Journal of Mobile Multimedia, Vol. 8, No. 4 (2013) 253–264

FUZZY LOGIC AND TEMPORAL INFORMATION APPLIED TO VIDEO QUALITY ASSESSMENT

CARLOS DANILO MIRANDA REGIS
Institute for Advanced Studies in Communications, Federal University of Campina Grande and Federal Insitute of Education, Science and Technology of Paraíba
Campina Grande, Paraíba, 58429-900, Brazil
regis.danilo@ieee.org

JOSÉ VINÍCIUS DE MIRANDA CARDOSO
Institute for Advanced Studies in Communications, Federal University of Campina Grande
Campina Grande, Paraíba, 58429-900, Brazil
josevinicius@iecom.org.br

ÍTALO DE PONTES OLIVEIRA
Federal Insitute of Education, Science and Technology of Paraíba
Campina Grande, Paraíba, 58432-300, Brazil
italoifpb@gmail.com

MARCELO SAMPAIO DE ALENCAR
Institute for Advanced Studies in Communications, Federal University of Campina Grande
Campina Grande, Paraíba, 58429-900, Brazil
malencar@iecom.org.br

Received January 18, 2013
Revised February 8, 2013

Video Quality Assessment (VQA) plays an important role for video communications systems and services, mainly to determine, accurately, the ratio between the provided quality and the resource demand. The objective VQA is a fast and viable methodology to determine the video quality for video service providers, although it presents an unsatisfactory correlation with the scores of quality given by the Human Visual System (HVS). The authors propose a novel *full reference* objective video quality metric considering spatial and temporal analysis. The spatial analysis used an algorithm, based on fuzzy logic, to classify the regions in three components. Temporal analysis was performed by means of the perceptual weighted structural similarity index (PW-SSIM) between the frames that contained the differences of pixels in the same spatial position and in subsequent frames. To validate the proposed VQA algorithm, the correlation coefficients between the objective measures and the subjective scores provided by the LIVE Video Quality Database were computed, considering the following distortions: H.264 and MPEG-2 encoding and transmission of H.264 bit-streams over IP and wireless networks. The results demonstrate that the proposed algorithm is a competitive alternative when compared with the classical objective algorithms such as MOVIE.

Keywords: Objective Video Quality Assessment, Structural Similarity Index, Fuzzy Logic, Visual Attention, Temporal Analysis
Communicated by: J.-H. Ma & T. Shih

1 Introduction

Applications involving the transmission, reproduction, storage or processing of digital videos in communication networks are present in many technological devices. They often cause loss of information, introducing various artifacts in the video, such as blurring, blocking and noise, which cause loss of visual quality. Estimate the quality of the digital videos quickly and accurately has been a major challenge for communication systems and video services. The video quality measurement is an important factor to establish the quality of video communication and processing systems such as: broadcasting and mobile television. It affects directly the video on demand and the video services providers, since they need to monitor broadcasting quality, to compute optimal parameters that ensure a satisfactory quality level.

Video quality assessment (VQA) methodologies are subdivided into two categories: objective and subjective. The objective methods, also called objective metrics or objective algorithms, are designed from mathematical models that, in general, compare statistical features of the video to estimate a quality measure. The objective metric can be briefly classified according to the availability of the original video as: *full reference*, in which the video quality is computed considering an original video as reference; *reduced reference*, whenever only the features of the original video are available for comparison with the distorted video and *no reference*, in which only the distorted video is used to evaluate the video quality. Subjective methodologies assess the video quality via psychophysical experiments with human observers. The observer watches the video sequences and evaluates the video according to a personal concept of quality.

The subjective approach is the natural way to assess the video quality. Nevertheless, subjective experiments are complex and time-consuming. Objective metrics are faster and has lower cost than the subjective metrics, because their results may be applied automatically to video systems.

Recently, studies developed objective metrics considering mainly the fact that the efficiency of a metric is highly dependent on the attributes of the human visual system that it emulates. In this sense, one of the spatial features that is intensively investigated is the visual attention. The salience map model is often used as a visual attention predictor. Studies reported that the Image Quality Assessment (IQA) algorithms, when combined with salience map models, presented significant improvement on its performance [1], [2].

