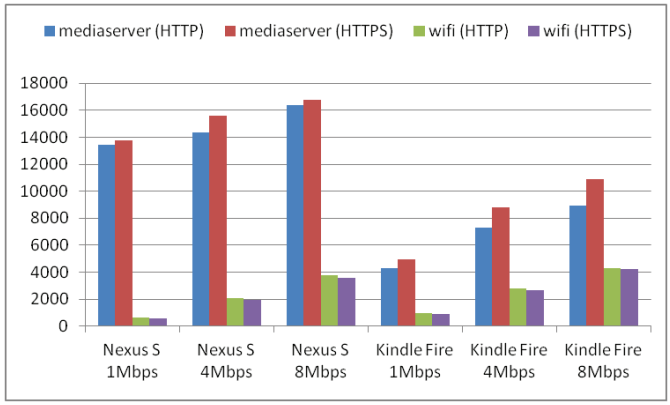
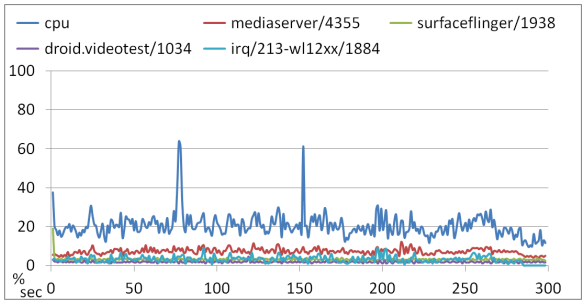
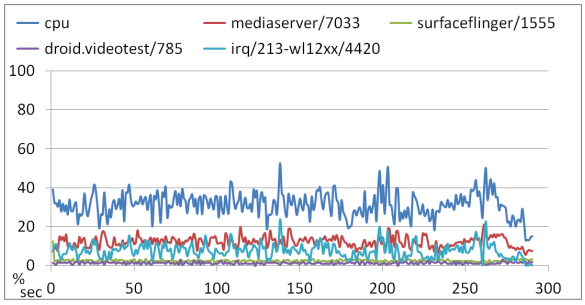


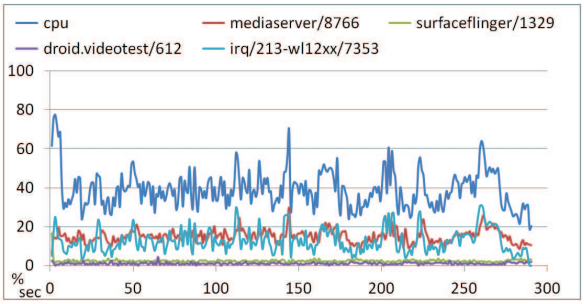
Figure 44: Total CPU ticks at HTTPS streaming



(a) 1 Mbps

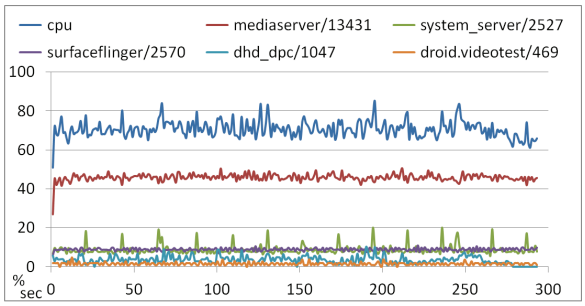


(b) 4 Mbps

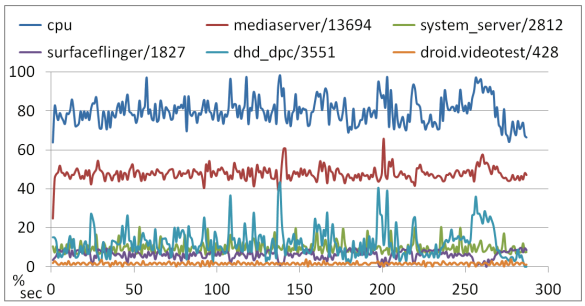


(c) 8 Mbps

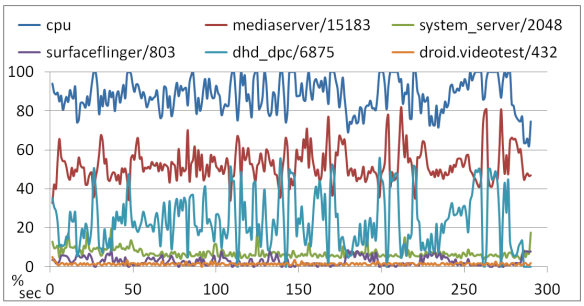
Figure 45: Kindle Fire, HTTP streaming through IPSec tunnel



(a) 1 Mbps

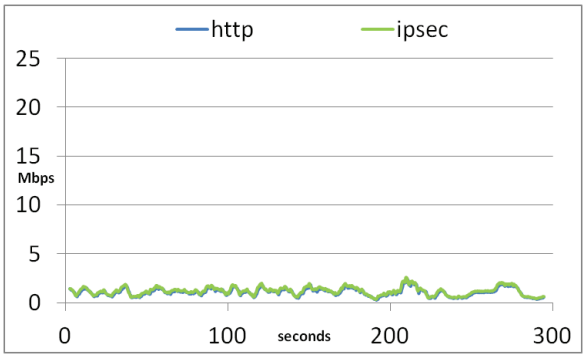


(b) 4 Mbps

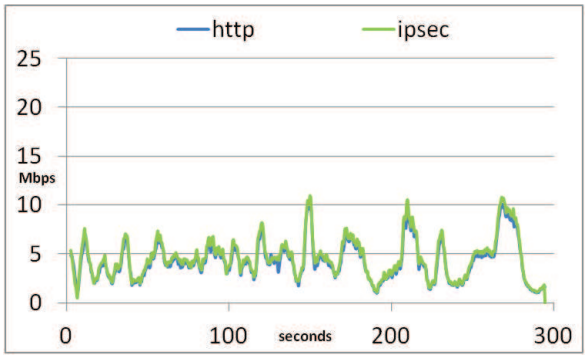


(c) 8 Mbps

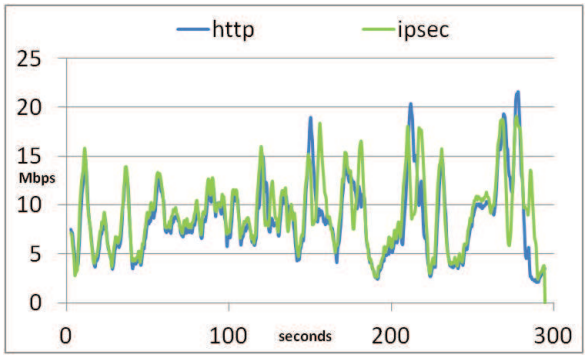
Figure 46: Nexus S, HTTP streaming through IPSec tunnel



(a) 1 Mbps



(b) 4 Mbps



(c) 8 Mbps

Figure 47: Nexus S, Video stream bandwidth

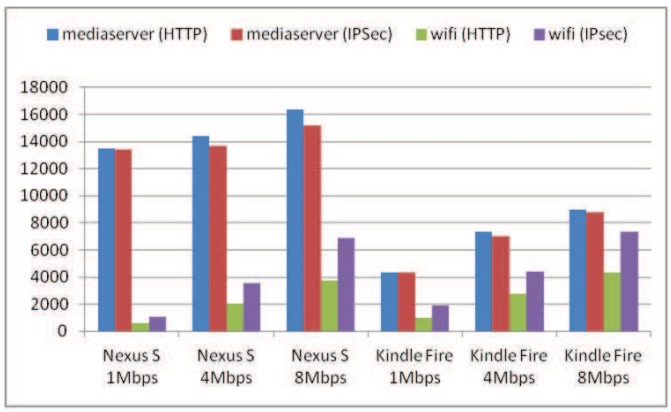


Figure 48: Total CPU ticks at HTTP streaming through IPSec tunnel

7 Reduced Reference Image and Video Quality Assessment

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 only 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 reduces reference 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 reduced-reference metric.

7.1 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 [98], or on a combination of relevant parameters tuned with subjective tests [149, 122].

It is also important to evaluate the quality of the received video sequence with minimal reference to the transmitted one [152]. For closed loop optimisation of video transmission, the video quality measure can be provided as feedback information to a system controller [77]. 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 49 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 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 reduced reference (RR) quality metric.

We propose here a reduced reference 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 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 [76].

Some works in the literature proposed considering edge structure information. For instance in [165] 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 [98] 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 [98] we consider here the Sobel operator [159] 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 subsection.

A few reduced-reference 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 Tele- communications and Information Administration) has developed a general Video Quality Model (VQM) [98] 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

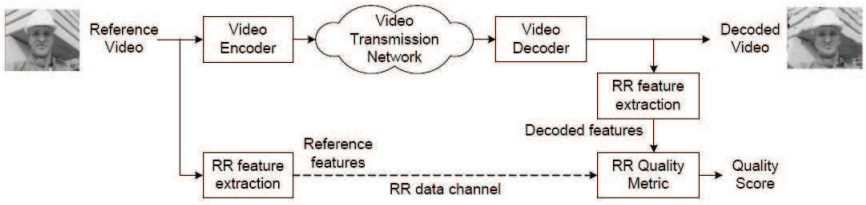


Figure 49: Reduced reference scheme.

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 [157] in order to provide a trade-off between the correlation with subjective values and the overhead for side-information. Later on a low-rate reduced reference metric based on the full reference metric [158] (10 kbits/s VQM) was developed by the same authors. A subjective data set was used to determine the optimal linear combination of the 8 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 [158].

The quality index in [152] 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 reduced-reference objective picture quality measurement tool of compressed video in [32] 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 [2] [157] 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 [32] 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 [18] relies on the extraction, from an image represented in a perceptual space, of visual features that can be compared to those used by the human visual system (HVS): perceptual color space, contrast sensitivity function (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 [26] a reduced-reference 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 human visual system (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 [152].

The metric in [68] 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 [152] and with PSNR.

In this contribution we propose a low complexity reduced-reference metric based on edge preservation which can be calculated in real time in practical image/video processing and transmission systems, performs comparably

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

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

Figure 50: Sobel masks

with the mostly used full reference metrics and requires a limited overhead for the transmission of side information.

The remainder of this section is organized as follows. Edge detection methodologies are introduced in Subsection 7.2. Subsection 7.3 presents the proposed reduced reference image and video quality metric. Simulation set-up and results are reported in Subsection 7.4. Conclusions about the novelty and performance of the metric are then reported in Subsection 7.6.

7.2 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 two-dimensions case. The Sobel operator performs a 2-D 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 3x3 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:

$$S_x = f(x+1, y-1) + 2f(x+1, y) + f(x+1, y+1) + \tag{29}$$
$$-[f(x-1, y-1) + 2f(x-1, y) + f(x-1, y+1)]$$

and

$$S_y = f(x-1, y+1) + 2f(x, y+1) + f(x+1, y+1) + \tag{30}$$
$$-[f(x-1, y-1) + 2f(x, y-1) + f(x+1, y-1)]$$

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 50: the two masks are used to calculate S_y and S_x, respectively.

7.3 Proposed metric

Since structural distortion is tightly linked with edge degradation, we propose a reduced reference (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 51. For instance, if images have size 512 x 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 51 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 addition, sub-windows are non coincident with macroblocks, 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 visual attention are often developed and validated by visual fixation patterns through eye tracking experiments [74] [100]. In [27] a framework is proposed in order to extend existing image quality metrics with a simple VA model. A subjective ROI (Region Of Interest) experiment was performed, with 7 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 51 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 $i = 1, ..., m, j = 1, ..., n$, and $SC_l = \mathcal{S}(C_l)$, with elements $sc_{i,j}$, with $i = 1, ..., m, j = 1, ..., n$. 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_l, associated to the filtered version of the block in the original image, and SC_l, 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

$$I_{s,l} = n_l / p_l \tag{31}$$

where

$$n_l = p_l - \sum_{i=1}^{m} \sum_{j=1}^{n} |sc_{l,ij} - so_{l,ij}| \tag{32}$$

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

$$I_s = \frac{1}{N_b} \sum_{l=1}^{N_b} I_{s,l} \tag{33}$$

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

$$I_s = \frac{1}{N_b} \sum_{l=1}^{N_b} \left(1 - \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} |sc_{l,ij} - so_{l,ij}|}{mn} \right) \tag{34}$$

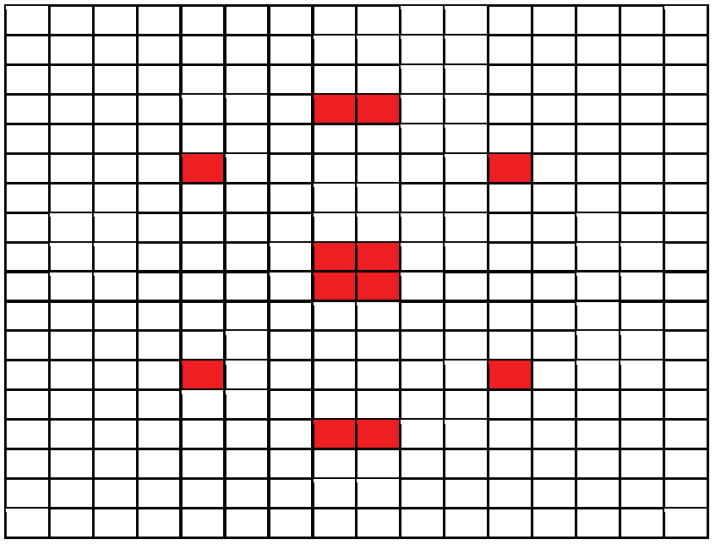


Figure 51: Example of block pattern selected based on visual attention models

7.3.1 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 52 reports the correlation coefficient of our proposed metric and DMOS values in the LIVE [123] 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.

7.3.2 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 [86] 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 real-time performance is required, such as for real-time volume rendering systems, and video assisted transportation systems [170] [50]. 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.

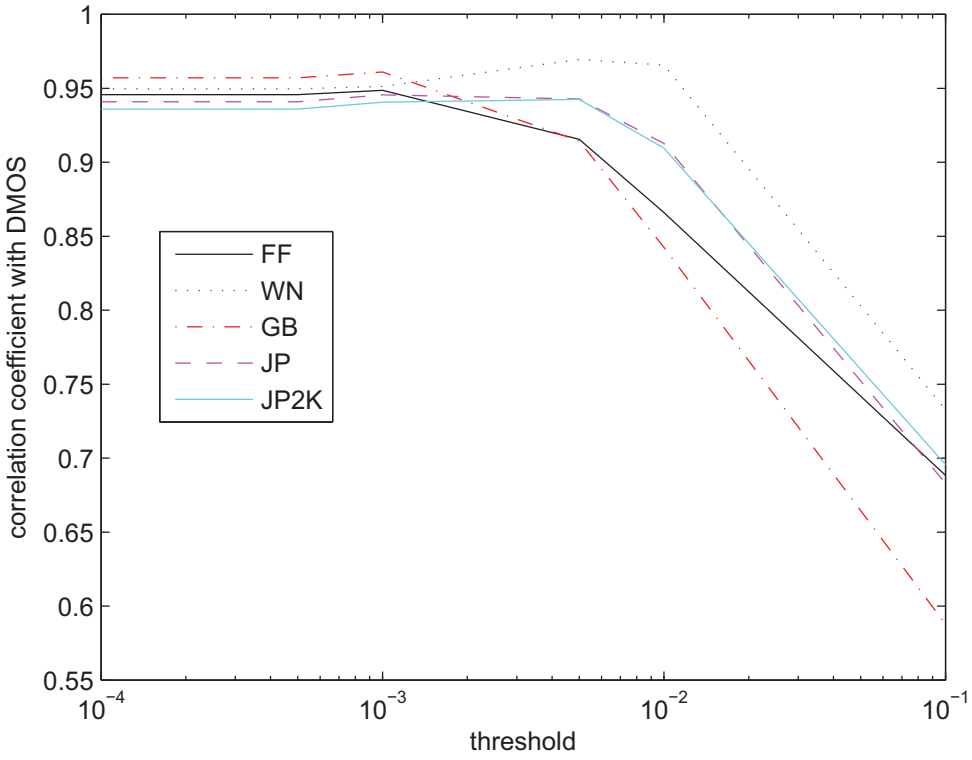


Figure 52: Correlation coefficient (proposed metric - DMOS) versus threshold value in Sobel filtering, LIVE [123] image database

7.3.3 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 51, this would result for images of resolution 512×768 in the transmission of $19\times32\times12 = 7.29$ 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 [32] reduces it of a factor 1024:1 and the metric in [2] 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.

7.4 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 53), altered with five types of distortions at different distortion levels: besides the original images, images corrupted with JPEG2000 and JPEG compression, white-noise, Gaussian blur and JPEG2000 compression and subsequent transmission over

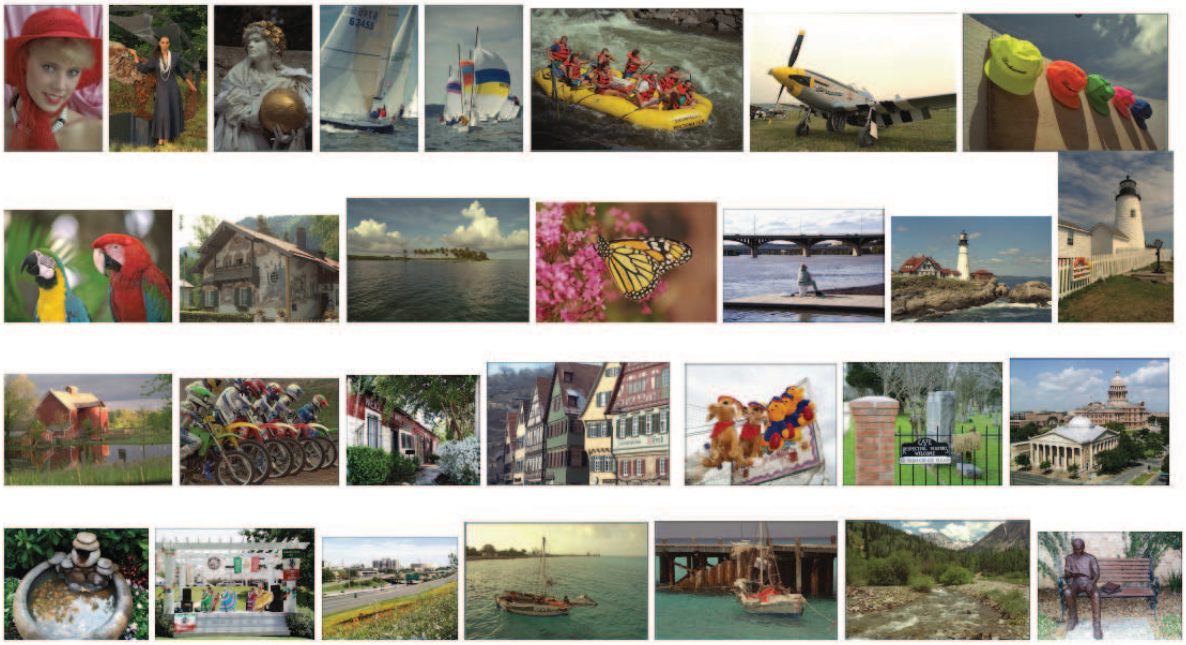


Figure 53: Images in the LIVE [123] database.

a fast fading (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 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 [138], 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 [17], was developed by the *Institut de Recherche en Communications et Cybernétique de Nantes*. It is a 512×512 pixels color images database. This database is composed by 10 original images and 235 distorted images generated by 4 different processing methods / impairments (JPEG, JPEG2000, LAR coding and blurring). Subjective evaluations were made at a viewing distance of 6 times the screen height, by using a Double Stimulus Impairment Scale (DSIS) method with 5 categories and 15 observers. The images in the database are reported in Figure 54.

Finally, for video we consider the database in [117] [118] [119]. 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 seconds long, are 10 seconds long. The frame rate is 25 frames per second for seven sequences and 50 frames per second for three sequences. 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 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 results 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

Figure 54: Images in the IRCCyN/IVC [17] database.

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 reduced reference metrics with the best performance and whose results are directly comparable or reproducible.

Namely, we consider:

- MSSIM [149] (full reference);
- peak signal-to-noise-ratio - PSNR (full reference)
- [152] (reduced reference);
- [68] (reduced reference);
- [18] (reduced reference);
- Proposed Sobel-based metric (reduced reference).

To apply the MSSIM metric, the images have been modified according to [150].

We report our results in terms of scatter plots, where each symbol in the plot refers to a different image: Figures 56, 57, 58, and 59 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 [123] in terms of norms of residuals is reported in Table 11. Table 12 reports 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. In the latter we have also reported results for two slightly different versions - (a) and (b) - of a the recent RR metric [68], whose performance results available in the literature can be compared with our ones for some of the impairments included in the LIVE database.

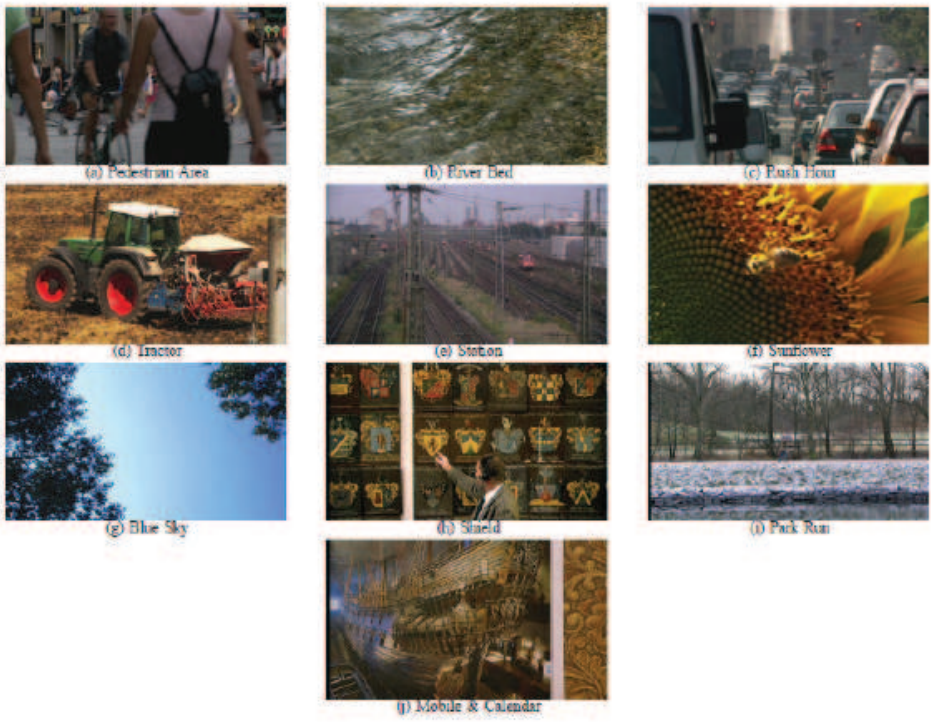


Figure 55: Sample frames from video sequences in the LIVE video database [117].

We can observe that our metric well correlates with subjective tests, with results comparable to those achieved by full reference metrics. For the images in the LIVE database our metric outperforms the considered state-of-the-art reduced reference metrics in all the considered scenarios, except for the case of white noise, where the metric in [68] performs better at the expense of a higher complexity, and the case of JPEG2000 where the benchmark reduced reference metric [152], 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 [17] are considered. The relevant results are reported in Table 14 and Table 15 while Figures 60, 61, 62, and 63 present in detail the relevant results for the case of JPEG compression.

Figures 64-67 report example results for the LIVE video database [117], where our metric is applied for all video frames. Figure 64 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 65 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 66 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 67 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 13 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 Gaussian noise, and the RR metric in [152] for all the reported cases except the case of fast fading. The more complex RR metric in [68] is outperformed in the case of Gaussian blur.

Tables 14 and 15 report the results for the IVC image database in terms of norm of residuals and correlation coefficient, respectively. We observe that our metric outperforms the full reference metric PSNR and the reduced reference metric in [152] in all cases. Considering the Spearman rank, reported in Table 16, our metric outperforms

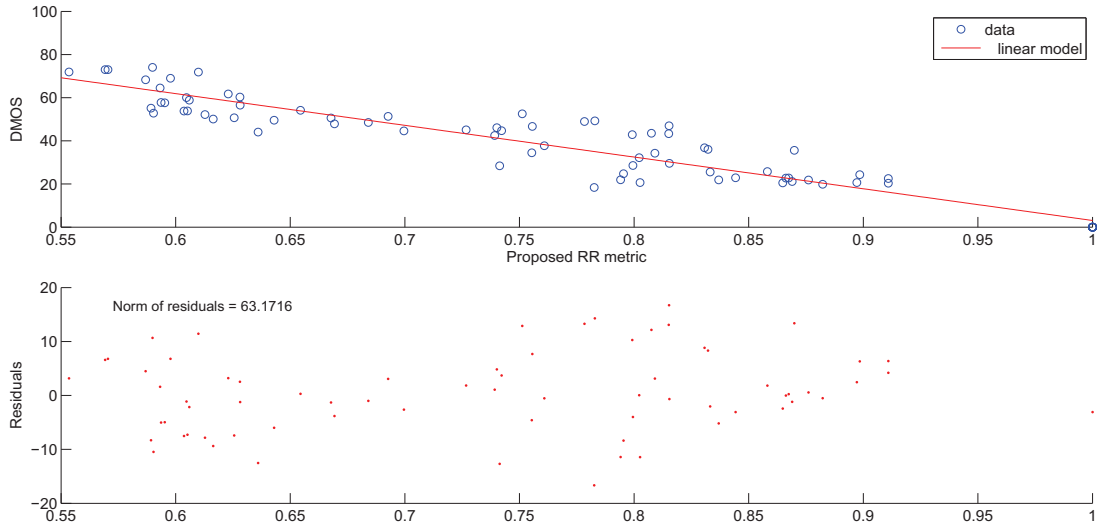


Figure 56: Fast fading, LIVE image database [123] - Proposed metric. Above: scatter plot between difference mean opinion score and proposed metric. Below: residuals for the linear approximation and norm of residuals.

both the full reference PSNR metric and the RR metric in [152] 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 [152] was tailored to the LIVE database. We reported for completeness the results in terms of correlation coefficient for the metric [18]. 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 17 for the correlation coefficient and in Table 18 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, 50 frames. The necessity of a more or less frequent calculation of the metric depends on the motion characteristics of the video sequence.

