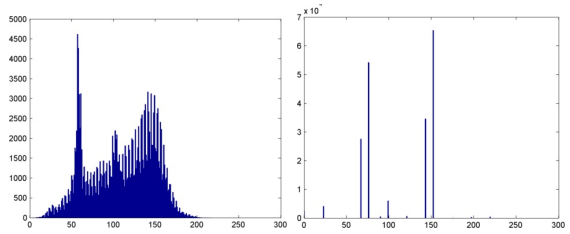
The above-mentioned features were found incapable to detect distortions caused by quantisation noise. Fig. 2(a) and 2(b) depict the histograms of an undistorted and distorted image, respectively. It can be seen that the histogram of a distorted image is generally more sparse. This observation can be exploited in order to improve performance, using three features as follows: (i) Q1: Difference between largest and smallest non-zero bins, (ii) Q2: Number of bins containing non-zero amplitudes, divided by Q1, (iii) Q3: Highest histogram amplitude divided by the total number of pixels. Although these features were primarily designed to capture quantisation noise, it can be shown that they are also beneficial to characterise other distortions such as contrast change.

B. Temporal and Motion features

The success of natural statistics in IQA has led to the hypothesis that they can also be applicable in the temporal dimension. As shown in Fig. 3, the histogram of frame difference coefficients of an undistorted natural sequence is considerably different than those of distorted versions, and thus can be suitable to evaluate the quality of the video. The distribution also reveals that the coefficients are symmetrically distributed and that varying levels of peakedness and spread are exhibited according to the distortions present. In general, as the amount of distortion increases, the histogram becomes wider. Hence, frame differences are modelled using the GGD given in (1), which encompasses a range of tail behaviors [19]. Specifically, in order to capture local statistics, each frame difference was partitioned into 5×5 blocks and the mean for each block was found. The GGD was then fitted on these values using the method in [23] to yield three parameters: σ^2, γ and μ.

It was noted that the γ values were able to capture significant changes in quality and hence the absolute difference between consecutive γ values was computed to capture these sharp changes. Measures such as the highest peak and mean and standard deviation of the differenced γ were used to quantify these distortions based on the fact that humans typically penalise even few quantities of high distortions heavily [4], [18]. Obviously, even good quality regions affect video quality perception and thus measures such as the mean of peaks below the mean of all the differenced γ were also implemented. The frequency of occurrence of high distortions was also considered by finding the mean distance between large peaks, as shown in Table I.

Similar measures were also applied on the σ^2 and μ parameters since they were also found to capture high amounts of distortion, although they emphasise different types of artefacts. For example, σ^2 seems mostly suited to capture large relatively homogeneous artefacts. The geometric mean between motion-related features and the statistical parameters was also computed to capture motion information. The geometric mean was chosen because it is able to find the central tendency of sets of numbers that have different numeric ranges, and hence all features combined have an equal contribution to the final result. Since at most two features are combined in the proposed metric, the general equation for the geometric mean may be simplified as follows:



(a) (b)

Fig. 2. Comparison between an original image and a quantisation-noise-distorted version of it, acquired from the TID2008 database [20]: (a) Histogram of an undistorted image; (b) Histogram of an image distorted with quantisation noise. Histograms computed on luminance channels.

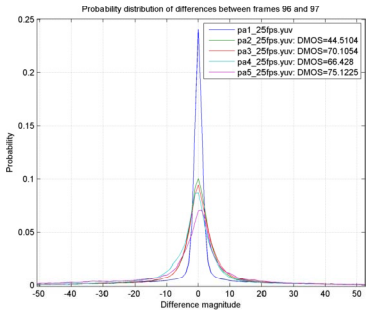


Fig. 3. Empirical probability distributions of the difference between two frames of videos in the LIVE Video database [1], [24], where the two frames selected correspond to a sharp change in quality for the distorted videos. 'pa1_25fps.yuv' denotes the reference; others are distorted versions of it.

$$GMM(\mathbf{f}_1, \mathbf{f}_2) = \sqrt{\mathbf{f}_1 \times \mathbf{f}_2} \tag{4}$$

where $\times$ represents the element-by-element multiplication between the two vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ whose dimensionality depends on the duration of the video sequence.