Another way to consider spatial characteristics as a visual attention predictor are the classification algorithms. In general, the classification algorithms compare, individually, the spatial characteristics of the pixels, which are generally given by the gradient vectors, on the same spatial position in the reference and distorted videos [3]. Works reported that VQA algorithms are enhanced when combined with classification algorithms [4]. The benefit of the classification algorithms with respect to the salience map model is the low computational complexity, which is a key feature when combined with the VQA algorithms. Nevertheless, the computational or mathematical formulation of the visual attention is very limited, since the visual attention is an imprecise measurement.

In this paper the authors propose an algorithm based on fuzzy logic to the task of classification of the image in the three components: edge regions, smooth regions and texture regions. Fuzzy logic is a mathematical tool that allows, among other features, obtaining accurate measurements from a set of inaccurate measurements.

The principal contribution of this article is a *full reference* VQA algorithm that includes spatial analysis, as an application of the proposed classification algorithm, and temporal analysis. Temporal analysis plays an important role in VQA, since that the perception of the distortions by the HVS is attenuated or evidenced depending of the level of temporal activity. The subjective scores, available in the LIVE Video Quality Database were used to validate the proposed VQA algorithm [5], [6]. The results indicate a competitive performance of the proposed and the classical VQA algorithms.

The remainder of this paper is organized as follows. Section II briefly reviews the existing three-component classification model and its application in IQA algorithms and cites the VQA algorithms that obtained an improvement on the performance by means of the analysis of temporal perceptual factors. The gradient vector computation on the image space and a brief review of fuzzy logic are presented in Section III. Section IV describes the proposed three-component classification algorithm that uses the fuzzy logic and its application to estimate the spatial video quality. The aspects of the temporal perceptual information and its impact on the overall video quality are described in Section V. The discussion of the performance of the proposed VQA algorithm is presented in Section VI. Finally, Section VII presents the conclusion.

2 Related Works

Visual attention is a cognitive ability of the Human Visual System (HVS) that involves search, selection and focus of relevant stimuli [7]. The visual attention is little explored by computer systems, since the knowledge of this feature is scarce, which makes it hard to estimate it with analytical models.

One way to identify the visual attention is by means of classification algorithms. The purpose of classification algorithms is to distinguish regions according to the visual interest by the HVS. The classification algorithms have been used in image and video quality metrics to obtain the importance, i. e. the visual interest, of the different regions and, thus, to improve the objective measures. Regis *et al* [8] proposed an algorithm that uses the average magnitude of the gradient vectors (MGV) to identify relevant areas on the video, i. e. the MGV is used as a visual attention predictor, combined with a quality algorithm.

A model is presented in [9] in which the image is segmented is three regions: edge, smooth and texture. Based on that model, Li *et al* [3] presented an algorithm to recognize the pixels in the same three classes of regions. The classification algorithm presented in [3] uses the MGV as input. The pseudo-code of this algorithm is presented in Algorithm 1.

Based on Algorithm 1, the works of [10] and [11] proposed image quality metrics in which a quality index is computed for each segmented region and an overall quality index is obtained as the average weighted among the scores for each region.

The motivation to use the vector-gradient approach as a prediction of visual attention is the low computational complexity and simplicity, which are important specifications when incorporated into the metrics of video quality [12]. Moreover, the vector-gradient approach also provides a good approximation of the visual interest of the HVS, since the gradient vector provides the rate and the direction of variation of the pixels luminance levels.

Another fundamental point in the objective VQA is the temporal quality analysis. The

Algorithm 1: Pseudo-code for the classification algorithm proposed by Li [3].

Data: Original image $f = (x, y, f(x, y))$ and processed image $h = (x, y, h(x, y))$
Result: Classification of the pixels in three regions (edge, texture and smoothness)
begin

1) Apply the Sobel operator to the original and degraded videos to obtain the magnitude of the gradient vectors (MGV) in each (x, y) position.
2) Determine the thresholds $th_1 = 0.12 \cdot g_{max}$ and $th_2 = \frac{th_1}{2}$, in which g_{max} is the highest MGV in the image.
3) Assuming that $||f(x, y)||$ e $||h(x, y)||$ denote the MGV, in the position (x, y), apply the following rules:

R1: If $||f(x, y)|| > th_1$ **or** if $||h(x, y)|| > th_1$,
then the pixel is classified in the edge region.

R2: If $||f(x, y)|| < th_2$ **and** if $||h(x, y)|| \leq th_1$,
then the pixel is classified in the smooth region.

R3: If $th_1 \geq ||f(x, y)|| \geq th_2$ **and** if $||h(x, y)|| \leq th_1$,
then the pixel is classified in the texture region.

