

RESEARCH

Open Access

A reduced-reference perceptual image and video quality metric based on edge preservation

Maria G Martini^{1*}, Barbara Villarini² and Federico Fiorucci²

Abstract

In image and video compression and transmission, it is important to rely on an objective image/video quality metric which accurately represents the subjective quality of processed images and video sequences. In some scenarios, it is also important to evaluate the quality of the received video sequence with minimal reference to the transmitted one. For instance, for quality improvement of video transmission through closed-loop optimisation, the video quality measure can be evaluated at the receiver and provided as feedback information to the system controller. The original image/video sequence—prior to compression and transmission—is not usually available at the receiver side, and it is important to rely at the receiver side on an objective video quality metric that does not need reference or needs minimal reference to the original video sequence. The observation that the human eye is very sensitive to edge and contour information of an image underpins the proposal of our reduced reference (RR) quality metric, which compares edge information between the distorted and the original image. Results highlight that the metric correlates well with subjective observations, also in comparison with commonly used full-reference metrics and with a state-of-the-art RR metric.

1 Introduction

For recent and emerging multimedia systems and applications, such as modern video broadcasting systems (including DVB/DVB-H, IPTV, webTV, HDTV,...) and telemedical applications, user requirements are going beyond requirements on connectivity, and users now expect the services to meet their requirements on quality. In recent years, the concept of quality of service (QoS) has been augmented towards the new concept of quality of experience (QoE), as the first only focuses on the network performance (e.g., packet loss, delay, and jitter) without a direct link to the perceived quality, whereas the QoE reflects the overall experience of the consumer accessing and using the provided service. The main target in the design of modern multimedia systems is thus the improvement of the (video) quality perceived by the user. For the provision of such quality improvement the availability of an objective quality metric well representing the human perception is crucial. Objective quality assessment methods based on subjective measurements are based either on a perceptual model of the

human visual system (HVS) [1], or on a combination of relevant parameters tuned with subjective tests [2,3].

It is also important to evaluate the quality of the received video sequence with minimal reference to the transmitted one [4]. For closed loop optimisation of video transmission, the video quality measure can be provided as feedback information to a system controller [5]. The original video sequence—prior to compression and transmission—is not usually available at the receiver side and it is important to rely at the receiver side on an objective video quality metric that does not need reference or needs minimal reference to the original video sequence. Figure 1 reports a schematic representation of an image/video processing system, consisting of a video encoder and/or a transmission network, with the calculation of a reduced reference (RR) quality metric. Reference features are extracted from the original image/video sequence and these are then compared with the same features extracted from the impaired video to obtain the RR quality metric.

We propose here a RR video quality metric well correlated with the perceived quality, based on the comparison of the edge information between the distorted image and the original one. The human eye is in fact very sensitive to the edge and contour information of an

* Correspondence: m.martini@kingston.ac.uk

¹SEC Faculty, School of Computing and Information Systems, Kingston University London, Penrhyn road, Kingston upon Thames KT1 2EE, UK
Full list of author information is available at the end of the article

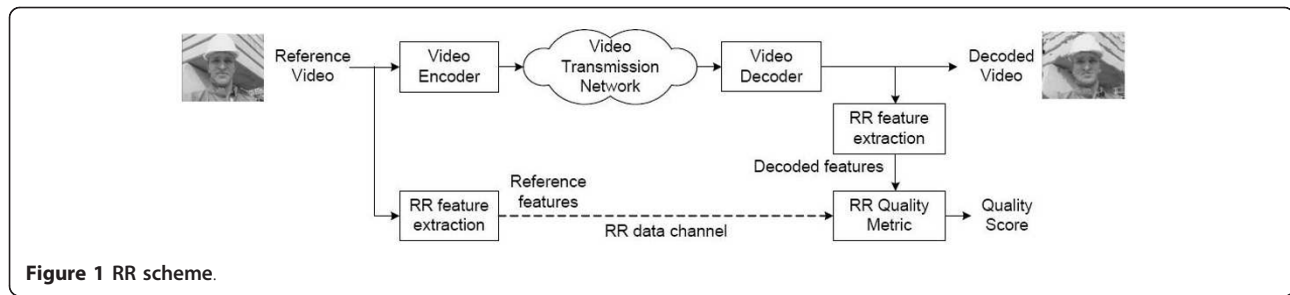


Figure 1 RR scheme.

image, i.e., the edge and contour information gives a good indication of the structure of an image and it is critical for a human to capture the scene [6].

Some works in the literature proposed considering edge structure information. For instance in [7] the structural information error between the reference and the distorted image is computed based on the statistics of the spatial position error of the local modulus maxima in the wavelet domain. In [1] a parameter is considered to detect a decrease or loss of spatial information (e.g., blurring). This parameter uses a 13 pixel spatial information filter (SI13) to measure edge impairments rather than Sobel filtering. Differently from [1] we consider here the Sobel operator [8] for edge detection, since this is one of the most used methodologies to obtain edge information due to its simplicity and efficiency. Further details on this choice are reported in the following section.

A few RR metrics have been proposed, with different characteristics in terms of complexity, of correlation with subjective quality and of overhead associated to the transmission of side information.

The ITS/NTIA (Institute for Telecommunication Sciences/National Telecommunications and Information Administration) has developed a general video quality model (VQM) [1] that was selected by both ANSI and ITU as a video quality assessment standard based on its performance. This general model requires however a bit-rate of several Mbps (more than 4 Mbps for 30 fps, CIF size video) of quality features for the calculation of the VQM value, which prevents its use as a RR metric in practical systems. The possibility to use spatial-temporal features/regions was considered in [9] in order to provide a trade-off between the correlation with subjective values and the overhead for side-information. Later on a low-rate RR metric based on the full reference metric [10] ("10 kbits/s VQM") was developed by the same authors. A subjective data set was used to determine the optimal linear combination of the eight video quality parameters in the metric. The performance of the metric was presented in terms of a scatter plot with respect to subjective data, although numerical performance results are not provided in [10].

The quality index in [4] is based on features which describe the histograms of wavelet coefficients. Two parameters describe the distribution of the wavelet coefficients of the reference image using a generalized Gaussian density (GGD) model, hence only a relatively small number of RR features are needed for the evaluation of image quality.

The RR objective picture quality measurement tool of compressed video in [11] is based on a discriminative analysis of harmonic strength computed from edge-detected pictures to create harmonics gain and loss information that could be associated with the picture. The results achieved are compared by the authors with a VQEG RR metric [9,12] and the performance of the proposed metric is shown to be comparable to the latter, with a reduction in overhead with respect to it and a global reduction of overhead with respect to full reference metrics of 1024:1. The focus is on the detection of blocking and blurring artifacts. This metric considers edge detection as our proposed metric, but in [11] edge detection is performed over the whole image and edge information is not used as side information, but just as a step for further processing of the image for the extraction of different side information.

The quality criterion presented in [13] presents relies on the extraction, from an image represented in a perceptual space, of visual features that can be compared to those used by the HVS (perceptual color space, CSF, psychophysical subband decomposition, masking effect modeling). Then a similarity metric computes the objective quality score of a distorted image by comparing the features extracted from this image to features extracted from its reference image. The performance is evaluated with the aid of three different databases with respect to three full reference metrics. The size of the side information is flexible. The main drawback of this metric is its complexity, since the HVS model (which is an essential part of the proposed image quality criterion) requires a high computation complexity.