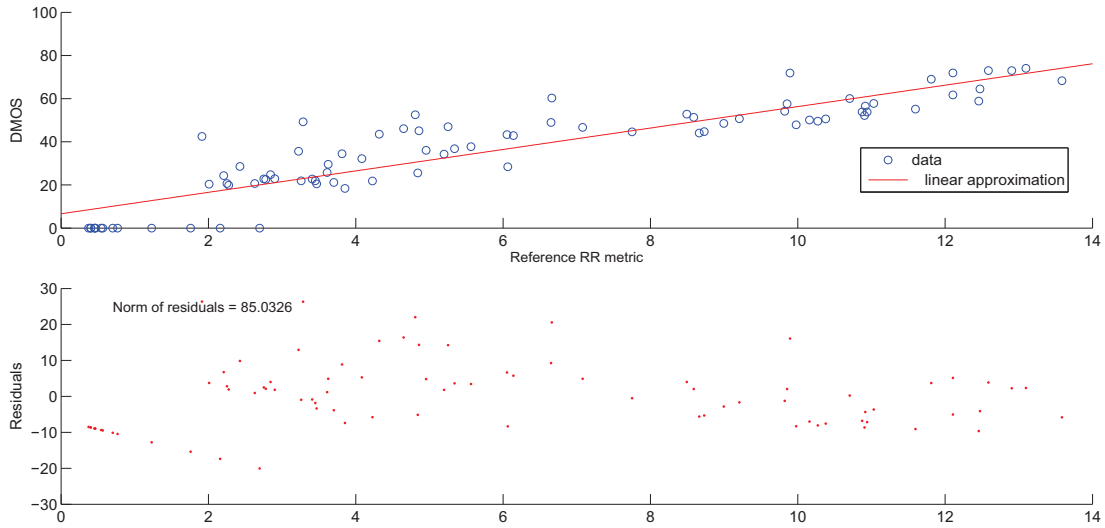


Figure 57: Fast fading, LIVE image database [123] - Reduced reference metric in [152]. Above: scatter plot between difference mean opinion score and metric in [152]. Below: residuals for the linear approximation and norm of residuals.

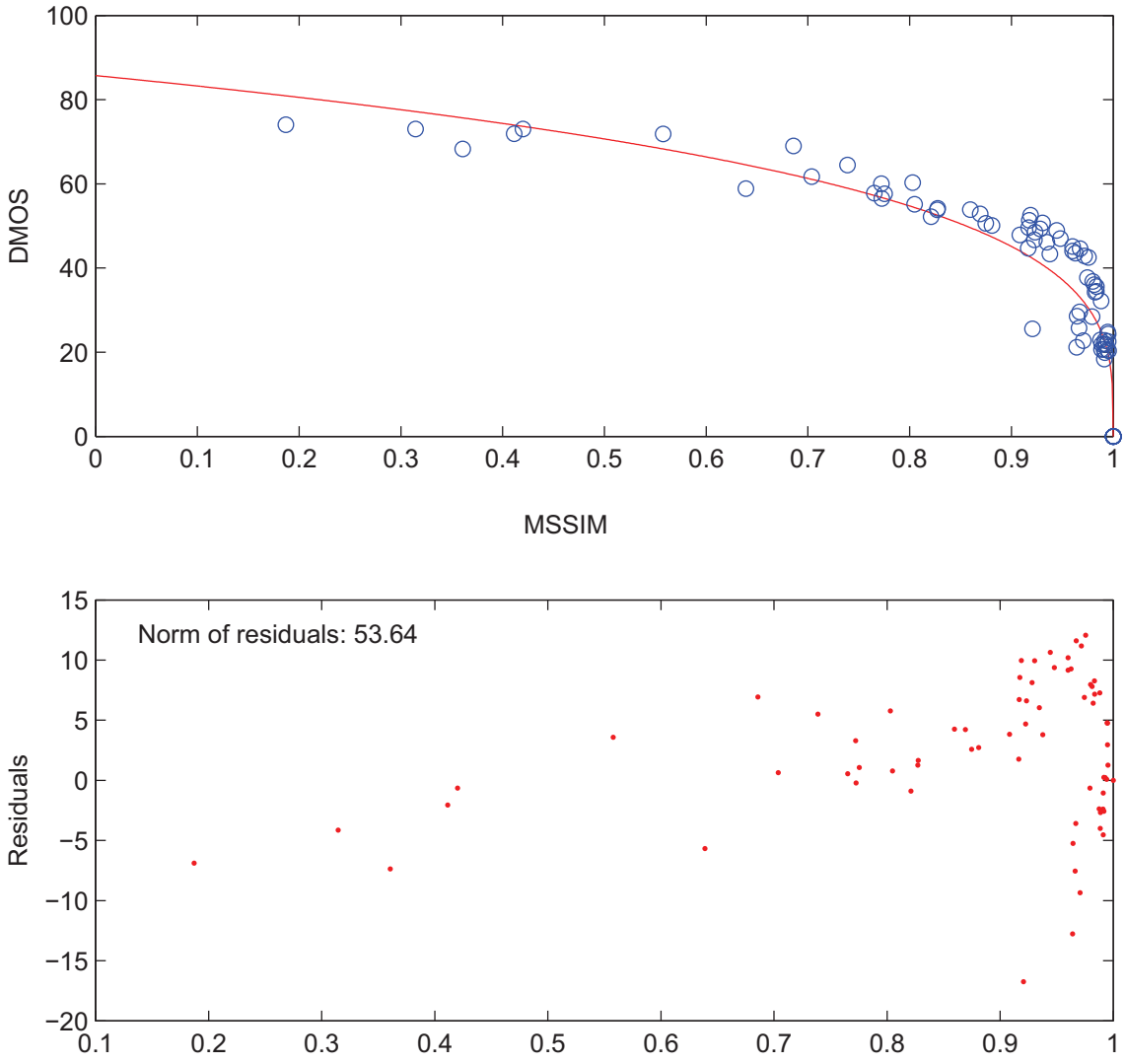


Figure 58: Fast fading, LIVE image database [123] - MSSIM. Above: scatter plot between difference mean opinion score and MSSIM. Below: residuals for the linear approximation and norm of residuals.

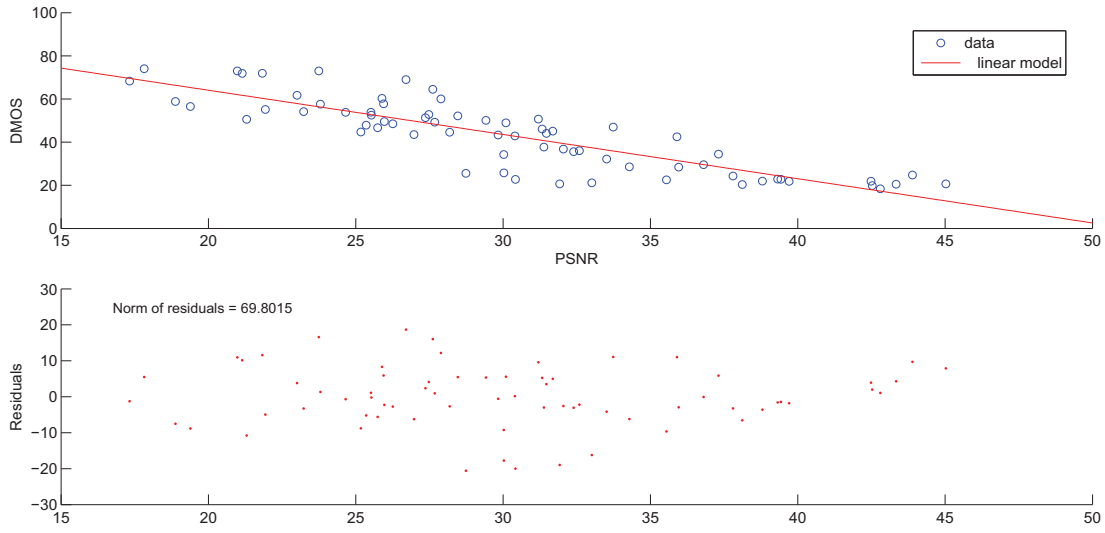


Figure 59: Fast fading, LIVE image database [123] - PSNR. Above: scatter plot between difference mean opinion score and PSNR. Below: residuals for the linear approximation and norm of residuals.

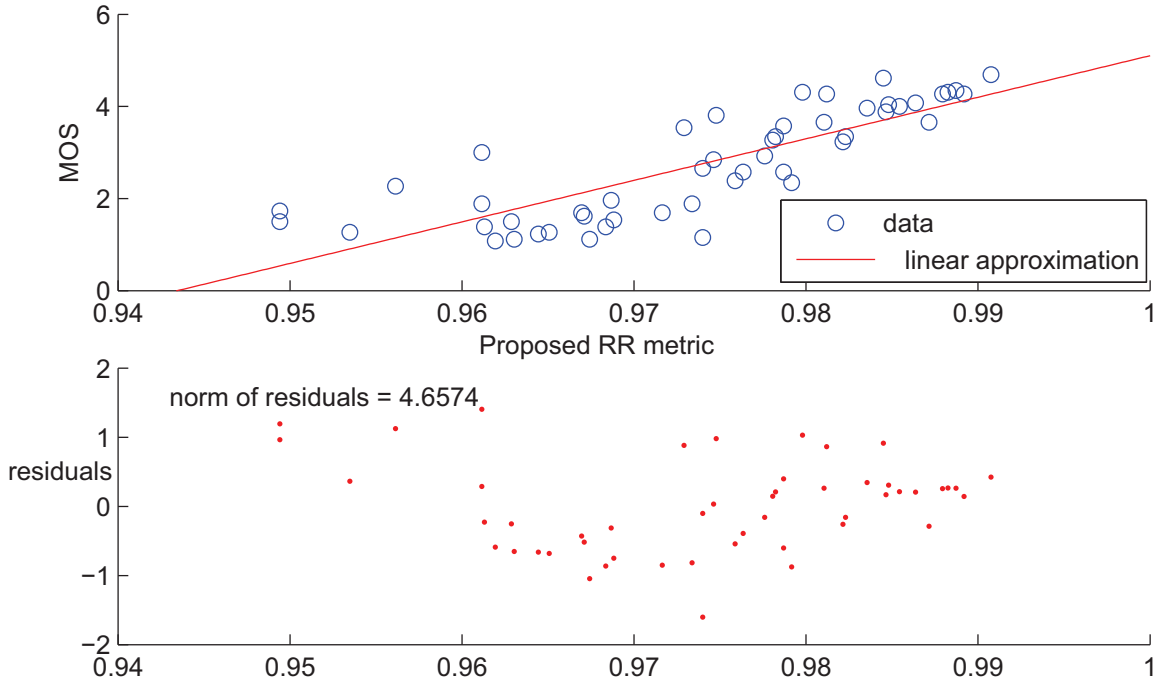


Figure 60: JPEG compression, IRCCyN/IVC image database [17] - Proposed metric. Above: scatter plot between mean opinion score and proposed metric. Below: residuals for the linear approximation and norm of residuals.

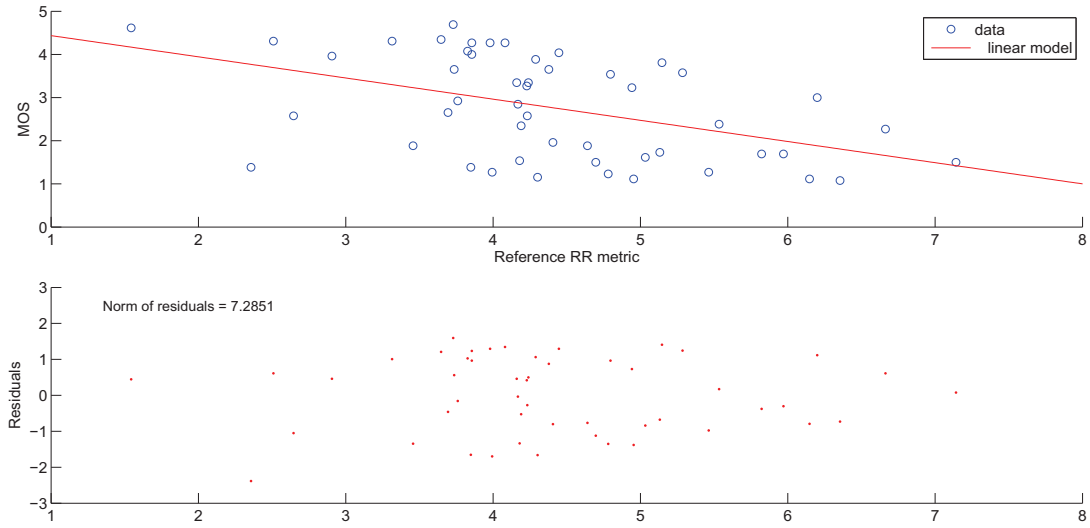


Figure 61: JPEG compression, IRCCyN/IVC image database [17] - Reduced reference metric in [152]. Above: scatter plot between mean opinion score and metric in [152]. Below: residuals for the linear approximation and norm of residuals.

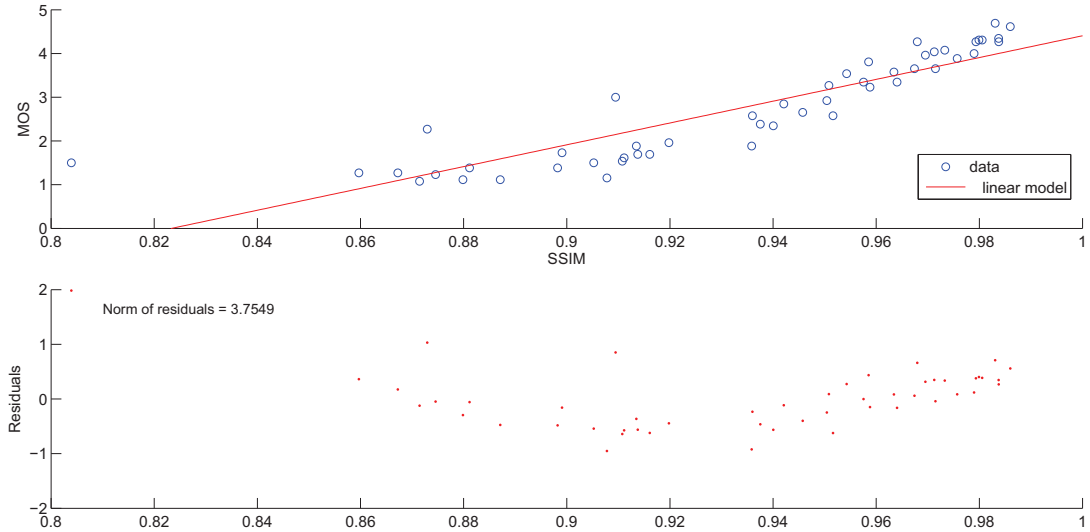


Figure 62: JPEG compression - IRCCyN/IVC image database [17] - MSSIM. Above: scatter plot between mean opinion score and MSSIM. Below: residuals for the linear approximation and norm of residuals.

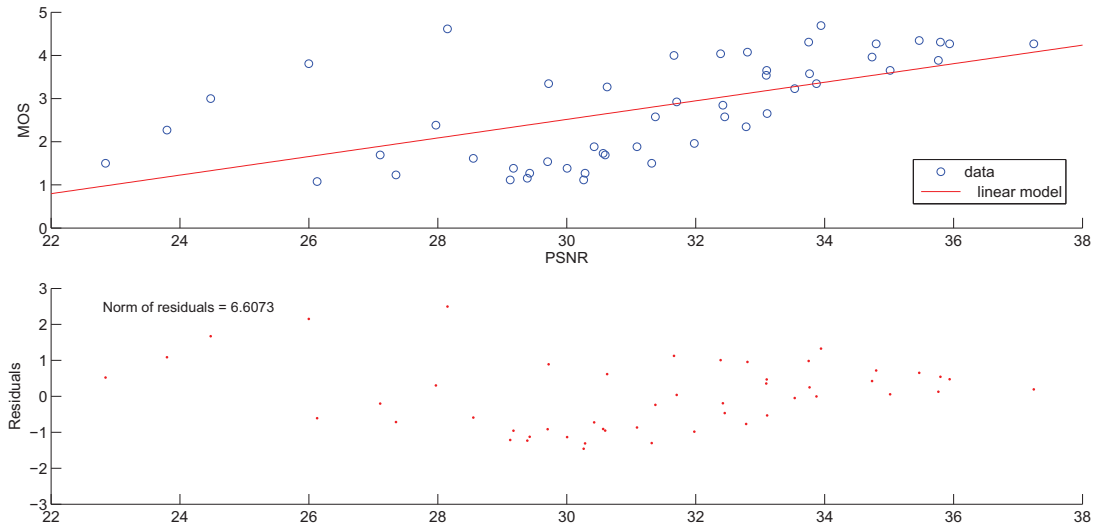


Figure 63: JPEG compression - IRCCyN/IVC image database [17] - PSNR. Above: scatter plot between mean opinion score and PSNR. Below: residuals for the linear approximation and norm of residuals.

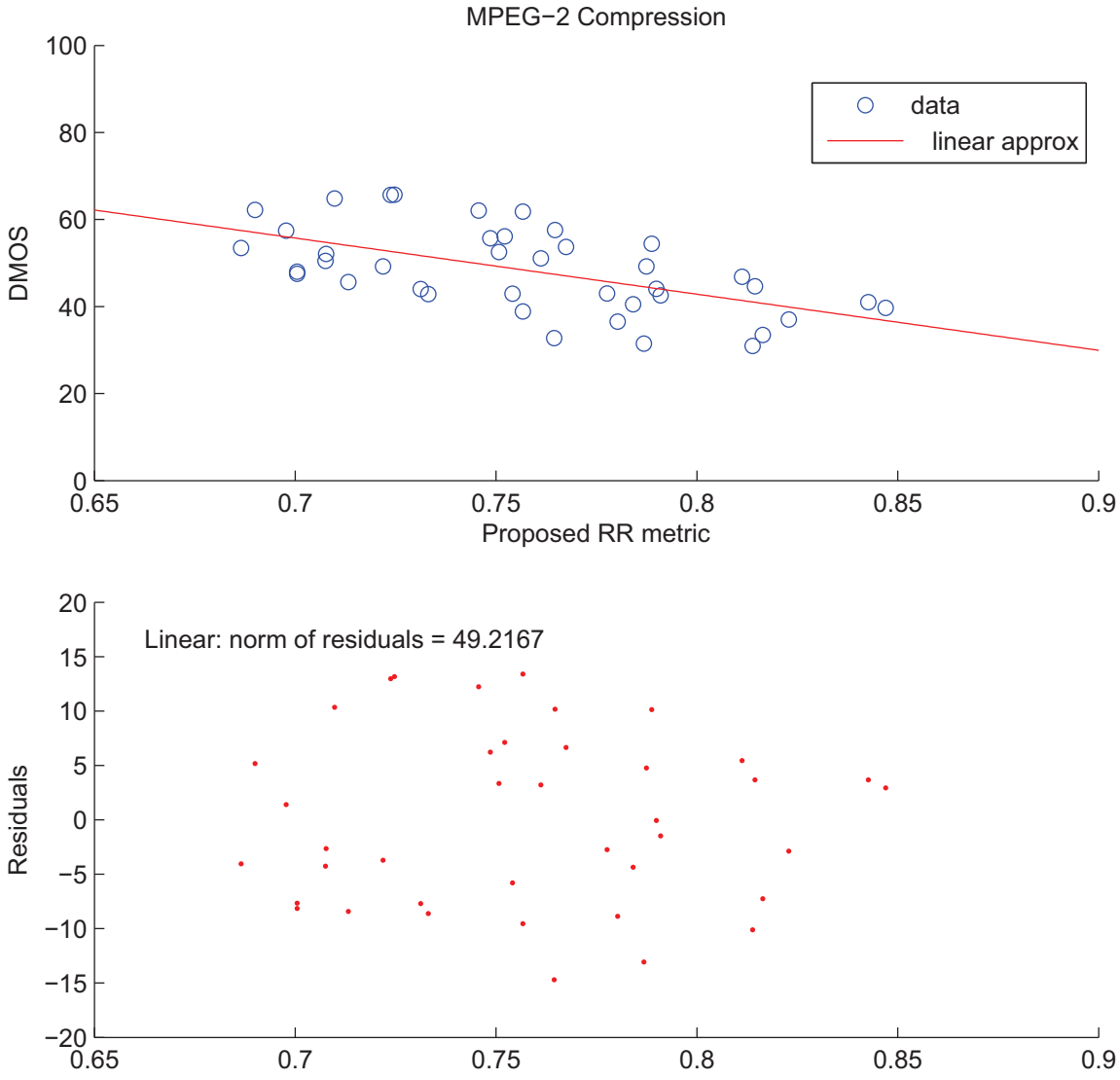


Figure 64: MPEG-2 compression - LIVE video database [117]. Above: scatter plot between diff. mean opinion score and proposed metric. Below: residuals for the linear approximation and norm of residuals.

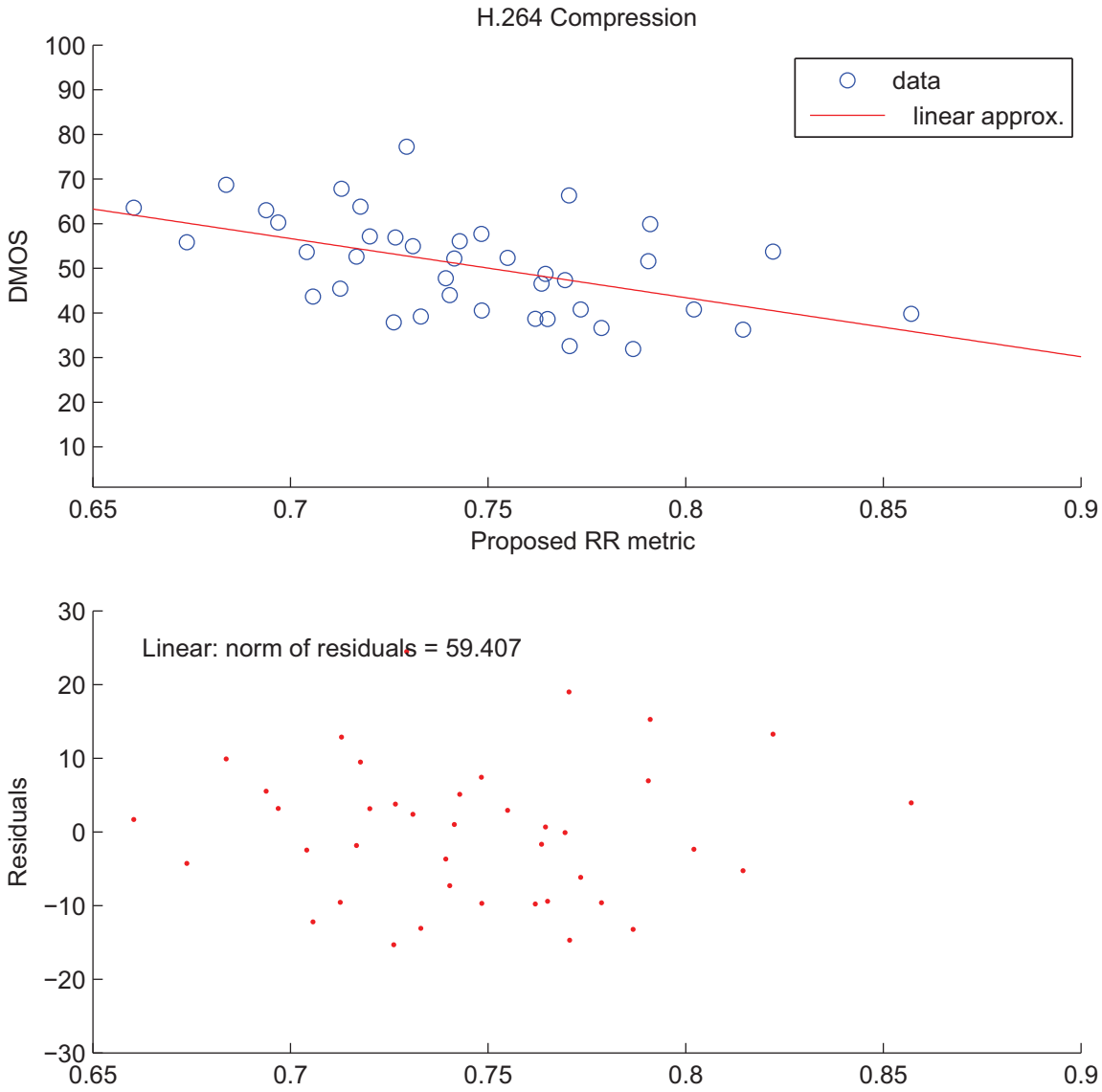


Figure 65: H.264 compression - LIVE video database [117]. Above: scatter plot between diff. mean opinion score and proposed metric. Below: residuals for the linear approximation and norm of residuals.