ITU-T P.910 [13] defines the temporal perceptual information (TI) as a measure of the differences of the pixels in the same spatial position but in successive frames. Vu *et al* [14] incorporate the temporal variation by computing the MS-SSIM for two new frames, one that is the difference between the current reference frame and the next reference frame, and another that is the difference between the current reference frame and the next distorted frame. In [15] the temporal quality variations are measured via the singular value decomposition, assuming that the difference between the singular vectors is a good estimate for temporal variations.

3 Gradient Vector and Fuzzy Logic Review

3.1 Gradient Computation

Let $\mathcal{V} = \{(x, y, n, \nu(x, y, n)) \in \mathbb{Z}^4 \mid 0 \leq x \leq I - 1,\ 0 \leq y \leq J - 1,\ 0 \leq n \leq N - 1,\ 0 \leq \nu(x, y, n) \leq 2^b - 1\}$, be a video signal with spatial resolution $I \times J$, N the number of frames, $\nu(x, y, n)$ the luminance level and b is the number of bits used to discriminate the luminance levels.

The differential operation on a video $\mathcal{V}$, defined as

$$O_s : (x, y, n, \nu(x, y, n)) \longrightarrow (x, y, n, \nabla\nu(x, y, n)), \tag{1}$$

associates, for each spatial coordinate (x, y), a gradient vector

$$\vec{\nabla}\nu(x, y) = \frac{\partial\nu(x, y)}{\partial x}\vec{i} + \frac{\partial\nu(x, y)}{\partial y}\vec{j} \tag{2}$$

with angle of direction

$$\theta = \arctan\left(\frac{\partial\nu(x, y)/\partial y}{\partial\nu(x, y)/\partial x}\right) \text{rad.} \tag{3}$$

The magnitude of the gradient vector is defined as

$$||\nabla \nu(x,y)|| = \left[\left(\frac{\partial \nu(x,y)}{\partial x}\right)^2 + \left(\frac{\partial \nu(x,y)}{\partial y}\right)^2\right]^{1/2}, \tag{4}$$

that corresponds to the rate of change of the luminance.

Digitally, the magnitude of the gradient vector is computed by means of finite differences, as

$$||\nabla \nu(x,y)|| \approx \left[(\mathcal{V}(x,y) * O_1)^2 + (\mathcal{V}(x,y) * O_2)^2\right]^{1/2}, \tag{5}$$

$$O_1 = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix},\ O_2 = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix},$$

in which O_1 and O_2 are the Sobel operators and $*$ is the linear filtering operation [16].

3.2 Fuzzy Logic

Fuzzy logic is a logical system that generalizes the classical logic among sets for reasoning under uncertainty, in which the elements of the system (inputs) are qualitatively interpreted [17].

A fuzzy set is an extension of the classical set, in which the elements have a membership degree that indicates the inclusion of the element in the set. The membership degree is expressed by a real number between $[0,1]$, in which 0 and 1 denote, respectively, the total incompatibility or compatibility in the fuzzy set. The membership degree is measured according a membership function. Mathematically, a fuzzy set $\mathcal{A}$, in an universe of discourse U, is completely determined by its membership function,

$$\mu_{\mathcal{A}} : U \longrightarrow [0,1]. \tag{6}$$

A membership function can be designed as an interview with those who are familiar with the underlying concept and later adjust it based on a tuning strategy [17].

Using fuzzy logic, it is possible to manipulate imprecise measurements by means of linguistic variables, linguistic terms and linguistic rules. A linguistic variable is defined as the quadruple $(N, T(N), U, F)$ in which N is the name of the linguistic variable, $T(N)$ is a set of the linguistic terms that represent the state of N, U is the universe of discourse and F is a function that associates for each $T(N)$ a membership function $\mu_{\mathcal{A}}$. The purpose of the linguistic variable is to provide a systematic method to the approximate concept of complex phenomena [17]. The linguistic term is a qualitative way to represent the state of a linguistic variable and to express concepts and knowledge in human communication, where as the membership function is useful for processing numeric input data. The linguistic rules are a way of the inference to combine the linguistic terms of two or more linguistic variables in IF...THEN propositions.

In general, the algorithm of fuzzy rule-based inference consists of three basic steps and an optional step [17]:

- Fuzzification: Calculate the membership degree to which the input data, given by measurements or observations, for each fuzzy sets associated to the linguistic variable.