In [14] an RR objective perceptual image quality metric for use in wireless imaging is proposed. Specifically, the normalized hybrid image quality metric (NHIQM) and a perceptual relevance weighted Lp-norm

are designed, based on the observation that the HVS is trained to extract structural information from the viewing area. Image features are identified and measured based on the extent by which individual artifacts are present in a given image. The overall quality measure is then computed as a weighted sum of the features. The authors did not rely on public databases for performance evaluation, but performed their own subjective tests. The performance of this metric is evaluated with respect to full reference metrics and the metric in [14].

The metric in [15] is based on a divisive normalization image representation. No assumptions are made about the type of impairment. This metric requires training: before applying the proposed algorithm for image quality assessment, five parameters need to be learned from the data. These parameters are cross-validated with different selections of the training and testing data. Results are compared with the RR metric in [14] and with peak signal-to-noise ratio (PSNR).

In this article we propose a low complexity RR metric based on edge preservation which can be calculated in real time in practical image/video processing and transmission systems, performs comparably with the mostly used full reference metrics and requires a limited overhead for the transmission of side information.

The remainder of this article is organized as follows. Edge detection methodologies are introduced in Section 2. Section 3 presents the proposed RR image and video quality metric. Simulation set-up and results are reported in Section 4. Conclusions about the novelty and performance of the metric are then reported in Section 5.

2 Edge detection

There are many methods to perform edge detection. The majority of these may be grouped into two categories: gradient and Laplacian. The gradient method detects the edges by finding the maximum and minimum in the first derivative of the image. This method is characteristic of the gradient filter family of edge detection and includes the Sobel method. A pixel location is declared an edge location if the value of the gradient exceeds a threshold. Edges will have higher pixel intensity values than those surrounding it. Once a threshold is set, the gradient value can be compared to the threshold value and an edge is detected when the threshold is exceeded. When the first derivative is at a maximum, the second derivative is zero. As a result, an alternative to finding the location of an edge is to locate the zeros in the second derivative. This method is known as the Laplacian.

The aforementioned methods can be extended to the 2D case. The Sobel operator performs a 2D spatial gradient measurement on an image. Typically it is used to

find the approximate absolute gradient magnitude at each point in an input grayscale image. The Sobel edge detector uses a pair of 3×3 convolution masks, one estimating the gradient in the x -direction (columns) and the other estimating the gradient in the y -direction (rows). The mask is then slid over the image, manipulating a square block of pixels at a time.

The Sobel operator can detect edges by calculating partial derivatives in 3×3 neighborhood. The main reason for using the Sobel operator is that it is relatively insensitive to noise and it has relatively smaller masks than other operators such as the Roberts operator and the two-order Laplacian operator.

The partial derivatives in x and y directions are given as:

$$S_x = f(x + 1, y - 1) + 2f(x + 1, y) + f(x + 1, y + 1) - [f(x - 1, y - 1) + 2f(x - 1, y) + f(x - 1, y + 1)] \quad (1)$$

and

$$S_y = f(x - 1, y + 1) + 2f(x, y + 1) + f(x + 1, y + 1) - [f(x - 1, y - 1) + 2f(x, y - 1) + f(x + 1, y - 1)] \quad (2)$$

The gradient of each pixel is calculated according to $g(x, y) = \sqrt{S_x^2 + S_y^2}$ and a threshold value t is selected. If $g(x, y) > t$, this point is regarded as an edge point.

The Sobel operator can also be expressed in the form of two masks as shown in Figure 2: the two masks are used to calculate S_y and S_x , respectively.

3 Proposed metric

Since structural distortion is tightly linked with edge degradation, we propose a RR quality metric which compares edge information between the distorted image and the original one. We propose to apply Sobel filtering locally, only for some blocks of the entire image, after subsampling the images.

Images are divided in sub-windows, as shown in Figure 3. For instance, if images have size 512×768 we could subsample of a factor of 2 and consider 16×16 macroblocks of size 16×24 each, or we can subsample of a factor 1.5 and consider 18×16 macroblocks with size 19×32 each. The example in Figure 3 reports the second option. The block size is chosen such that it is sufficiently large to account for vertical and/or horizontal activities within each block, but small enough to reduce complexity and the size of side information. In

-1	-2	-1	-1	0	1
0	0	0	-2	0	2
1	2	1	-1	0	1

Figure 2 Sobel masks.

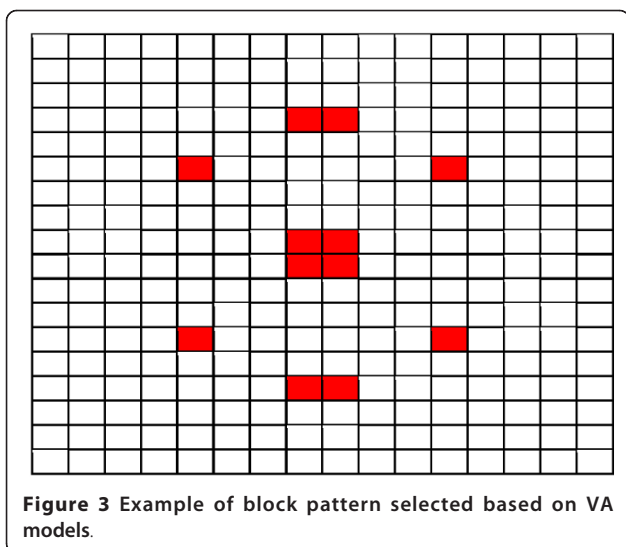


Figure 3 Example of block pattern selected based on VA models.

addition, sub-windows are non coincident with macro-blocks, to enable a better detection of DCT artifacts in the case of DCT compressed images and video.

In order to reduce the overhead associated with the transmission of side information, only 12 blocks are selected to represent the different areas of the images. The block pattern utilized for our tests is chosen after several investigations based on visual attention (VA). Various experiments have been proposed in the literature for VA modeling and salient region identification, aiming at the detection of salient regions in an image. Models on VA are often developed and validated by visual fixation patterns through eye tracking experiments [16,17]. In [18] a framework is proposed in order to extend existing image quality metrics with a simple VA model. A subjective region of interest (ROI) experiment was performed, with seven images, in which the viewers' task was to select within each image the region that drew most of their attention. For simplicity, in this experiment only rectangular-shaped ROIs were allowed. Considering the obtained ROI as a random value, it is possible to calculate the mean value and the standard deviation. It was observed that the ROI's center coordinates are around the image center for most of the images, and the mean of the ROI dimensions are very similar in both x and y directions. This confirms that the salient region, which include the most important informative content of the image, is often placed in the center of the picture.

Following these guidelines we have chosen the block pattern as a subset of the ROI with a central symmetry, minimizing the number of blocks to reduce the overhead associated to the transmission of side information. Figure 3 shows an example of block pattern.

For the assessment of the quality of the corrupted image, the edge structure of the blocks of the corrupted

image should be compared to the structure of the correspondent blocks in the original image. For the identification of edges we use Sobel filtering, which is applied locally in these selected blocks.