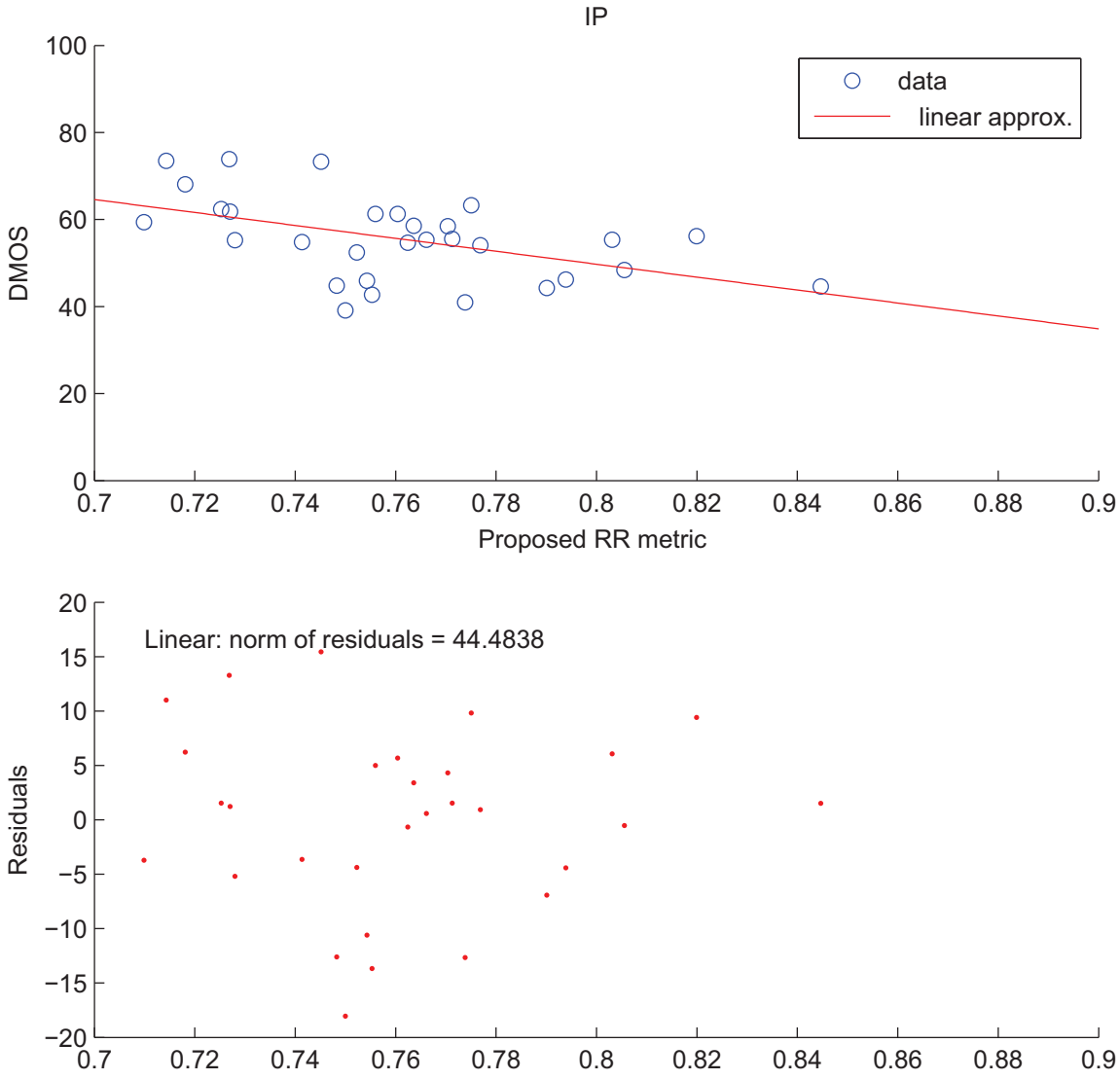


Figure 66: IP distortion - LIVE video database [117]. Above: scatter plot between diff. mean opinion score and proposed metric. Below: residuals for the linear approximation and norm of residuals.

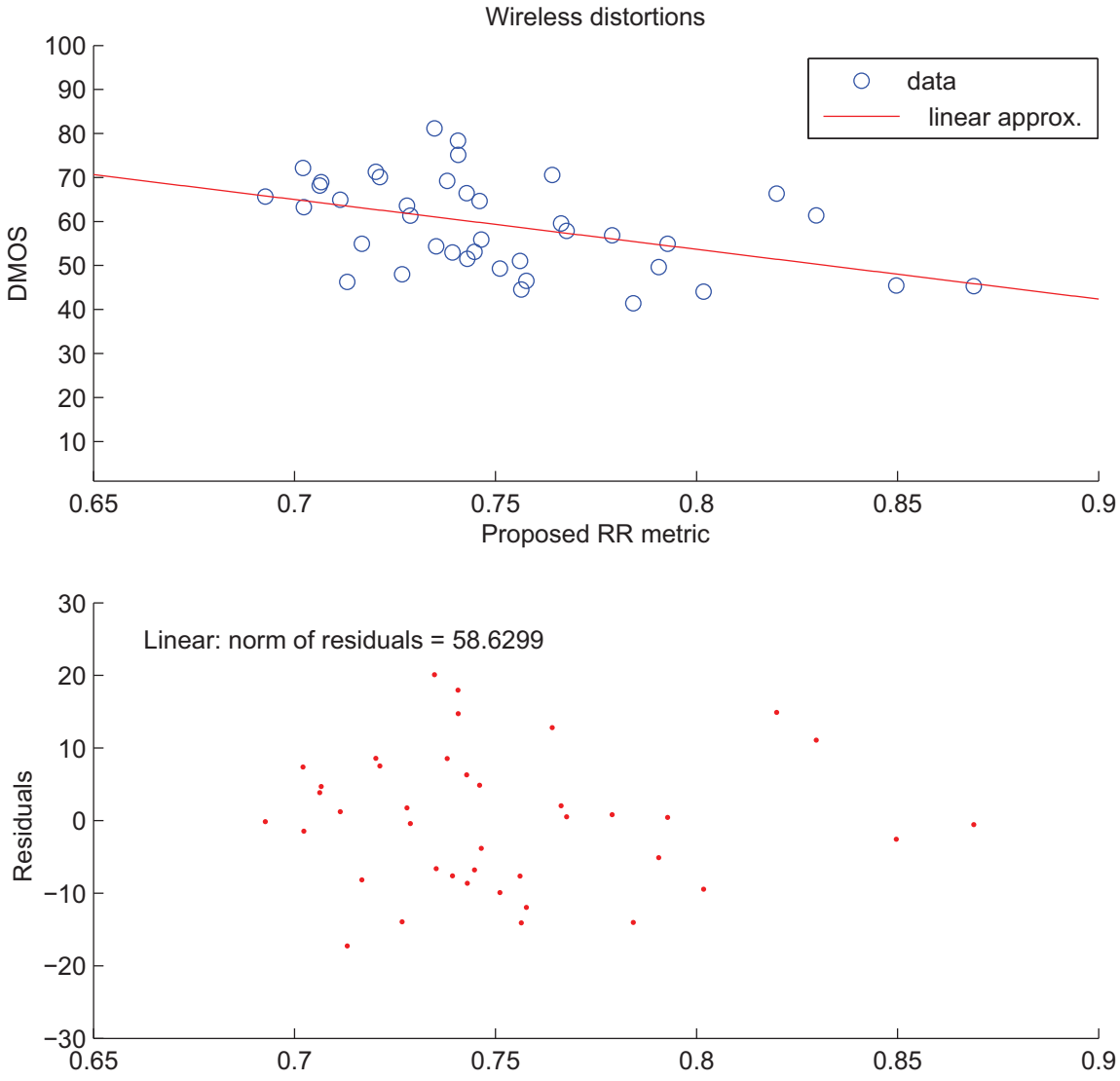


Figure 67: Wireless distortion - LIVE video database [117]. Above: scatter plot between diff. mean opinion score and proposed metric. Below: residuals for the linear approximation and norm of residuals.

Table 11: Norm of residuals versus DMOS, LIVE image database [123].

	PSNR	RR [152]	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 12: Correlation coefficient versus DMOS, LIVE image database [123].

	PSNR	RR [152]	RR in [68] (a)	RR in [68] (b)	Proposed RR	MSSIM
Fast fading	0.8556	0.9175	0.9243	0.9464	0.9478	0.9439 [152]
White noise	0.981	0.8889	0.9401	0.9654	0.9513	0.9706 [152]
Gaussian blur	0.79491	0.8872	0.8773	0.9561	0.9611	0.9361 [152]
JPEG comp.	0.8245	-	-	-	0.9456	0.958 [18]
JPEG2000 comp.	0.8703	-	-	-	0.9407	0.942 [18]

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 of 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.

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

$$I_{fr} = \frac{1}{N_{tot}} \sum_{l=1}^{N_{tot}} \left(1 - \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} |sc_{l,ij} - so_{l,ij}|}{mn} \right) \tag{35}$$

where the notation used is defined in Subsection 7.3, and N_{tot} is the total number of blocks in the image.

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 19. This confirms that the selected pattern well represents the region of interest of the image and enables a reliable quality assessment, although with a very limited overhead for the transmission of side information.

7.5 Alternative metric for image quality evaluation based on edge detection

Similar as above, we propose to divide images in sub-windows, as shown in Figure 68. For instance, if images have size 512 x 768, after subsampling of a factor of two, we could consider 16×16 macroblocks of size 16×24 pixels each, or, after subsampling of a factor 1.5, we could consider 18×16 macroblocks with size 19×32 pixels each. The example in Figure 68 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

Table 13: Spearman rank versus DMOS, LIVE image database [123].

	PSNR	RR [152]	RR in [68] (a)	RR in [68] (b)	Proposed RR	MSSIM
Fast fading	0.8770	0.9162	0.9237	0.9443	0.8988	0.9435 [152]
White noise	0.9881	0.8639	0.9316	0.9559	0.8962	0.9581 [152]
Gaussian blur	0.7585	0.9145	0.8608	0.9584	0.9260	0.9705 [152]
JPEG comp.	0.850	-	-	-	0.920	0.964 [18]
JPEG2000 comp.	0.896	-	-	-	0.919	0.965 [18]

Table 14: Norm of residuals versus MOS, IRCCyN/IVC image database [17].

	PSNR	RR [152]	Proposed RR	MSSIM
JPEG compression	6.60	7.29	5.11	3.75
JPEG2000 compression	5.30	5.42	5.27	5.28

Table 15: Correlation coefficient versus MOS, IRCCyN/IVC image database [17].

	PSNR	RR [152]	Proposed RR	MSSIM	C4 [131] [18]
JPEG compression	0.5957	0.4644	0.7837	0.8897	0.92
JPEG2000 compression	0.8143	0.8043	0.8163	0.8149	0.925

Table 16: Spearman rank versus MOS, IRCCyN/IVC image database [17].

	PSNR	RR [152]	Proposed RR	MSSIM
JPEG compression	0.8794	0.4542	0.9162	0.9814
JPEG2000 compression	0.842	0.800	0.823	0.943

Table 17: Correlation coefficient versus DMOS, LIVE video database [117] [118] [119].

	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 18: Spearman rank versus DMOS, LIVE video database [117] [118] [119].

	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 19: Norm of residuals vs. DMOS, full reference vs. reduced reference 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

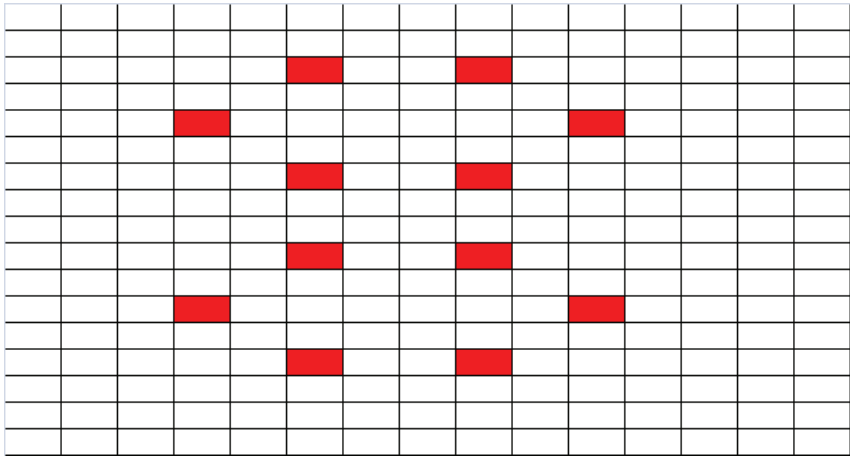


Figure 68: Example of block pattern selected based on visual attention models.

and the size of side information. In addition, sub-windows are non coincident with macroblocks, 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. 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 68 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 as above Sobel filtering, which is applied locally in these selected blocks.

Differently from what presented above, we propose here to measure the "displacement" of the edges. We define hence two threshold values, θ_1 and θ_2, where θ_1 and θ_2 represent two different distance values (in pixels) between original and displaced edges, for instance $\theta_1 = 2$ pixels and $\theta_2 = 5$ pixels; we define Δ_{θ_1} as the number of edge pixels with a displacement after processing lower than θ_1 and $\Delta_{\theta_1,\theta_2}$ the number of edge pixels with a displacement after processing θ, such that $\theta_1 \leq \theta \leq \theta_2$. In the quality assessment metric we consider all the edges displaced less than θ_2 pixels through the following weighted sum:

$$\Delta = w_1 \, \Delta_{\theta_1} + w_2 \, \Delta_{\theta_1,\theta_2} \tag{36}$$

The weights w_1 and w_2 and the thresholds θ_1 and θ_1 can be determined experimentally.

Hence, for each block l of image s the metric can be computed as

$$I_{s,l} = \Delta_l / p_l \tag{37}$$

where Δ_l is the weighted sum in (36) calculated for 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 finally defined here as

$$I_s = \frac{1}{N_b} \sum_{l=1}^{N_b} I_{s,l} \tag{38}$$

For instance, for the pattern considered below, $N_b = 12$. The lower N_b, the lower the overhead associated to the transmission of side information. We have observed - results are not reported here for brevity - that by selecting only some blocks in the image slightly reduces the performance with respect to the corresponding metric obtained with all the blocks, although the difference is not remarkable if the blocks are selected appropriately (e.g., according to visual attention studies, as in this work).

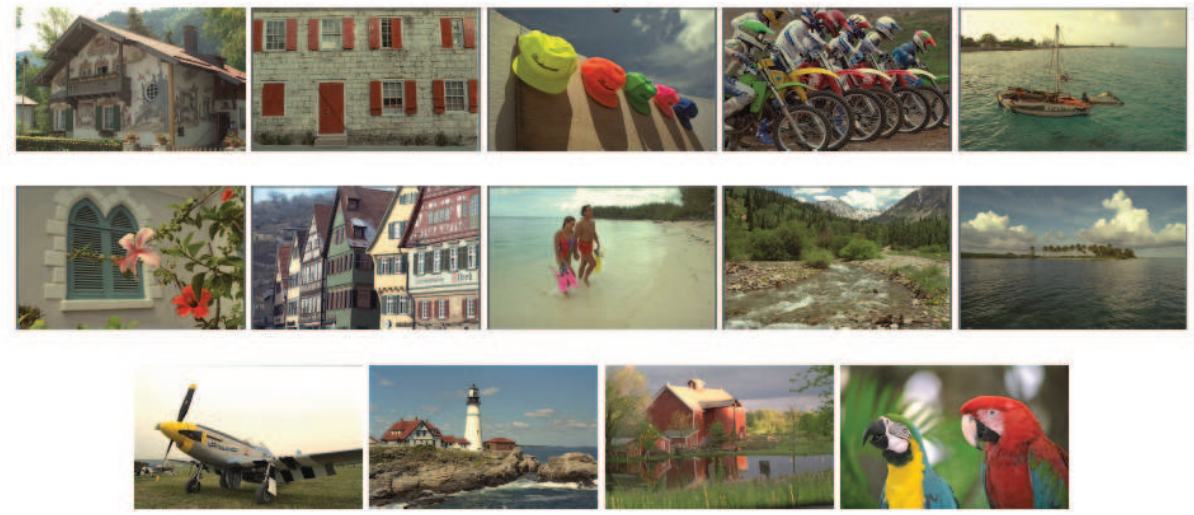


Figure 69: Images from the Toyama database [42].

7.5.1 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 $t = 0.001$.

7.5.2 Parameters selection

According to a number of tests performed on the different image databases, we have found that appropriate values for the metric parameters are $\theta_1 = 2$ pixels, $\theta_2 = 5$ pixels, $w_1 = 0.6$ and $w_2 = 0.4$. These are the values we consider in the following for the performance evaluation of the metric.

7.5.3 Performance evaluation

In addition to the image databases used above, we consider here the Toyama subjective database [42], that contains 182 images of 768x512 pixels, where 14 are original images (24 bit/pixel RGB). The rest of the images are JPEG and JPEG2000 coded images (84 compressed images for each type of distortion). Six quality scales and six compression ratios were respectively selected for the JPEG and JPEG2000 encoders. The subjective scores were collected using a calibrated CRT monitor in fixed viewing conditions. The subjective ratings were collected using a single stimulus absolute scaling. The overall ratings are presented in the form of MOS. The images in the database are reported in Figure 69.

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

Namely, we consider:

- MSSIM [149] (full reference);
- PSNR (full reference);
- [152] (reduced reference);

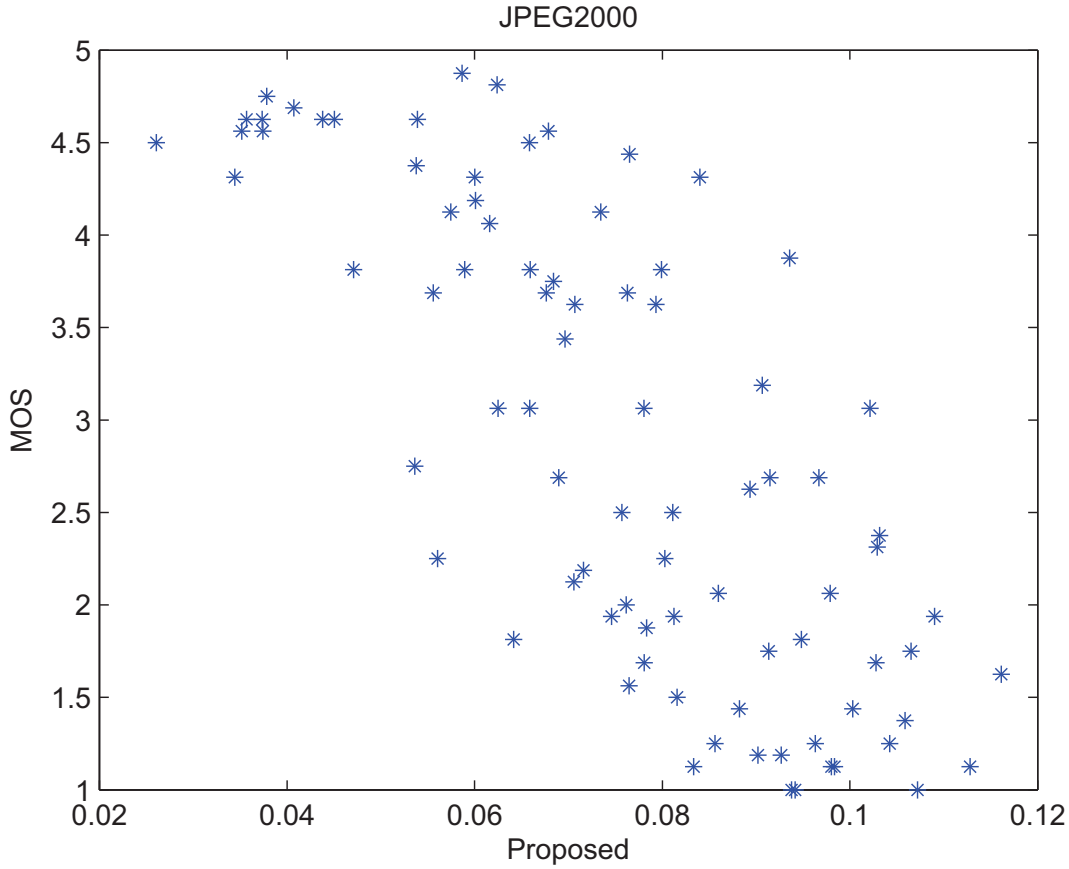


Figure 70: JPEG 2000 compression, Toyama image database [42] - Proposed metric.

- [18] (reduced reference);
- Proposed Sobel-based metric (reduced reference).

To apply the MSSIM metric, the images have been modified according to [150].

Example results for the Toyama database are reported in Figures 70, 71, 72, 73, 74, 75. In the scatter plots the objective quality metric is reported in the horizontal axis, whereas MOS/DMOS values are reported in the vertical axis. Each symbol in the plot refers to a different image in the database. In particular, Figure 70 reports the scatter plot for the proposed reduced-reference metric for the JPEG2000 images and Figure 71 reports the scatter plot for the MSSIM metric (full reference) for the JPEG2000 images. Figures 72, 73, 74, and 75 report the scatter plots for the JPEG images in the database. In particular Figure 72 refers to the proposed RR metric, Figure 73 refers to the RR metric in [152], Figure 74 refers to the MSSIM and Figure 75 refer to the PSNR metric. We can observe that for JPEG2000 compression our metric results in a scatter plot whose values are less dispersed not only with respect to the RR metric considered as a benchmark, but also with respect to the full reference SSIM and PSNR metrics.

Table 20 reports a summary of the results for the LIVE image database in terms of correlation coefficient, to enable an easy comparison with other metrics. We can observe that our metric well correlates with subjective tests, with results comparable to those achieved by full reference metrics. For the images in the LIVE database our metric outperforms the considered state-of-the-art reduced reference metric. Only for the case of JPEG2000 the benchmark reduced-reference metric from [152], which is based on the wavelet transform, provides a better performance in terms of norm of residuals. For the same type of impairment (JPEG2000 compression) our metric performs slightly worse than the benchmark one also when the images in the IRCCyN/IVC database [17] are considered, but it presents an evident improvement in the case of JPEG2000 compression. We should not neglect

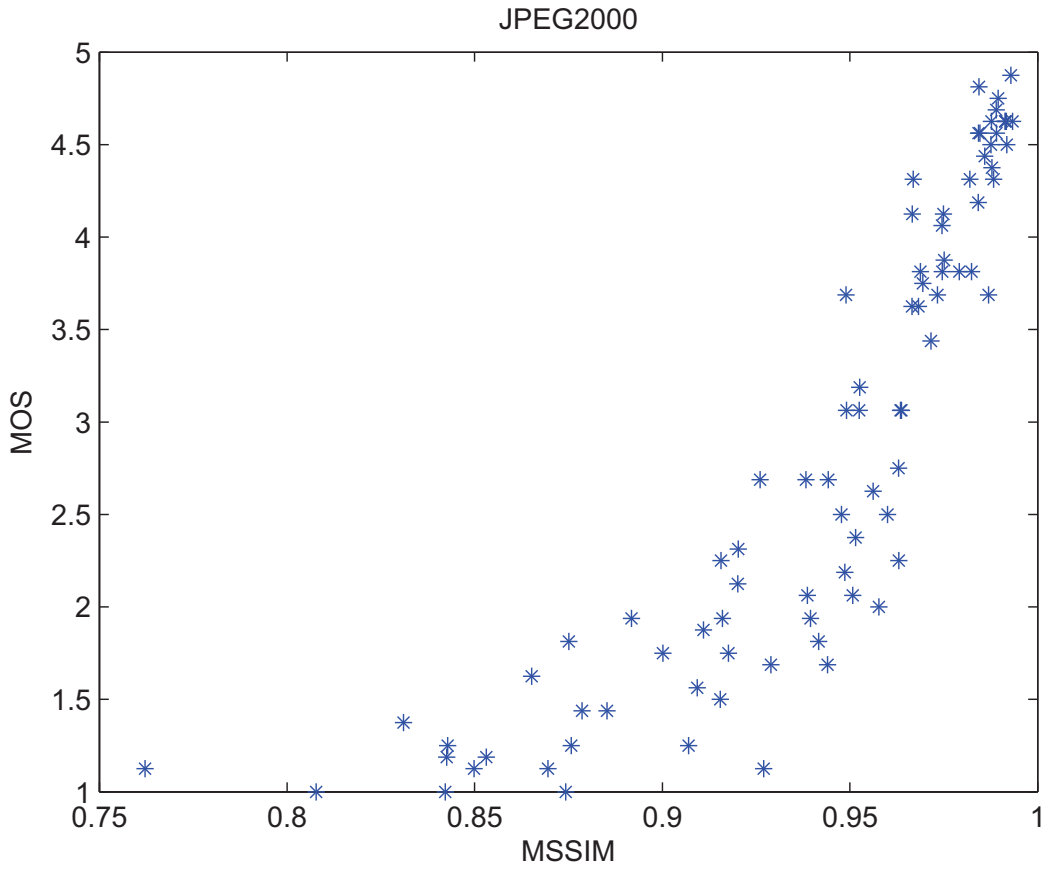


Figure 71: JPEG 2000 compression, Toyama image database [42] - SSIM.

that the metric [152] relies on the Wavelet transform, as well as the JPEG2000 compression scheme. Table 21 reports the results for the IVC image database in terms of correlation coefficient. We observe that our metric outperforms both the RR metric in [152] and the full-reference metric PSNR for JPEG compression, but not for JPEG2000. We reported for completeness the results in terms of correlation coefficient for the metric [18]. This metric has very high correlation with subjective results; it is however too complex when real time implementation is required. Finally, Table 22 reports the results for the Toyama image database in terms of correlation coefficient. We can observe that the results in terms of correlation coefficient confirm what observed based on the scatter plots.