- Inference: Calculate the conclusion based on its membership degree.
- Combination: Combine the conclusion inferred by all fuzzy rules into a final conclusion.
- Defuzzification: For applications that need a crisp value. It is used to convert a fuzzy conclusion into a crisp one.

4 Fuzzy Spatial Video Quality

For the task of segmentation in three regions (edge, texture and smoothness), the authors propose a fuzzy decision system considering the visual interest regions, described as follows.

Let $\mathbb{R}$ be the universe of discourse, in which $z \in \mathbb{R}$ in the interval $0 \leq z \leq 255$, f and h are the reference and the distorted video, respectively, and the linguistic variable *Visual Attention* (VA) as $(Visual\ Attention, T(Visual\ Attention), \mathbb{R}, F)$, that assumes the linguistic terms $T(Visual\ Attention) = \{low, mid, high\}$.

Initially, it defines a function $\texttt{average}_k(\cdot)$ that computes the average value of the magnitude of the gradient vectors in a block of index k and size 8×8 pixels. The value of $\texttt{average}_k(\cdot)$ represents an estimate of the visual interest in the k-th block.

In the Fuzzification process, the crisp values of $\texttt{average}_k(\cdot)$ are mapped according the membership functions of the fuzzy sets, that are illustrated in Fig. 1. This process is performed for all blocks of the videos.

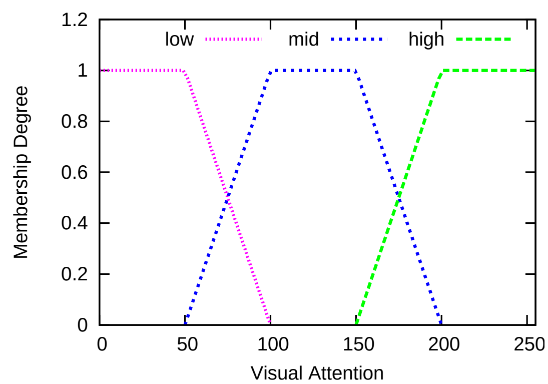


Fig. 1. Membership functions for the fuzzy sets *low*, *mid* and *high* associated to the linguistic variable *Visual Attention*.

The OR fuzzy logical operator, defined in [18] as $\mu_{\mathcal{A}\vee\mathcal{B}} = max\,(\mu_{\mathcal{A}}, \mu_{\mathcal{B}})$, was used for the fuzzification process. For example, considering the value $z = \texttt{average}_k(f) = 55$, then $\mu_{low}(z) = 0.9$, $\mu_{mid}(z) = 0.1$ and $\mu_{high}(z) = 0$, the OR operator among these fuzzy sets returns that z presents a greater membership degree to the *low* fuzzy set than others, therefore, the visual attention in this block is *low* with degree 0.9.

In the step of fuzzy inference used the IF-THEN implication rules presented in Table 1, in which the results of the fuzzification are compared. The AND operator is defined in [18] as $\mu_{\mathcal{A}\wedge\mathcal{B}} = min\,(\mu_{\mathcal{A}}, \mu_{\mathcal{B}})$. The propositions contained in TABLE 1 are performed between the membership degree of visual attention in the original and degraded videos.

In the proposed fuzzy system, the module of the decision region (Fig.2) informs the class to which the block k was marked according the output of the fuzzy inference.

Table 1. IF-THEN database rules.

IF			THEN
Rules	$VisualAttention(f)$	$VisualAttention(h)$	**Region**
1	*low*	*low*	Smooth
2	*low*	*mid*	Texture
3	*low*	*high*	Edge
4	*mid*	*low*	Smooth
5	*mid*	*mid*	Texture
6	*mid*	*high*	Edge
7	*high*	*low*	Edge
8	*high*	*mid*	Texture
9	*high*	*high*	Edge

Finally, the algorithm PW-SSIM (Perceptual Weighted Structural Similarity Index), proposed by Regis *et al* [19], was used to emit a video quality index for each region (edge, smooth and texture). The PW-SSIM is described as

$$\text{PW-SSIM}(f,h) = \frac{\sum_{k=1}^{K} \text{SSIM}(f_k,h_k)\cdot w_k}{\sum_{k=1}^{K} w_k}, \tag{7}$$

in which

$$w_k = \sqrt{\frac{1}{P-1}\sum_{p=1}^{P}\left(\mu_k\left(f\right) - ||\nabla f_p||\right)^2}, \tag{8}$$

is the local spatial perceptual information [19] in the k-th block and P is the number of gradient vectors in a block ($P = 64$ in the case of a block size 8×8).

