PAPER

Objective No-Reference Video Quality Assessment Method Based on Spatio-Temporal Pixel Analysis

Wyllian B. da SILVA[†a)], Keiko V. O. FONSECA[††b)], *Members*, *and* Alexandre de A. P. POHL[††c)], *Nonmember*

SUMMARY Digital video signals are subject to several distortions due to compression processes, transmission over noisy channels or video processing. Therefore, the video quality evaluation has become a necessity for broadcasters and content providers interested in offering a high video quality to the customers. Thus, an objective no-reference video quality assessment metric is proposed based on the sigmoid model using spatial-temporal features weighted by parameters obtained through the solution of a nonlinear least squares problem using the Levenberg-Marquardt algorithm. Experimental results show that when it is applied to MPEG-2 streams our method presents better linearity than full-reference metrics, and its performance is close to that achieved with full-reference metrics for H.264 streams.

key words: *H.264, MPEG-2, no-reference metric, objective video quality assessment*

1. Introduction

The demand for high-quality digital videos has put enormous pressure on Internet Service Providers (ISPs) and broadcast operators in recent years, which has led to an increasing offer of broadband access to customers. On offering video services, providers and broadcasters are also required to assess constantly the quality of videos being delivered, once these are affected by processing and transmission degradations over the network. In such an effort, several methods for video quality assessment have been developed, which can be broadly classified into Full-Reference (FR), Reduced-Reference (RR) and No-Reference (NR) metrics. At the same time, the need for local assessment, particularly ones far away from the network stations, has driven research to focus lately on NR methods, since FR and RR metrics require previous information about the video source, which put additional constraints to the evaluation process [1]. Although the NR metric needs only the received or processed video, the development of a technique that presents a high degree of correlation for all types of video with subjective measures is far from satisfactory.

Most of works available in the literature on the NR metrics are based on distortions or artifacts, such as blocking and blurring. For instance, in [2] the spatial distortion of each frame in a video is calculated using the differences between the corresponding regions of two adjacent frames in the video sequence. The predicted distortion is then weighted according to the temporal activities of the frame sequence. In [3] the NR quality score is obtained from the bitstream without the need for the complete video decoding. Three factors are taken into account: picture distortion caused by quantization, quality degradation due to packet loss and error propagation, and temporal effects based on the perception by the Human Visual System (HVS). In another work [4], the quality score is obtained from the coding error estimation computed in the transform domain (Discrete Cosine Transform – DCT), whose coefficients are corrupted by the quantization noise, followed by the perceptual weighting of the error. Moreover, the availability of the corrupted bitstream is assumed for the analysis, but in fact NR metrics do not take into account the additional encryption or processing by third-party decoders, where only the decoded pixel value is available. In a different approach, [5] describes an alternative method based on the pattern estimation of lost macroblocks, which assumes the knowledge of the pixels only. This information is then used in an NR quality monitoring system that delivers an estimate of the Mean Square Error (MSE) distortion caused by channel errors. In an earlier work [6], the NR video-quality estimation based on compressed videos through inter-frame prediction technique, using the activity value pixel information of decoded videos which indicates a variance of luminance for given-size pixel block. In addition, the authors use a blur level and a blockiness level estimation by analyzing pixel information. These works use the luminance component for video quality estimation. However, in other work [7], the authors propose a chroma component approach for NR image quality estimation that measures the blocking artifact level based on analysis of color discontinuities in YUV color space. In addition, this approach explores the color-shifting and color-disappearing areas through gradient differences across the block boundaries in U and Y components to obtain the blocking artifact score.

In this work we present a simple analytical NR method, which takes into account spatial and temporal descriptors, such as blurring and blocking artifacts (A, B and Z activity measures) [8], plus the Temporal perceptual Information (TI) [9], the average Mean Absolute Difference ($\overline{MAD}$) and

Manuscript received May 15, 2014.
Manuscript revised January 15, 2015.
Manuscript publicized April 3, 2015.

[†]The author is with Center of Mobility Engineering (CEM), Federal University of Santa Catarina (UFSC), Joinville, Santa Catarina, Brazil.

[††]The authors are with Graduate Program in Electrical and Computer Engineering (CPGEI), Federal University of Technology – Paraná (UTFPR), Curitiba, Paraná, Brazil.

a) E-mail: wyllian.bs@ufsc.br
b) E-mail: keiko@utfpr.edu.br
c) E-mail: pohl@utfpr.edu.br

DOI: 10.1587/transinf.2014EDP7155

average weighted Mean Absolute Difference ($MADw$) between successive frames. Such descriptors are weighted by values obtained through the solution of the nonlinear least squares method using the Levenberg-Marquardt (LM) algorithm [10]–[12], which takes as input the different video sequences from a database.

This way, the proposed No-Reference Video Quality Assessment (NRVQA) combines detectors of blocking and blurring distortions (spatial domain) and incorporates differences between successive frames (temporal domain). The weights of descriptors (parameters of our model) are obtained based on empirical studies involving the mathematical manipulation through statistical analysis during training phase, whose values depend on the type of video encoding or video artifacts, *i.e.*, our method is specialized because it considers the specific artifacts or video encoding type. Once the weights are available, they are used in the proposed sigmoid mathematical model to assess the video quality of encoded streams, such as those encoded by H.264 and MPEG-2 systems.

The paper is divided as follows. Section 2 describes the proposed NRVQA method and Sects. 3, 4 and 5 presents the features of the video database used, the detail about quality calibration and the statistical method for measure of the video quality prediction, respectively. Experimental results and their discussion are presented in Sect. 6, followed by the conclusion in Sect. 7.