Table 20: Correlation coefficient versus DMOS, LIVE image database [123].

	PSNR	RR [152]	Proposed RR	MSSIM
Fast fading	0.8556	0.9175	0.9418	0.9439 [152]
White noise	0.981	0.8889	0.9573	0.9706 [152]
Gaussian blur	0.79491	0.8872	0.9627	0.9361 [152]
JPEG comp.	0.8245	0.8927	0.9529	0.958 [18]
JPEG2000 comp.	0.8703	0.9663	0.9536	0.942 [18]

7.6 Conclusion

We proposed in this section a perceptual reduced reference image and video quality metric which compares edge information between portions of the distorted image and the original one by using Sobel filtering. The algorithms

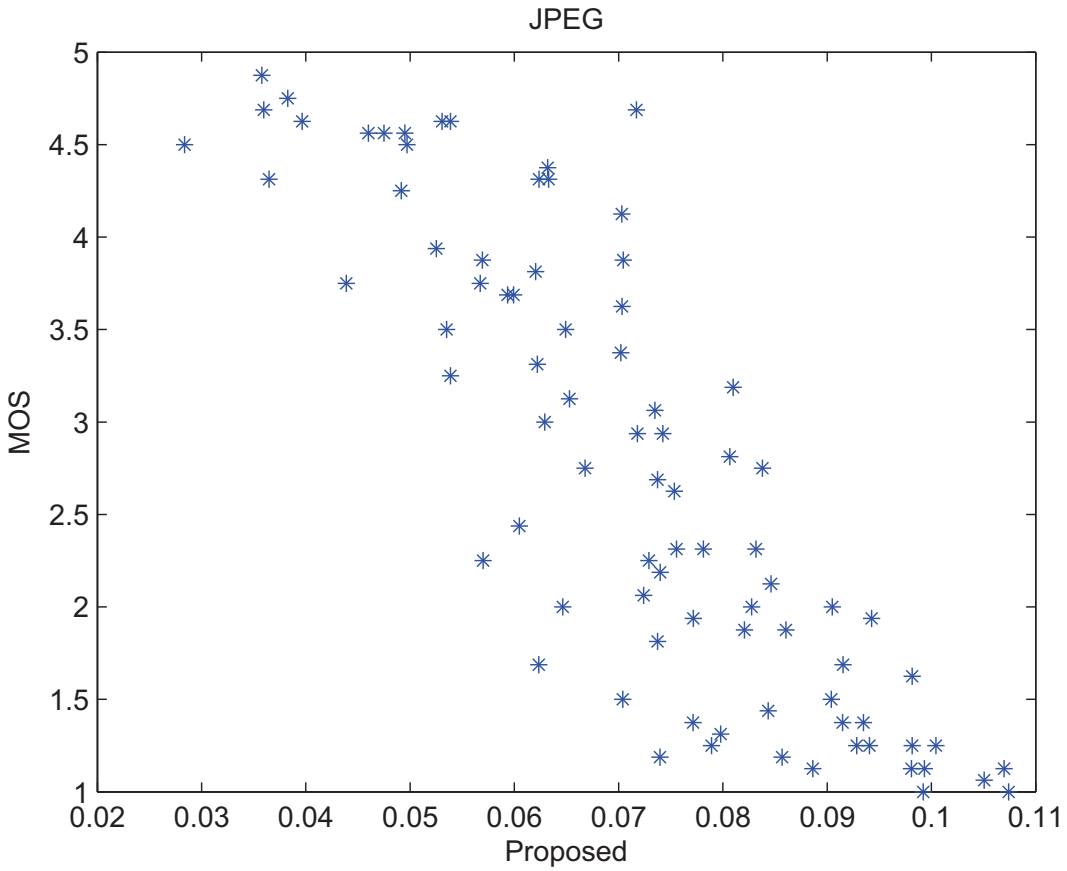


Figure 72: JPEG compression - Toyama image database [42] - Proposed metric.

Table 21: Correlation coefficient versus MOS, IRCCyN/IVC image database [17].

	PSNR	RR [152]	Proposed RR	MSSIM	C4 [131] [18]
JPEG compression	0.5957	0.4644	0.66	0.8897	0.92
JPEG2000 compression	0.8143	0.8043	0.72	0.8149	0.925

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 reduced-reference metrics. Note that this work has been co-published in [78] and [79].

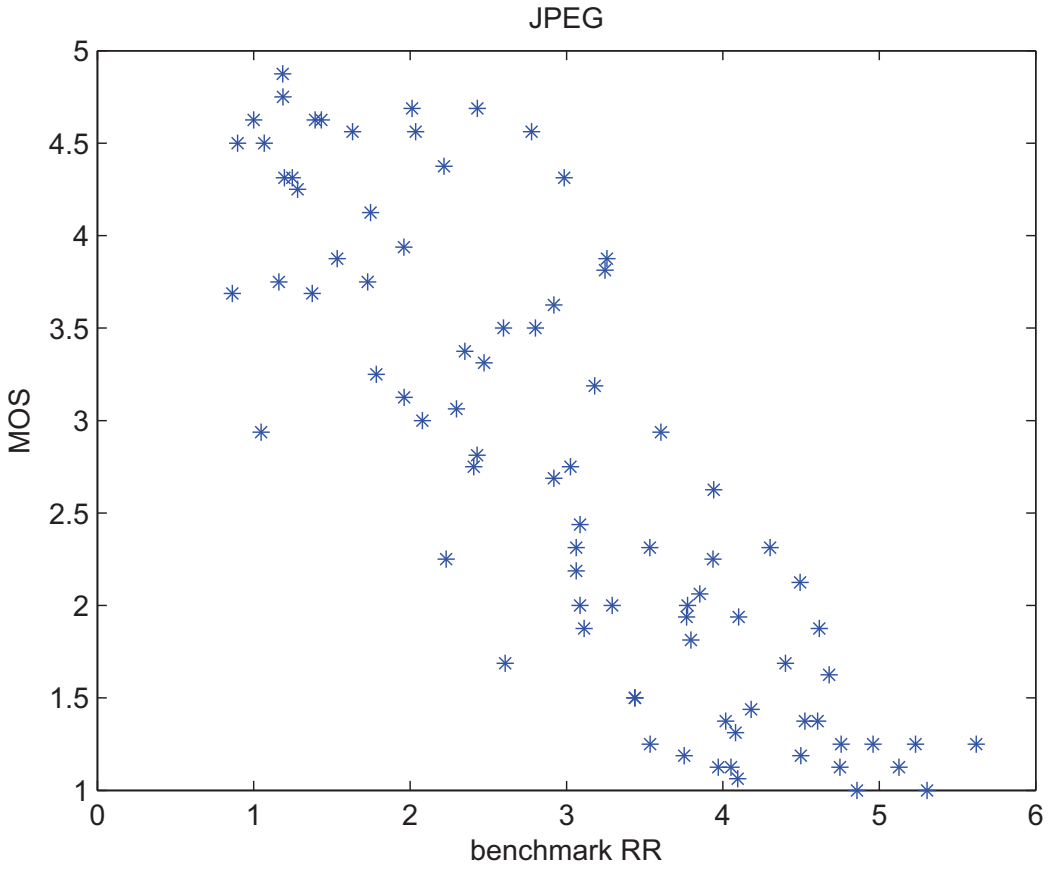


Figure 73: JPEG compression - Toyama image database [42], Benchmark RR metric [152].

Table 22: Correlation coefficient versus MOS, Toyama image database [42].

	PSNR	Reduced Reference [152]	Proposed RR	MSSIM	C4 [131] [18]
JPEG compression	0.61	0.8486	0.8502	0.8141	0.887
JPEG2000 compression	0.82	0.9108	0.7629	0.8581	0.934

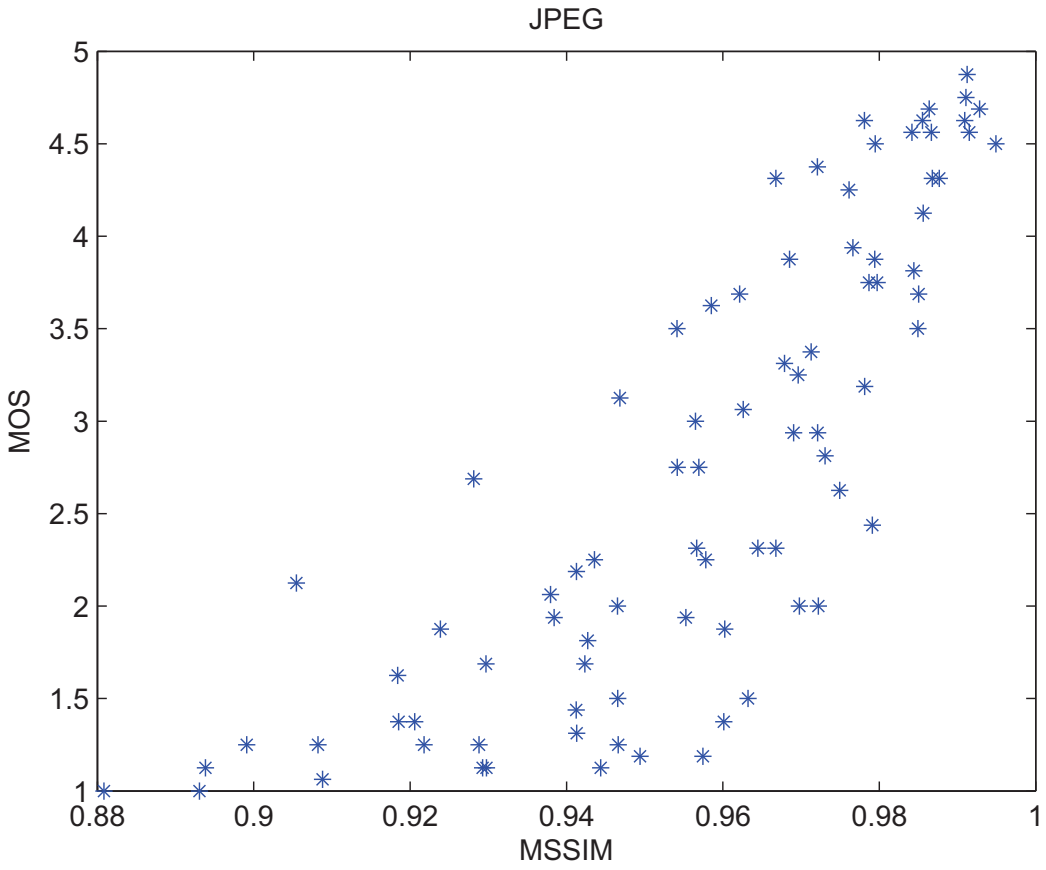


Figure 74: JPEG compression - Toyama image database [42] - SSIM.

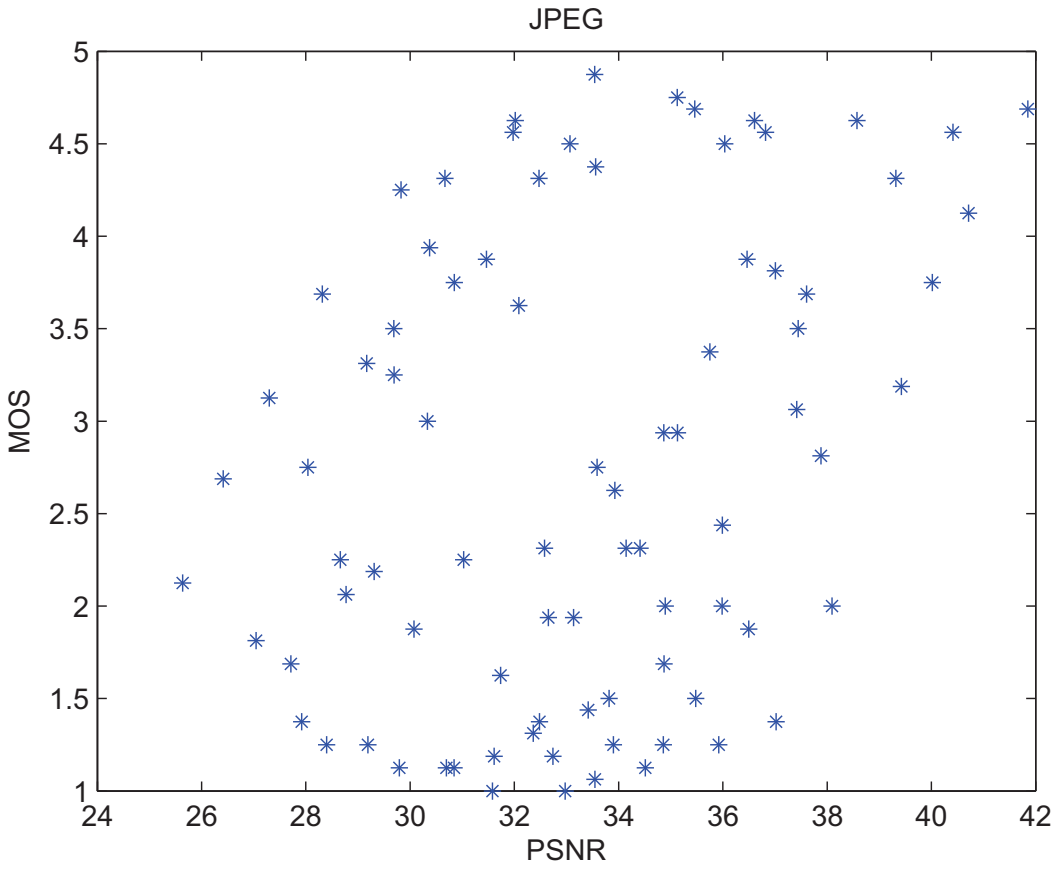


Figure 75: JPEG compression - Toyama image database [42] - PSNR.

8 Reduced Reference Quality Assessment for 3D Video

3D video applications are delivered over a range of different transmission systems. In order to provide demanding customers with a better service over unreliable communication channels, compression and transmission system parameters can be changed on the fly. For interactive 3D video services, video compression can be adapted (e.g., it can be made more robust and/or rate adaptive) based on the quality measured at the receiver. It has been shown that measuring the (3D) video quality at the receiver-side, and using this information as a feedback to fine tune the system parameters, will result in improved performance in such systems. However, measuring 3D video quality using Full-Reference (FR) quality metrics is not feasible due to the need of the original 3D video sequence at the receiver-side for comparison. Therefore, this contribution proposes a Reduced-Reference (RR) quality metric for colour plus depth 3D video compression and transmission, using the extracted edge information of colour plus depth map 3D video. This work is motivated by the fact that the edges/contours of the depth map can represent different depth levels and this can be considered for measuring structural degradations. Since depth map boundaries are also coincident with the corresponding colour image object boundaries, edge information of the colour image and of the depth map is compared to obtain a quality index (structural degradation) for the corresponding colour image sequence. The performance of the method is evaluated for different compression ratios and network conditions. The proposed method achieves good results compared to its counterpart FR quality metric, with a lower overhead for side-information.

8.1 Introduction

3D video processing technologies and related standardization efforts have been evolving fast to cater for diverse needs of future 3D video services [92]. Even though much effort has been spent on developing novel 3D video processing algorithms (e.g., MVC compression methods), the adaptation of existing technologies have not been studied in depth. For instance, the effects of time variant channel conditions and different user contexts can be addressed by changing compression parameters on the fly.

Such adjustment of compression and transmission system parameters in real time can be achieved by utilizing the video quality measured at the receiver-side as feedback information [77]. However, measuring the 3D video quality at the receiver-side is a challenge due to:

- the complex nature of 3D video quality [144];
- operational difficulties of using Full-Reference quality metrics, due to the need to have the original image sequence at the receiver to measure the quality.

Therefore, this section addresses how we can measure the 3D video quality at the receiver-end by using a RR quality metric to better adapt compression and transmission parameters for emerging 3D video services.

3D video quality can be described as a collection of different perceptual attributes such as depth perception, presence, eye strain, etc. Therefore, measuring the effect of different system parameter changes (e.g., compression level, Packet Loss Rate (PLR)) on 3D quality is a complex procedure. Even though several studies are present on subjective 3D video quality evaluations under different system parameter settings [95][43], these have a number of limitations and can only be utilized for certain types of image artifacts (e.g., compression artifacts). Objective quality metrics are also proposed for 3D video quality evaluations [53][11][54] even though their accuracy is inferior to subjective quality assessments. Due to this diversity and to the unavailability of accurate objective quality metrics for 3D video, rigorous and time consuming subjective test campaigns (e.g., according to recommendations ITU-R BT.1438 [114] and ITU-R BT. 500-11 [115]) are the only accurate and feasible method of measuring 3D video quality at present.

However, researchers have found out that there is a strong correlation between subjective 3D video quality ratings and candidate objective quality measures (e.g., video quality metric (VQM) [99], peak signal to noise ratio (PSNR), structure similarity metric (SSIM) [149]) of individual image components of 3D video, e.g., PSNR values of colour image and corresponding depth map or average PSNR of left and right view images [41][40][162]. This means that we can use individual objective quality ratings of 3D video components in place of time consuming subjective test procedures for most of the system parameter changes with a reasonable accuracy. Most of the candidate objective quality metrics (e.g., PSNR, SSIM) are Full-Reference methods. Therefore, the original 3D video sequence should be available at the receiver-side to evaluate the quality on the fly. This is not a viable

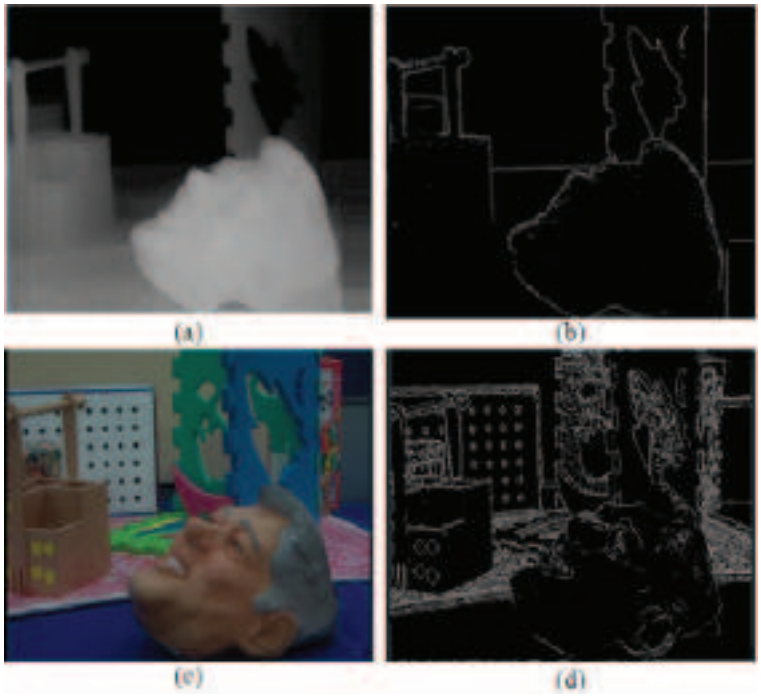


Figure 76: The *Orbi* sequence: (a) depth map, (b) extracted binary edge mask of the depth map using Sobel filtering, (c) corresponding colour image, and (d) extracted binary edge mask of the colour image using Sobel filtering.

solution, especially for bandwidth demanding 3D video applications. The only alternative is to resort to No-Reference (NR) and/or RR quality metrics for 3D video. NR quality metrics do not need any information of the original sequence at the receiver-side to evaluate the quality, whereas Reduced-Reference metrics only exploit a subset of the information of the transmitted image in order to perform a comparison. A number of NR and RR quality metrics are proposed for conventional 2D video (e.g., [10][151][164][158]). Even though the proposed NR/RR metrics for 2D video can be individually applied for different 3D video components separately (e.g., by applying a RR metric for left and right views of stereoscopic video separately), an effective metric can be invented by exploiting the common features of individual 3D video components. In addition, the overhead would be kept to a minimum by an approach of this nature. Studies focusing on 3D video quality evaluation (NR and RR methods) through the exploitation of feature correlations between individual 3D video streams (e.g., correlations between the colour image sequence and the corresponding depth map sequence) are not common at the moment. The study carried out in [9] proposed a NR quality metric for stereo images. However, this study is conducted using only left and right based stereoscopic images and colour plus depth map stereo images are not considered [9]. Furthermore, quality evaluation of stereo image sequences (i.e., stereoscopic video) is not considered. The work described in this section proposes a RR quality metric for colour plus depth 3D video transmission based on their corresponding objective quality measures.

The depth map of a colour plus depth 3D video sequence determines the position of the corresponding colour image in the 3D space (Depth Image-Based-Rendering, DIBR). Therefore, the quality of depth maps is crucial since they are used at the receiver-side to render novel views (e.g., adjacent views of free-view point video) [39][130]. Associated depth levels and individual objects of the depth map can be represented by the edges or the contour information (i.e., binary edge mask) of the depth maps (see Fig. 76). The proposed Reduced-Reference quality metric therefore compares the extracted edge information from the original and distorted depth maps to rate structural degradation due to distortions. More details about the proposed metric with edge detection using filtering are described in Subsection 8.2.

In addition, this edge information generated from depth maps identifies image objects in the corresponding colour image (see Fig. 76). Therefore, the extracted edge information of the processed colour image and original depth map are compared in the proposed method to quantify the structural degradation of the corresponding colour image. Subsection 8.2 elaborates more on colour image quality evaluation using the edges extracted from the colour image.

This section is organized as follows: Subsection 8.2 reports the proposed RR quality metric for colour plus depth 3D video transmission. This Subsection also investigates the overhead reduction of the proposed method compared to FR quality metrics. The experimental setup, results and discussion are presented in Subsection 8.3. Subsection 8.4 concludes the section.

8.2 Proposed quality metric

As described in Subsection 8.1, edges/contours extracted from a depth map can represent different depth planes and describe the basic structure of depth map objects. The importance of accurate depth map edges in novel view rendering has been discussed in [25]. This binary edge information is utilized in our proposed metric as side-information to evaluate the structural degradation of a depth map, with a reduced overhead compared to that of Full-Reference metrics. This structural comparison, together with other parameters related to luminance and contrast information from both the original and processed depth maps are employed/sent to rate the quality of the received depth map over error prone channels. The depth map also describes the basic structure of the corresponding colour image objects, i.e., boundaries of colour image and depth map objects are coincident, see Fig. 76. By relying on this observation, the edge information extracted from the corrupted colour image is compared with the edge information generated from the corresponding original depth map (i.e., side-information sent by the sender for depth map quality evaluation) to quantify the structural degradation of the processed colour image. The use of the extracted edge information to quantify the structural degradation in the proposed method is further justified by the fact that colour images with distorted edges will influence the rendered 3D video quality if depth discontinuities occur around the edges of the foreground colour image objects. The importance of having coincident edges for better novel view rendering with DIBR is discussed in [25]. Therefore, the use of colour edge information would also provide a better indication about the structural degradation of the rendered left and right views with the DIBR method. In order to evaluate the colour image quality with the proposed method, the only additional side-information necessary are the statistics related to luminance and contrast details of the original colour image (details are provided in the next few paragraphs of Subsection 8.2), since the side-information sent for the depth map quality (i.e., edge information of the original depth map) is being used to evaluate the structural degradation of the colour image as well.