Although most features were extracted using the luminance channel, the chrominance channels were found to provide complementary information where sharp variations in intensity are present in distorted sequences. The same features described above were computed on the two chrominance channels which were combined such that the highest values (representing the largest distortions) captured by either channel are used.

The mean absolute value of consecutive frames was also computed on the luminance channel to capture sudden global changes resulting from local flicker. The absolute mean of all these values over an entire video sequence is used as another feature for prediction. The first feature considering motion is obtained by finding the number of blocks whose mean absolute values are less than the global mean absolute value. This feature is henceforth referred to as the Local and Global Motion Indicator (LGMI), such that frame differences which contain a few areas of motion and several stationary areas have a high LGMI value (indicating that viewers are likely to focus on these few areas of motion) whereas frame differences which are composed mostly of either high motion or low motion are given a low LGMI value. In the latter case, if the majority of the blocks in a frame difference have similar levels of motion, then eye fixations may focus on any part of the frame.

To consider global motion, the Cross-Correlation Coefficient (XCC) was also considered by finding the highest

cross-correlation value between successive frames. For frames containing low motion and are thus similar to each other, the value is expected to be high and thus indicates the global amount of motion in contrast to LGMI that primarily measures local motion. The temporal and motion features adopted in the proposed methods are summarised in Table I.

TABLE I. FEATURES USED FOR TEMPORAL QA. THE TERM 'DIFFERENCE' REFERS TO COMPUTING DIFFERENCES BETWEEN ADJACENT ELEMENTS OF *F*

Feature #	Pooling scheme	Feature Vector (*f*)
1	Highest peak of *f*	Absolute difference of γ
2	Mean of *f*	
3	Standard deviation of *f*	
4	Number of peaks < Feature B	Absolute difference of GMM(γ,LGMI)
5	Number of peaks > Feature A	
6	Feature A / Feature B	
7	Mean distance of the locations of the peaks of *f*>Feature A	
A	Mean of peaks>the mean of *f*	
B	Mean of peaks<the mean of *f*	
8	Mean distance of the locations of the peaks of *f*>Feature A	Difference of γ
9	Maximal value after obtaining the highest peak of both chroma channels from *f*	Absolute Difference of γ on both chroma channels
10	Highest peak of *f*	Absolute difference of σ^2
11	Mean of *f*	
12	Standard deviation of *f*	σ^2
13	Highest peak of *f*	Absolute difference of μ
14	Mean of *f*	
15	Standard deviation of *f*	
16	Highest peak of *f*	Absolute difference of GMM(γ, σ^2)
17	Standard deviation of *f*	
18	Mean of *f*	Absolute difference of average luminance
19	Highest peak of *f*	GMM(absolute difference of GMM(γ,XCC), absolute difference of GMM(σ^2,XCC))
20	Lowest value after obtaining the highest peak of both chroma channels from *f*	Absolute difference of GMM(γ, σ^2) on both chroma channels
21	Highest value after obtaining the highest peak of both chroma channels from *f*	Absolute difference of σ^2 on both chroma channels
22	Highest value after obtaining the highest peak of both chroma channels from *f*	Absolute difference of μ on both chroma channels
23	Number of peaks<Feature B	Absolute difference of GMM(μ,LGMI)
24	Mean of peaks>Feature A	