2. No-Reference Assessment Method for Video Quality

The proposed method explores features in the spatial-temporal domain and is based on the detection of artifacts and differences between successive frames of a video sequence. The method is devised and optimized to assess the quality of H.264 and MPEG-2 streams, which are still the most used encoding systems for video delivery.

The method combines the detectors of blocking and blurring artifacts (spatial features) and temporal features, such as the TI and the MAD. The descriptors of blurring and blocking artifacts are represented by the features A, B, and the Z activity measure [8] adapted for video assessment. Thus, a video signal is composed of luminance frames $y(f, i, j)$ with $i \in [1, M]$ and $j \in [1, N]$, where M is the number of rows and N the number of columns in the frame f, whose difference frames, along each horizontal and vertical line is determined by

$$d_h(f, i, j) = y(f, i, j+1) - y(f, i, j), j \in [1, N-1], \quad (1)$$

$$d_v(f, i, j) = y(f, i+1, j) - y(f, i, j), i \in [1, M-1]. \quad (2)$$

The blocking effect can be estimated by the average differences between the edges of DCT blocks in the horizontal and vertical directions, as given in Eqs. (3)–(4), with $\tau{\times}\tau$ being the DCT block size, where $\tau = 8$ and total number of frames F.

$$B_h = \frac{1}{FM\left(\lfloor \frac{N}{\tau} \rfloor - 1\right)} \sum_{f=1}^{F} \sum_{i=1}^{M} \sum_{j=1}^{\lfloor \frac{N}{\tau} \rfloor - 1} |d_h(f, i, \tau j)|, \quad (3)$$

$$B_v = \frac{1}{FN\left(\lfloor \frac{M}{\tau} \rfloor - 1\right)} \sum_{f=1}^{F} \sum_{i=1}^{\lfloor \frac{M}{\tau} \rfloor - 1} \sum_{j=1}^{N} |d_v(f, \tau i, j)|. \quad (4)$$

The combination between B_h and B_v produces the descriptor for the blocking artifacts B:

$$B = \frac{B_h + B_v}{2}. \quad (5)$$

The measure of blurring can be obtained by calculating the reduction of activity, combined with the detection of blocking in the vertical and horizontal directions. Eqs. (6)–(7) shows average absolute difference in horizontal and vertical directions, respectively.

$$A_h = \frac{\tau}{FM(\tau-1)(N-1)} \sum_{f=1}^{F} \sum_{i=1}^{M} \sum_{j=1}^{N-1} |d_h(f, i, j)| - B_h, \quad (6)$$

$$A_v = \frac{\tau}{FN(\tau-1)(M-1)} \sum_{f=1}^{F} \sum_{i=1}^{M-1} \sum_{j=1}^{N} |d_v(f, i, j)| - B_v. \quad (7)$$

The combination between A_h and A_v produces the descriptor A for the blurring artifacts:

$$A = \frac{A_h + A_v}{2}. \quad (8)$$

The second factor that contributes to the detection of blurring effects is the rate zero-crossing (ZC) in the horizontal and vertical directions.

$$Z_h = \frac{1}{FM(N-2)} \sum_{f=1}^{F} \sum_{i=1}^{M} \sum_{j=1}^{N-2} z_h(f, i, j), \quad (9)$$

$$Z_v = \frac{1}{FN(M-2)} \sum_{f=1}^{F} \sum_{i=1}^{M-2} \sum_{j=1}^{N} z_v(f, i, j), \quad (10)$$

where z_h and z_v are expressed as

$$z_h(f, i, j) = \begin{cases} 1, & \text{horizontal } ZC \text{ at } d_h(f, i, j) \\ 0, & \text{otherwise} \end{cases}, \quad (11)$$

$$z_v(f, i, j) = \begin{cases} 1, & \text{vertical } ZC \text{ at } d_v(f, i, j) \\ 0, & \text{otherwise} \end{cases}. \quad (12)$$

The combination between Z_h and Z_v produces the descriptor of blurring artifacts Z, given by

$$Z = \frac{Z_h + Z_v}{2}. \quad (13)$$

We used a slightly modified version of the TI measure from ITU-T [9] that apply the motion difference feature between the luminance pixel values, at the same space location in subsequent frames, which is expressed as

$$TI = \frac{1}{F-1} \sum_{f=2}^{F} \sigma\left[m\left(f, i, j\right)\right], \quad (14)$$

where $\sigma\left[m\left(f,i,j\right)\right]$ is the standard deviation of the luminance difference between the present frame, $y(f,i,j)$, and the previous one, $y(f-1,i,j)$. The MAD feature represents the temporal difference between successive frames. The $\overline{MAD}$ corresponds to the average of MAD for all frames in the video sequence with $f > 1$, as follows

$$\overline{MAD} = \frac{1}{MN(F-1)}\sum_{f=2}^{F}\sum_{i=1}^{M}\sum_{j=1}^{N}\left|y(f,i,j)-y(f-1,i,j)\right|. \quad (15)$$

The $MADw$ describes the motion of the actual frame (f) relative to the previous frame $(f-1)$ and is denoted by

$$MADw = \frac{1}{F-1}\sum_{f=2}^{F}\frac{MAD_f}{MAD_{f-1}}. \quad (16)$$

Finally, in order to employ all descriptors we propose a nonlinear sigmoid mathematical model to establish the relationship among the A, B, Z, TI, $\overline{MAD}$, $MADw$ features and subjective scores DMOS (Difference Mean Opinion Scores). This model is based on empirical studies of the authors involving the mathematical manipulation of such descriptors through statistical analysis (box-plot, as describes Sect. 6). Thus, the analytical expression of our method is based on a sigmoid mathematical model is then given as

$$\mathrm{NRVQA} = \frac{1}{1+e^{\left(\beta_1 B+\beta_2 Z+\beta_3 A+\beta_4 TI+\beta_5\overline{MAD}+\beta_6 MADw+\beta_7\right)}}, \quad (17)$$

where β_1 to β_7 are optimized with the LM method [10]–[12], used to solve the nonlinear least squares problem.

