Spatial and Temporal Analysis Considering Relevant Regions Applied to Video Quality Assessment

Ítalo de Pontes Oliveira, *IFPB Campus Campina Grande*, José Vinícius de Miranda Cardoso, Carlos Danilo Miranda Regis, *IFPB Campus João Pessoa*, and Marcelo Sampaio de Alencar, *IEEE Senior Member*

Abstract— **Video Quality Assessment (VQA) is an important factor to establish the quality of video communication and processing systems, such as broadcasting and mobile television. Video quality measurements affect directly video on demand and video services providers, since they need to monitor the broadcasting quality to compute optimal parameters that ensure a satisfactory quality level. The objective video quality assessment is a quick and low cost alternative compared with the subjective evaluation. However, the VQA is not as reliable, because their results are not always according to the perceived quality of the Human Visual System (HVS). This fact occurs due the incapacity of the objective algorithms in simulate the HVS characteristics. This paper describes a novel objective algorithm that includes spatial and temporal measurements, considering relevant regions. To validate the proposed algorithm, was computed the correlations coefficients between the objective measures and the subjective scores, provided by LIVE Video Quality Database, considering the following scenarios: H.264 and MPEG-2 encoding and transmission of H.264 bitstreams over IP and wireless networks. The simulation results suggest a significant enhance of the evaluation capacity of the objective algorithms when they were combined with spatial and temporal information and considering blocks of relevance. Furthermore, the results indicate that the proposed algorithm is a competitive alternative when compared with classical objective algorithms such as MOVIE.**

Keywords— **Objective Video Quality Assesment, Structural Similarity, Spatial Quality, Temporal Quality, Human Visual System.**

I. Introduction

Digital video bitstreams transmitted over error-prone channels, such as wireless channels, are subjected to transmission impairments. The packet loss during TCP/IP transmission causes distortions in the received video. Knowledge of the video quality is important to maintain, control and possibly enhance, the quality of the received video [1].

The 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 stastistical features of the video to estimate a quality measure. Subjective methodologies, in turn, evaluate the video quality via psychophysical experiments, in which the observers watch video sequences and evaluate the video quality according to a personal concept of quality.

Federal Institute of Education, Science and Technology of Paraíba – IFPB, Federal University of Campina Grande – UFCG, Institute for Advanced Studies in Communications – Iecom, Campina Grande – PB, Brazil, E-mails: italo.oliveira@ieee.org, josevinicius@iecom.org.br, regis.danilo@ieee.org, malencar@iecom.org.br.

The subjective approach is the natural way to assess of the video quality [2]. 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, to detect imperceptible degradations to the human eye. Objective video quality assessment constitutes an important sector for video services and processing systems, such as: vigilance systems [3], video on demand [4], spatial transcoding systems [5] and video conferecing [6]. However, the classical objective metrics, such as MSE (Mean Squared Error) and PSNR (Peak Signal to Noise Ratio), present an unsatisfactory correlation with the results provided by subjective evaluation, compromising the reliability of their measures [7].

Visual attention is a cognitive ability that involves search, selection and focus of relevant stimuli [8]. Experiments indicate that the human visual attention is not equally distributed throughout the image environment, but concentrates in a few regions [9]. It is estimated that the inclusion of methods that can identify the visual attention of a scene, i.e., assign a weight to the visual importance of regions on the image, tends to enhance the measures provided by the objective metrics.

The authors propose a new objective metric, for full reference video quality assessment, derived from the Structural Similarity Index with Perceptual Weighting (PW-SSIM), which includes a visual attention model based on the blocks of relevance from the edge detection and temporal information. The proposed metric was compared with another objective metrics, and presented results very competitive.

This paper is organized as follow. Section II describes the Structural Similarity Index approach. Section III describes the objective metric PW-SSIM. Section IV describes the algorithm VAA-PW-SSIM that is based on blocks of relevance. Section V is seen the description the metric TP-VQI that regards temporal information using the metric PW-SSIM, and in this section is discussed the proposed metric that combining evaluation of spatial and temporal information, by means of the indexes VAA-PW-SSIM, PW-SSIM and TP-VQI. Section VI shows the simulations results and section VII presents the conclusions.