For each pixel in each block we obtain a bit value, where one represents an edge and zero means that there are no edges. If m and n are the block dimensions, we denote the corresponding blocks l in the original and the possibly corrupted image as the $m \times n$ matrices O_l and C_l respectively, and the Sobel-filtered version of blocks l as the $m \times n$ binary matrices $SO_l = \mathcal{S}(O_l)$, with elements $so_{i,j}$, with $i = 1, \dots, m, j = 1, \dots, n$, and $SC_l = \mathcal{S}(C_l)$, with elements $sc_{i,j}$, with $i = 1, \dots, m, j = 1, \dots, n$. We denoted above with \mathcal{S} the Sobel operator. The similarity of two images can be assessed based on the similarity of the edge structures, i.e., by comparing the matrices SO_b , associated to the filtered version of the block in the original image, and SC_b , associated to the filtered version of the block in the possibly corrupted image.

We can check if the edges of the reference image are kept, simply by counting the zeros and ones which are unchanged after compression or lossy transmission of the image. Hence, for each block l of image s the similarity index can be computed as

$$I_{s,l} = n_l/p_l \quad (3)$$

where

$$n_l = p_l - \sum_{i=1}^m \sum_{j=1}^n |sc_{l,ij} - so_{l,ij}| \quad (4)$$

is the number of zeros and ones unchanged in the l -th block and $p_l = m \times n$ is the total number of pixels in the l -th block.

If N_b is the number of blocks in the selected block pattern, the similarity index I_s for image s is defined here as

$$I_s = \frac{1}{N_b} \sum_{l=1}^{N_b} I_{s,l} \quad (5)$$

For images decomposed in blocks of equal size, as considered here, the proposed quality index is thus:

$$I_s = \frac{1}{N_b} \sum_{l=1}^{N_b} \left(1 - \frac{\sum_{i=1}^m \sum_{j=1}^n |sc_{l,ij} - so_{l,ij}|}{mn} \right) \quad (6)$$

3.1 Threshold selection

The threshold value is an important parameter that depends on a number of factors, such as image brightness, contrast, level of noise, and even edge direction. The selection of the threshold in Sobel filtering is

associated to the sensitivity of the filter to edges. In particular, the lower the value of the threshold, the higher the sensitivity to edges. Too high values of the threshold do not detect edges which are important for quality assessment. On the other side, if the value of the threshold is too small, large parts of the image are considered as edges, whereas these are irrelevant for quality assessment. The threshold can be selected following an analysis of the gradient image histogram. Based on this consideration and on the analysis of Sobel filtering performance for the images of the considered databases, the selected threshold value is $t = 0.001$.

Figure 4 reports the correlation coefficient of our proposed metric and DMOS values in the LIVE [19] image quality assessment database. The correlation coefficient is calculated for different selections of the threshold, for the different types of impairments considered in the database: fast fading (FF), white noise (WN), Gaussian blur (GB), JPEG compression (JP), JPEG2000 compression (JP2K). We can observe that the performance drops after a threshold value of approximately 0.005. For lower values, the dependence of the performance on the threshold is very limited.

3.2 Complexity

The selection of Sobel filtering results in a low complexity metric. The Sobel algorithm is characterized, in fact, by a low computational complexity and consequently high calculation speed. In [20] some edge detection techniques are compared for an application which uses a DSP implementation: the Sobel filter exhibits the best performance in terms of edge detection time in comparison with the other wavelet-based edge detectors. Sobel filtering has been implemented in hardware and used in

different areas, often when realtime performance is required, such as for real-time volume rendering systems, and video assisted transportation systems [21,22]. This makes the proposed metric suitable for real-time implementation, an important aspect when an image/video metric is used for the purpose of “on the fly” system adaptation as in the scenario considered here.

3.3 Overhead

In order to perform the proposed edge comparison, we should transmit the matrices composed of one’s and zeros’s in the reference blocks. By considering the pattern in Figure 3, this would result for images of resolution 512×768 in the transmission of $19 \times 32 \times 12 = 7.29$ kbits per image. Note that the size of the original image (not compressed) is $3 \times 512 \times 768 \times 8 = 9.4$ Mbits.

In the worst case (side information not compressed) our metric reduces thus the needed reference with respect to FR metrics of a factor 1290:1. As a comparison, the RR metric in [11] reduces it of a factor 1024:1 and the metric in [12] of 64:1.

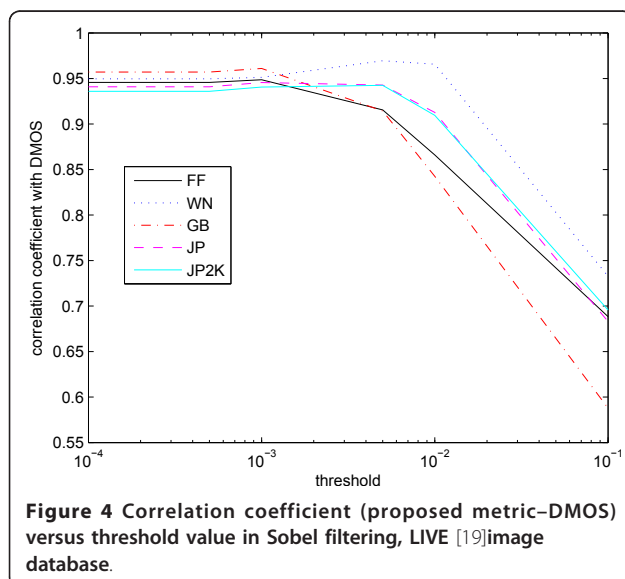
Since side information is in our case composed of a large number of zeros appearing in long runs, it is possible to further reduce the overhead by compressing the relevant data, e.g., through run-length encoding, or to transmit only the positions of ones in the matrix.

Furthermore, in the case of video, quality assessment can be performed only on a fraction of the transmitted frames (e.g., five frames per second) in order to reduce the side information overhead needed for the calculation of the quality metric.

4 Simulation set-up and results

In order to test the performance of our quality assessment algorithm, we considered publicly available databases.

The first one is provided by the Laboratory for Image & Video Engineering (LIVE) of the University of Texas Austin (in collaboration with The Department of Psychology at the same University). An extensive experiment was conducted to obtain scores from human subjects for a number of images distorted with different distortion types. The database contains 29 high-resolution (typically 768×512) original images (see Figure 5), altered with five types of distortions at different distortion levels: besides the original images, images corrupted with JPEG2000 and JPEG compression, white-noise, GB and JPEG2000 compression and subsequent transmission over a FF Rayleigh channel are considered. The latter set of images is in particular interesting since it enables to assess the quality of images impaired by both compression and transmission errors. Our quality metric is tested versus the subjective quality values provided in



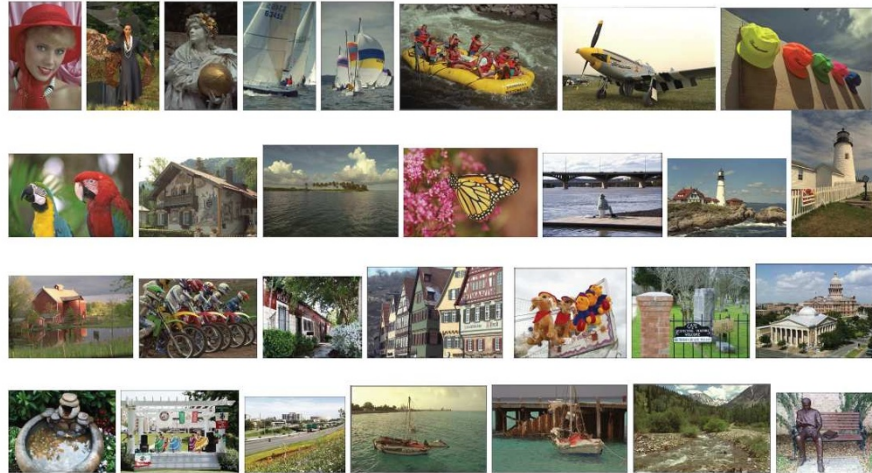


Figure 5 Images in the LIVE [19]database.