Figure 1 shows a diagram of the proposed technique. Initially, during the training phase, the particular video database (H.264 or MPEG-2 subset) is selected with its subjective DMOS normalized between 0 and 1. Other inputs are also selected, such as the analytical expression, where the spatial-temporal descriptors are outlined and their corresponding initial weights (initial guess) are chosen, described as β parameters. The LM algorithm is then applied to solve the nonlinear least squares problem posed by the analytical expression in order to optimize the values attributed to the different β's. Once this step is completed, the testing phase can be initiated, where the optimized β's are loaded and the corresponding quality score of the distorted video sequence is computed. The LM algorithm has been applied in several fields, such as applied mathematics [2], [3] and neural networks [13].

The LM method combines the advantage of speeding up the process of finding the minimum of a nonlinear function with its operating stability based on the steepest gradient descent method [14], and assumes its accelerated convergence in the minimum vicinity from the Gauss-Newton method [15]. In addition, several works on image and video quality assessment [4], [16]–[18] employ the LM method for solution of the nonlinear least squares problem. The LM algorithm, detailed in Eqs. (18) through (27), aims to minimize the error vector $(\mathbf{d} = \mathbf{s} - \widehat{\mathbf{s}})$ between all actual output $\mathbf{s}$ and the desired output $\widehat{\mathbf{s}}$ given by

$$\underset{(\beta_1,\beta_2,\ldots,\beta_7)}{\arg\min}\ \mathbf{d}^{\mathrm{T}}\mathbf{d}, \quad (18)$$

where $\mathbf{d} = \left[\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_\phi\right]$ and $d_\phi = \left[\mathrm{Subj}_\phi - f\left(A_\phi, B_\phi, Z_\phi, TI_\phi, \overline{MAD}_\phi, MADw_\phi; \beta_1, \beta_2, \ldots, \beta_7\right)\right]$, with Subj_ϕ denoting the DMOS values for all samples of a particular video set (ϕ).

The function that describes the product $\mathbf{d}^{\mathrm{T}}\mathbf{d}$ is given by

$$\begin{aligned}\mathbf{d}^{\mathrm{T}}\mathbf{d} &= f\left(A_\phi, B_\phi, Z_\phi, TI_\phi, \overline{MAD}_\phi, MADw_\phi;\right.\\ &\qquad \left.\beta_1, \beta_2, \ldots, \beta_7\right), \end{aligned} \quad (19)$$

and $\mathbf{s}$ is the vector that contains all the objective scores from the selected video database, expressed as

$$\mathbf{s} = \left(\mathrm{NRVQA}^{\mathrm{T}}_{1:\phi}\right)^{\mathrm{T}}. \quad (20)$$

The Gauss-Newton method is used for minimizing performance functions with the recurrence formula

$$\mathbf{s}_{k+1} = \mathbf{s}_k - \mathbf{H}^{-1}\nabla f, \quad (21)$$

where k is the iteration number, the Hessian ($\mathbf{H}$) and is the square matrix of second-order partial derivatives of the function, which can be approximated by

$$\mathbf{H} = \mathbf{J}^{\mathrm{T}}\mathbf{J}, \quad (22)$$

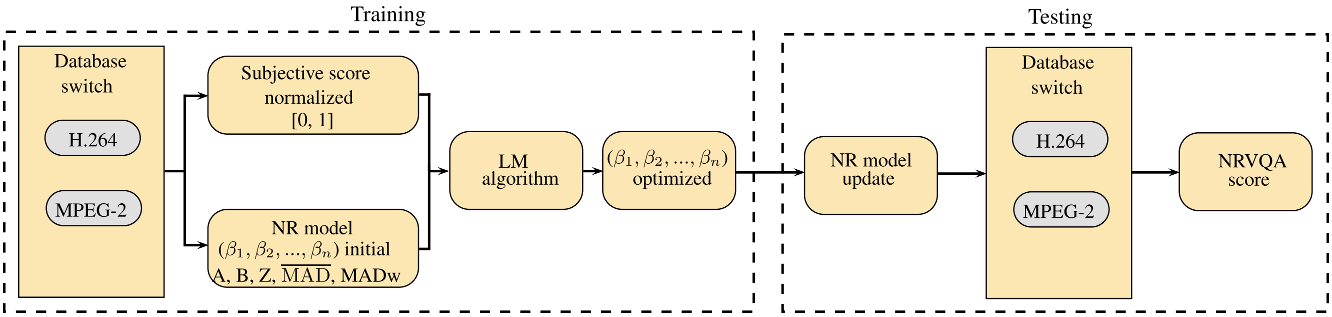
Training
Database switch
H.264
MPEG-2
Subjective score normalized [0, 1]
NR model $(\beta_1, \beta_2, ..., \beta_n)$ initial A, B, Z, $\overline{\text{MAD}}$, MADw
LM algorithm
$(\beta_1, \beta_2, ..., \beta_n)$ optimized
Testing
NR model update
Database switch
H.264
MPEG-2
NRVQA score

Fig. 1 NRVQA framework method with database source switched during training phase and video quality score optimized through of the LM algorithm.