Fig. 77 shows the effectiveness of using colour and depth map edge information to quantify the structural degradation in the proposed method. The $Ballet$ 3D video sequence is encoded using Quantization Parameter (QP) = 30 according to the H.264/AVC standard and corrupted with error traces generated for an IP core network (20% PLR) [155]. According to this figure, it is clear how effective is the use of the extracted depth map edge information as side-information for measuring the depth map structural degradation and later on for evaluating the structural degradation of the colour image. Since the edge comparison only accounts for the structural degradation between the original and processed images, luminance and contrast comparisons are also taken into account in the proposed method. In this way, distortions occurring at every location of the image (not only around the edges) are going to be covered and a performance similar to a Full-Reference method can be achieved.

In order to quantify structural comparison, luminance comparison and contrast comparison parameters for the depth map and corresponding colour image, the commonly used SSIM metric is adopted [149]. Equations (39), (40), (41), (42) and (43) describe the calculation of SSIM metric.

The Structural SIMilarity (SSIM) index between signal x and y is defined by:

$$SSIM(x,y) = [l(x,y)]^{\alpha}.[c(x,y)]^{\beta}.[s(x,y)]^{\gamma} \tag{39}$$

where $SSIM(x,y)$ represents the similarity map, $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ and $l(x,y)$, $c(x,y)$ and $s(x,y)$ represent luminance, contrast and structural comparisons respectively. The $l(x,y)$, $c(x,y)$ and $s(x,y)$ components are given by (40), (41) and (42) respectively

$$l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \tag{40}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \tag{41}$$

$$s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \tag{42}$$

where x and y are block vectors (i.e., local 8 x 8 square window) of images X and Y respectively, and are the mean of vector x and y respectively, and are the standard deviation of vector x and y respectively, is the covariance of vector x and y, and C_1, C_2 and C_3 are small constants to avoid the denominators being zero. Mean and standard deviation are calculated per block (each block is 8x8). The luminance comparison mainly provides a rating for the average intensity of the image (using the mean value) whereas the contrast comparison provides an indication about the variation of the pixel intensity of the image frame (using the standard deviation of individual images). The structural comparison provides an indication about the structural degradation compared to the original image by calculating the covariance between the original and processed images. The mean SSIM (MSSIM) index for overall image quality evaluation is defined as

$$MSSIM(X,Y) = \frac{1}{M}\sum_{j=1}^{M} SSIM(x_j, y_j) \tag{43}$$

where X and Y are the reference and the distorted images, respectively; x_j and y_j are the image contents at the j^{th} local window; and M is the number of local windows in the image. The MSSIM index provides a measurement of how similar the images X and Y are.

In the proposed method, the luminance and contrast comparisons are calculated based on the corresponding colour and depth map images, as described in (40) and (41). However, the structural comparison is performed on the extracted edge information (i.e., gradient image) of colour and depth map images rather than on the pixel-domain. In case of depth map quality evaluation, the edge information of the original and processed depth maps is compared to obtain an index for structural degradation of the depth map using (42) (i.e., $s_{Depth}(x', y')$). Then the proposed SSIM based quality index for the depth map Q_{Depth} can be described as follows:

$$Q_{Depth}(x,y) = [l(x,y)]^{\alpha}.[c(x,y)]^{\beta}.[s_{Depth}(x^{'}, y^{'})]^{\gamma} \tag{44}$$

where $l(x,y)$ and $c(x,y)$ are luminance and contrast comparisons performed on original and processed depth maps respectively, and $S_{Depth}(x', y')$ is the structural comparison between the gradient/edge maps of original and processed depth map images. Then the overall depth map quality is calculated as

$$MQ_{Depth}(X,Y) = \frac{1}{M}\sum_{j=1}^{M} Q_{Depth}(x_j, y_j) \tag{45}$$

where MQ_{Depth} is the mean SSIM quality rating for the depth map with the proposed method.

A similar SSIM metric based on edge information to calculate the structural degradation has been proposed for conventional 2D images in [19].

In case of the corresponding colour image quality evaluation, the edge information of the original depth map (i.e., side-information sent by the transmission-side) and the edge information extracted from the processed/received colour image are compared to obtain the index for structural comparison for the colour image using (42) (i.e., $s_{Colour}(x', y')$). Similar to (44), a structural comparison is performed for the edge map of the corresponding original depth map and the processed colour image instead of considering the actual pixel values of the original and processed colour images. Hence the proposed SSIM based quality index for the colour image Q_{Colour} can be described as follows:

$$Q_{Colour}(x,y) = [l(x,y)]^{\alpha}.[c(x,y)]^{\beta}.[s_{Colour}(x^{'}, y^{'})]^{\gamma} \tag{46}$$

where $l(x,y)$ and $c(x,y)$ are luminance and contrast indexes evaluated by comparing the original and processed colour images respectively, and $s_{Colour}(x', y')$ is the structural index evaluated by comparison between the gradient maps of the corresponding original depth map and the processed colour image. Hence the overall colour image quality is calculated as

$$MQ_{Colour}(X,Y) = \frac{1}{M}\sum_{j=1}^{M} Q_{Colour}(x_j, y_j) \tag{47}$$

where MQ_{Colour} is the mean SSIM quality rating for the depth map with the proposed method.

Since the structural comparison with the proposed method depends on the accuracy of edge detection, a suitable edge detection scheme has to be utilized [76]. In this work, the Sobel operator is selected to obtain edge information (i.e., the binary edge mask) due to its simplicity and efficiency [159]. Initially the horizontal and vertical edges (i.e., $E_{Horizontal}(x,y)$ and $E_{Vertical}(x,y)$) are calculated by applying the Sobel filter in each image dimension. Then the gradient image $G(x,y)$ is calculated as follows

$$G(x,y) = \sqrt{|E_{Horizontal}(x,y))|^2 + |E_{Vertical}(x,y))|^2)} \tag{48}$$

The extracted gradient information $G(x,y)$ from the processed/received depth maps are employed to quantify the structural degradation (i.e., $s_{Depth}(x',y')$) in (44). In case of colour image structural comparison (i.e., $s_{Colour}(x',y')$), gradient images obtained from the original depth map and the corresponding processed colour images are compared.

However, due to the abstract level of information (i.e., extracted binary edge masks) being used for the structural comparison in the proposed method, the results may not always follow its counterpart Full-Reference method, where the original and processed depth maps/colour image information are compared to evaluate the respective structural degradations. Therefore, relationships could be derived between the Full-Reference and the proposed method using experimental findings for both colour and depth map sequences separately. The relationships between these two methods (i.e., FR and the proposed RR methods) could be given by (49) and (50) for depth map and colour video sequences respectively.

$$MSSIM_{Depth} \approx f_{Depth}(MQ_{Depth}(X,Y)) \tag{49}$$

$$MSSIM_{Colour} \approx f_{Colour}(MQ_{Colour}(X,Y)) \tag{50}$$

where $MSSIM_{Depth}$ and $MSSIM_{Colour}$ refer to Full-Reference quality ratings (mean values) achieved with SSIM for the depth map and the corresponding colour image sequences respectively.

The process for the evaluation of the proposed RR quality metric using edge information and statistics related to luminance and contrast comparisons as side-information is shown in Fig. 78. Initially, side-information (i.e., edge information) is generated from the original depth map by using Sobel filtering. Then, the binary edge map, together with the calculated statistics required for luminance and contrast comparisons at the receiver-side (see (40) and (41)), is transmitted over the RR channel to the receiver. Ideally, this RR channel should be lossless. In the case of in-band transmission of the side-information, a high protection through unequal error protection can be provided. At the receiver-side also the edge information is obtained from the processed/received depth map as well as from the received corresponding colour image. Then binary edge information/masks for the original and processed depth maps are compared to perform structural comparison for the depth map. For the corresponding colour image sequence, the edge map information sent from the sender-side as side-information and the edge map extracted from the corresponding received colour image are then compared to obtain a rating for the structural degradation of the colour image sequence. Afterwards, luminance and contrast comparisons are performed for both colour and depth map sequences using the statistics generated from both the original and processed colour and depth map sequences. Then (44) and (46) are employed to obtain the overall quality rating for the respective colour and depth map sequences. In this study the case of transmission of the compressed colour plus depth map 3D video over an IP core network is assumed, in order to evaluate the quality under different PLRs.

Even though the proposed metric provides SSIM based ratings for the colour and depth map sequences separately, the relationship between these objective measures and subjective quality ratings (e.g., overall image quality, depth perception) are studied in our previous research [41][40][162], as elaborated in Subsection 8.1. These studies show strong correlation between subjective and objective quality ratings (including SSIM) for a range of compression rates and PLRs.

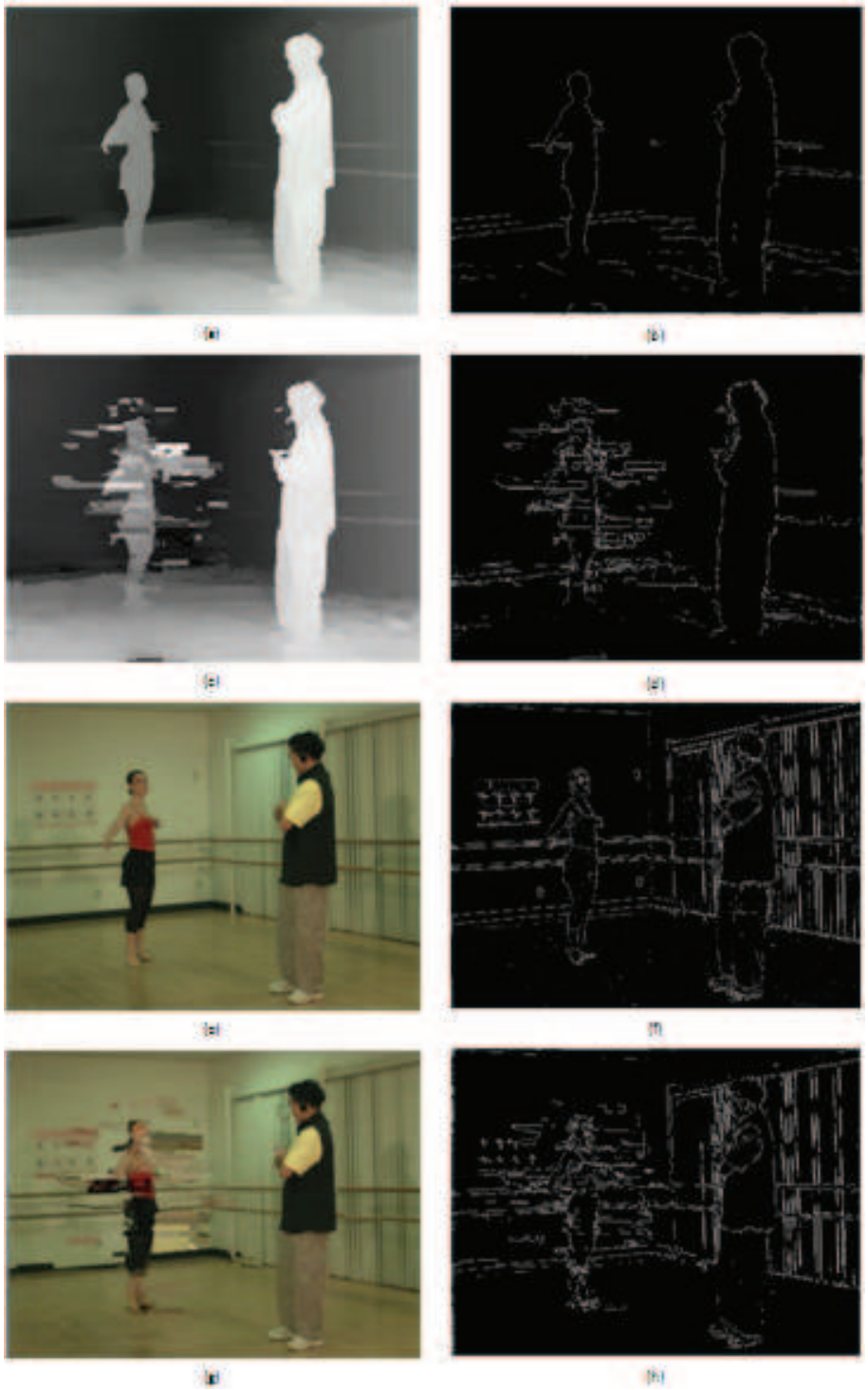
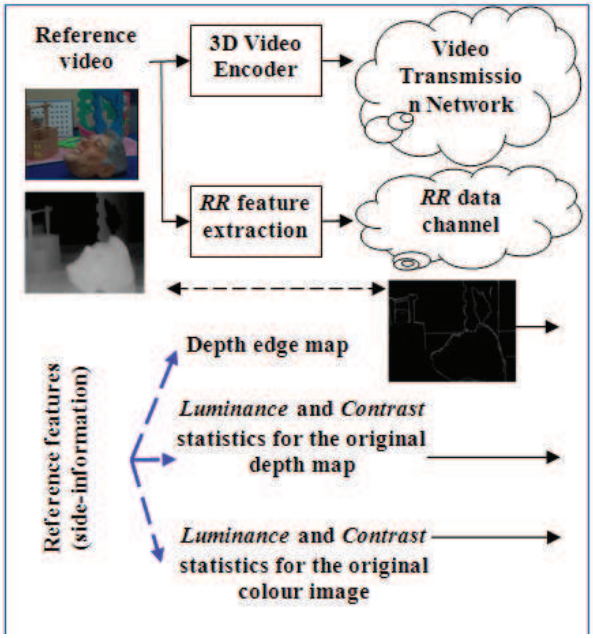
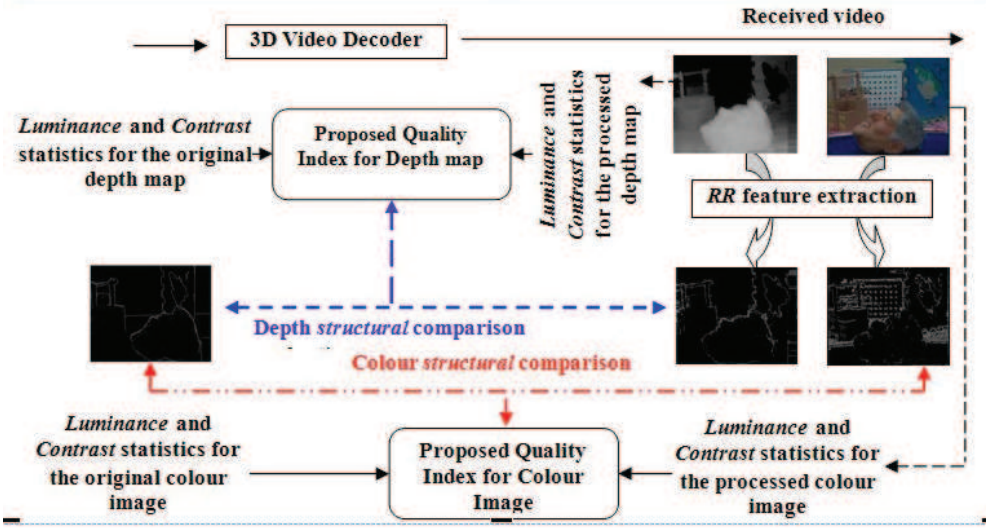


Figure 77: Graphical illustration of the proposed quality metric using the *Ballet* sequence (30th frame) coded with QP 30 and transmitted over a network with PLR 20%: (a) original depth map, (b) extracted edge information from the original depth map, (c) processed depth map, (d) extracted edge information from the processed depth map, (e) original colour image, (f) extracted edge information from the original colour image, (g) processed colour image, and (h) extracted edge information from the processed colour image.



(a)



(b)

Figure 78: Block diagram of the proposed Reduced-Reference quality metric: (a) sender-side, (b) receiver-side.

Since the reference edge information needs to be transmitted over the channel, either in-band or on a dedicated connection, the overhead associated to this side-information should be kept at a minimum level. In the case of the proposed method, only the binary edge information of the original depth map (i.e., ones and zeros of the edge map/1-bit per pixel) and statistics required to perform luminance and contrast comparisons (mean and standard deviation) will be transmitted and hence require a lower bitrate than the Full-Reference methods.

If we assume a depth map with resolution 720x576 is represented with 8 bits per pixel, the FR method generates 8 x 720 x 576 = 3317760 bits (3.31776 Mbit) per depth map. For a colour image of YUV 4:4:4 (24-bits per pixel) format and 720x576 resolution, Full-Reference metrics may require up to 3 x (8 x 720 x 576) = 9953280 bits (9.95328 Mbit) per colour image. Therefore, the FR metric may require up to 13.27104 Mbit per colour plus depth map image. Since the proposed method transmits only the binary depth edge mask as side-information, it requires only 720 x 576 = 414720 bits (414.720 kbit) per image. This is a significant bitrate saving compared to that of the FR method. However, the number of bits required for the binary depth edge mask is still high. Therefore, in order to further reduce the overhead, the reference data (i.e., binary edge mask) can be compressed, e.g., through run-length encoding or by considering techniques for binary maps encoding such as those used for MPEG-4 [93][8]. Since the binary edge mask of the depth map is composed of a high number of zero values, a high compression rate is achievable. Moreover, the possibility of selecting a fewer number of blocks from the binary depth edge mask will be studied in our future work. Also, by assessing the quality of a video sequence through a reduced number of frames, the sending rate of side-information is reduced (e.g., side-information can be sent by every 25 frames). This approach can further reduce the bitrate requirement of the proposed method.

8.3 Experimental setup, results and discussion

In order to evaluate the performance of the proposed Reduced-Reference quality metric for colour plus depth 3D video, experiments are performed assuming compression and transmission of the video sequences over a network. Different PLRs (i.e., 0%, 3%, 5%, 10% and 20%) and compression levels (i.e., with different QP values) are considered. The $Orbi$, $Interview$, $Ballet$ and $Breakdance$ 3D video test sequences are encoded using the H.264/AVC video coding standard (JM reference software version 16.0). The selected test sequences cover a range of texture and motion characteristics and the depth maps of these sequences are obtained using a depth-range camera ($Orbi$ and $Interview$) and a stereo matching algorithm ($Ballet$ and $Breakdance$). Ten-second long sequences (i.e., 250 frames from $Orbi$ and $Interview$ sequences and 150 frames from $Ballet$ and $Breakdance$ sequences) are encoded with IPPPIPPP... format, using QP values 1, 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50. An I frame is encoded every second. Slices (one row of MBs = one slice) are also introduced in order to make the decoding process more robust to errors. The transmission of the encoded bit-stream over an IP core network is simulated by using IP error patterns generated for Internet experiments [155]. In order to obtain average results, random starting positions are used for the error pattern files. The corrupted bit-streams with different PLRs are later decoded using the JM reference software decoder. Slice copy is enabled during decoding to conceal the missing slices of the corrupted bit-stream. At each PLR and QP value, the quality is measured using both Full-Reference (i.e., SSIM) and the proposed method for both the colour and depth map images. Furthermore, to understand the relationship between the corrupted colour and depth map images and rendered left and right views, the proposed quality metrics for colour and depth maps are compared with the quality of the rendered left and right view as well. In order to obtain the left and right views using the corrupted colour and depth maps, the DIBR technique is utilized with standard settings. Simple hole filling algorithm is employed to conceal the occluded areas of the rendered images. The rendered left and right image quality is measured with the PSNR metric and the average PSNR is calculated to represent the rendered quality in general.

In order to map proposed RR quality ratings into true 3D video perception, subjective tests were performed using the DSCQS method with the participation of 16 observers. The corrupted colour and depth map sequences were first rendered into left and right image sequences using the DIBR algorithm, and subsequently arranged into side-by-side format and fed into a LG-47" passive stereoscopic display. Tests were conducted under normal room lighting condition and all the subjects were screened prior to the tests for normal visual acuity (using Snellen chart), stereo vision (using Randot stereo test) and colour blindness (using Ishihara chart). The subjective quality (overall 3D video quality) was measured for all the test sequences under different PLRs. The sequences generated with QP = 30 were selected for these tests. The Opinion Scores (OS) of individual subjects were averaged to obtain the Mean Opinion Score (MOS) for all the test cases.

Subsection 8.3.1 presents the performance of the proposed RR quality metric for depth maps, whereas sub-

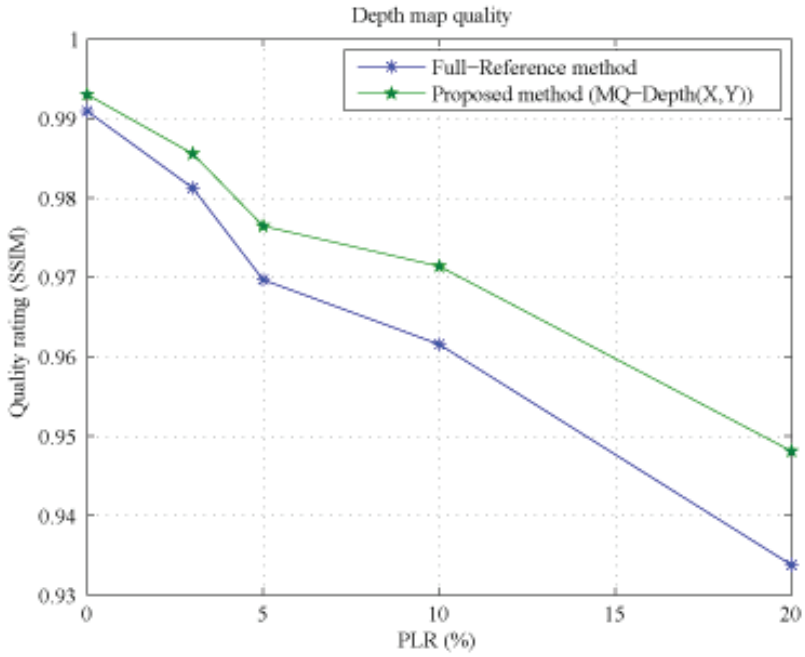


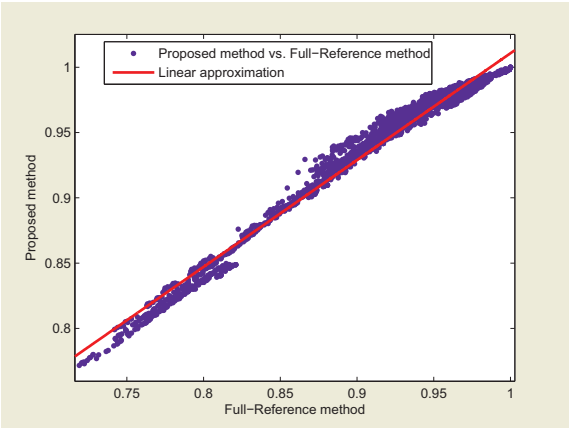
Figure 79: Measured image quality of the *Ballet* sequence at QP = 30 using FR and RR methods.

subsection 8.3.2 elaborates on the performance of the proposed RR metric for the corresponding colour image sequence.

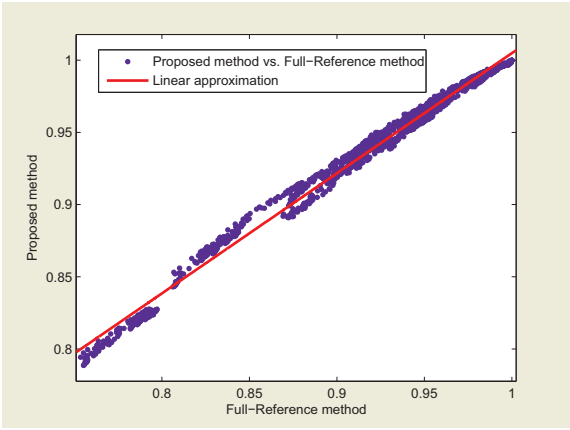
8.3.1 Depth Map Quality Evaluation ($MQ_{Depth}(X,Y)$)

Fig. 79 shows the quality measured with the Full-Reference and the proposed methods for the *Ballet* depth map image sequence at QP = 30 and all the considered PLRs. It can be seen that the proposed method closely follows the Full-Reference method with an insignificant offset between them. However this offset could be minimized by finding an approximation between the Full-Reference and the proposed methods based on experimental results as described in (49) and (50).

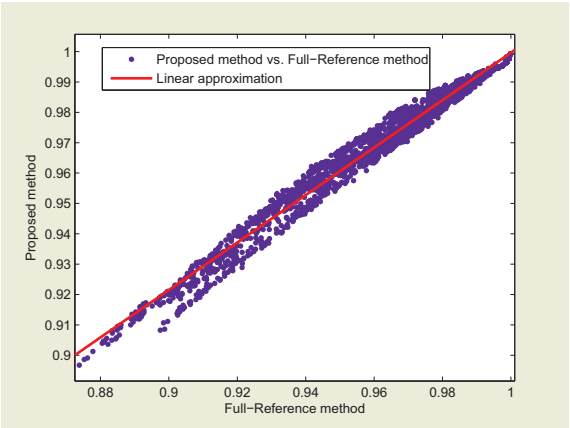
Fig. 80 shows the scatter plots for the measured depth map quality using both FR and proposed methods for *Orbi*, *Interview*, *Ballet* and *Breakdance* and all the sequences in general. Each point of this plot corresponds to the measured quality of a depth map frame using both the Full-Reference and the proposed RR methods for a given compression level (QPs ranging from 1 to 50) and PLR (0%, 3%, 5%, 10% and 20% PLRs). According to these figures it is evident that there is a close correlation between the quality ratings of these methods regardless of the compression level and PLR being used. In order to analyze the correlation between the reference and the proposed methods, the quality ratings are approximated by a linear polynomial. Linear approximations between the proposed and the Full-Reference methods are shown in each of the scatter plots in Fig. 80 to show the good fitting of the approximation.