the database. Subjective results reported in the database were obtained with observers providing their quality score on a continuous linear scale that was divided into five equal regions marked with adjectives bad, poor, fair, good, and excellent. Two test sessions, with about half of the images in each session, were performed. Each image was rated by 20-25 subjects. No viewing distance restrictions were imposed, and normal indoor illumination conditions were provided. The observers received a short training before the session. The raw scores were converted into difference scores (between the test and the reference) and then converted to Z-scores [23], scaled back to 1-100 range, and finally a difference mean opinion score (DMOS) for each distorted image was obtained.

The second database, IRCCyN/IVC [24], was developed by the *Institut de Recherche en Communications et Cyberntique de Nantes*. It is a 512×512 pixels color images database. This database is composed by ten original images and 235 distorted images generated by four different processing methods/impairments (JPEG, JPEG2000, LAR coding, and blurring). Subjective evaluations were made at a viewing distance of six times the screen height, by using a double stimulus impairment scale (DSIS) method with five categories and 15 observers. The images in the database are reported in Figure 6.

Finally, for video we consider the database in [25-27]. The database is composed of ten video sequences. These are high definition (HD) YUV 4:2:0 format sequences downsampled to a resolution of 768×432 pixels. All videos, except one 8.68 s long, are 10 s long. The frame rate is 25 frames per second for seven sequences and 50 frames per second for three sequences. Example frames from the video sequences in the database are reported in Figure 7. For each video

sequence, 15 distorted versions are present, with four types of distortion: wireless distortion, IP distortion, H.264 compression, MPEG-2 compression. For MPEG-2, the reference software available from the International Organization for Standardization (ISO) was used to compress the videos. Four compressed MPEG-2 videos spanning the desired range of visual quality were selected for each reference video. For H.264 the JM reference software (version 12.3) was used. The procedure for selecting the videos was the same as that used to select MPEG-2 compressed videos, with compression rates varied from 200 Kbps to 5 Mbps. For "IP distortion", three IP videos corresponding to each reference are present in the database, created by simulating IP losses on an H.264 compressed video stream. Four IP error patterns supplied by the Video Coding Experts Group (VCEG), with loss rates of 3, 5, 10, and 20%, were used. Since losses in different portions of the video stream may result in different visual effects, the authors viewed and selected a diverse set of videos suffering from different types of observed artifacts. For the "wireless" scenario, the video streams were encoded according to the H.264 standard using multiple slices per frame, where each packet contained one slice. Errors in the wireless environments were simulated using bit error patterns with packet error rates varied between 0.5-10%. The differential MOS (DMOS) value is provided for each impaired video sequence, in a scale from 1 to 100.

With the aid of the databases above, we compare the performance versus subjective tests of our metric with respect to the most popular full reference metrics and to the RR metrics with the best performance and whose results are directly comparable or reproducible.

Namely, we consider:

- MSSIM [2] (full reference);



Figure 6 Images in the IRCCyN/IVC [24]database.

- PSNR (full reference);
- [14] (reduced reference);
- [15] (reduced reference);
- [13] (reduced reference);
- Proposed Sobel-based metric (reduced reference).

To apply the MSSIM metric, the images have been modified according to [28].

We report our results in terms of scatter plots, where each symbol in the plot refers to a different image: Figures 8,9, 10, and 11 report scatter plots for the metrics above in the case of compression according to the JPEG2000 standard and subsequent transmission over a fast fading channel.

The figures report, besides scatter plots, the linear approximation best fitting the data using the least-squares method, the residuals and the norm of residuals L for the linear model, i.e., $L = \sqrt{\sum_{i=1}^N (d_i)^2}$, where the residual d_i is the difference between the predicted quality value and

the experimental subjective quality value for image i , and N is the number of the considered images. The values of the norms of residuals enable a simple numerical comparison among the different metrics. Note that in the case of the MSSIM metric we have provided a non-linear approximation, better fitting the data.

A summary of the results for the LIVE image database [19] in terms of norms of residuals is reported in Table 1. Tables 2 and 3 report a summary of the results for the LIVE image database in terms of correlation coefficient, since this is more commonly used and enables an easier comparison with other metrics, and of Spearman rank. We have also reported results for two slightly different versions—(a) and (b)—of a the recent RR metric [15], whose performance results available in the literature can be compared with our ones for some of the impairments included in the LIVE database.

We can observe that our metric well correlates with subjective tests, with results comparable to those

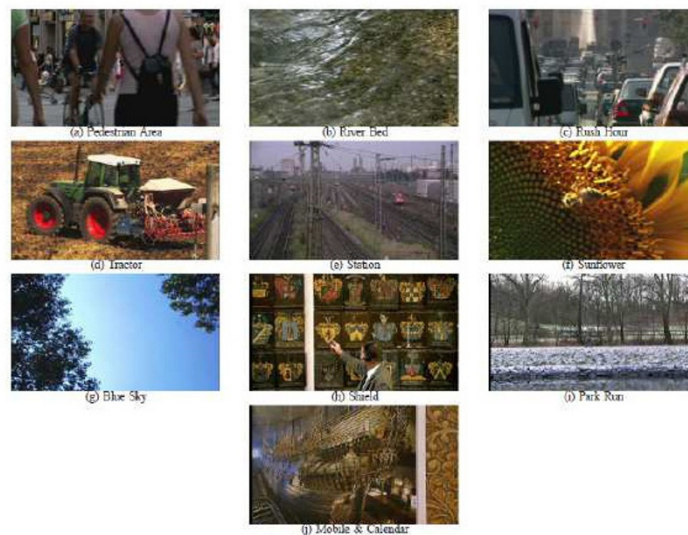
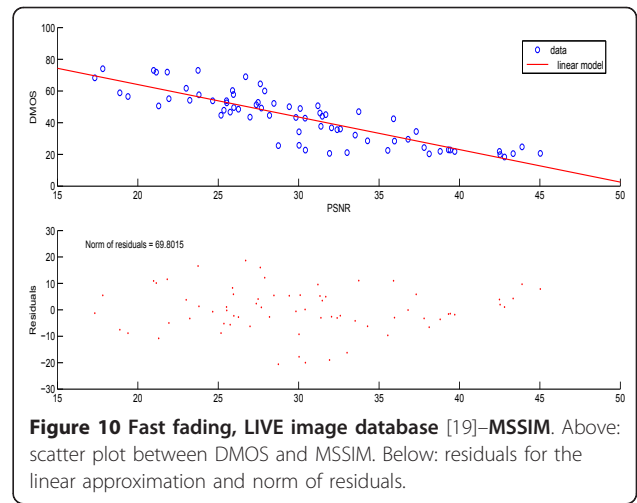
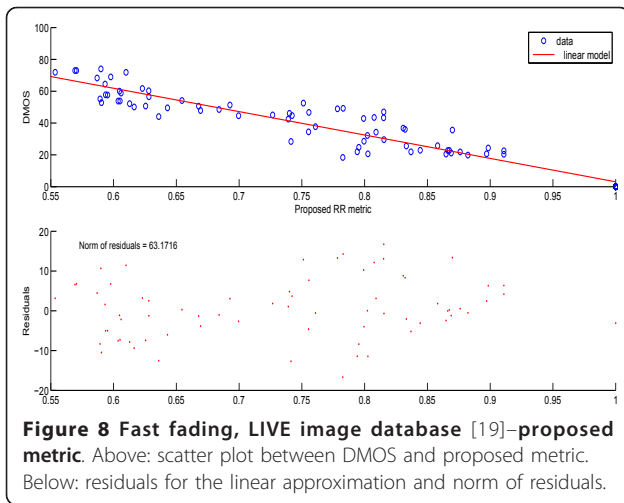