and ∇f is the gradient of the function, expressed as

$$\nabla f = \mathbf{J}^{\mathrm{T}}\mathbf{d}, \tag{23}$$

where $\mathbf{J}$ is the Jacobian matrix that contains the first derivatives of the errors ($\mathbf{d}$) as

$$\mathbf{J} = \begin{pmatrix} \frac{\partial d_1}{\partial \beta_1} & \frac{\partial d_1}{\partial \beta_2} & \cdots & \frac{\partial d_1}{\partial \beta_7} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial d_\phi}{\partial \beta_1} & \frac{\partial d_\phi}{\partial \beta_2} & \cdots & \frac{\partial d_\phi}{\partial \beta_7} \end{pmatrix}. \tag{24}$$

The problem of inverting the Hessian is overcome through the introduction of a modified $\mathbf{H}$, as follows

$$\mathbf{H} = \mathbf{J}^{\mathrm{T}}\mathbf{J} + \lambda_k \mathbf{I}, \tag{25}$$

where $\mathbf{I}$ is the identity matrix, and λ, called combination coefficient, is always positive. This makes $\mathbf{H}$ positive definite and therefore invertible. If λ trend to zero, this gives the Gauss-Newton method with the approximate Hessian. However, if λ is large, this becomes gradient descent using a small step size [19]. Initially λ is set equal to 10^{-4}, but at each iteration the λ value is changed.

Thus, considering the approximate Hessian solution, Eq. (21) can be written as

$$\mathbf{s}_{k+1} = \mathbf{s}_k - \left(\mathbf{J}^{\mathrm{T}}\mathbf{J} + \lambda_k \mathbf{I}\right)^{-1} \nabla f. \tag{26}$$

Finally, the LM algorithm through the update rule of Eq. (26) is stated as

$$\mathbf{s}_{k+1} = \mathbf{s}_k - \left[\mathbf{J}^{\mathrm{T}}\mathbf{J} + diag\left(\mathbf{J}^{\mathrm{T}}\mathbf{J}\right)\lambda_k\right]^{-1} \nabla f, \tag{27}$$

where the identity matrix $\mathbf{I}$ is replaced by the diagonal matrix of the elements $\mathbf{J}^{\mathrm{T}}\mathbf{J}$ [10]–[12].

3. Video Database

The LIVE video quality database is used in this work. This database includes 150 videos from 10 reference video contents, as shown in Fig. 2.

This database includes distorted videos by MPEG-2 compression with rates varied from 700 Kbps to 4 Mbps and H.264 compression with rates varied from 200 Kbps to 5 Mbps, error-prone wireless networks, and IP networks [20]. The first seven video sequences (from left to right and from top to bottom) have a frame rate of 25 frames per second (fps), while the remaining three (Mobile and Calendar, Park Run, and Shields) have a frame rate of 50 fps. All video files do not contain headers and have 8-bit planar YUV 4:2:0 chroma format, whose resolution is 768 × 432 pixels. The LIVE video database only contains DMOS subjective samples. This video database was chosen because it has several video distortions, such as the ones originated at the MPEG-2 and H.264 encoding process. In addition, the LIVE database is currently the most used for the objective Video Quality Assessment (VQA)

The diagram of Fig. 3 shows the relationship between Temporal perceptual Information (TI) and Spatial perceptual Information (SI) for MPEG-2 and H.264 processed video sequences of the LIVE database, both with 40 video samples each.

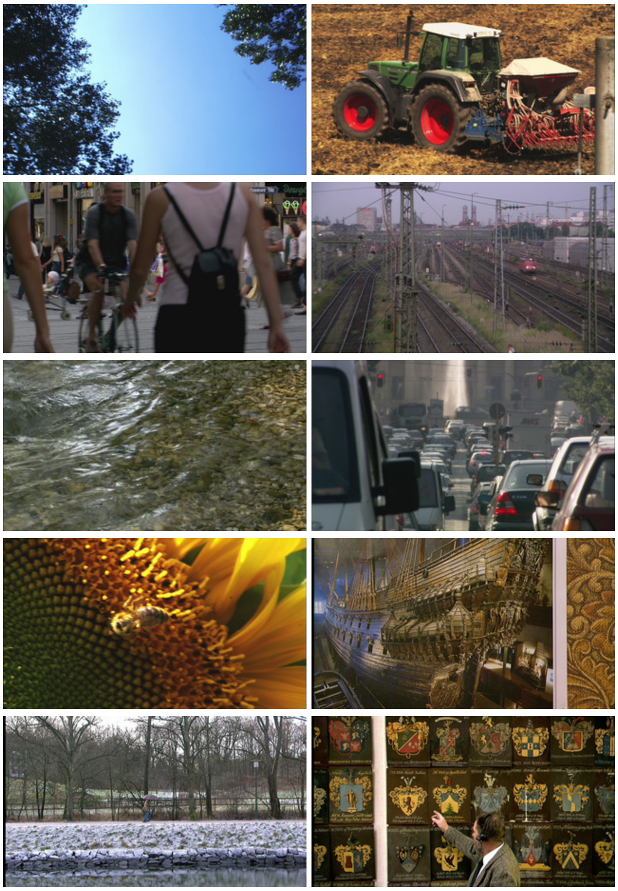

Fig. 2 Pictures of the video samples from the LIVE database [20].