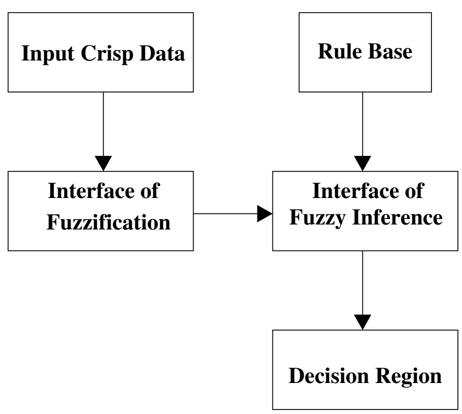


Fig. 2. Architecture of the fuzzy system.

The combination of the PW-SSIM scores results in the Fuzzy Spatial Video Quality Index, defined as

$$\text{FS-VQI} = \sum_{i=1}^{3} \text{PW-SSIM}_i \cdot \lambda_i. \tag{9}$$

in which λ_i are weights associated for each region and $\sum_{i=1}^{3} \lambda_i = 1$.

5 Temporal Perceptual Quality

Many algorithms of image quality assessment are also used for predicting the video quality. However, the video shows a temporal component which is not considered in such algorithms, which present an unsatisfactory correlation with the mean opinion scores obtained from subjective evaluations.

The rate of the temporal changes in the video is quantified by the differences of the pixels in the same spatial position of successive frames [13]. A similar approach to the proposed by Vu *et al* [14] was used to estimate the quality on the temporal component, using the MS-SSIM index between the differences of the subsequent frames, are computed as follows:

$$\begin{aligned} \mathcal{D}_{f,n} &= ||\mathcal{F}_{n+1} - \mathcal{F}_n||, \\ \mathcal{D}_{h,n} &= ||\mathcal{H}_{n+1} - \mathcal{F}_n||, \end{aligned} \tag{10}$$

in which $\mathcal{F}$ and $\mathcal{H}$ are the original and distorted frames, respectively, and n is the frame number.

In the proposed algorithm, the temporal quality is estimated by mean of the PW-SSIM index between the differences of the frames ($\mathcal{D}_{f,n}$ and $\mathcal{D}_{h,n}$), i.e.

$$\text{TP-VQI} = \frac{1}{N-1} \sum_{n=0}^{N-2} \text{PW-SSIM}\left(\mathcal{D}_{f,n}, \mathcal{D}_{h,n}\right). \tag{11}$$

The PW-SSIM index uses regions with large perceptual changes and presents a better correlation than the MS-SSIM [19].

The overall quality index is the average between the spatial and the temporal perceptual quality indices,

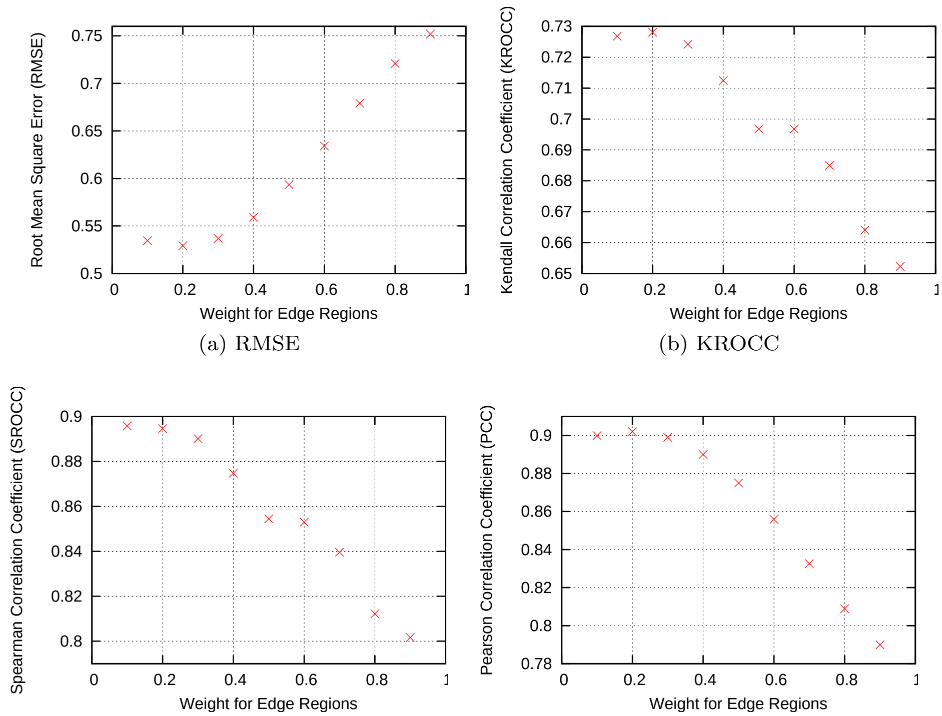
$$\text{O-SSIM} = \frac{\text{FS-VQI} + \text{TP-VQI}}{2}. \tag{12}$$