Figure 7 Sample frames from video sequences in the LIVE video database [25].



achieved by full reference metrics. For the images in the LIVE database our metric outperforms the considered state-of-the-art RR metrics in all the considered scenarios, except for the case of WN, where the metric [15] performs better at the expense of a higher complexity, and the case of JPEG2000 where the benchmark RR metric [4], based on the wavelet transform, provides a better performance in terms of norm of residuals.

However, for the same type of impairment (JPEG2000 compression) our metric performs slightly better than the benchmark one when the images in the IRCCyN/IVC database [24] are considered. The relevant results are reported in Tables 4 and 5 and Figures 12, 13, 14,

and 15 present in detail the relevant results for the case of JPEG compression.

Figures 16, 17, 18, and 19 report example results for the LIVE video database [25], where our metric is applied for all video frames. Figure 16 reports the scatter plot for our metric versus MOS in the case the video sequences in the database are compressed according to the MPEG-2 standard; Figure 17 reports the scatter plot for our metric versus MOS in the case the video sequences in the database are compressed according to the H.264 standard; Figure 18 reports the scatter plot for our metric versus MOS in the case the video sequences in the database are compressed according to the H.264 standard and affected by IP distortions; Figure 19 reports the scatter plot for our metric versus MOS in the case the video sequences in the database are compressed according to the H.264 standard and transmitted over a wireless channel. In all cases our metric well matches the subjective results.

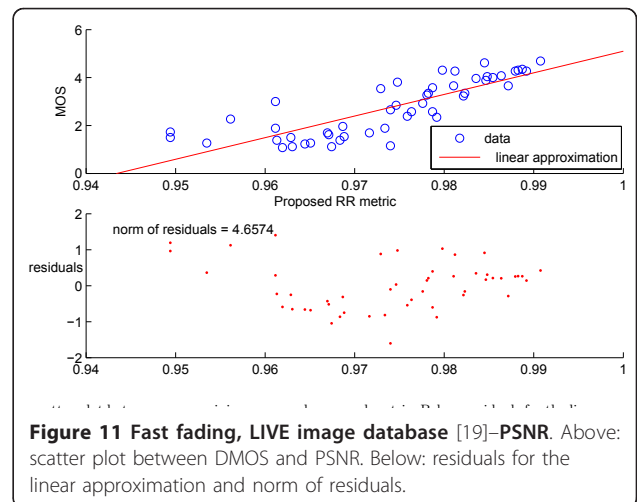
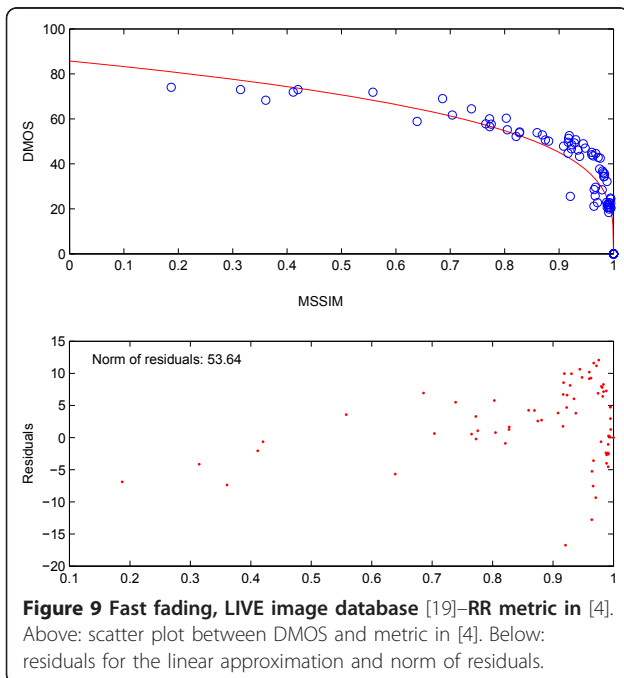


Table 1 Norm of residuals versus DMOS, LIVE image database [19]

	PSNR	RR [4]	Proposed RR	MSSIM
Fast fading	69.80	85.03	63.17	53.64
White noise	25.25	63.68	60.82	31.44
Gaussian blur	90.08	70.63	55.96	34.06
JPEG compression	82.83	115.06	83.04	65.58
JPEG2000 compression	70.23	63.94	84.33	74.06

Table 2 Correlation coefficient versus DMOS, LIVE image database [19]

	PSNR	RR [4]	RR in [15](a)	RR in [15](b)	Proposed RR	MSSIM
Fast fading	0.8556	0.9175	0.9243	0.9464	0.9478	0.9439 [4]
White noise	0.981	0.8889	0.9401	0.9654	0.9513	0.9706 [4]
Gaussian blur	0.79491	0.8872	0.8773	0.9561	0.9611	0.9361 [4]
JPEG compression	0.8245	-	-	-	0.9456	0.958 [13]
JPEG2000 compression	0.8703	-	-	-	0.9407	0.942 [13]

Table 3 Spearman rank versus DMOS, LIVE image database [19]

	PSNR	RR [4]	RR in [15](a)	RR in [15](b)	Proposed RR	MSSIM
Fast fading	0.8770	0.9162	0.9237	0.9443	0.8988	0.9435 [4]
White noise	0.9881	0.8639	0.9316	0.9559	0.8962	0.9581 [4]
Gaussian blur	0.7585	0.9145	0.8608	0.9584	0.9260	0.9705 [4]
JPEG compression	0.850	-	-	-	0.920	0.964 [13]
JPEG2000 compression	0.896	-	-	-	0.919	0.965 [13]

Table 4 Norm of residuals versus MOS, IRCCyN/IVC image database [24]

	PSNR	RR [4]	Proposed RR	MSSIM
JPEG compression	6.60	7.29	5.11	3.75
JPEG2000 compression	5.30	5.42	5.27	5.28

Table 5 Correlation coefficient versus MOS, IRCCyN/IVC image database [24]

	PSNR	RR [4]	Proposed RR	MSSIM	C4 [13,29]
JPEG compression	0.5957	0.4644	0.7837	0.8897	0.92
JPEG2000 compression	0.8143	0.8043	0.8163	0.8149	0.925