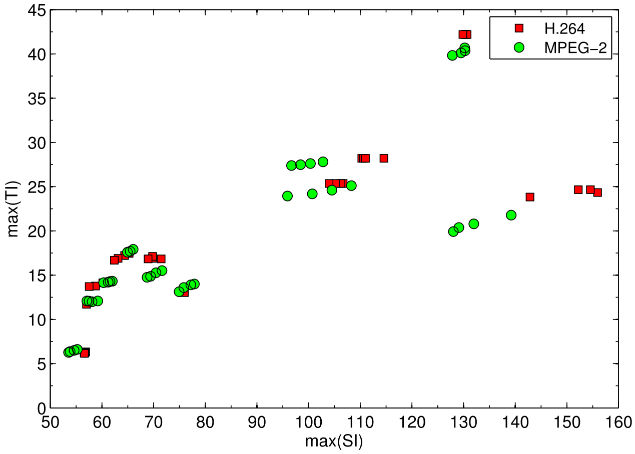

Fig. 3 Temporal perceptual information vs. spatial perceptual information complexity of the LIVE database for H.264 and MPEG-2, both with 40 samples video sequences.

These measures are defined by Recommendation ITU-T P.910 [9], whose comparison between TI and SI measures shows that most H.264 video samples have greater spatial and temporal activity than MPEG-2 video samples. Although this database also contains other distortions our focus is on H.264 and MPEG-2 encoding distortions that are widely used in video transmission over wireless and IP networks.

4. Quality Calibration

The mapping of the objective score scale into the subjective score scale of DMOS can be performed using either a nonlinear logistic function [21], [22] or nonlinear polynomial functions, according to the Video Quality Experts Group (VQEG) recommendation [23]. This mapping must provide a simple empirical prediction and shall not cause an overfitting of data points [24]. In this work, the mapping between DMOS and NRVQA (expressed as x) was performed using a cubic polynomial function [23], [25], [26], defined as

$$\text{DMOSp} = ax^3 + bx^2 + cx + d, \tag{28}$$

where DMOSp is the predicted DMOS, *i.e.*, NRVQA expressed on the DMOS scale.

The cubic polynomial function is better suited as it does not cause overfitting of data points at the low extreme, as it happens with the monotonic logistic function [24]. This way, the Pearson Linear Correlation Coefficient (PLCC) were computed after performing the nonlinear regression using the cubic polynomial function, according to VQEG recommendations.

5. Statistical Method for Linearity Assessment of Video Quality Prediction

The perceptual significance of the metric is determined by the PLCC index (linearity), which is one of the most used for this purpose. If the correlation coefficient approaches 1, the relationship between the scores of the objective metric and the perceptual quality perceived by the HVS is strongly developed. PLCC is calculated using a set of ξ data pairs (μ_k, ν_k) that can be quantified as [21]–[23], [25]–[27]:

$$\text{PLCC} = \frac{\sum_{k=1}^{\xi}(\mu_k - \overline{\mu})(\nu_k - \overline{\nu})}{\sqrt{\sum_{k=1}^{\xi}(\mu_k - \overline{\mu})^2}\sqrt{\sum_{k=1}^{\xi}(\nu_k - \overline{\nu})^2}}, \tag{29}$$

where μ_k and ν_k are the objective and the subjective scores related to the k^{th} frame, respectively; $\overline{\mu}$ and $\overline{\nu}$ are the means of the respective data sets.

6. Results and Discussion

The PLCC correlation coefficient [27] is used as the statistical method to measure the performance (linearity) between our objective metric and the subjective scores (DMOS) of the LIVE database [20]. The experimental procedure for cross-validation occurs in two steps: a) the calculation of the coefficients β for each video subset followed by b) the calculation of the PLCC coefficients for the performance check. First, the H.264 and MPEG-2 video from LIVE database were employed, named as video set categories. Then, in the training phase, each one of these two categories was further divided in three subsets, named Group 1 ($G1_k$), Group 2 ($G2_k$) and S, where this last one represents the union between $G1_k$ and $G2_k$, *i.e.*, $G1_k \cup G2_k$. The groups $G1_k$ and $G2_k$ have the same number of video samples, but both are different in contents through randomized training-test divisions, using 50% with 20 samples for each subset $G1_k$ and $G2_k$ for training, while other 50% also with 20 samples for testing and S with 40 samples for both H.264 and MPEG-2.

We adopt a robust approach for performance analysis of video quality methods through K-fold cross-validation method [28], [29] that randomizes the statistical video groups repeatedly and splitting the available spatial-temporal features in a training-test pairs sets. In the cross-validation process, we use a subset pair as training and the other as testing, *e.g.*, $G1_k - G2_k$, where the first as training $G1_k$ and the second as the testing pattern $G2_k$ for $k = 1, 2, \ldots, K$ different empirical sequences, whose performance is analyzed by PLCC distribution using the box-plot statistical distribution. Literature on image and video quality evaluation does not adopt the box-plot for performance analysis and it uses a small K-value for cross-validation process, for instance, in [8], [30], [31]–[33], [34], and [35]–[37] the K-value is equal to 1, 2, 5, 6, and 10, respectively, while we use a large random permutation of training-test pairs sets with $K = 1{,}000$, *i.e.*, one thousand distinct training-test set partitions evaluated in the cross-validation process using the box-plot statistical tool to measure the PLCC distribution.

The performance results based on the calculation of the PLCC coefficients of the proposed method are compared with the results of two other metrics: Peak Signal-to-Noise Ratio (PSNR) and Multi-Scale Structural SIMilarity index (MS-SSIM) [38], which are FR metrics. Data shown in bold type in Table 1 give results for the median of PLCC and point out to the score for H.264 and MPEG-2 using disjoint training-test sets that characterizes real-world environmental problem situations involving video quality applications. For these cases, the results show that the proposed method achieves better linearity, whose PLCC is greater than 0.88 for MPEG-2 in comparison with PSNR and MS-SSIM metrics.