II. Structural Similarity Index

The Structural SIMilarity (SSIM) [10] is a full-reference approach to image and video quality assessment based on the assumption that the Human Visual System (HVS) is highly adapted to recognize structural information in the visual environment and, therefore, the changes in the structural information provide a good approximation to the quality perceived by the HVS.

Let $f(x, y, n)$ and $h(x, y, n)$ scalar functions that represent 2D video sequences. In which, x and y represents the rectangular spatial coordinates and n represents the frame number. The SSIM(f, h) is computed as a product of three measures over the luminance plane: luminance comparison $l(f, h)$, the contrast comparison $c(f, h)$ and the structural comparison $s(f, h)$:

$$l(f, h) = \frac{2\mu_f \mu_h + C_1}{\mu_f^2 + \mu_h^2 + C_1}, \tag{1}$$

$$c(f, h) = \frac{2\sigma_f \sigma_h + C_2}{\sigma_f^2 + \sigma_h^2 + C_2}, \tag{2}$$

$$s(f, h) = \frac{\sigma_{fh} + C_3}{\sigma_f \sigma_h + C_3}, \tag{3}$$

that μ is the sample average, σ is the sample standard deviation, σ_{fh} is the covariance, $C_1 = (0.01 \cdot 255)^2$, $C_2 = (0.03 \cdot 255)^2$ and $C_3 = \frac{C_2}{2}$.

The structural similarity index is described as

$$\text{SSIM}(f, h) = [l(f, h)]^\alpha \cdot [c(f, h)]^\beta \cdot [s(f, h)]^\gamma, \tag{4}$$

in which usually $\alpha = \beta = \gamma = 1$.

In practice the SSIM is computed for an 8×8 sliding squared window or for an 11×11 Gaussian-circular window. The first approach is used in this paper. Then, for two videos which are subdivided into J blocks, the SSIM is computed as

$$\text{SSIM}(f, h) = \frac{1}{J}\sum_{j=1}^{J} \text{SSIM}(f_j, h_j). \tag{5}$$

III. Perceptual Weighted Structural Similarity Index

Regis *et al.* [11] proposed a technique called Perceptual Weighting (PW), which combines the local Spatial Perceptual Information (SI), as a visual attention estimator, with the SSIM, since experiments indicate that the quality perceived by the HVS is more sensitive in areas of intense visual attention [12]. The SI is computed using the Sobel differential operator, which estimates the magnitude of the gradient vectors of the video.

The PW technique uses the local SI to weigh the most visually important regions. This weighting is obtained as follows: compute the magnitude of the gradient vectors in the original video by means of the Sobel masks, then generate a perceptual map in which the pixel values are the magnitude of the gradient vectors. The frame is partitioned into blocks 8×8 pixels, and the local SI in each block is computed as

$$\text{SI}(f_j) = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}(\mu_j \; - \; |\nabla f(k)|)^2}, \tag{6}$$

in which, μ_j represents the sample average of the perceptual map in a j-block and K is a total of gradient vectors in j-th block. For the case that the frames are partitioned uniformly in squares 8×8, $K = 64$. The Perceptual Weighted Structural Similarity Index (PW–SSIM) is computed as

$$\text{PW–SSIM}(f, h) = \frac{\sum_{j=1}^{J} \text{SSIM}(f_j, h_j) \cdot \text{SI}(f_j)}{\sum_{j=1}^{J} \text{SI}(f_j)}. \tag{7}$$

IV. Block Detection with Perceptual Weighted Structural Similarity

Experiments indicated that the human visual attention is not equally distributed throughout the image space, but concentrates on a few regions of interest [13], and the visual attention is an important feature of the HVS to indicate the quality of that region [14].

Based on the assumption that the edge information is closely related to the visual attention, and taking into account that the HVS is more sensitive to asses the quality for areas of higher visual attention, a new approach to objective assessment of video quality is proposed using the PW technique proposed in [11] and a segmentation algorithm presented in [15].