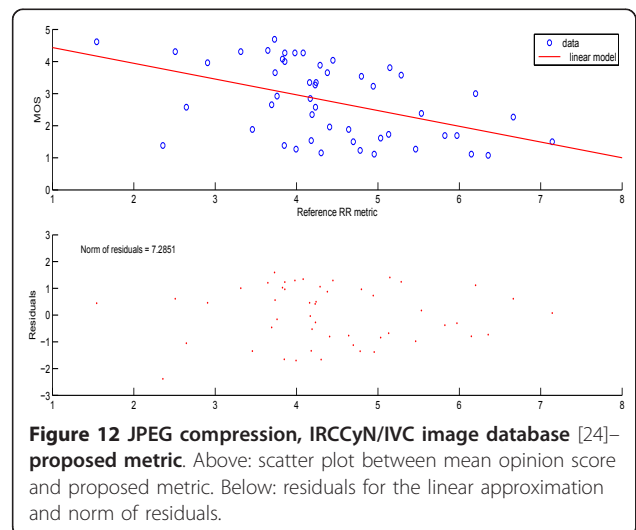


Figure 12 JPEG compression, IRCCyN/IVC image database [24]–proposed metric. Above: scatter plot between mean opinion score and proposed metric. Below: residuals for the linear approximation and norm of residuals.

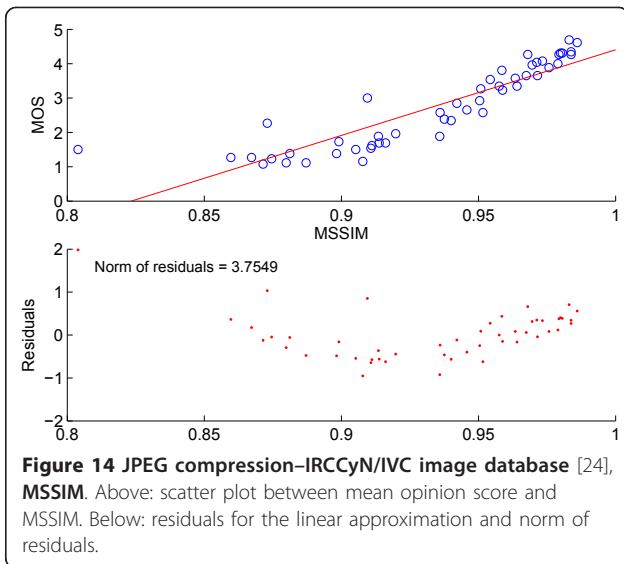
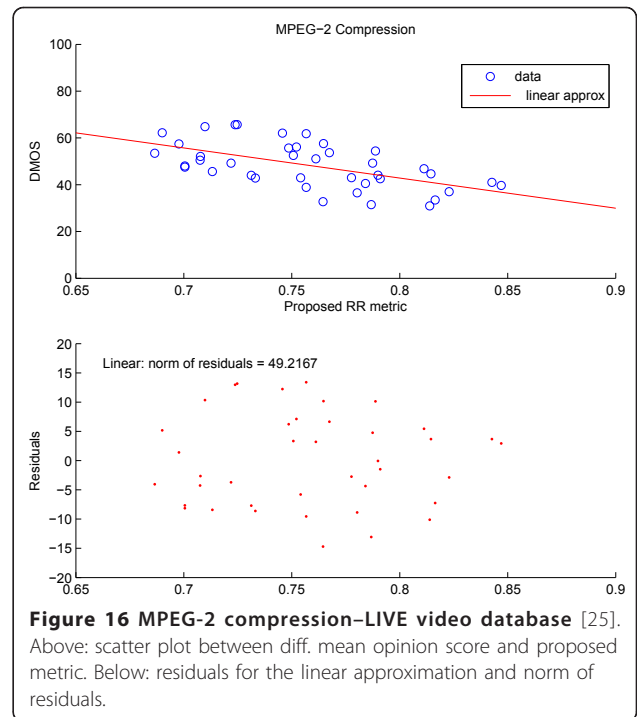
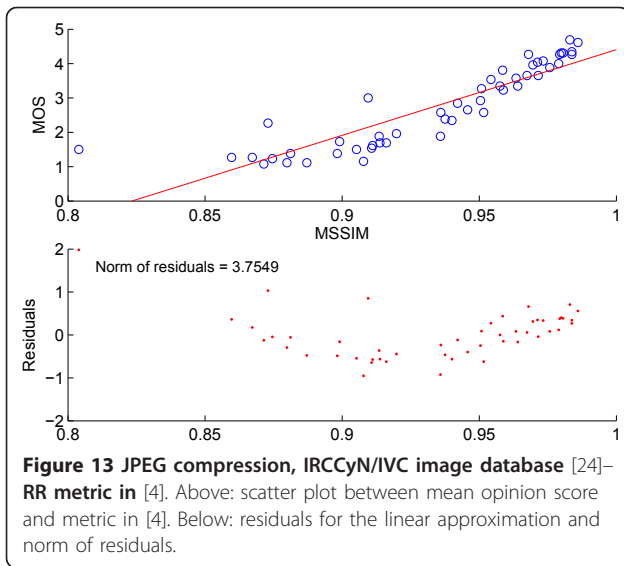
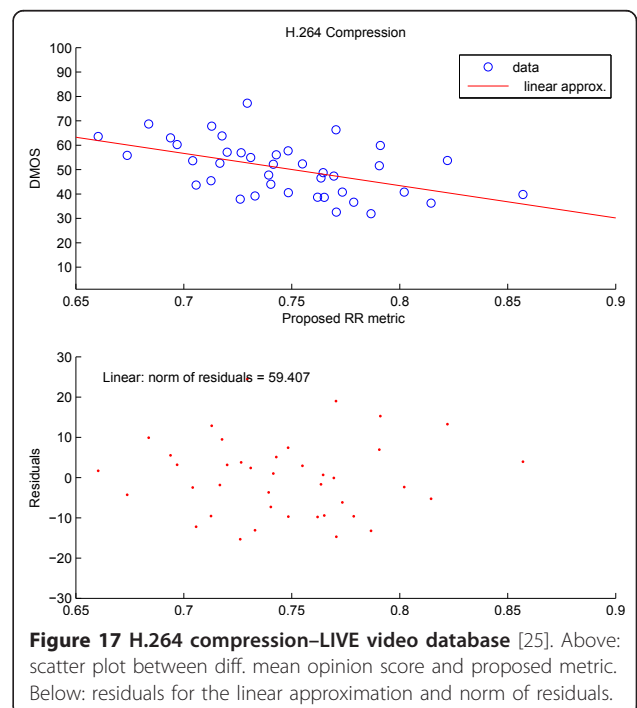
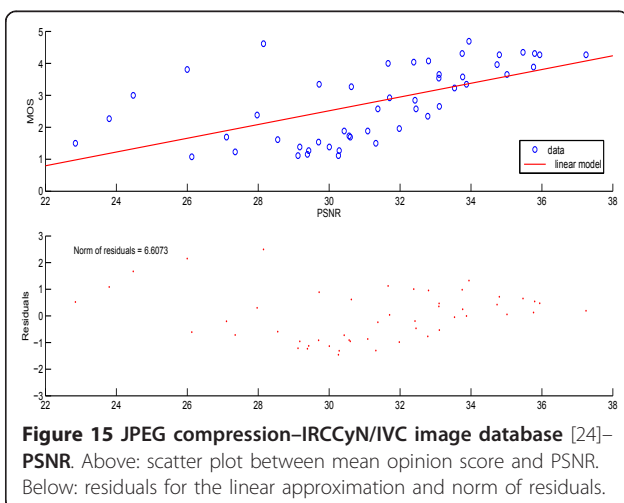