However, when our method is applied to H.264, we obtain a performance close to the PSNR metric (between 0.8% and 2% greater) and somewhat lower than MS-SSIM met-

Table 1 Comparison of the cross-validation with linearity (PLCC) for second quartile or median (50% video quality score distribution) between full-reference and proposed method for LIVE database.

Metric	Training	Testing using H.264			Testing using MPEG-2		
		G1	G2	S	G1	G2	S
PSNR		0.6252	0.6163	0.5886	0.4765	0.4713	0.4252
MS-SSIM		0.7647	0.7671	0.7305	0.7357	0.7300	0.6851
NRVQA	G1	0.8497	**0.6286**	0.7080	0.9506	**0.8862**	0.9079
	G2	**0.6300**	0.8528	0.7110	**0.8875**	0.9514	0.9095
	S	0.7637	0.7648	0.7451	0.9367	0.9380	0.9317

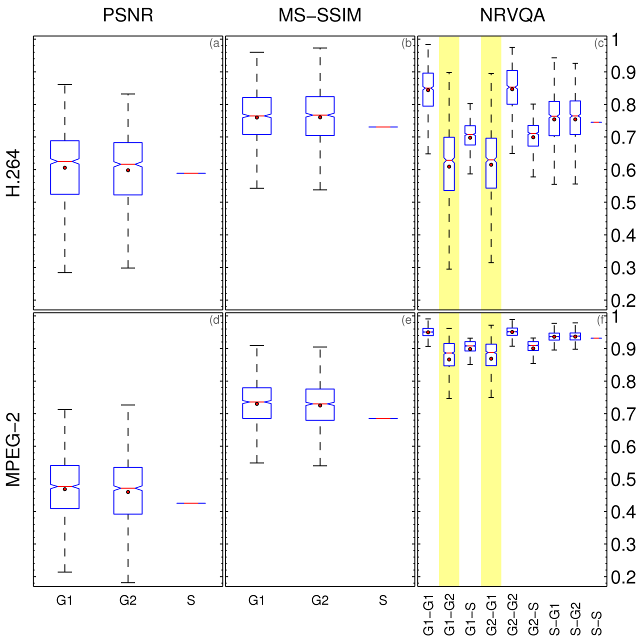

Fig. 4 Comparison of linearity distributions using box plots with mean (red circle) and notches between FR metrics and NRVQA composed of 1,000 different cross-validation experiments for $G1$ and $G2$ and one S with H.264 and MPEG-2 video samples from LIVE database.

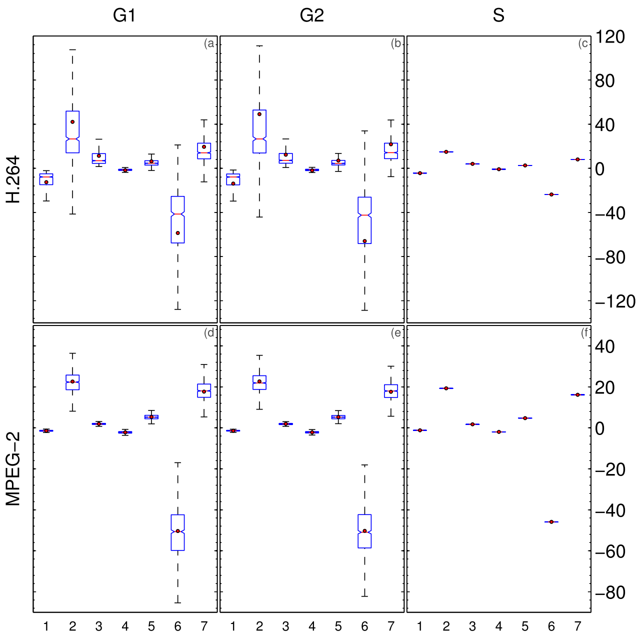

Fig. 5 Box plots distributions with mean (red circle) and notches of β_1 to β_7 from Eq. (17) for 1,000 different $G1_k$, $G2_k$ and S training groups involving H.264 and MPEG-2 from LIVE video database.

ric, approximately 18%, which is fairly acceptable for objective video quality evaluation using an NR method. The video quality prediction with 1,000 samples comprising data distribution can be analyzed by the box plot technique [39], which is a powerful tool for providing graphical support to display and compare video quality data sets and their statistics. The box plot summarizes the distributions of video quality data and allow visual comparisons of centers and spread through the six-number summary named as minimum (10%), lower quartile (first quartile or $Q1$ with 25%), mean (red circle), median (second quartile or $Q2$ represented by red line with 50%), upper quartile (third quartile or $Q3$ with 75%) and maximum (90%), which divides the data into four segments. Thus, the Fig. 4 shows the linearity distributions of PSNR, MS-SSIM and proposed method for H.264 and MPEG-2 video sequences from LIVE. The FR metrics contains three testing patterns named as $G1_k$, $G2_k$ and S has nine different training-test combinations, necessary for the cross-validation process. These results are confirmed by the visual inspection of Fig. 4 which shows the higher performance of the proposed method when it is applied to MPEG-2 video sequences, even it uses disjoint sets for the highlight training-test pair, such as $G1$-$G2$ and $G1$-$G2$.

Figure 5 shows the distribution of the β parameters of NRVQA method, as used in Eq. (17) for seven β parameters in the $G1_k$, $G2_k$ and S training sets. The coefficients β_2, β_6 and β_7 for both H.264 and MPEG-2 showed higher variation in the distribution as can be observed from the visual inspection of the box plot distribution of β's.