Let $\mathcal{T} = \{\tau_i \mid \tau_i \in Z\ [0, 255]$ and $i = 0, 1, 2, \ldots, P\}$ be a video signal with 2^8 luminance levels and P is the image dimension. The magnitude of the gradient vector of $\mathcal{T}$ is defined as [16]

$$|\nabla\mathcal{T}| \;=\; \left[\left(\frac{\partial\mathcal{T}}{\partial x}\right)^2 + \left(\frac{\partial\mathcal{T}}{\partial y}\right)^2\right]^{\frac{1}{2}} \tag{8}$$

in which x and y denotes horizontal and vertical direction, respectively.

Digitally, the magnitude of the gradient is approximated by

$$|\nabla\mathcal{T}| \;=\; [(\mathbf{O}_1 * \mathcal{T})^2 + (\mathbf{O}_2 * \mathcal{T})^2]^{1/2}, \tag{9}$$

in which $*$ denotes the liner filtering operation and

$$\mathbf{O}_1 = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}, \qquad \mathbf{O}_2 = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix},$$

are the Sobel operators.

The gradient vectors are computed over the original video and the resulting video is subdivided into D blocks, 8×8 pixels. The average amplitude of the gradient vectors are computed for each block and the highest average among all the blocks is found. The blocks classified as of higher visual attention are those that present an average amplitude of the gradient vectors obeying the condition

$$\begin{aligned} \textbf{if}\ \mu_{j,n} \;&\geq\; \phi_n \;\Rightarrow\; (j, n) \in \mathcal{U}, \\ \phi_n \;&=\; \frac{\mu_{max,n}}{\mathcal{R}}, \end{aligned} \tag{10}$$

in which $\mu_{j,n}$ is average gradient vectors in the block j and in the window n, $\mu_{max,n}$ is the largest value of average in all blocks j, $\mathcal{R}$ is a constant, n is the frame number, j is the index of the block and $\mathcal{U}$ is a set that contains the blocks with higher visual attention. In the experiments $\mathcal{R} = 2.1$ was the optimal value that maximize the correlation for these sequences.

Only blocks that are below the threshold of Inequality 10 are considered in the calculation of the metric, i.e., the blocks for which the average magnitude of the gradient vectors do not obey the criteria are ignored. Figure 1 shows an example

(a) Original frame (b) Gradient frame (c) Relevant blocks

Fig. 1: Result of the segmentation techinique presented in [15].

of this method for the video Glasgow [17], in which the white squares in Figure 1c represent the blocks for which the PW-SSIM metric is computed. The structural similarity with the proposed segmentation algorithm is called VAA-PW-SSIM (Visual Attention Areas into PW-SSIM Index). In [15], is shown the an algorithm which computes using the objective metric SSIM, in this work, it is computed using the PW-SSIM, because this metric has shown better correlation results [11].

Mathematically, the local statistics μ, σ and σ_{fh} are computed using the previously selected blocks, resulting in

$$\text{VAA-PW-SSIM} = \frac{1}{J}\sum_{j\in\mathcal{U}}^{J} \text{PW-SSIM}(f_j, h_j), \qquad (11)$$

and, Block Detection with PW-SSIM (called BD-PW-SSIM), can be calculated making the combination of metrics VAA-PW-SSIM and PW-SSIM, which considers the spatial information in two different ways, combining blocks of relevance with highest visual attention and spatial information over the whole frame:

$$\text{BD-PW-SSIM} = \frac{\text{VAA-PW-SSIM + PW-SSIM}}{2}. \qquad (12)$$

V. Temporal Information with Perceptual Weighted Structural Similarity

Many algorithms of image quality assessment are also used for predicting the video quality. However, the video presents 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 sucessive frames [18]. A similar approach to the proposed by Vu *et al* [19] was used to estimate the quality on the temporal component, using the Multi-Scale Structural SIMilarity (MS-SSIM [20]) index between of the differences of the subsequent frames. The differences of the subsequent frames are computed as follow:

$$\begin{aligned}\mathcal{D}_{f,n} &= ||\ f_{n+1} - f_n\ ||,\\ \mathcal{D}_{h,n} &= ||\ h_{n+1} - f_n\ ||,\end{aligned} \qquad (13)$$

in which f and h are the original and distorted frames, respectively, and n is the frame number.