6 Simulation Results

The performance of an objective VQA algorithm is validated by means of the correlation between the objective measure, i.e. the predicted quality, and the mean opinion scores obtained in subjective evaluations. The Pearson Linear Correlation Coefficient (PCC) and the Spearman Rank-order Correlation Coefficient (SROCC) and Kendall Rank-order Correlation Coefficient (KROCC) were used to validate and compare the proposed algorithm, assessing the accuracy and the monotonicity of the objective model prediction with respect to human subjective scores.

6.1 Optimal Weights

Experiments were performed to find optimal weights that maximize the correlations coefficients and minimize the root mean square error among the objective values and the subjective scores, in which the weights λ_1, associated to the edge region, were changed in an interval $[0.1, 0.9]$ in steps of 0.1 and the weights associated to the smooth and texture regions had the same value. The experiment used the subjective evaluation performed by [20], in which video sequences in CIF (352×288) format were encoded using the H.264/AVC standard. Information about the subjective video quality database used in those experiments can be found in [20].



(a) RMSE (b) KROCC (c) SROCC (d) PCC

Fig. 3. Correlation coefficients and root mean square error used as criteria to find optimal weights.

The results of the experiments suggest that values around the triple $(0.2, 0.4, 0.4)$ are adequate. Fig. 3 shows the results of the correlation coefficients by Pearson, Spearman and Kendall, and the root mean square error. It is noted that the weight $\lambda_1 = 0.2$ is the best one for KROCC (Fig. 3b), PCC (Fig. 3d) and RMSE (Fig. 3a).

6.2 Performance Evaluation using the LIVE Video Quality Database

The LIVE Video Quality Database (LIVE) [5], [6] was used to compare the performance of the proposed algorithm with the classical objective metrics, considering videos with the following degradations: H.264 and MPEG-2 compression, simulated transmission of H.264 compressed bit-streams over error-prone IP and wireless networks. The videos from LIVE were: “Blue Sky”, “River Bed”, “Pedestrian area”, “Tractor”, “Sunflower”, “Rush hour”, “Station”, “Park run”, “Shields” and “Mobile & Calender”. For each video 15 test videos were produced, with the degradations cited previously. They were evaluated using the Absolute Category Rating (ACR) with a continuous scale. Information about the parameters used to distort the video, the conditions of the subjective experiments and the processing of subjective scores can be found in [6].

The PCC and SROCC were computed after performing a non-linear regression on the objective video quality assessment algorithmic measures, using a four parameter monotonic logistic function to fit the objective prediction to the subjective quality scores. The function is the following [6]

$$Q_k^{'} = \beta_2 + \frac{\beta_1 - \beta_2}{1 + e^{-(Q_k - \beta_3)/|\beta_4|}}, \tag{13}$$

in which Q_k represents the level that a video quality assessment algorithm predicts for the k-th video in the LIVE Video Quality Database.

The non-linear least squares optimization is performed using the MATLAB function `nlinfit` to find the optimal parameter β that minimizes the least squares error between the subjective scores (DMOS_k) and the fitted objective scores ($Q_k^{'}$). The initial estimates of the β parameter were [21]: $\beta_1 = \max(\mathrm{DMOS}_k)$, $\beta_2 = \min(Q_k^{'})$, $\beta_3 = \mathrm{mean}(Q_k^{'})$ and $\beta_4 = 1$. The MATLAB function `nlpredci` was used to obtain the DMOS, after the least squares optimization.

Tables 2 and 3 present the performance of the algorithms, in terms of PCC and SROCC, for each distortion provided by the LIVE Video Quality Database. The boldface correlation coefficients represent the two best performances.