7. Conclusion

This work proposes a new no-reference video quality assessment method based on a sigmoid model approach, where the spatial-temporal features are weighted by values obtained in the training phase through the LM algorithm, which is employed to solve the nonlinear least squares problem. The experimental results show that although the comparison between FR and NR metrics is unfair due to the NR metric being a blind method (absence of the video-reference), our method presents best performance in terms of linearity (PLCC) in comparison with FR metrics, such as PSNR and MS-SSIM, when our method is applied to MPEG-2 and when it is applied to H.264 video sequences, it presents equivalent performance to the PSNR.

The comparison between the results of the proposed technique and those from other NR metrics available in the literature is however difficult due to the use of different mapping functions and different available databases, which are obtained under different conditions and video parameters. The cubic mapping function is used, as recommended by VQEG, while most works used the logistic mapping function that has been overtaken. Our method presents best performance, in comparison with FR metrics, for any training-test sets for MPEG-2, and when it is applied to H.264 video sequences has shown performance close to that achieved MS-SSIM. However, the method can still be applied with success to the evaluation of videos, whose degradations are originated through other mechanisms and encoding processes. As the proposed method does not require information of the video reference, it is suited for monitoring the video quality at the receiver side. For instance, in digital TV broadcast or mobile systems (where an increasing video content is being transmitted to devices, such as smartphones, tablets, mobile PCs, and Wireless Display – WiDi), the video quality scores can be sent back to the central station, via a return channel for further analysis and possible local corrections of the video distortion whenever feasible.

Acknowledgments

The authors thank Fundação Araucária for the financial support – Convênio 098/2014.

References

[1] H.R. Wu, K.R. Rao, and A.A. Kassim, "Digital video image quality and perceptual coding," J. Electronic Imaging, vol.16, no.3, 2007.

[2] F. Yang, S. Wan, Y. Chang, and H.R. Wu, "A novel objective no-reference metric for digital video quality assessment," IEEE Signal Process. Lett., vol.12, no.10, pp.685–688, 2005.

[3] F. Yang, S. Wan, Q. Xie, and H.R. Wu, "No-reference quality assessment for networked video via primary analysis of bit stream," IEEE Trans. Circuits Syst. for Video Technology, vol.20, no.11, pp.1544–1554, Nov. 2010.

[4] T. Brandão and M.P. Queluz, "No-reference quality assessment of H.264/AVC encoded video," IEEE Trans. Circuits Syst. Video Technol., vol.20, no.11, pp.1437–1447, Nov. 2010.

[5] G. Valenzise, S. Magni, M. Tagliasacchi, and S. Tubaro, "No-reference pixel video quality monitoring of channel-induced distortion," IEEE Trans. Circuits Syst. for Video Technology, vol.22, no.4, pp.605–618, April 2012.

[6] T. Yamada and T. Nishitani, "No-reference quality estimation for compressed videos based on inter-frame activity difference," IEICE Trans. Fundamentals, vol.E95-A, no.8, pp.1240–1246, 2012.

[7] L. Li, H. Zhu, J. Qian, and J.S. Pan, "No-reference quality metric of blocking artifacts based on color discontinuity analysis," IEICE Trans. Inf. & Syst., vol.E97-D, no.4, pp.993–997, 2014.

[8] Z. Wang, H.R. Sheikh, and A.C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," Proc. IEEE International Conference on Image Processing (ICIP'02), New York, pp.I–477–I–480, Sept. 2002.

[9] ITU-T P.910, "Subjective video quality assessment methods for multimedia applications," Tech. Rep. Recommendation ITU-T P.910, ITU Telecom, Standardization Sector of ITU, 1999.

[10] K. Levenberg, "A method for the solution of certain problems in least squares," Quarterly Applied Math, vol.2, pp.164–168, 1944.

[11] D.W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," SIAM J. Applied Mathematics, vol.11, no.2, pp.431–441, 1963.

[12] J. Moré, "The Levenberg-Marquardt algorithm: Implementation and theory," in Numerical Analysis, ed. G.A. Watson, Lecture Notes in Mathematics, vol.630, ch. 10, pp.105–116, Springer, Berlin, 1977.

[13] Y. Xie and C. Ma, "A smoothing Levenberg-Marquardt algorithm for solving a class of stochastic linear complementarity problem," Applied Mathematics and Computation, vol.217, no.9, pp.4459–4472, 2011.

[14] B. Widrow and J.M.E. Hoff, "Adaptive switching circuits," IRE WESCON Convention Record, vol.4, pp.96–104, 1960.

[15] J.D. Hoffman, Numerical Methods for Engineers and Scientists, 2nd ed., Taylor & Francis, 2001.

[16] M. Ries, J. Kubanek, and M. Rupp, "Video quality estimation for mobile streaming applications with neuronal networks," Proc. MESAQIN'06, Prague, Czech Republic, 2006.

[17] M. Shahid, A. Rossholm, and B. Lovstrom, "A reduced complexity no-reference artificial neural network based video quality predictor," 4th International Congress on Image and Signal Processing (CISP'11), pp.517–521, 2011.

[18] K. Kipli, M.S. Muhammad, S.M.W. Masra, N. Zamhari, K. Lias, and D.A.A. Mat, "Performance of Levenberg-Marquardt backpropagation for full reference hybrid image quality metrics," Lecture Notes in Engineering and Computer Science, vol.2195, no.1, pp.704–707, 2012.

[19] M.T. Hagan and M.B. Menhaj, "Training feedforward networks with the Marquardt algorithm," IEEE transactions on neural networks, vol.5, no.6, pp.989–993, Jan. 1994.