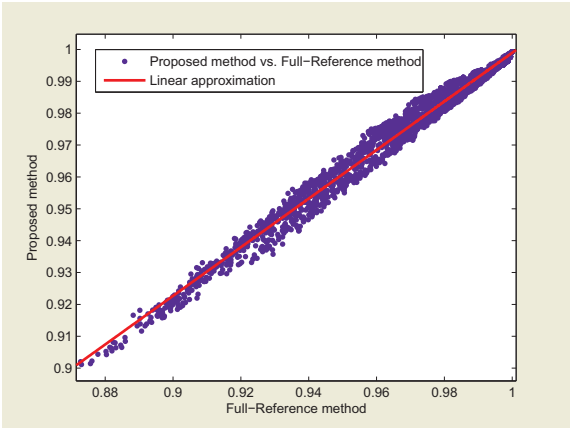
(a)



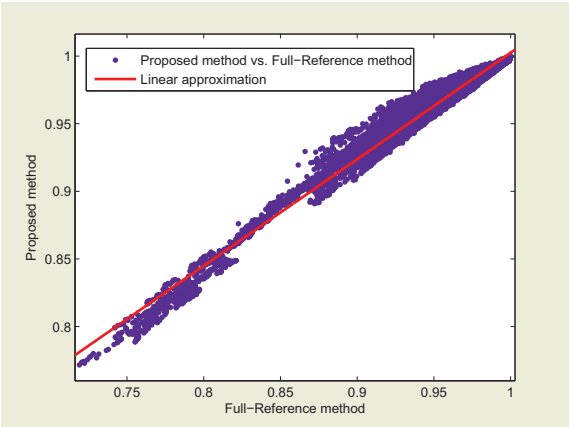
(b)



(c)



(d)



(e)

Figure 80: Scatter plots of the proposed RR method versus FR method for the (a) *Orbi*, (b) *Interview*, (c) *Ballet*, (d) *Breakdance*, and (e) all depth map sequences at different compression levels and PLRs.

Fig. 80 (e) shows the scatter plot for the measured depth map quality using both FR and the proposed RR methods for all the depth map sequences (i.e., *Orbi*, *Interview*, *Ballet* and *Breakdance*). Each point of this plot corresponds to the measured quality of a depth map frame using both the Full-Reference and proposed methods for a given sequence, compression level and PLR. The linear approximation between these methods is also plotted on the scatter plot. This shows that the proposed depth map quality evaluation method closely follows the Full-reference method regardless of the sequence type, compression level and packet loss rate being employed.

Coefficients measuring the validity of the approximation (i.e., Sum of Squares for Error (SSE), Pearson correlation of coefficient (R-square), and Root-mean-square error (RMSE) for the *Orbi*, *Interview*, *Ballet* and *Breakdance* depth map sequences and for all the sequences in general are listed in Table 23. According to Table 23, it is clear that the proposed RR method for depth map quality evaluation (i.e., $MQ_{Depth}(X,Y)$) is highly correlated with the FR method at all the considered QP levels, and PLRs. This suggests that we can use the proposed methods for depth map quality evaluation of colour plus depth map 3D video in place of the Full-Reference methods with a reasonable accuracy at all PLRs and compression levels regardless of the complexity of the depth map sequence.

Table 23: Correlation coefficients

Sequence	SSE	R-square	RMSE
Orbi	0.1452	0.9824	0.007268
Interview	0.05706	0.991	0.005095
Ballet	0.02693	0.9779	0.0033
Breakdance	0.01795	0.9848	0.002694
All sequences	0.3997	0.9795	0.006355

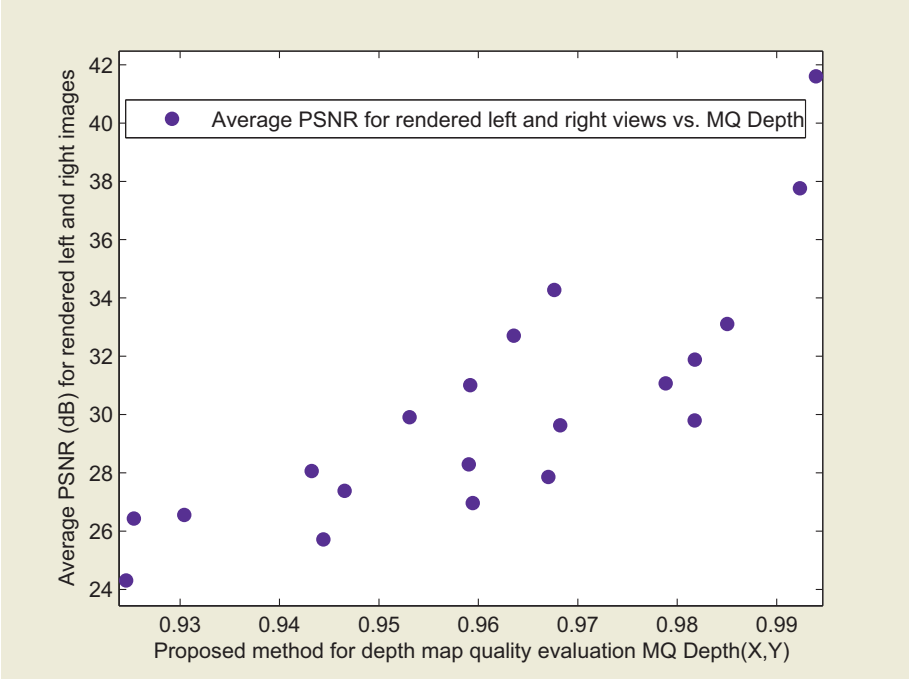


Figure 81: Scatter plot of rendered left and right image quality (average PSNR) versus depth image quality obtained with the proposed $MQ_{Depth}(X,Y)$ method.

The scatter plot in Fig. 81 shows the relationship between the proposed depth map quality evaluation method and the rendered left and right view quality. The Orbi and Interview sequences compressed using QP = 10 and QP = 30 are used to generate this scatter plot. Fig. 81 shows the degree of correlation between the proposed depth map quality measures and the average quality of the rendered left and right images.

8.3.2 Colour Image Quality Evaluation ($MQ_{Colour}(X,Y)$)

Fig. 82 shows the colour image quality measured with the Full-Reference and the proposed methods for the $Ballet$ sequence at QP = 30 and all the considered PLRs. It can be seen that the proposed method closely follows the Full-Reference method with an offset. As discussed in Subsubsection 8.3.1, this offset could be minimized by finding an approximation between the Full-Reference and the proposed methods based on experimental results as described in (49) and (50).

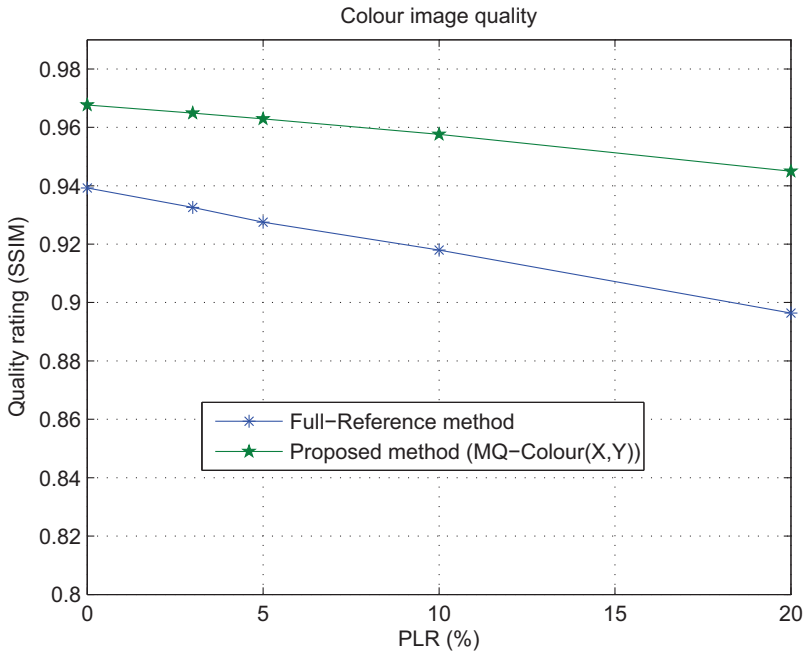


Figure 82: Measured colour image quality of the $Ballet$ image sequence using FR and RR methods at QP = 30.

Fig. 83 shows the scatter plots for the measured colour image quality using both the FR and the proposed methods for the sequences $Orbi$, $Interview$, $Ballet$, $Breakdance$, and for all the sequences in a single plot. Each point of this plot corresponds to the measured quality of a colour image frame using both the Full-Reference and the proposed RR methods for a given compression level (QPs ranging from 1 to 50 in steps of 5) and PLR (0%, 3%, 5%, 10% and 20% PLRs). According to these plots it is evident that there is a close correlation between the quality ratings of these methods regardless of the compression level and PLR being used. Linear approximations between the proposed and Full-Reference methods are shown in each of these scatter plots to show the degree of correlation.

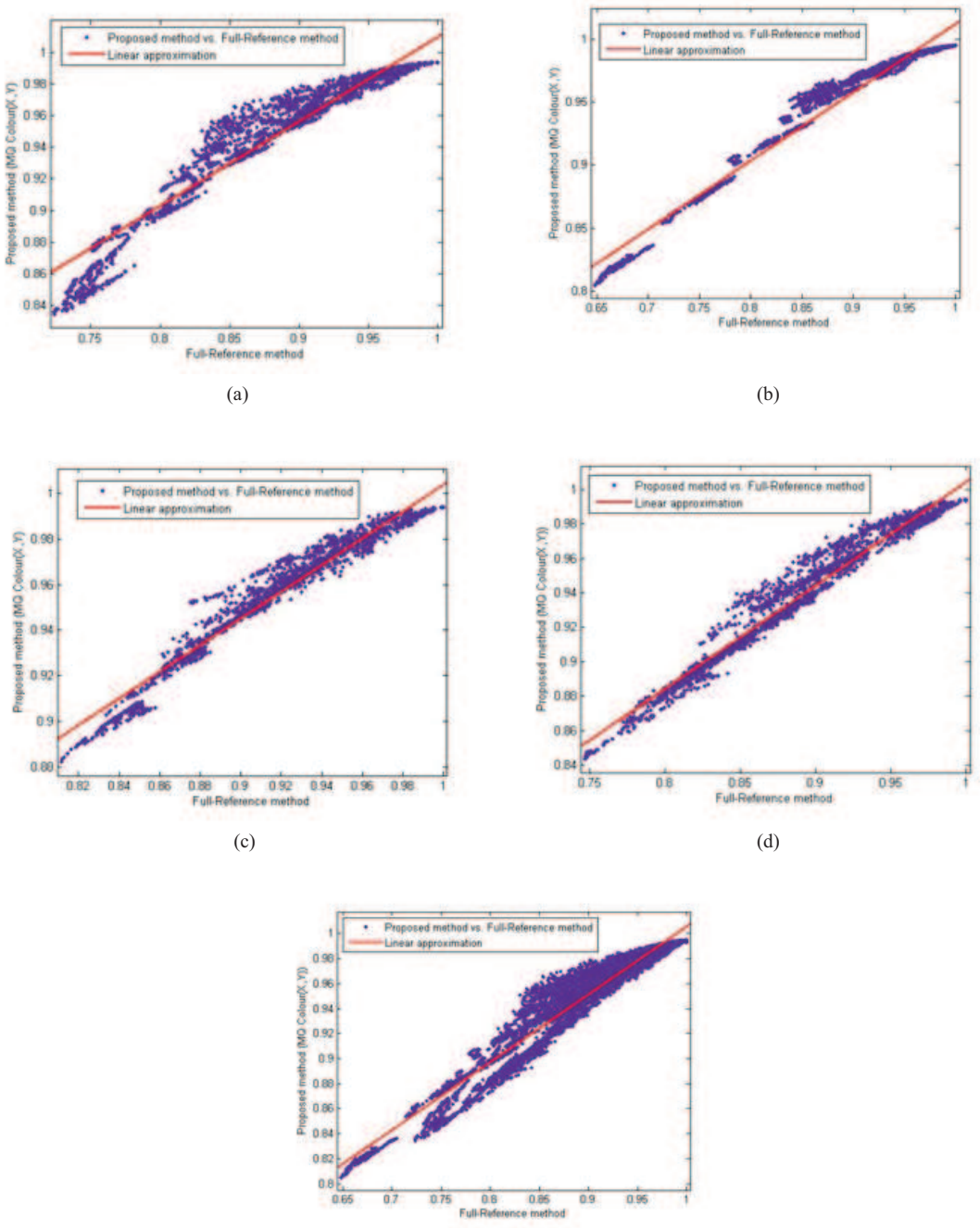


Figure 83: Scatter plot of proposed RR method versus FR method for (a) *Orbi*, (b) *Interview*, (c) *Ballet*, (d) *Breakdance*, and (e) all colour image sequences at different compression levels and PLRs.

Fig. 83 (e) shows the scatter plot for the measured colour image quality using both FR and the proposed RR methods for all the colour image sequences (i.e., *Orbi*, *Interview*, *Ballet* and *Breakdance*) considered in the experiments. Each point of this plot corresponds to the measured quality of a colour image frame using both the Full-Reference and proposed methods for a given sequence, compression level, and PLR. Linear approximation between these methods is also plotted on the scatter plot. This shows that the proposed colour image quality evaluation method closely follows the Full-Reference method regardless of the sequence type, compression level, and packet loss rate being employed.

The individual correlation coefficients (i.e., SSE, R-square, and RMSE) for the *Orbi*, *Interview*, *Ballet* and *Breakdance* colour image sequences and for all colour image sequences in general are listed in Table 24. According to Table 24, it is clear that the proposed RR method for colour image quality evaluation (i.e., $MQ_{Colour}(X,Y)$) is highly correlated with the FR method at all the considered QP levels, and PLRs. This suggests that we can use the proposed methods for the corresponding colour image quality evaluation of colour plus depth map 3D video in place of the Full-Reference methods with a reasonable accuracy at all PLRs and compression levels.

Table 24: Correlation coefficients

Sequence	SSE	R-square	RMSE
Orbi	0.368	0.9161	0.01157
Interview	0.2612	0.9654	0.009749
Ballet	0.08401	0.9528	0.005829
Breakdance	0.1316	0.9607	0.007294
All sequences	1.266	0.9273	0.01101

The scatter plot in Fig. 84 shows the relationship between the proposed colour image quality evaluation method ($MQ_{Colour}(X,Y)$) and the rendered left and right view quality. Orbi and Interview sequences compressed using QP = 10 and QP = 30 were used to generate this scatter plot. This scatter plot shows the degree of correlation between the proposed colour image quality measures and rendered left and right image quality.

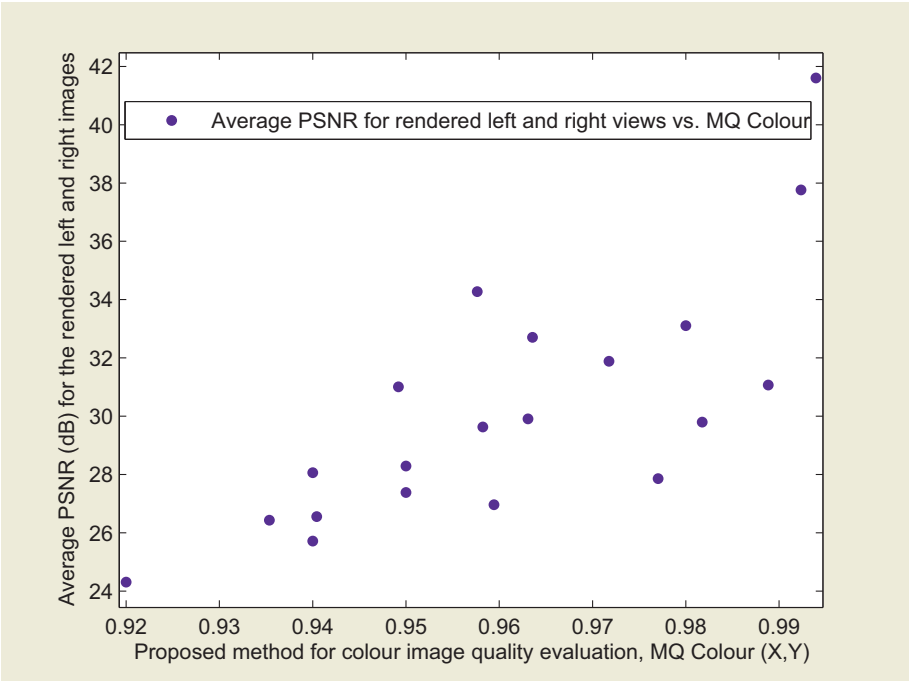


Figure 84: Scatter plot of rendered left and right image quality (average PSNR) versus colour image quality obtained with the proposed $MQ_{Depth}(X,Y)$ method.

In order to better understand the relationship between true 3D perception (i.e., subjective quality ratings) and

results obtained for colour image quality evaluation with the proposed RR method $MQ_{Colour}(X,Y)$, results were mapped into subjective MOS values, which were obtained using subjective quality tests. The mapping between the proposed RR metric for colour images (i.e., $MQ_{Colour}(X,Y)$ ratings) and subjective quality ratings for the overall 3D image quality is shown in Fig. 85. This scatter plot shows the subjective (MOS) and the proposed objective rating for the colour image results under different PLRs for all four sequences (compressed with QP = 30). Since average quality ratings are shown for all the test points only a few test points are visible in the scatter plot. This is mainly due to the fact that the subjective quality ratings were obtained for the whole sequence and not measured continuously (i.e., not marked for each frame). The relationship is approximated by a linear polynomial as shown in Fig. 85. This shows that the proposed method produces matching results in terms of true user perception (i.e., subjective ratings for overall 3D video quality) as well.

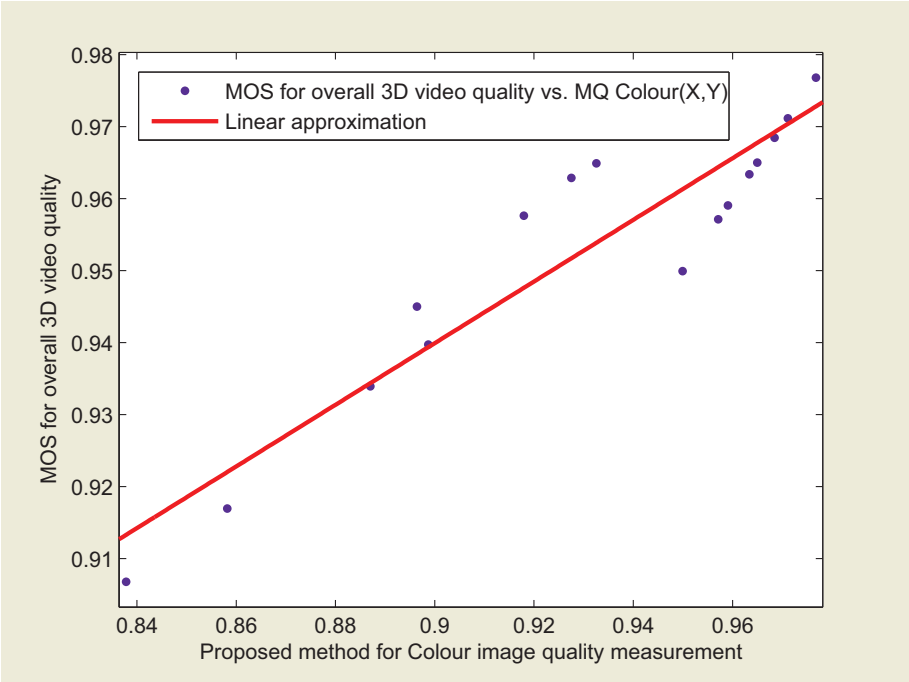


Figure 85: Scatter plot of subjective MOS (i.e., MOS for overall 3D image quality) versus colour image SSIM obtained with the proposed $MQ_{Colour}(X,Y)$ method. Each sample point represents one test image sequence.

8.4 Conclusion

On the fly change of compression and transmission parameters are the key for delivering robust and high quality 2D/3D video services over unreliable communication links. The measured quality at the receiver-side can be fed to the video processing module to change the source coding parameters which match the prevailing user context. In order to measure 3D video quality at the receiver-end, this section proposed a Reduced-Reference quality metric for colour plus depth 3D video transmission based on edge detection of depth maps. Since the edges of the depth map represent different depth levels and the basic structure of the depth map, edge information is used as the side-information for depth map quality evaluation (i.e., to evaluate the structural degradation of the depth map). The generated edge information (i.e., side-information for depth map quality evaluation) only contains binary zeros and ones and therefore requires less bandwidth compared to the FR method. On the other hand, edge information of the depth map sequence can also be utilized to describe the main foreground objects (i.e., shapes of colour image objects) of the corresponding colour image sequence. Therefore, the extracted edge information of the received colour image sequence is compared against the edge information generated for the corresponding original depth map in the proposed method for colour image quality measurement to quantify the structural degradation of colour images. In this work the SSIM metric is adopted to compare the structural differences using edge information (i.e., the obtained binary edge masks). Results show that the proposed RR method is a good approximation for the FR

quality metric for all the considered PLRs and compression levels. This suggests that due to the practical problems associated with the Full-Reference method (i.e. need for bandwidth), Reduced-Reference quality metrics as described in this contribution are an acceptable compromise for the 3D video research and development community. Note that this work has been co-published in [38].

9 Analysis of QoE Measurement Distortions Caused by Preconceptions of Test Subjects

9.1 Introduction

As the true value of a service rather lies in user satisfaction (Quality of Experience - QoE) than pure numerical system properties (Quality of Service - QoS), QoE provision has become essential in modern telecommunications: QoE monitoring plays an essential role in designing, initializing and maintaining services. Therefore service providers inevitably require user feedback to reach a cost effective level of QoE. Monitoring this phenomenon primarily benefits service providers, but on the other hand, it improves reception for the subscriber. However, the results of such measurements are affected by the subjects prior knowledge on the present technology (called here as the Level of Comprehension - LoC), especially if one is aware of the QoS parameters. In some cases, the awareness of such parameters cannot be avoided; thus the results are preordained to be shaped and distorted. The direction and power of these effects are quite far from triviality, yet it hasn't been circumspectly analyzed so far. Medical quality 2D/3D video services demand accurate measurements and the avoidance of such distortions. Our goal in this section is to enhance QoE evaluation schemes by dealing with this untended topic: we study how the combination of aforementioned QoS parameters and different LoC levels of test subjects distort the results of a QoE measurement. The QoE measurements were made on a real-life 3G HSDPA network. The objective of the test subjects was to grade the experienced quality of a real-life video transmission, while possessing the parameters of the mobile connection. We show how subjects' evaluations were affected and investigate the identified phenomenon in terms of Mean Opinion Score (MOS) deviations and the overall QoE result distortion.

The standard techniques for QoE measurements are defined by the recommendation [136] of the International Telecommunication Union. It contains all the important parameters that can be involved in the configuration of a QoE measurement. Subjective determination of transmission quality can be achieved by four different clusters of methods. The most popular ones are considered to be the conversation-opinion tests, since they are designed to replicate actual service usage situations. Listening-opinion tests rather focus on ones perception, which makes them excellent to measure basic usability and acceptance. Interview and survey tests are efficient methods to extract information beyond a numerical judgment. A group labeled "other tests" is also defined. We decided to use conversation-opinion tests in our researches, with minor additions from interview tests methodology [135]. An excellent example for conversation-opinion tests is the measurement [73] achieved by Yue Lu and his colleagues in the Netherlands. They focused their work on the astute choice of the video conference client. For listening-opinion tests, an absolutely noteworthy example is the measurement series [52] done by István Ketykó and his colleagues in 2010. Their topic was user experience and service quality acceptance in different environments.

The field of subjective determination of transmission quality has well defined standards, intensely detailed recommendations, and countless of exceptional papers sharing the experiences of researches and measurements. The ITU-T P series [134] provide a wide range of recommendations relating to the topic. For those who would rather generally approach the subject matter, the publication of Telenor [37] is advised, or the book of David Soldani, Man Li and Renaud Cuny [125], which also gives a view of great extent into the world of QoS and QoE Management in mobile networks. When mentioning measurement examples involving not only audio, but video quality evaluation, it is meritorious to separate those configured on a wired [85, 58, 106, 116] and a wireless [97, 23, 83, 84, 51] networks. It needs to be mentioned that while the preceding instances were based on human evaluation, there are automated evaluation methods [22] as well, excluding the human factor and the possible distortions caused by human test subjects. However, subjective tests performed by humans, if done accurately, are still considered as the best way to measure user experience / QoE. That is why it is crucial to analyze potential corruptions in human based QoE measurements and be aware of likely deviations and their effects.