Table 2. Pearson correlation coefficients.

Algorithms	**H.264**	**IP**	**wireless**	**MPEG-2**	**All**
PSNR	0.5492	0.4645	0.6690	0.3891	0.5621
SSIM [22]	0.6656	0.5119	0.5401	0.5491	0.5444
MS-SSIM [23]	0.6919	0.7219	0.7170	0.6604	0.7441
Speed SSIM	0.7206	0.5587	0.5867	0.6270	0.5962
VSNR [24]	0.6216	0.7341	0.6992	0.5980	0.6896
VQM	0.6459	0.6480	0.7325	0.7860	0.7236
V-VIF [25]	0.6911	0.5102	0.5488	0.6145	0.5756
S-MOVIE [26]	0.7252	0.7378	0.7883	0.6587	0.7451
T-MOVIE [26]	0.7920	0.7383	0.8371	**0.8252**	**0.8217**
MOVIE [26]	0.7902	**0.7622**	**0.8386**	0.7595	0.8116
S-ViMSSIM [14]	0.7834	0.7503	0.7837	0.7515	0.7796
T-ViMSSIM [14]	**0.8810**	0.6890	0.8219	0.7909	0.8122
ViMSSIM [14]	0.8117	0.7322	0.8327	0.7978	**0.8260**
O-SSIM (Proposed)	**0.8229**	**0.7623**	**0.8568**	**0.8034**	0.7649

The correlation coefficients indicated that the proposed algorithm is adequate for wireless transmission of H.264 bit-streams. On the transmission over IP networks the proposed algorithm is equivalent to the MOVIE index. For distortion generated at the H.264 encoding the Temporal-ViMSSIM (T-ViMSSIM) is the best one. Finally, the T-MOVIE showed the best correlation for the distortion created by the MPEG-2 encoding. However, T-MOVIE takes approximately five hours to compute the quality of a video with 250 frames and spatial resolution of 768 × 432 [14].

7 Conclusions

The authors proposed an algorithm for Video Quality Assessment (VQA) that divided the quality evaluation into in spatial analysis and temporal analyses. The overall quality assessment is an average of these two analysis. A classification algorithm based on fuzzy logic, which uses the magnitude of the gradient vectors in a block of size 8 × 8 pixels was proposed for spatial quality prediction. Unlike the pixel-by-pixel classification algorithms proposed in the literature, the use of 8 × 8 blocks enables to consider the spatial characteristics provided by neighborhood pixels, that improves the classification. The PW-SSIM index between pixels

Table 3. Spearman correlation coefficients.

Algorithms	**H.264**	**IP**	**wireless**	**MPEG-2**	**All**
PSNR	0.4585	0.4167	0.6574	0.3862	0.5398
SSIM [22]	0.6514	0.4550	0.5233	0.5545	0.5257
MS-SSIM [23]	0.7051	0.6534	0.7285	0.6617	0.7361
Speed SSIM	0.7086	0.4727	0.5630	0.6185	0.5849
VSNR [24]	0.6460	0.6894	0.7019	0.5915	0.6755
VQM	0.6520	0.6383	0.7214	0.7810	0.7026
V-VIF [25]	0.6807	0.4736	0.5507	0.6116	0.5710
S-MOVIE [26]	0.7066	0.7046	0.7927	0.6911	0.7270
T-MOVIE [26]	0.7797	**0.7192**	**0.8114**	**0.8170**	**0.8055**
MOVIE [26]	0.7664	**0.7157**	0.8109	**0.7733**	0.7890
S-ViMSSIM [14]	0.7713	0.6521	0.7340	0.7694	0.7690
T-ViMSSIM [14]	**0.8580**	0.6650	0.7951	0.7499	0.7984
ViMSSIM [14]	**0.8559**	0.6774	0.8111	0.7630	**0.8211**
O-SSIM (Proposed)	0.7812	0.7126	**0.8349**	0.7634	0.7576

in the same spatial position in subsequent frames is used to predict the temporal perceptual quality. The proposed algorithm was validated using the LIVE Video Quality Database. It showed satisfactory correlation and is an alternative to VQA.

Acknowledgments

This research was supported by the National Council for Scientific and Technological Development (CNPq/PIBITI), by the Federal University of Campina Grande (UFCG/COPELE), and by the Institute for Advanced Studies in Communications (Iecom).