C. Pooling

Two distinct pooling stages have been implemented, one of which is dedicated for the spatial features. Specifically, when implemented in the VQA metric, the frame-level scores are pooled to obtain a single score for each video sequence. Two pooling schemes are used, the first of which is a moving average window which exploits the fact that humans typically focus on the worst distortions [18] and is computed as follows:

$$S_1 = \frac{1}{p}\sum_{i=1}^{p} M_i \tag{5}$$

and for n=1, 2, ..., N-p:

$$S_{n+1} = \varepsilon M_{n+p} + (1-\varepsilon)S_n \tag{6}$$

where M_i are the spatial scores to be pooled, p determines the number of frames to be considered and ε is a smoothing factor selected using:

$$\varepsilon = \eta/(p+1) \tag{7}$$

In the proposed system, p was chosen to represent 500 ms of the video similar to the approach used by the authors of [18] and η was selected to be 0.05. The maximum S_n, corresponding to the lowest quality, is chosen to represent the pooled spatial score. The second feature is the mean of the frame-level scores, to consider the quality of the entire video sequence as used in most other metrics.

The second and final pooling stage combines the two pooled spatial features and the 24 temporal/motion features. In this case, a logarithmic function is first applied on all 26 features to cater for the non-linearity of the HVS and supra-threshold effects [7], [12], [13], [19], such that a feature X becomes $\log(1+X)$. The resultant features are then passed through a SVR module to yield the Natural Video Statistical Model. The features from a subset of videos are used for training while the features of the remaining videos are used for testing, where the predicted objective quality scores of the latter set are obtained.

III. SYSTEM EVALUATION

To evaluate the proposed system, 50% of videos in a given database are used for training while the rest are used for testing and 100 random train/test combinations were considered. The metric is compared to some of the most popular and best-performing metrics currently available in the literature using SROCC and Pearson/Linear Correlation Coefficient (PLCC). Only the median correlations over all the train/test results are reported in accordance to what has been done previously in the literature. The PLCC was obtained after non-linear regression was performed using the model in [24] for IQA metrics and the model in [1], [25], [26] for VQA metrics.

Note that results for individual distortion categories were obtained by first performing training/testing on the entire dataset and then categorizing each video to a particular distortion. The performance on each distortion category was finally evaluated individually. Hence, training was not performed on each distortion type separately and as a result the algorithm does not have implicit knowledge of the distortion to be expected. Unless otherwise stated, reference videos are not used in the computation of the results since FR and RR algorithms have access to the original content and consequently have a significant advantage compared to NR metrics. Lastly, the Analysis of Variance (ANOVA) test is also used to determine if differences in SROCC correlations between the various algorithms are statistically significant, at the 95% confidence level.

In order to evaluate the performance of the spatial features, the spatial features in Section II-A were combined using a SVR to yield a NR-IQA metric. The performance of the proposed NR-IQA metric is summarised in Table II where it is compared to several state-of-the-art IQA metrics on the TID2008 database [20] and LIVE Image database [6], [26] containing 1700 distorted images over 17 distortion types and 779 distorted images over 5 distortion types, respectively. This proposed metric is statistically superior to all metrics considered on the TID2008 database except MS-SSIM, to which it is statistically equivalent, and $FSIM_C$, to which it is inferior. On the LIVE Image database, the proposed evaluator statistically outperforms all metrics except BRISQUE, MS-SSIM and $FSIM_C$, to which it is statistically equivalent. The performance of the proposed metric is remarkable considering that, in contrast to MS-SSIM and $FSIM_C$, no information from the reference is used. Although the NR BRISQUE achieves high correlation on the LIVE Image database, this is mainly because the data provided was a result of training the metric on the entire database. Moreover, the inclusion of the DIIVINE features in the SVR module in addition to spatial features yielded a statistically significant gain in correlation with subjective data of around 0.036, validating the fact that combining spatial and frequency-based features is beneficial.

The performance of the proposed NR-VQA metric was evaluated using the LIVE Video quality database [1], [24], where ten reference videos are distorted with four distortions, where each distortion type is represented by 30-40 videos. There are 15 distorted versions of each reference video for a total of 150 distorted videos. The metric is also compared to state-of-the-art algorithms proposed in the literature, including the RR-VQA STRRED 'single number' (STRREDsn) metric which is almost NR since only one scalar from the reference is required [17]. The results are summarised in Tables III-V.