9.2 Testbed and configuration

In order to study the QoE measurement distortions, we set up a real-life testbed where a video conference between the measurement guide and the test subject was emulating a typical mobile video service. The tests were performed on the test network (see Figure 86) of BME Mobile Innovation Centre [13]. In order to acquire the preliminary results, twenty subjects participated in the series of measurements, all of them with different levels of prior technical knowledge, ranging from simple inexperienced user to engineer with PhD degree. Just before the measurement itself, the Level of Comprehension of each subject was revealed by asking a set of questions related to

the technical background of telecommunication activities. These conversations, each taking approximately thirty minutes, were recorded for further analysis to precisely determine the LoC of the subjects. Ten different levels were distinguished: level one introduces the lowest, while level ten represents the highest level of technical comprehension. To preserve the purity of LoC determination, the subjects were given no information about the nature of the measurement before it had begun. The variety of technical competence was not the only aspect during the selection of the test subjects, but it was also necessary to only select people who have never seen each other before in order to prevent information leak between measurements. The subjects have not even met each other during the series of measurements due to the different dates of the measurements. If any subject had received even the slightest information about the measurement before its date, it could have and probably would have resulted in LoC overestimation.

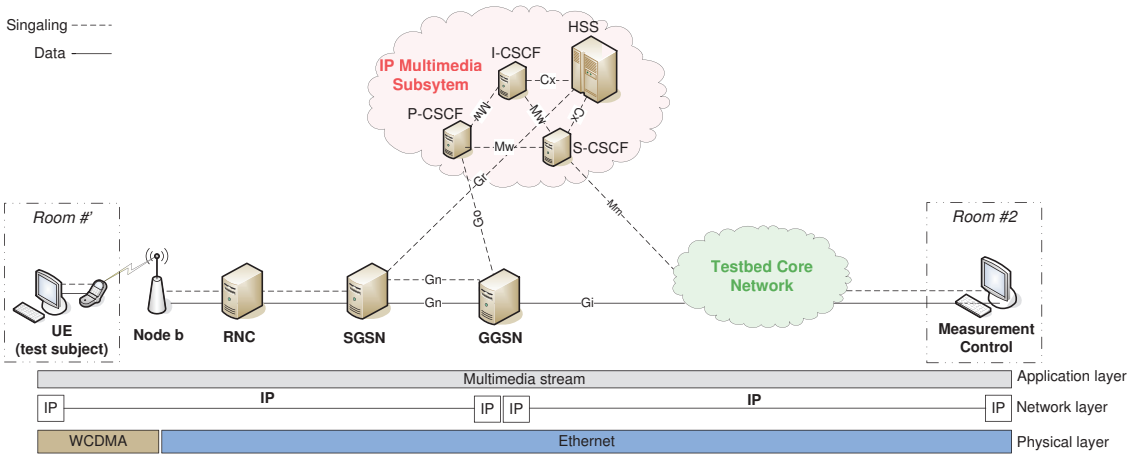


Figure 86: Testbed architecture and setup

The complete process of a measurement was divided into four sections, following each other without delay. The first part is considered to be the LoC level determination conversation, as mentioned before. This was followed by questions on general user behavior, involving the quality of previously experienced video conferences. After the basic instructions, began the third and most important part of the process, namely the mobile video teleconference and its written evaluation. This was concluded by an oral re-evaluation of the experienced quality, which was also recorded like the first two conversations. The measurement guide was the same in each and every part of the process and for all subjects.

Table 25: Fixed Parameters of the QoS Matrix

Delay: 133 ms	Resolution: 640x480
Jitter: 30 ms	Video codec: MPEG-4
Packet loss: 0%	Audio codec: speex

During the video conference, the guide used a terminal in the laboratory (Room #2 in Figure 86), while the subject was isolated in the conference room of the laboratory. The audiovisual connection was established by a `Linphone 3.2.1` client [70] on an `Ubuntu 10.04` operation system. Both end terminals shared the same hardware and software, including multimedia equipment such as web camera and headset. Connection to the test network, however, was different. While the terminal at the laboratory was connected via Ethernet, the computer at the conference room (Room #1 in Figure 86) used a Huawei 3G HSDPA wireless modem. IP Multimedia Subsystem (IMS) [4] was in control of the mobile multimedia traffic over the UMTS network. The video conversation took approximately one hour for every subject. Although it was divided into twenty subsections (test cases), the conversation itself was fluent and natural. Every subsection had a different artificial one-way QoS parameter load

in terms of delay, jitter and packet loss, in addition to the real QoS values of the network. To achieve this, we used the command line based `netem` application [88] in order to change the output traffic of the laboratory terminal without the interruption or pause of the video conversation. The parameter values were given to the subject before commencing the conversation, in a form of a QoS parameter matrix (see Table 26), together with the fixed parameters of the measurement (see Table 25), such as video resolution. The objective of the subject was to separately evaluate the audio and video quality of the twenty different test cases on an MOS scale from one to ten.

Table 26: Variable Parameters of the QoS Matrix

Test case	***Delay***	***Jitter***	***Packet loss***
1	0 ms	0 ms	0%
2	50 ms	10 ms	0.5 %
3	200 ms	40 ms	2%
4	800 ms	180 ms	8%
5	0 ms	180 ms	8%
6	0 ms	0 ms	8%
7	0 ms	180 ms	0%
8	800 ms	0 ms	0%
9	800 ms	100 ms	1.2 %
10	400 ms	100 ms	1.2 %
11	200 ms	100 ms	1.2 %
12	100 ms	100 ms	1.2 %
13	100 ms	180 ms	0.5 %
14	100 ms	100 ms	0.5 %
15	100 ms	40 ms	0.5 %
16	100 ms	20 ms	0.5 %
17	200 ms	20 ms	0.5 %
18	200 ms	20 ms	2%
19	200 ms	20 ms	4%
20	200 ms	20 ms	8%

The initial step during the processing of the measurement results was to match all the subjects to their most relevant LoC levels. This was achieved by analyzing the conversations recorded during the first section of the measurement process. The completion of the LoC based subject list generated the preferred results from the raw data set (see Figure 87) of the measurements which was the ground base for our evaluation.

9.3 Results and discussion

First we took a closer look at the MOS values, independently from the LoC levels. As can be seen on the parameter matrix, while the first test case was free of any additional load, case number 8 suffered 800 ms of further delay. It would be expected for number 1 to have a better MOS score, however, the outcome showed the opposite. Even though the difference is rather minor, it cannot be denied that test case number 8 achieved a higher score in terms of video quality. By relying only on the MOS results, it would be quite exigent to give an accurate explanation to this phenomenon. The source of these MOS values can be found by dividing the results into two sets based on LoC levels: the first eight and the last two levels (see Figure 88). This subtracts those from the results who possess the lowest prior technical knowledge and provides an unequivocal explanation. While the subjects of the first set were commonly controlled by the fact that delay is noxious to experienced quality and thus such measurement case cannot achieve a better score, the rest was not aware of this. In fact, as heard on the recorded summary conversations, some people even think delay is beneficial and produces a higher level of quality. The subjects of the first set were not affected by such misbelieves so not even a single subject gave number 8 a better score, even though there is no major difference between the video quality of the two cases. The opinions of the people in the bottom two levels on the quality of these two cases were enough to create a distortion large enough to significantly affect the overall MOS results. It needs to be noted that the evaluation set of the top eight levels is also distorted, since there were people who barely distinguished the quality of the two cases but made a difference in evaluation

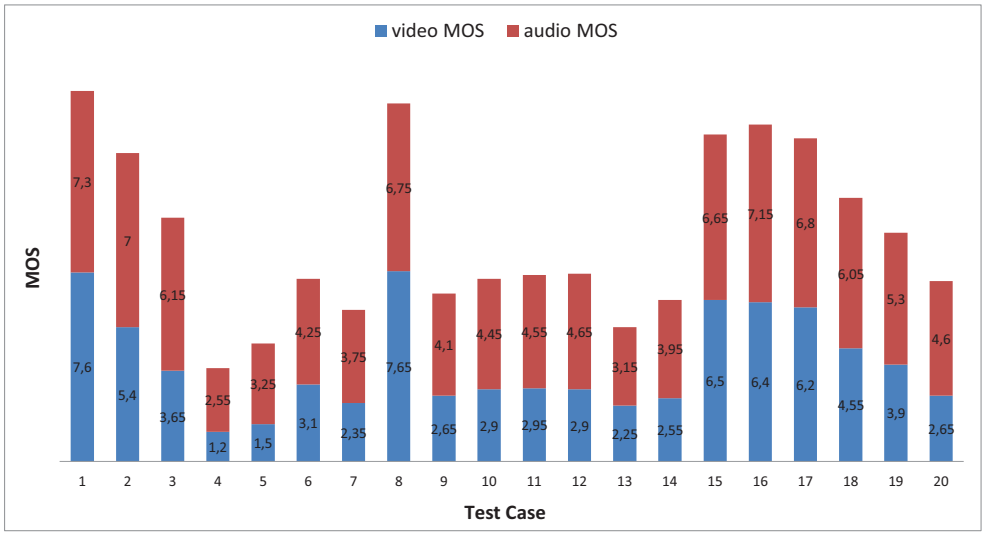


Figure 87: Raw data set of the measurement results

because of their preconceptions. Another noteworthy phenomenon in overall video MOS is the progress from case thirteen to sixteen. Jitter reduction describes these four cases as seen in the QoS parameters. Here we would expect an obvious rise in MOS, but the last case decreases instead. The previous explanation applies to this problem as well. If we take a look at the separated MOS diagram (see Figure 88), it can be seen that the subjects in the lowest two LoC levels are responsible again. In addition, the score of case 1 should be compared to case 15 on these lowest levels. There is indeed an immense difference between the quality case 14 and 15. While in case 14 the video images are barely recognizable, case 15 presents a quite enjoyable view. Because of the lack of evaluation control provided by prior knowledge and preconceptions, case number 15 received a stupendous score compared to number 1, the case where no additional load was applied.

The first four measurement cases represent a general decrement in QoS values, both delay, jitter and packet loss increases. In this progress, the overall MOS results provide nothing unusual, experienced quality decreases as it should. However, after we removed the lowest two LoC levels, we obtained something that was beyond our expectations. The average evaluation of the remaining subjects is mathematically equable, the scores show perfect uniformity, the line connecting these cases is straight. In some ways, this can be considered as one of the greatest distortions we have experienced so far. Before even encountering these cases, most subjects had a prior idea of how they should evaluate the quality, because they were already aware of the upcoming changes. This does not mean that the lowest two levels represent actual experienced quality values, since they are less likely to be controlled by preconceptions. As mentioned before, they are just as controlled. It is just enough to compare the audio scores of the first two cases on these levels. Number 2 suffers loads in all three QoS parameters, but still received a higher score because of technical misbelieves. Audio evaluation holds a few interesting phenomena as well (see Figure 89). The focus is now on the progress from case number 9 to 12. These four cases endure delay reduction while preserving a notable constant jitter. Presuming the experienced quality tendencies in the progress is not a trivial task. It is beneficial to have a smaller delay, however, the ratio of jitter and delay increases. The audio MOS shows a definite raise in these four cases, even though none of the subjects thought it that way. After analyzing the LoC levels from one to seven, we could not find any obvious behavior pattern. In fact, there were not even two levels showing the same relations between the adjacent cases, since there was no major overall difference in

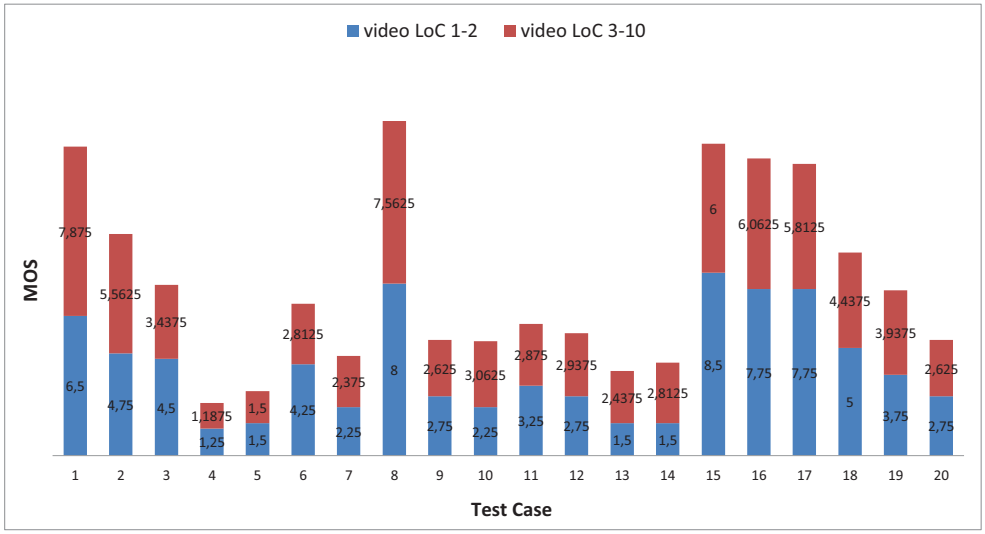


Figure 88: Video MOS values for LoC levels 1-2 and 3-10

experienced audio quality. On one hand, mutual speech interruptions were fewer, but on the other hand, audio quality was less enjoyable to some extent. The scores given by the subjects were based on the personal decision whether the first or the second effect was more dominant. However, what we have discussed so far deals only with the lowest seven levels of LoC. The top three levels produced a marvelous result. The scores on these four cases were constant. It means that preconceptions had such a high level impact on evaluation that these subjects ignored any lesser differences that they experienced between cases. They considered the opposing effects nearly equal, which supposes an unvarying overall experience.

9.4 Conclusions

Audio and video quality evaluation in our measurements achieved more detailed and accurate information considering the distortions of MOS due to our novel approach. We showed, how seriously preconceptions of test subjects can harm the scoring, and we also managed to explain several phenomena with the help of LoC that would have been unaccountable and ignored otherwise. The aim of this section is to highlight the hazard that lies within preconception based distortions in human-based QoE evaluation measurements, and to introduce the level of impact it can attain on the overall evaluation values. The results showed that with our methodology it is possible to further improve QoE evaluation of mobile video services, and such provide better feedback to service providers of the rapidly developing, progressing world of mobile Internet. In our future work we will separate the LoC from the initial configuration. In this case the same subjects would first evaluate the experienced audio and video quality without access to the QoS parameters, and then repeat the measurements with parameter awareness. This could result in a greater insight to distortions and would allow defining more detailed behavior models. We also plan to use this approach for modem or end terminal equipment comparison, involve a full duplex delay load in the measurements, compare 2D/3D transmissions in fixed and mobile environments, and perform QoE based analysis of different mobility management solutions. The number of varying parameters, test cases, or test subjects could also be increased, but these are bounded by the cost effectiveness of the measurements.

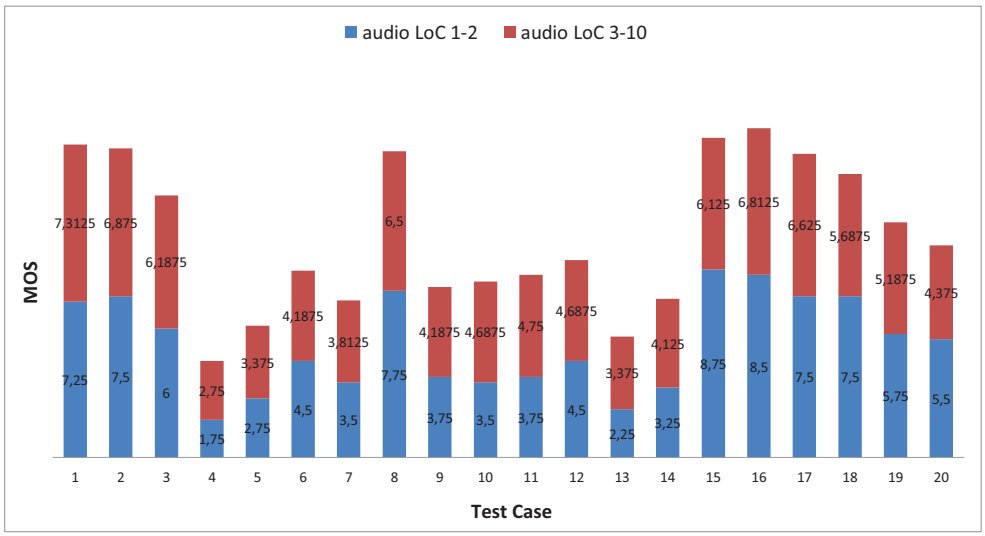


Figure 89: Audio MOS values for LoC levels 1-2 and 3-10

10 Conclusion

This deliverable presented the work conducted in CONCERTO on image/video coding and transmission, including the aspects of adaptation, security, and quality assessment. Several novel methods and algorithms were proposed, many of which show a clear advantage over the state of the art.

The work on secure QoE-aware image/video coding reported in this deliverable will continue throughout the Work Package 3 lifetime until M32. Even after the termination of WP3 work, the research related to standardization activities will continue until the end of the project and will be executed within Work Package 7 and reported in Deliverable D7.6 ("Final report on dissemination and standardization") due in M36.

Furthermore, the work in this work package will be the basis for the implementation of the simulator and the demonstrator, both to be developed within the scope of Work Package 6, where a selection of the components developed in this work package will be applied.

References

[1] The SVT high definition multi format test set. Sveriges Television. `ftp://vqeg.its.bldrdoc.gov/`.

[2] Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, phase II. Video Quality Expert Group, Aug. 2003.

[3] K. J. Lee, H. J Sung, E. S. Park and I. K Lee. Joint optimization for one and two-way MIMO AF multiple-relay systems. *Wireless Communications, IEEE Transactions on*, 9(12):3671 –3681, december 2010.

[4] 3GPP TS 23.228. IP Multimedia Subsystem (IMS); Stage 2. Rel-8, 2008.

[5] A. J. Aljohani, S. X. Ng, R. G. Maunder and L. Hanzo. Joint source-coding, channel-coding and modulation for two-way relaying. *IEEE Transactions on Vehicular Technology*, submitted.

[6] A. Papadogiannis, A. G. Burr and M. Tao. On the maximum achievable sum-rate of interfering two-way relay channels. *IEEE Communications Letter*, 16(1):72–75, January 2012.

[7] A. Sendonaris, E. Erkip and B. Aazhang. User cooperation diversity part I: System description. *IEEE Transactions on Communications*, 51(11):1927–1938, 2003.

[8] S.M. Aghito and S. Forchhammer. Efficient coding of shape and transparency for video objects. *Image Processing, IEEE Transactions on*, 16(9):2234–2244, Sep. 2007.

[9] R. Akhter, Z. M. P. Sazzad, Y. Horita, and D. Meek. No reference stereoscopic image quality assessment. In *Proceedings of SPIE: Stereoscopic Displays and Applications XXI vol: 7524*, volume 7524, 2010.

[10] R. V. Babu, A. S. Bopardikar, A. Perkis, and O. I. Hillestad. No-reference metrics for video streaming applications. In *Proc. International Packet Video Workshop*, 2004.

[11] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau. Quality assessment of stereoscopic images. *EURASIP Journal on Image and Video Processing*, 2008:1–13, 2009.

[12] A. Biryukov. Block ciphers and stream ciphers: The state of the art. *IACR Cryptology ePrint Archive*, 2004:94, 2004.

[13] BME-MIK. Budapest University of Technology and Economics - Mobile Innovation Centre, Official Website. `https://www.mik.bme.hu/home/aboutus/`.

[14] F. Bossen. Common test conditions and software reference configurations. ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document JCTVC-H1100, San Jose, CA, USA, Feb. 2012.

[15] M. Brandas, M. Uitto, M. Martini, and J. Vehkaperä. Quality assessment and error concealment for SVC transmission over unreliable channels. In *Proceedings of the Multimedia & Expo (ICME) - Workshop on Acoustics and Video Coding and Communication (AVCC)*, 2011.

[16] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, and T. Wiegand. High efficiency video coding (HEVC) text specification draft 6. ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document JCTVC-H1003, San Jose, CA, USA, Feb. 2012.

[17] P. Le Callet and F. Autrusseau. Subjective quality assessment IRCCyN/IVC database, 2005. http://www.irccyn.ec-nantes.fr/ivcdb/.

[18] M. Carnec, P. Le Callet, and D. Barba. Objective quality assessment of color images based on a generic perceptual reduced reference. *Signal Proc: Image Communication*, 23(4):239–256, Apr. 2008.

[19] G. H. Chen, C. L. Yang, L. M. Po, and S. L. Xie. Edge-based structural similarity for image quality assessment. In *International conference on - Acoustics Speech and Signal Processing*, pages 933–936, 2006.

[20] L. Crauss Daronco, V. Roesler, J. Valdeni de Lima, and R. Balbinot. Quality analysis of scalable video coding on unstable transmissions. *Springer Multimedia Tools and Applications*, pages 1–27, 2011.

[21] L.C. Daronco, V. Roesler, and J. V. de Lima. Subjective video quality assessment applied to scalable video coding and transmission instability. In *Proceedings of the 2010 ACM Symbosium on Applied Computing*, pages 1898–1904, 2010.

[22] D. De Vera, P. Rodriguez-Bocca, and G. Rubino. Automatic quality of experience measuring on video delivering networks. *SIGMETRICS Perform. Eval. Rev.*, 36(2):79–82, August 2008.

[23] Dialogic. Quality of Experience for Mobile Video Users. TU Berlin / Deutsche Telekom Laboratories, 2010.

[24] A. Eichhorn and P. Ni. Pick your layers wisely - a quality assesment of H.264 scalable video coding for mobile devices. In *Proceedings of the IEEE International Conference on Communications, ICC'09*, 2009.

[25] E. Ekmekcioglu, M. Mrak, S. T. Worrall, and A. M. Kondoz. Edge adaptive upsampling of depth map videos for enhanced free-viewpoint video quality. *Electronics Letters*, 45(7):353–354, 2009.

[26] U. Engelke, M. Kusuma, H. Zepernick, and M. Caldera. Objective quality assessment of color images based on a generic perceptual reduced reference. *Signal Proc: Image Communication*, 24, Aug. 2009.

[27] U. Engelke and H.J. Zepernick. Framework for optimal region of interest-based quality assessment in wireless imaging. *Journal of Electronic Imaging*, 19(1):1 – 13, Jan. 2010.

[28] T. A. Fouad and M. M. Hira. Joint optimization of physical layer parameters and routing in wireless mesh networks. In *The 9th IFIP Annual Mediterranean, Ad Hoc Networking Workshop (Med-Hoc-Net), 2010* , pages 1–8, June 2010.

[29] W. Gao, M. Jiang, Y. He, J. Song, and H. Yu. A Lossless Coding Solution for HEVC. ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document JCTVC-G664, Geneva, Switzerland, Nov. 2011.

[30] W. Gao, M. Jiang, and H. Yu. AHG13: Improved Binarization Scheme for Intra Luma Prediction Residuals in Lossless Coding. ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document JCTVC-I0311, Geneva, Switzerland, May 2012.

[31] S. Gueron. Intel® Advanced Encryption Standard (AES) instructions set (white paper).

[32] I. Gunawan and M. Ghanbari. Reduced-reference video quality assessment using discriminative local harmonic strength with motion consideration. *IEEE Trans. on Circuits and Systems for Video Technology*, 18(1):71–83, Jan. 2008.

[33] H, Ochiai, P. Mitran and V. Tarokh. Design and analysis of collaborative diversity protocols for wireless sensor networks. In *Proceedings of IEEE VTC Fall*, pages 4645 – 4649, Los Angeles, USA, 26-29 September 2004.

[34] L. Hanzo, M. Münster, B.J. Choi, and T. Keller. *OFDM and MC-CDMA for Broadband Multi-user Communications, WLANs and Broadcasting*. John Wiley & Sons, May 2003.

[35] A. Haseeb, M. G. Martini, S. Cicalo, and V. Tralli. Rate and distortion modeling for real-time MGS coding and adaptation. In *Wireless Advanced (WiAd), 2012*, pages 85–89, London, U.K., 2012.

[36] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6):409–436, 1952.

[37] B. Hestnes, P. Brooks, and S. Heiestad. QoE (Quality of Experience) – measuring QoE for improving the usage of telecommunication services. In *Telenor Research & Innovation Research Report, Telenor Research & Innovation-R21/2009*, 2009.

[38] C. Hewage and M.G. Martini. Edge based reduced-reference quality metric for 3d video compression and transmission. *IEEE Journal of Selected Topics in Signal Processing*, 2012.

[39] C. T. E. R. Hewage, S. Worrall, S. Dogan, H. Kodikaraarachchi, and A. M. Kondoz. Stereoscopic TV over IP. In *Visual Media Production, 2007. IETCVMP. 4th European Conference on*, pages 1–7, Nov. 2007.

[40] C. T. E. R. Hewage, S. T. Worrall, S. Dogan, and A. M. Kondoz. Prediction of stereoscopic video quality using objective quality models of 2-D video. *Electronics Letters*, 44(16):963–965, 2008.

[41] C. T. E. R. Hewage, S. T. Worrall, S. Dogan, S. Villette, and A. M. Kondoz. Quality evaluation of color plus depth map-based stereoscopic video. *Selected Topics in Signal Processing, IEEE Journal of*, 3(2):304–318, Apr. 2009.