References

1. Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *Image Processing, IEEE Transactions on*, vol. 20, no. 5, pp. 1185 –1198, may 2011.
2. M. Farias and W. Akamine, "On performance of image quality metrics enhanced with visual attention computational models," *Electronics Letters*, vol. 48, no. 11, pp. 631 –633, 24 2012.
3. J. Li, G. Chen, Z. Chi, and C. Lu, "Image coding quality assessment using fuzzy integrals with a three-component image model," *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 1, pp. 99 – 106, feb. 2004.
4. C. Li and A. C. Bovik, "Content-weighted video quality assessment using a three-component image model," *J. Electronic Imaging*, vol. 19, no. 1, p. 011003, 2010.
5. 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*, 2010.
6. ——, "Study of subjective and objective quality assessment of video," *IEEE Transactions on Image Processing*, pp. 1427–1441, 2010.
7. M. Corbetta, "Frontoparietal cortical networks for directing attention and the eye to visual locations: Identical, independent, or overlapping neural systems?" *Proceedings of the National Academy of Sciences USA*, vol. 95, pp. 831 – 838, 1998.
8. C. D. M. Regis, J. V. M. Cardoso, and M. S. Alencar, "Effect of visual attention areas on the objective video quality assessment," in *Proceedings of 18th Brazilian Symposium on Multimedia and the Web (WebMedia'12)*, October 2012.
9. X. Ran and N. Farvardin, "A perceptually motivated three-component image model-part i: description of the model," *IEEE Transactions on Image Processing*, vol. 4, no. 4, pp. 401 –415, apr 1995.
10. C. Li and A. C. Bovik, "Content-partitioned structural similarity index for image quality

assessment," *Image Communications*, vol. 25, no. 7, pp. 517–526, Aug. 2010. [Online]. Available: http://dx.doi.org/10.1016/j.image.2010.03.004

11. S. Gu, F. Shao, G. Jiang, and M. Yu, "A new four-component gradient-based structural similarity metric using adaptive weights," in *Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, vol. 2, july 2011, pp. 970 –973.
12. C. D. M. Regis, J. V. M. Cardoso, I. P. Oliveira, and M. S. Alencar, "Performance of the objective video quality metrics with perceptual weighting considering first and second order differential operators," in *Proceedings of 18th Brazilian Symposium on Multimedia and the Web (WebMedia'12)*, October 2012.
13. ITU-T, "ITU-T recommendation P.910, subjective video quality assessment methods for multimedia applications," April 2008.
14. C. Vu and S. Deshpande, "Vimssim: from image to video quality assessment," in *Proceedings of the 4th Workshop on Mobile Video*, ser. MoVid '12. New York, NY, USA: ACM, 2012, pp. 1–6. [Online]. Available: http://doi.acm.org/10.1145/2151677.2151679
15. M. Narwaria, W. Lin, and A. Liu, "Low-complexity video quality assessment using temporal quality variations," *IEEE Transactions on Multimedia*, vol. 14, no. 3, pp. 525 –535, june 2012.
16. R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.
17. J. Yen and R. Langari, *Fuzzy Logic: Intelligence, Control and Information*. Prentice-Hall, 1999.
18. L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.
19. C. D. M. Regis, J. V. M. Cardoso, and M. S. Alencar, "Video quality assessment based on the effect of the estimation of the spatial perceptual information," in *Proceedings of XXX Brazilian Symposium of Telecommunications (SBrT'12)*, 2012.
20. T. Brandão and M. Queluz, "No-reference quality assessment of h.264/avc encoded video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 11, pp. 1437 –1447, nov. 2010.
21. VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment, phase II (fr-tv2," April 2000, available at http://www.vqeg.org/.
22. Z. Wang, A. C. Bovik, H. R. Sheikh, S. Member, and E. P. Simoncelli, "Image quality assessment: From error measurement to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, 2004.
23. Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proceedings of IEEE Asilomar Conference on Signals, Systems, and Computers*, 2003, pp. 1398–1402.
24. D. M. Chandler and S. S. Hemami, "Vsnr: A wavelet-based visual signal-to-noise ratio for natural images." *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007.
25. H. R. Sheikh and A. C. Bovik, "Image information and visual quality," in *IEEE Transactions on Image Processing*, 2004, pp. 430–444.
26. K. Seshadrinathan and A. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *Image Processing, IEEE Transactions on*, vol. 19, no. 2, pp. 335 –350, feb. 2010.