When all videos excluding MPEG-2 are considered, the proposed metric is statistically equivalent to the top-performing FR- and RR-VQA metrics. The proposed metric also achieves the highest median correlation on simulated wireless transmission in terms of both SROCC and PLCC. These are remarkable achievements since, in contrast to the FR and RR algorithms, the proposed metric does not utilise any information from the reference.

The proposed metric is also compared to the NR-VQA Video-BLIINDS [19] metric. However, at the time that the research in this paper was carried out, the full implementation of Video-BLIINDS was unavailable. As a result, a similar train/test methodology adopted in [19] was implemented instead. Specifically, 80% of content was used for training, the remaining content was used for testing, and the process was performed for each distortion type both separately and on all videos mixed together. Reference videos were also used. From the results in Table VI, it is evident that the proposed metric approaches the performance of Video-BLIINDS and exceeds its performance on the MPEG-2 and H.264 distortion types.

In terms of computation time, the proposed IQA and VQA metrics are more computationally intensive than the other metrics considered except DIIVINE. However, since the code was not optimised, any similar computations performed by the algorithms considered are carried out multiple times.

TABLE II. MEDIAN CORRELATIONS FOR STATE-OF-THE-ART METRICS AND THE PROPOSED NR-IQA METRIC ON ALL DISTORTIONS. THE TOP 3 RESULTS FOR EACH PERFORMANCE METRIC ARE HIGHLIGHTED IN BOLDFACE.

Metric #	Metric Name	TID2008		LIVE	
		SROCC	PLCC	SROCC	PLCC
1	PSNR	0.5517	0.5706	0.8742	0.8662
2	SSIM	0.7762	0.7737	0.9475	0.8586
3	MS-SSIM	**0.8531**	**0.8450**	0.9507	0.9463
4	$FSIM_C$	**0.8844**	**0.8771**	**0.9642**	**0.9594**
5	RRED	0.8233	0.7383	0.9507	0.9353
6	BRISQUE	0.3231	0.4082	**0.9654**	**0.9667**
7	DIIVINE	0.2721	0.4104	0.8562	0.8439
8	NIQE	0.2442	0.2860	0.9062	0.7114
9	Proposed	**0.8577**	**0.8609**	**0.9593**	**0.9476**

IV. CONCLUSION

A NR-VQA metric has been proposed, which evaluates quality in the spatial and temporal domains without requiring any information from the reference. The proposed metric not only combines both spatial- and frequency-domain features but also utilises three novel features that improve performance on images distorted with quantisation noise and contrast change. A set of features capturing temporal and motion information were implemented based on the statistics of frame differences of natural videos. In contrast to the majority of work published in the literature, chrominance information was also considered since it provides a noticeable gain in correlation with subjective data. The proposed NR-VQA metric is statistically identical to the state-of-the-art full-reference and reduced-reference VQA metrics, which is a considerable achievement given that no-reference metrics are unable to use any information from the original content. This is all the more remarkable when considering that the metric does not require any knowledge regarding the distortion type and is able to achieve high performance even when using considerably less videos for training than other NR training-based metrics in the literature.

TABLE III. MEDIAN SROCC FOR STATE-OF-THE-ART METRICS AND THE PROPOSED NR-VQA METRIC ON (I) ALL VIDEOS AND (II) ALL VIDEOS EXCEPT MPEG-2 CODED VIDEOS IN THE LIVE VIDEO DATABASE. THE TOP 3 RESULTS FOR EACH PERFORMANCE METRIC HIGHLIGHTED IN BOLDFACE.