Table 3 reports results in terms of Spearman rank, an indicator of monotonicity, for the LIVE image database. With this criterion, our metric outperforms the full reference PSNR metric for all impairments except



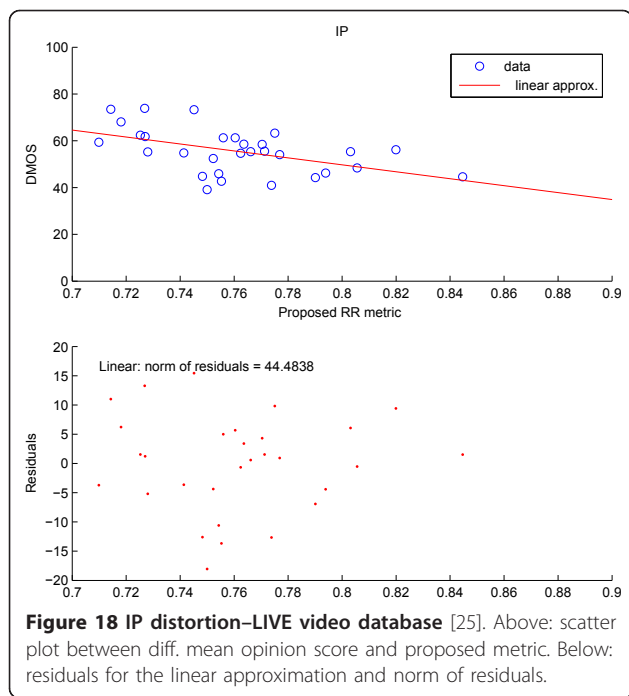


Figure 18 IP distortion–LIVE video database [25]. Above: scatter plot between diff. mean opinion score and proposed metric. Below: residuals for the linear approximation and norm of residuals.

Gaussian noise, and the RR metric in [4] for all the reported cases except the case of fast fading. The more complex RR metric in [15] is outperformed in the case of GB.

Tables 4 and 5 report the results for the IVC image database in terms of norm of residuals and correlation coefficient, respectively. We observe that our metric

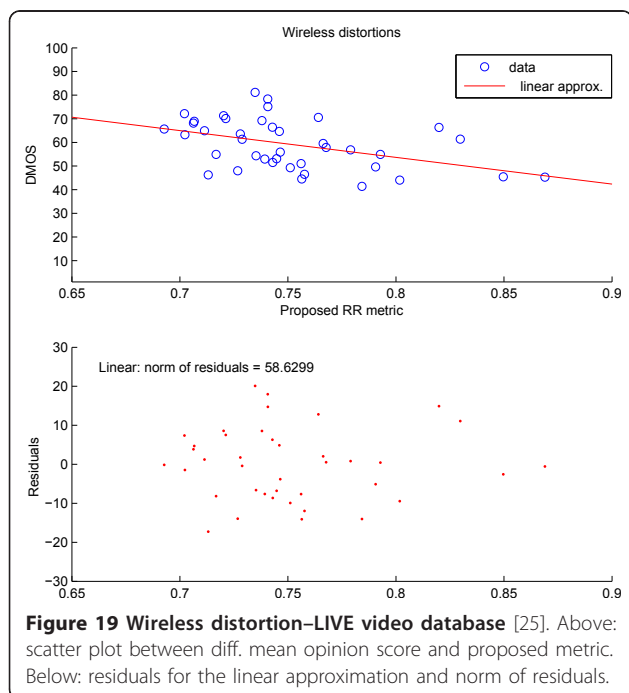


Figure 19 Wireless distortion–LIVE video database [25]. Above: scatter plot between diff. mean opinion score and proposed metric. Below: residuals for the linear approximation and norm of residuals.

outperforms the full reference metric PSNR and the RR metric in [4] in all cases. Considering the Spearman rank, reported in Table 6, our metric outperforms both the full reference PSNR metric and the RR metric in [4] in all cases except for PSNR in the case of JPEG2000 compression. Note that with this database, the gain obtained with our metric with respect to the others is higher, probably due to the fact that the metric in [4] was tailored to the LIVE database. We reported for completeness the results in terms of correlation coefficient for the metric [13]. This metric has very high correlation with subjective results; it is however too complex when real time implementation is required.

The results obtained for the case of video sequences in the LIVE video database are summarised in Table 7 for the correlation coefficient and in Table 8 for the Spearman coefficient. We can observe that our metric outperforms the full reference PSNR metric in most cases.

Note that for video sequences, in order to reduce the overhead, it is possible to apply the metric only for selected frames, for instance by every 5, 10, 25, and 50 frames. The necessity of a more or less frequent calculation of the metric depends on the motion characteristics of the video sequence.

We can observe that the performance of our metric is comparable with the considered full reference metrics, and our metric outperforms PSNR in the case of both MPEG2 and H.264 compression and also in the case “IP distortion”, i.e., the case of H.264 video transmitted over a network. Our metric outperforms also the MSSIM metric in terms of correlation coefficient with subjective data for the case of MPEG2 compressed video.

4.1 Comparison between full reference edge-based metric and RR one

We found interesting to perform a comparative evaluation of our metric, where edges are compared for a selected set of blocks (RR), and the metric obtained through the comparison of full edge maps (Sobel based full reference metric), that we define as below:

$$I_{fr} = \frac{1}{N_{tot}} \sum_{l=1}^{N_{tot}} \left(1 - \frac{\sum_{i=1}^m \sum_{j=1}^n |scl_{i,j} - sol_{i,j}|}{mn} \right) \quad (7)$$

where the notation used is defined in Section 3, and N_{tot} is the total number of blocks in the image.

Table 6 Spearman rank versus MOS, IRCCyN/IVC image database [24]

	PSNR	RR [4]	Proposed RR	MSSIM
JPEG compression	0.8794	0.4542	0.9162	0.9814
JPEG2000 compression	0.842	0.800	0.823	0.943

Table 7 Correlation coefficient versus DMOS, LIVE video database [25,27]

	PSNR	MSSIM	Proposed RR metric
Wireless distortion	0.4675	0.5401	0.4392
H.264 compression	0.4385	0.6656	0.5012
MPEG2 compression	0.3856	0.5491	0.5778
IP distortion	0.4108	0.5119	0.4971

Table 8 Spearman rank versus DMOS, LIVE video database [25,27]

	PSNR	MSSIM	Proposed RR metric
Wireless distortion	0.4334	0.5233	0.439
H.264 compression	0.4296	0.6514	0.516
MPEG2 compression	0.3588	0.5545	0.582
IP distortion	0.3206	0.4550	0.458

Table 9 Norm of residuals versus DMOS, full reference versus RR edge-based metric, LIVE image database

	WN	Gaussian blur	fast fading	JPEG	JPEG2000
FR Sobel	58.36	53.67	61.27	73.68	76.05
RR Sobel	60.82	55.96	63.17	83.04	84.33

We found that, although the correlation with subjective results is higher for the full reference metric, the difference with our proposed metric is very small. The results are reported in Table 9. This confirms that the selected pattern well represents the ROI of the image and enables a reliable quality assessment, although with a very limited overhead for the transmission of side information.