[42] Y. Horita, Y. Kawayoke, and Z. M. Parvez Sazzad. Image quality evaluation database. In *http:///160.26.142.130/toyama*.

[43] W. A. I. Jsselsteijn, H. de Ridder, and J. Vliegen. Subjective evaluation of stereoscopic images: effects of camera parameters and display duration. *Circuits and Systems for Video Technology, IEEE Transactions on*, 10(2):225–233, Mar. 2000.

[44] International Telecommunication Union - Radiocommunication sector. *Methodology for the subjective assessment of the quality of television picture*, 2002.

[45] International Telecommunication Union - Standardization sector. *Subjective video quality assessment methods for multimedia applications*, 1999.

[46] T. Ishiguro, S. Kiyomoto, and Y. Miyake. Latin dances revisited: New analytic results of Salsa20 and ChaCha. In *ICICS*, pages 255–266, 2011.

[47] J. Kliewer, S. X. Ng and L. Hanzo. Efficient computation of EXIT functions for non-binary iterative decoding. *IEEE Transactions on Communications*, 54(12):2133–2136, December 2006.

[48] Jianguo Jiang, Yang Liu, Zhaopin Su, Guofu Zhang, and Shiyi Xing. An improved selective encryption for H.264 video based on intra prediction mode scrambling. pages 464–472, 2010.

[49] S. Jumisko-Pyykkö and J. Häkkinen. Evaluation of subjective video quality of mobile devices. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 535–538, 2005.

[50] N. Kazakova, M. Margala, and N. G. Durdle. Sobel edge detection processor for a real-time volume rendering system. *Proc. of the 2004 International Symposium on Circuits and Systems (ISCAS '04)*, pages 913–916, May 2004.

[51] I. Ketykó, K. De Moor, T. De Pessemier, A. Juan Verdejo, K. Vanhecke, W. Joseph, L. Martens, and L. De Marez. QoE measurement of mobile YouTube video streaming. In *Proceedings of the 3rd workshop on Mobile video delivery*, MoViD '10, pages 27–32, New York, NY, USA, 2010. ACM.

[52] I. Ketykó, K. De Moor, W. Joseph, L. Martens, and L. De Marez. Performing QoE-measurements in an actual 3G network. In *Broadband Multimedia Systems and Broadcasting (BMSB), 2010 IEEE International Symposium on*, pages 1 –6, march 2010.

[53] J. Kilner, J. Starck, J.Y. Guillemaut, and A. Hilton. Objective quality assessment in free-viewpoint video production. *Signal Processing: Image Communication*, 24(12):3–16, 2009. Special issue on advances in three-dimensional television and video.

[54] D. Kim, D. Min, J. Oh, S. Jeon, and K. Sohn. Depth map quality for three-dimensional video. In *Proceedings of SPIE: Stereoscopic Displays and Applications XX vol: 7237*, 2010.

[55] K. J. Kim, B. Kim, R. Mantiuk, T. Richter, H. Lee, H.-S. Kang, J. Seo, and K. H. Lee. A comparison of three image fidelity metrics of different computational principles for JPEG2000 compressed abdomen CT images. *IEEE Transactions on Medical Imaging*, 29(8):1496–1503, June 2010.

[56] S.-H. Kim and A. Segall. Improved CABAC for lossless compression. ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document JCTVC-I0311, Geneva, Switzerland, May 2012.

[57] S.-H. Kim and A. Segall. Simplified CABAC for lossless compression. ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document JCTVC-H0530, San Jose, CA, USA, Feb. 2012.

[58] A. Kwon, J.-M. Kang, S.-S. Seo, S.-S. Kim, J. Y. Chung, J. Strassner, and J. W.-K. Hong. The design of a quality of experience model for providing high quality multimedia services. In *Proceedings of the 5th IEEE international conference on Modelling autonomic communication environments*, MACE'10, pages 24–36, Berlin, Heidelberg, 2010. Springer-Verlag.

[59] L. Hanzo and P. Cherriman and J. Streit. *Wireless Video Communications: From Second to Third Generation Systems, WLANs and Beyond*. IEEE Press, 2001. (For detailed contents please refer to http://www-mobile.ecs.soton.ac.uk.).

[60] L. Hanzo, J. Akhtman, M. Jiang and L. Wang . *MIMO-OFDM for LTE, WIFI and WIMAX: coherent versus non-coherent and cooperative turbo-transceivers*. IEEE Press – John Wiley, 2011.

[61] L. Hanzo, R. G. Maunder, J. Wang and L-L. Yang . *Near-capacity variable-length coding: regular and EXIT-chart-aided irregular designs*. Wiley-IEEE Press, 2010 .

[62] L. Hanzo, T. H. Liew, B. L. Yeap, R. Y. S. Tee and S. X. Ng. *Turbo coding, turbo equalisation and space-time coding: EXIT-chart aided near-capacity designs for wireless channel*. John Wiley IEEE Press, New York, USA, 2011.

[63] L. Kong, S. X. Ng, R. G. Maunder and L. Hanzo. Maximum-throughput irregular distributed space-time code for near-capacity cooperative communications. *IEEE Transactions on Wireless Communications*, 59(3):1511–1517, March 2010.

[64] C. Lamy-Bergot, R. Fracchia, M. Mazzotti, S. Moretti, E. Piri, T. Sutinen, J. Zuo, J. Vehkaperä, G. Fehér, G. Jeney, G. Panza, and P. Amon. Optimisation of multimedia over wireless IP links via X-layer design: an end-to-end transmission chain simulator. pages 261–288, 2011.

[65] J.-S. Lee, F. De Simone, and T. Ebrahimi. Subjective quality assesment of scalable video coding: A survey. In *Proceedings of the third Workshop on Quality of Multimedia Experience*, pages 25–30, 2011.

[66] J.-S. Lee, F. De Simone, et al. Subjective evaluation of scalable video coding for content distribution. In *Proceedings of the ACM International Conference on Multimedia*, pages 65–72, 2010.

[67] C. LeRouge, M. J. Garfield, and A. R. Hevner. Quality attributes in telemedicine video conferencing. In *Proceedings of the 35th Hawaii International Conference on System Sciences*, 2002.

[68] Q. Li and Z. Wang. Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE Journal of Selected Topics in Signal Processing*, 3(9):202–211, Apr. 2009.

[69] X. Li and M. T. Orchard. Edge-directed prediction for lossless compression of natural images. *IEEE Transactions on Image Processing*, 10(6):813–817, 2001.

[70] Linphone. Official Website. `http://www.linphone.org/`.

[71] F. Liu and H. König. A survey of video encryption algorithms. pages 3–15, 2010.

[72] R. Loose, R. Braunschweig, E. Kotter, P. Mildenberger, R. Simmler, and M. Wucherer. Compression of digital images in radiology - results of a consensus conference. *RöFo : Fortschritte auf dem Gebiete der Röntgenstrahlen und der Nuklearmedizin*, 181(1):32–7, January 2009.

[73] Y. Lu, Y. Zhao, F. Kuipers, and P. Van Mieghem. Measurement study of multi-party video conferencing. In *Proceedings of the 9th IFIP TC 6 international conference on Networking (NETWORKING'10)*, page 96–108, 2010.

[74] A. L.Yarbus. *Eye Movements and Vision*. Plenum Press, New York, USA, 1967.

[75] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[76] D. Marr and E. Hildreth. Theory of edge detection. *In Proc. of the Royal Society of London*, 207(1167):187–217, 1980.

[77] M. G. Martini, M. Mazzotti, C. Lamy-Bergot, J. Huusko, and P. Amon. Content adaptive network aware joint optimization of wireless video transmission. *IEEE Communications Magazine*, 45(1):84–90, Jan. 2007.

[78] M.G. Martini, C.T. Hewage, and B. Villarini. Image quality assessment based on edge preservation. *Signal Processing: Image Communication*, 2012.

[79] M.G. Martini, B. Villarini, and F. Fiorucci. A reduced-reference perceptual image and video quality metric based on edge preservation. *EURASIP Journal on Advances in Signal Processing*, 2012(1):66, 2012.

[80] A. Massoudi, F. Lefèbvre, C. De Vleeschouwer, B. M. Macq, and J.-J. Quisquater. Overview on selective encryption of image and video: Challenges and perspectives. 2008.

[81] I. Matsuda, N. Ozaki, Y. Umezu, and S. Itoh. Lossless coding using variable block-size adaptive prediction optimized for each image. In *13th European Signal Processing Conference*, page 30–33, September 2005.

[82] K. McCann, B. Bross, W.-J. Han, S. Sekiguchi, and G. J. Sullivan. HM6: High efficiency video coding (HEVC) test model 6 encoder description. ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document JCTVC-H1002, San Jose, CA, USA, Feb. 2012.

[83] A. Mehmood, S. Agarwal, C. Sengul, and A. Feldmann. Mobile Video QoE in Future Mobile Communications. White Paper, Dec 2009.

[84] V. Menkovski, G. Exarchakos, A. Liotta, and A. C. Sandnchez. Measuring quality of experience on a commercial mobile TV platform. In *Advances in Multimedia (MMEDIA), 2010 Second International Conferences on*, pages 33 –38, june 2010.

[85] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang. Measuring the quality of experience of HTTP video streaming. In Nazim Agoulmine, Claudio Bartolini, Tom Pfeifer, and Declan O'Sullivan, editors, *Integrated Network Management*, pages 485–492. IEEE, 2011.

[86] Z. Musoromy, F. Bensaali, S. Ramalingam, and G. Pissanidis. Comparison of real-time DSP-based edge detection techniques for license plate detection. In *Sixth International Conference on Information Assurance and Security*, pages 323 – 328, Atlanta, GA, Aug. 2010.

[87] N. Laneman, D. N. C. Tse and G. W. Wornell. Cooperative diversity in wireless networks: efficient protocols and outage behavior. *IEEE Trans. on Information Theory*, 50(12):3062–3080, 2004.

[88] netem. Linux Network Emulation Official Website. http://www.linuxfoundation.org/collaborate/workgroups/networking/netem.

[89] S. X. Ng and L. Hanzo. On the mimo channel capacity of multi-dimensional signal sets. *IEEE Transactions on Vehicular Technology*, 55(2):528–536, March 2006.

[90] P. Ni. *Towards Optimal Quality of Experience via Scalable Video Coding*. PhD thesis, Mälardalen University, Sweden, August 2009.

[91] P. Ni, A. Eichhorn, C. Griwodz, and P. Halvorsen. Fine-grained scalable streaming from coarse-grained videos. In *Proceedings of the 18th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, pages 103–108, 2009.

[92] L. Onural. Television in 3-D: What are the prospects? *Proceedings of the IEEE*, 95(6):1143–1145, June 2007.

[93] J. Ostermann. Efficient encoding of binary shapes using MPEG-4. In *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, volume 1, pages 295–298, Oct. 1998.

[94] P. Robertson and T. Wörz. Bandwidth-efficient turbo trellis-coded modulation using punctured component codes. *IEEE Journal of Selected Areas in Communications*, 16(2):206–218, Feb 1998.

[95] L. Meesters P. Seuntians and W. IJsselsteijn. Perceptual evaluation of jpeg coded stereoscopic images. In *Proceedings of SPIE vol: 5006*, pages 215–226, 2003.

[96] S.-W. Park and S.-U. Shin. Efficient selective encryption scheme for the H.264/scalable video coding (SVC). In *NCM (1)*, pages 371–376, 2008.

[97] Piamrat. Quality of experience measurements for video streaming over wireless networks. In *Information Technology: New Generations, 2009. ITNG '09. Sixth International Conference on*, pages 1184 –1189, april 2009.

[98] M. H. Pinson and S. Wolf. A new standardized method for objectively measuring video quality. *IEEE Transactions on Broadcasting*, 50(3):312–322, Sep. 2004.

[99] M. H. Pinson and S. Wolf. A new standardized method for objectively measuring video quality. *Broadcasting, IEEE Transactions on*, 50(3):312–322, Sep. 2004.

[100] C. M Privitera and L. W. Stark. Algorithms for defining visual regions-of-interest: comparison with eye fixations. *IEEE Trans. Pattern Anal. Mach. Intell*, 22(9):970–982, 2000.

[101] J. G. Proakis. *Digital Communications*. New York: Mc-Graw Hill International Editions, 3rd edition, 1995.

[102] R. Ahlswede, N. Cai, S.-Y. R. Li and R. W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46:1204–1216, July 2000.

[103] R. Bauer, J. Hagenauer. On variable length codes for iterative source/channel decoding. In *IEEE Data Compression Conference*, pages 273–282, UT, USA, 27-29 March 2001.

[104] R. G. Maunder, J. Wang, S. X. Ng and L. Hanzo. On the performance and complexity of irregular variable length codes for near-capacity joint source and channel coding. *IEEE Transactions on Wireless Communications*, 7(4):1338–1347, April 2008.

[105] C. N. Raju, G. Umadevi, K. Srinathan, and C. V. Jawahar. Fast and secure real-time video encryption. In *ICVGIP*, pages 257–264, 2008.

[106] D. Rodrigues, E. Cerqueira, and E. Monteiro. Quality of service and quality of experience in video streaming. In *Proceedings of the International Workshop on Traffic Management and Traffic Engineering for the Future Internet (FITraMEn2008)*, pages 1184 –1189, Dec 2008.

[107] J. Ruminski. Telemedicine methods for low bit rate communication. In *1st International Conference on Information Technology*, page 1–4, 2008.

[108] S. ten Brink. Convergence behaviour of iteratively decoded parallel concatenated codes. *IEEE Transactions on Communications*, 49(10):1727–1737, October 2001.

[109] S. X. Ng, F. Guo, J. Wang, L-L. Yang and L. Hanzo. Joint source-coding, channel-coding and modulation schemes for AWGN and Rayleigh fading channels. *Electronics Letters*, 39(17):1259 – 1261, aug 2003.

[110] S. X. Ng, F. Guo, J. Wang, L-L. Yang and L. Hanzo. Jointly optimised iterative source-coding, channel-coding and modulation for transmission over wireless channels. In *IEEE Vehicular Technology Conference Spring*, pages 313–317, Milan, Italy, 17-19 May 2004.

[111] S. X. Ng, K. Zhu and L. Hanzo. Distributed source-coding, channel-coding and modulation for cooperative communications. In *IEEE Vehicular Technology Conference Fall*, pages 1–5, Ottawa, Canada, September 2010.

[112] S. X. Ng, Y. Li and L. Hanzo. Distributed turbo trellis coded modulation for cooperative communications. In *Proceedings of International Conference on Communications (ICC)*, pages 1–5, Dresden, Germany, 14-18 June 2009.

[113] H. Schwarz, D. Marpe, and G. J. Sullivan. Overview of the scalable video coding extension of the H.264/AVC standard. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1103–1120, 2007.

[114] International Telecommunication Union/ITU Radio Communication Sector. Subjective assessment of stereoscopic television pictures. *ITU-R BT.1438*, 2000.

[115] International Telecommunication Union/ITU Radio Communication Sector. Methodology for the subjective assessment of the quality of television pictures. *ITU-R BT.500-11*, 2002.

[116] R. Serral-Gracià, E. Cerqueira, M. Curado, M. Yannuzzi, E. Monteiro, and X. Masip-Bruin. An overview of quality of experience measurement challenges for video applications in IP networks. In *Proceedings of the 8th international conference on Wired/Wireless Internet Communications*, WWIC'10, pages 252–263, Berlin, Heidelberg, 2010. Springer-Verlag.

[117] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack. LIVE video quality assessment database, 2010. http://live.ece.utexas.edu/research/quality/live_video.html.

[118] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack. Study of subjective and objective quality assessment of video. *IEEE Transactions on Image Processing*, 19(6):1427–1441, June 2010.

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

[120] J. Shah and V. Saxena. Video encryption: A survey. 2011.

[121] Z. Shahid, M. Chaumont, and W. Puech. Fast protection of H.264/AVC by selective encryption of CAVLC and CABAC for I and P frames. pages 565–576, 2011.

[122] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Processing*, 15(11):3440–3451, Nov. 2006.

[123] H. R. Sheikh, Z.Wang, L. Cormack, and A. C. Bovik. Live image quality assessment database. http://live.ece.utexas.edu/research/quality, 2008.

[124] J. Shukla, M. Alwani, and A. K. Tiwari. A survey on lossless image compression methods. In *2nd International Conference on Computer Engineering and Technology*, volume 6, page 136–141, 2010.

[125] D. Soldani, M. Li, and R. Cuny. *QoS and QoE Management in UMTS Cellular Systems*. John Wiley and Sons, August 2006.

[126] W. Song, D. Tjondronegoro, and S. Azad. User-centered video quality assesment for scalable video coding of h.264/AVC standard. In *Proceedings of the 16th international conference on Advances in Multimedia Modeling*, 2009.

[127] T. Stütz and A. Uhl. A survey of H.264 AVC/SVC encryption. *IEEE Trans. Circuits Syst. Video Techn.*, pages 325–339, 2012.

[128] T. Sutinen, J. Vehkaperä, E. Piri, and M. Uitto. Towards ubiquitous video services through scalable video coding and cross-layer optimization. *EURASIP J. Wireless Communications and Networking*, pages 1–15, 2012.

[129] J. Taquet and C. Labit. Hierarchical oriented predictions for resolution scalable lossless and near-lossless compression of CT and MRI biomedical images. *IEEE Transactions on Image Processing*, 21(5):2641–52, May 2012.

[130] A. Tikanmaki, A. Gotchev, A. Smolic, and K. Miller. Quality assessment of 3D video in rate allocation experiments. In *Consumer Electronics, 2008. ISCE 2008. IEEE International Symposium on*, pages 1–4, Apr. 2008.

[131] S. Tourancheau, F. Autrusseau, Z. M. Parvez Sazzad, and Y. Horita. Impact of subjective dataset on the performance of image quality metrics. In *IEEE International Conference on Image Processing (ICIP)*, San Diego, CA, 2008.

[132] M. Uitto and J. Vehkaperä. Video quality assurance for SVC in peer-to-peer streaming. In *Proceedings of the AP2PS 2011, The Third International Conference on Advances in P2P Systems*, 2011.

[133] M. Uitto, J. Vehkaperä, and P. Amon. Impact of slice size for SVC coding efficiency and error resilience. In *Proceedings of the Workshop on Impact of Scalable Video Coding on Multimedia Provisioning. SVCVision. MOBIMEDIA. 6th International Mobile Multimedia Communications Conference*, 2010.

[134] International Telecommunication Union. P series: Terminals and subjective and objective assessment methods. ITU-T Recommendations, P series.

[135] International Telecommunication Union. Methods for evaluation of service from the standpoint of speech transmission quality. ITU Recommendation - Was deleted on 2009-06-30 since the four-point MOS scale that it describes is obsolete and replaced by the five-point scales defined in ITU-T P.800, P.800.1 and P.805, April 1989.

[136] International Telecommunication Union. Methods for subjective determination of transmission quality. ITU Recommendation P.800 (08/96), August 1996.

[137] V. B. Balakirsky. Joint source-channel coding with variable length codes. In *IEEE International Symposium on Information Theory*, page 419, Ulm, Germany, 29 June – 4 July 1997.

[138] A. M. van Dijk, J. B. Martens, and A. B. Watson. Quality assessment of coded images using numerical category scaling. *Proc. SPIE*, 2451:99–101, 1995.

[139] A. C. A. van Tilborg and A. Jajodia, editors. *Encyclopedia of Cryptography and Security, 2nd Ed.* Springer, 2011.

[140] R. Vaze and R. W. Heath. On the capacity and diversity-multiplexing tradeoff of the two-way relay channel. *IEEE Transactions on Information Theory*, 57:4219–4234, July 2011.

[141] A. Vetro, T. Wiegand, and G. J. Sullivan. Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard. *Proceedings of the IEEE*, pages 626–642, 2011.

[142] H. Yu M. Zhou W. Gao, M. Jiang. AHG19: A lossless coding solution for HEVC. ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document JCTVC-H0530, San Jose, CA, USA, Feb. 2012.

[143] W. Liang, S. X. Ng, L. Hanzo. TTCM-aided SDMA-based two-way relaying. In *IEEE VTC 2011-Fall*. IEEE, 2011.

[144] L. Stelmach W. Tam and P. Corriveau. Psychovisual aspects of viewing stereoscopic video sequences. In *Proceedings of SPIE vol: 3295*, pages 226–235, 1998.

[145] D. Wang, F. Speranza, A. Vincent, T. Martin, and P. Blancghfield. Towards optimal rate control: A study of the impact of spatial resolution, frame rate and quantization on subjective video quality and bit rate. In *Proceedings of the SPIE*, pages 198–209, 2003.

[146] G. Wang, M. W. Vannier, M. W. Skinner, M. G. P. Cavalcanti, and G. W. Harding. Spiral CT image deblurring for cochlear implantation. *IEEE Transactions on Medical Imaging*, 17(2):251–62, April 1998.

[147] H. Wang and C.-W. Xu. A new lightweight and scalable encryption algorithm for streaming video over wireless networks. In *ICWN'07*, pages 180–185, 2007.

[148] K. Wang, M. Barkowsky, et al. Subjective evaluation of HDTV stereoscopic videos in IPTV scenarios using absolute category rating. In *Stereoscopic Displays and Applicaitons XXII, SPIE 2011*, 2011.

[149] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, Apr. 2004.

[150] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. The SSIM index for image quality assessment. http://www.ece.uwaterloo.ca/ z70wang/research/ssim/#usage, 2008.

[151] Z. Wang, H. R. Sheikh, and A. C. Bovik. No-reference perceptual quality assessment of jpeg compressed images. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I–477 – I–480, 2002.

[152] Z. Wang and E. P. Simoncelli. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. *Human vision and Electronic Imaging*, pages 149–159, Mar. 2005.

[153] A. Wegener, N. Chandra, Y. Ling, R. Senzig, and R. Herfkens. Effects of fixed-rate CT projection data compression on perceived and measured CT image quality. In David J. Manning and Craig K. Abbey, editors, *Medical Imaging 2010: Image Perception, Observer Performance, and Technology Assessment*, volume 7627. SPIE, 2010.

[154] M. J. Weinberger, G. Seroussi, and G. Sapiro. The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS. *IEEE Transactions on Image Processing*, 9:1309–1324, Aug. 2000.

[155] S. Wenger. Error patterns for internet video experiments. *ITU-T SG16 Document Q15-I-16-R1*, 1999.

[156] E. Wige, P. Amon, A. Hutter, and A. Kaup. In-loop denoising of reference frames for lossless coding of noisy image sequences. In *IEEE International Conference on Image Processing (ICIP)*, pages 461–464, Sept. 2010.

[157] S. Wolf and M. Pinson. In-service performance metrics for mpeg-2 video systems. In *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, Nov. 2009.

[158] S. Wolf and M. H. Pinson. Low bandwidth reduced reference video quality monitoring system. In *First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2005.

[159] J. Woods. *Multidimensional Signal, Image and Video Processing and Coding*. Elsevier, 2006.

[160] X. Wu and N. Memon. Context-based, adaptive, lossless image coding. *IEEE Transactions on Communications*, 45(4):437–444, April 1997.

[161] Y. Takishima, M. Wada and H. Murakami. Reversible variable length codes. *IEEE Transactions on Communications*, COM-43(2/3/4):158–162, 1995.

[162] S. L. P. Yasakethu, C. Hewage, W. Fernando, and A. Kondoz. Quality analysis for 3D video using 2D video quality models. *Consumer Electronics, IEEE Transactions on*, 54(4):1969–1976, Nov. 2008.

[163] R. Yeung, C. Lin, A. Lam, and K.K. Choy, Accessed: 06/09/2012. Real-time Transmission of High Definition (HD) 3D Video and HD Audio in Gigabit-LAN.

[164] E. Simoncelli Z. Wang. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *Proceedings of SPIE: Human Vision and Electronic Imaging X, vol: 5666*, volume 5666, 2005.

[165] M. Zhang and X. Mou. A psychovisual image quality metric based on multi-scale structure similarity. In *Proc. IEEE International Conference on Image Processing (ICIP)*, pages 381–384, San Diego, CA, Oct. 2008.

[166] Yong Zhang and Donald a Adjeroh. Prediction by partial approximate matching for lossless image compression. *IEEE Transactions on Image Processing*, 17(6):924–35, June 2008.

[167] M. Zhou. AHG22: Sample-based angular prediction (SAP) for HEVC lossless coding. ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document JCTVC-G093, Geneva, Switzerland, Nov. 2011.

[168] M. Zhou. AHG13: Sample-based angular prediction for HEVC lossless coding. ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document JCTVC-I0311, Geneva, Switzerland, May 2012.

[169] M. Zhou. AHG19: Method of frame-based lossless coding mode for HEVC. ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document JCTVC-H0083, San Jose, CA, USA, Feb. 2012.

[170] W. Zhou, Z. Xie, C. Hua, C. Sun, and J. Zhang. Research on edge detection for image based on wavelet transform. In *Proceedings of the 2009 Second International Conference on Intelligent Computation Technology and Automation*, pages 686–689, Washington, DC, USA, 2009.

[171] T. Zhu, L. Karam, and T.T. Lam. *Subjective assessment of compressed 3D video*. Intel Corporation, 2010.