Metric #	Metric Name	All videos		All videos except MPEG-2	
		SROCC	PLCC	SROCC	PLCC
1	PSNR	0.5361	0.5747	0.5802	0.6230
2	MS-SSIM	0.7453	0.6901	0.7347	0.6845
3	$FSIM_C$	0.7148	0.6810	0.7081	0.6682
4	T-ViMSSIM	**0.7976**	**0.8083**	**0.8063**	**0.8123**
5	S-ViMSSIM	0.7664	0.4765	0.7394	0.4173
6	ViMSSIM	**0.8111**	0.7677	**0.8040**	0.7295
7	SRRED	0.7557	0.7738	0.7621	0.7828
8	TRRED	0.7760	**0.7882**	**0.8062**	**0.8207**
9	STRRED	**0.7944**	**0.8055**	0.7943	0.8067
10	STRREDsn	0.7292	0.7367	0.6828	0.6986
11	BRISQUE	0.0941	0.1602	0.0689	0.1693
12	DIIVINE	0.1121	0.1889	0.0916	0.1871
13	NIQE	0.0573	0.1845	0.1104	0.0906
14	Proposed	0.6913	0.7310	0.7991	**0.8161**

TABLE IV. MEDIAN SROCC FOR STATE-OF-THE-ART METRICS AND THE PROPOSED NR-VQA METRIC ON INDIVIDUAL DISTORTIONS IN THE LIVE VIDEO DATABASE. THE TOP 3 RESULTS FOR EACH PERFORMANCE METRIC HIGHLIGHTED IN BOLDFACE.

Metric #	Metric Name	Wireless	IP	H.264	MPEG-2
1	PSNR	0.6503	0.4212	0.4642	0.3902
2	MS-SSIM	0.7273	0.6663	0.7372	0.6762
3	$FSIM_C$	0.7158	0.7099	0.6715	0.6947
4	T-ViMSSIM	**0.7821**	0.6743	**0.8408**	**0.7633**
5	S-ViMSSIM	0.7261	0.6257	0.7722	**0.7694**
6	ViMSSIM	**0.8043**	0.6652	**0.8404**	**0.7434**
7	SRRED	0.7776	**0.7582**	0.7626	0.7197
8	TRRED	0.7724	**0.7418**	0.8164	0.5906
9	STRRED	0.7699	**0.7653**	0.8110	0.7245
10	STRREDsn	0.7184	0.4922	0.7300	0.7220
11	BRISQUE	0.1482	0.1414	0.1697	0.3131
12	DIIVINE	0.1575	0.1498	0.2530	0.2678
13	NIQE	0.1145	0.2235	0.2038	0.3632
14	Proposed	**0.8205**	0.6978	**0.8239**	0.2654

TABLE V. MULTI-COMPARISON ANOVA RESULTS ON THE LIVE VIDEO DATABASE WHEN ALL DISTORTIONS EXCEPT MPEG-2 ARE MIXED TOGETHER. 'NO.' REFERS TO METRIC NUMBERS AS SHOWN IN TABLE IV. '1', '0' AND '-1' INDICATE THAT THE METRIC IN THE ROW IS STATISTICALLY SUPERIOR, IDENTICAL AND INFERIOR TO THE METRIC IN THE COLUMN, RESPECTIVELY.

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	-1
2	1	0	0	-1	0	-1	0	-1	-1	1	1	1	1	-1
3	1	0	0	-1	-1	-1	-1	-1	-1	0	1	1	1	-1
4	1	1	1	0	1	0	1	0	0	1	1	1	1	0
5	1	0	1	-1	0	-1	0	-1	-1	1	1	1	1	-1
6	1	1	1	0	1	0	1	0	0	1	1	1	1	0
7	1	0	1	-1	0	-1	0	-1	0	1	1	1	1	-1
8	1	1	1	0	1	0	1	0	0	1	1	1	1	0
9	1	1	1	0	1	0	0	0	0	1	1	1	1	0
10	1	-1	0	-1	-1	-1	-1	-1	-1	0	1	1	1	-1
11	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1	-1
12	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	-1
13	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	0	0	-1
14	1	1	1	0	1	0	1	0	0	1	1	1	1	0

TABLE VI. MEDIAN CORRELATIONS FOR THE VIDEO-BLIINDS METRIC AND THE PROPOSED NR-VQA METRIC. RESULTS FOR VIDEO-BLIINDS OBTAINED DIRECTLY FROM [19].