5 Conclusion

We proposed in this article a perceptual RR image and video quality metric which compares edge information between portions of the distorted image and the original one by using Sobel filtering. The algorithm is simple and has a low computational complexity. Results highlight that the proposed metric well correlates with subjective observations, also in comparison with commonly used full-reference metrics and with state-of-the-art RR metrics.

Acknowledgements

This work was partially supported by the European Commission (FP7 projects OPTIMIX and CONCERTO).

Author details

¹SEC Faculty, School of Computing and Information Systems, Kingston University London, Penrhyn road, Kingston upon Thames KT1 2EE, UK ²DIEI, University of Perugia, Perugia, Italy

Authors' contributions

MM conceived the work, proposed the edge-based metric, processed the data and analyzed the results, supervised the whole work and wrote the article. BV, in the framework of her internship in Kingston University, finalized the metric definition by proposing all the details, including the selection of the block pattern and the selection of the threshold in Sobel filtering; she performed all the simulations in the article and realized all the scatter plots; she also contributed to the processing of the data and analysis of the results. FF contributed to the literature review and supported BV in the selection of the final block pattern by taking into account visual attention models. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 16 May 2011 Accepted: 16 March 2012

Published: 16 March 2012

References

- MH Pinson, S Wolf, A new standardized method for objectively measuring video quality. *IEEE Trans Broadcast*. **50**(3), 312–322 (2004). doi:10.1109/TBC.2004.834028
- Z Wang, A Bovik, H Sheikh, E Simoncelli, Image quality assessment: from error measurement to structural similarity. *IEEE Trans Image Process*. **13**(4), 600–612 (2004). doi:10.1109/TIP.2003.819861
- HR Sheikh, MF Sabir, AC Bovik, A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans Image Process*. **15**(11), 3440–3451 (2006)
- Z Wang, EP Simoncelli, Reduced-reference image quality assessment using a wavelet-domain natural image statistic model, in *Human vision and Electronic Imaging*, vol. 5666. (San Jose, CA, 2005), pp. 149–159
- MG Martini, M Mazzotti, C Lamy-Bergot, J Huusko, P Amon, Content adaptive network aware joint optimization of wireless video transmission, in *IEEE Commun Mag*, vol. 45. (San Jose, CA, 2007)(1), pp. 84–90
- D Marr, E Hildreth, Theory of edge detection. *Proc R Soc Lond Ser B*. **207**, 187–217 (1980). doi:10.1098/rspb.1980.0020
- M Zhang, X Mou, A psychovisual image quality metric based on multi-scale structure similarity, *Proc IEEE International Conference on Image Processing (ICIP)*, (San Diego, CA, 2008), pp. 381–384
- J Woods, *Multidimensional Signal, Image and Video Processing and Coding*, (Elsevier, Amsterdam, 2006)
- S Wolf, M Pinson, In-service performance metrics for mpeg-2 video systems, *Proc Made to Measure 98—Measurement Techniques of the Digital Age Technical Seminar, International Academy of Broadcasting (IAB)*, (ITU and Technical University of Braunschweig, Montreux, Switzerland, 1998), pp. 12–13
- S Wolf, MH Pinson, Low bandwidth reduced reference video quality monitoring system, *Video Processing and Quality Metrics for Consumer Electronics*, (Scottsdale, Arizona, 2005), pp. 23–25
- I Gunawan, M Ghanbari, Reduced-reference video quality assessment using discriminative local harmonic strength with motion consideration. *IEEE Trans Circ Syst Video Technol*. **18**(1), 71–83 (2008)
- Final report from the video quality experts group on the validation of objective models of video quality assessment, phase ii. Video quality expert group, (San Jose, CA, 2003)
- M Carnec, P Le Callet, D Barba, Objective quality assessment of color images based on a generic perceptual reduced reference. *Signal Process Image Commun*. **23**(4), 239–256 (2008). doi:10.1016/j.image.2008.02.003
- U Engelke, M Kusuma, H Zepernick, M Caldera, Objective quality assessment of color images based on a generic perceptual reduced reference. *Signal Process Image Commun*. **24**, 525–547 (2009). doi:10.1016/j.image.2009.06.005
- Q Li, Z Wang, Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE J Sel Top Signal Process*. **3**(9), 202–211 (2009)
- AL Yarbus, *Eye Movements and Vision*, (Plenum Press, New York, 1967)
- CM Privitera, LW Stark, Algorithms for defining visual regions-of-interest: comparison with eye fixations. *IEEE Trans Pattern Anal Mach Intell*. **22**(9), 970–982 (2000). doi:10.1109/34.877520

18. U Engelke, HJ Zepernick, Framework for optimal region of interest-based quality assessment in wireless imaging. *J Electron Imaging*. **19**(1), 1–13 (2010)
19. HR Sheikh, Z Wang, L Cormack, AC Bovik, Live image quality assessment database. <http://live.ece.utexas.edu/research/quality> (2008)
20. Z Musoromy, F Bensaali, S Ramalingam, G Pissanidis, Comparison of real-time DSP-based edge detection techniques for license plate detection, Sixth International Conference on Information Assurance and Security, (Atlanta, GA, 2010), pp. 323–328
21. W Zhou, Z Xie, C Hua, C Sun, J Zhang, Research on edge detection for image based on wavelet transform, Proceedings of the 2009 Second International Conference on Intelligent Computation Technology and Automation, (Washington, DC, USA, 2009), pp. 686–689
22. N Kazakova, M Margala, NG Durdle, Sobel edge detection processor for a real-time volume rendering system, Proc of the 2004 International Symposium on Circuits and Systems (ISCAS '04), (Vancouver, Canada, 2004), pp. 913–916
23. AM van Dijk, JB Martens, AB Watson, Quality assessment of coded images using numerical category scaling, in *Proc SPIE*, vol. 2451. (Amsterdam, 1995), pp. 99–101
24. P Le Callet, F Atrousseau, Subjective quality assessment IRCCyN/IVC database. <http://www.irccyn.ec-nantes.fr/ivcdb/> (2005)
25. K Seshadrinathan, R Soundararajan, LK Cormack, AC Bovik, LIVE video quality assessment database. <http://live.ece.utexas.edu/research/quality/live-video.html> (2010)
26. K Seshadrinathan, R Soundararajan, AC Bovik, LK Cormack, Study of subjective and objective quality assessment of video. *IEEE Trans Image Process*. **19**(6), 1427–1441 (2010)
27. K Seshadrinathan, R Soundararajan, AC Bovik, LK Cormack, A subjective study to evaluate video quality assessment algorithms. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series: Human Vision and Electronic Imaging. **7527** (2010)
28. Z Wang, AC Bovik, HR Sheikh, EP Simoncelli, The SSIM index for image quality assessment. <http://www.ece.uwaterloo.ca/z70wang/research/ssim/#usage> (2008)
29. S Tourancheau, S Atrousseau, ZMP Sazzad, Y Horita, Impact of subjective dataset on the performance of image quality metrics, IEEE International Conference on Image Processing (ICIP), (San Diego, CA, 2008), pp. 365–368

doi:10.1186/1687-6180-2012-66

Cite this article as: Martini et al.: A reduced-reference perceptual image and video quality metric based on edge preservation. *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:66.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