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

[21] Video Quality Experts Group (VQEG), "Final report from the video quality experts group on the validation of objective models of video quality assessment," tech. rep., VQEG, 2000.

[22] Video Quality Experts Group (VQEG), "Final report from the video quality experts group on the validation of objective models video quality assessment, Phase II," tech. rep., VQEG, 2003.

[23] Video Quality Experts Group (VQEG), "Final report from the video quality experts group on the validation of objective models of multimedia quality assessment, Phase I," tech. rep., VQEG, 2008.

[24] U. Engelke, M. Kusuma, H.J. Zepernick, and M. Caldera, "Reduced-reference metric design for objective perceptual quality assessment in wireless imaging," Image Commun., vol.24, no.7, pp.525–547, 2009.

[25] Video Quality Experts Group (VQEG), "Final report from the video quality experts group on the validation of reduced-reference and no-reference objective models for standard definition television, Phase I," tech. rep., VQEG, 2009.

[26] Video Quality Experts Group (VQEG), "Report on the validation of video quality models for high definition video content, version 2.0," tech. rep., VQEG, 2010.

[27] M.R. Spiegel and L.J. Stephens, Theory and problems of statistics, 3rd ed., Schaum's Outline Series, McGraw-Hill, New York, 1998.

[28] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification, 2nd ed., Wiley-Interscience, 2000.

[29] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed., Springer Series in Statistics, Springer, 2009.

[30] I. Zyl van Marais, W. Steyn, and J. du Preez, "Construction of an image quality assessment model for use on board an Leo satellite," IEEE International Geoscience and Remote Sensing Symposium (IGARSS'08), pp.II–1068 –II–1071, July 2008.

[31] H. Tong, M. Li, H.-J. Zhang, C. Zhang, J. He, and W.-Y. Ma, "Learning no-reference quality metric by examples," Proc. International Multi-Media Modelling Conference, pp.247–254, 2005.

[32] H. Liu, J. Wang, J. Redi, P.L. Callet, and I. Heynderickx, "An efficient no-reference metric for perceived blur," 3rd European Workshop on Visual Information Processing (EUVIP'11), pp.174 –179, July 2011.

[33] S. Decherchi, P. Gastaldo, R. Zunino, E. Cambria, and J. Redi, "Circular-ELM for the reduced-reference assessment of perceived image quality," Neurocomputing, vol.102, pp.78–89, Feb. 2013.

[34] P. Gastaldo and R. Zunino, "Neural networks for the no-reference assessment of perceived quality," J. Electronic Imaging, vol.14, no.3, p.033004, Sept. 2005.

[35] A. Lahouhou, E. Viennet, and A. Beghdadi, "Selecting low-level features for image quality assessment by statistical methods," J. Computing and Information Technology, vol.18, no.2, pp.183–189, 2010.

[36] N. Staelens, N. Vercammen, Y. Dhondt, B. Vermeulen, P. Lambert, R. Van de Walle, and P. Demeester, "VIQID: a no-reference bit stream-based visual quality impairment detector," Proc. 2010 Second International Workshop on Quality of Multimedia Experience (QoMEX 2010), Piscataway, USA, pp.206–211, IEEE, 2010.

[37] R. Herzog, M. Čadík, T.O. Aydın, K.I. Kim, K. Myszkowski, and

H.-P. Seidel, "NoRM: no-reference image quality metric for realistic image synthesis," Computer Graphics Forum, vol.31, no.2, pp.545–554, 2012.

[38] Z. Wang, E.P. Simoncelli, and A.C. Bovik, "Multiscale structural similarity for image quality assessment," Proc 37th Asilomar Conf on Signals, Systems and Computers, Pacific Grove, CA, pp.1398–1402, IEEE Computer Society, Nov. 9- 2003.

[39] J.W. Tukey, Exploratory Data Analysis, Addison-Wesley, New York, 1977.

Wyllian B. da Silva was born in 1978. He received the B.S. and M.S. degrees in physics and electrical engineering from the Federal University of Uberlândia, Brazil, in 2005 and 2008, respectively, and the Ph.D. degree in electrical engineering in 2013 from the Graduate Program in Electrical and Computer Engineering at the Federal University of Technology – Paraná (UTFPR). Since 2014 he is with the Center of Mobility Engineering (CEM), Federal University of Santa Catarina (UFSC), Joinville, Brazil, where he focused on the research and development of video quality assessment methods and computer networks.

Keiko V. O. Fonseca received the BS degree in electrical engineering from the Federal University, Paraná, Curitiba, in 1985, the M.S. (1988) and Ph.D. degree (1997) in electrical engineering from the State University of Campinas, Campinas, São Paulo and the Federal University of Santa Catarina, Florianópolis-SC, Brazil. She is a member of the IEEE Communication Society, the Brazilian Computer Society and the Institute of Electronics, Information, and Communication Engineers (Japan).

Alexandre de A. P. Pohl received the B.S. and M.S. degree in Physics in 1983 and 1987, respectively, from the State University of Campinas (Unicamp), Brazil, and the Ph.D. degree in electrical engineering in 1994 from the Technical University of Braunschweig, Germany. From 1987 to 1989 he was with the laser research division of the Brazilian Airspace Technical Center, São José dos Campos. From 1995 to 2000 he worked at the telecommunications division of Furukawa, Inc in Brazil. Since 2001 he is with the electrical engineering department of the Federal University of Technology – Paraná (UTFPR), Curitiba, Brazil, where he leads a research group working in the area of optical fiber communications and digital TV systems. He is a member of the Optical Society of America (OSA), the Brazilian Telecommunications Society (SBrT) and the Brazilian Microwave and Optoelectronics Society (SBMO).