Distortion	SROCC		PLCC	
	Video-BLIINDS	Proposed	Video-BLIINDS	Proposed
Wireless	0.815	0.750	0.951	0.873
IP	0.779	0.657	0.946	0.840
H.264	0.839	0.929	0.893	0.959
MPEG-2	0.869	0.881	0.924	0.944
All	0.759	0.703	0.881	0.732
All w/o MPEG-2	/	0.854	/	0.874

REFERENCES

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

[2] T. Ebrahimi, "Quality of multimedia experience: past, present and future," in *Proc. Int.Conf. on Multimedia*, 2009, pp. 3–4.

[3] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. on Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[4] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335–350, Feb. 2010.

[5] Z. Wang, H. R. Sheikh, and A. C. Bovik, "Objective video quality assessment," in *The Handbook of Video Databases: Design and Applications*, B. Furht and O. Marqure, Eds. CRC Press, Sep. 2003, pp. 1041–1078.

[6] Z. Wang, A. C. Bovik, H. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[7] A. J. Chan, A. Pande, E. Baik, and P. Mohapatra, "Temporal quality assessment for mobile videos," in *Proc. of the 18th annual international conference on Mobile computing and networking*, 2012, pp. 221–232.

[8] R. A. Farrugia, C. Galea, S. Zammit, and A. Muscat, "Objective video quality metrics for HDTV services: A survey," in *Proc. 2013 IEEE EUROCON*, Jul. 2013, pp. 170-176.

[9] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Trans. Broadcasting*, vol. 57, no. 2, pp. 165–182, Jun. 2011.

[10] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Conf. Record of the 37th Asilomar Conf. on Signals, Systems and Computers*, vol. 2, 2003, pp. 1398–1402.

[11] L. Zhang, D. Zhang, X. Mou, and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[12] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 202–211, 2009.

[13] R. Soundararajan and A. C. Bovik, "RRED Indices: Reduced Reference Entropic Differencing for Image Quality Assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, 2012.

[14] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, 2012.

[15] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, 2012.

[16] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[17] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 684–694, 2013.

[18] C. Vu and S. Deshpande, "ViMSSIM: from image to video quality assessment," in *Proc. 4th Workshop on Mobile Video*, 2012, pp. 1–6.

[19] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind prediction of natural video quality," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1352–1365, Mar. 2014.

[20] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008 - A database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, pp. 30–45, 2009.

[21] A. Hu, R. Zhang, X. Zhan, and Y. Dong, "Image Quality Assessment Incorporating the Interaction of Spatial and Spectral Sensitivities of the HVS," *Proc. 13th IASTED conf. Signal and Image Process. (SIP 2011)*, 2011.

[22] C. Chen and J. A. Bloom, "A blind reference-free blockiness measure," in *Proc. 11th Pacific Rim conf. Adv. multimedia inf. process.*: Part I, ser. PCM'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 112–123.

[23] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video," *IEEE Trans. Circuits Syst. Video Technol*, vol. 5, no. 1, pp. 52–56, 1995.

[24] H. Sheikh, M. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[25] I. T. Union, "Final Report from the Video Quality Experts Group on the Validation of Objective Quality Metrics for Video Quality Assessment," Video Quality Experts Group, Tech. Rep., Mar. 2000.

[26] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. Cormack, "A subjective study to evaluate video quality assessment algorithms," *SPIE Proc. Human Vision and Electron. Imaging*, Jan. 2010.