In the proposed algorithm, the temporal quality is estimated by means of the PW-SSIM index between the differences of the frames ($\mathcal{D}_{f,n}$ and $\mathcal{D}_{h,n}$), in which the Temporal Perceptual Video Quality Index (TP-VQI) is calculate as [21]:

$$\text{TP-VQI} = \frac{1}{N-1}\sum_{n=1}^{N-1} \text{PW-SSIM}\left(\mathcal{D}_{f,n}, \mathcal{D}_{h,n}\right), \qquad (14)$$

that N is the total number of frames.

The PW-SSIM index uses regions with large perceptual changes and presents a better correlation than the MS-SSIM [11]. The proposed objective video quality algorithm in this article combines the spatial analysis (using the metric BD-PW-SSIM, Eq. 12) and the temporal analysis (using the metric TP-VQI, Eq. 14) to produce an overall video quality estimative, computed as:

$$\text{BD-TPW-SSIM} = \frac{\text{BD-PW-SSIM} \ + \ \text{TP-VQI}}{2}. \qquad (15)$$

VI. Simulation Results

The LIVE Video Quality Database (LIVE) [22], [23] 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 used in 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 [23].

Tables I and II present the performance of the algorithms in terms of Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank Order Correlation Coefficient (SROCC), for each distortion provided by the LIVE Video Quality Database. The boldface correlation coefficients represent the two better performances.

The correlation coefficients indicated that the proposed algorithm is better mainly for wireless transmission of H.264 bit-streams, and for transmission over IP networks. For the distortions H.264 encoding the Temporal-ViMSSIM (T-ViMSSIM) is the better one. Finally, the MOVIE showed the better correlation for the distortions created by the MPEG-2 encoding [24]. However, MOVIE takes approximately five hours to compute the quality of a video with 250 frames and spatial resolution of 768 $\times$ 432 [19], because the complexity of the metric is much higher than the others. However, simulations were performed for these same videos, using the metric proposed BD-TPW-SSIM, coded in C programming language and performed on two machines: a desktop with Intel (R) Core 2 Duo E7400, 3GB of DDR2 main memory model 800 MHz, and operating system GNU/Linux 10.04.4 LTS, performed the simulations of BD-TPW-SSIM with an average time of 63.3 seconds for each video and a notebook with Intel (R) Core (TM) i5-2410M CPU $2,3$ GHz, and the operating system GNU/Linux 10.04.4 LTS, with the average time of 52 seconds for each video.

In the Table III presents the increase in percentage for metrics: PW-SSIM, TP-VQI, VAA-PW-SSIM, BD-PW-SSIM, and BD-TPW-SSIM, related to metric SSIM. The results in bold is the highest gain. Concludes that combining spatial and temporal information considering blocks of relevance from

TABLE III: Increase of the performance in percentage of the correlation coefficient in relation to SSIM in LIVE database.

(a) Increase for the PLCC

Algorithm	H.264	IP	MPEG-2	Wireless	All
PW-SSIM	9,28 %	24,30 %	21,91 %	18,72 %	17,98 %
TP-VQI	7,29 %	47,86 %	41,78 %	57,73 %	37,05 %
VAA-PW-SSIM	9,65 %	19,16 %	19,63 %	35,40 %	20,48 %
BD-PW-SSIM	9,07 %	19,83 %	24,95 %	34,98 %	20,46 %
BD-TPW-SSIM	**25,12 %**	**53,82 %**	**45,66 %**	**58,38 %**	**41,15 %**

(b) Increase for the SROCC

Algorithm	H.264	IP	MPEG-2	Wireless	All
PW-SSIM	11,90 %	23,65 %	16,90 %	17,85 %	20,53 %
TP-VQI	21,86 %	**66,40 %**	33,83 %	57,29 %	40,65 %
VAA-PW-SSIM	3,29 %	25,41 %	7,16 %	39,58 %	22,90 %
BD-PW-SSIM	4,18 %	27,96 %	26,46 %	38,10 %	22,77 %
BD-TPW-SSIM	**22,75 %**	59,74 %	**35,15 %**	**62,16 %**	**45,03 %**

TABLE I: PLCC considering the scenarios presented in LIVE [19].

Algorithm	H.264	IP	MPEG-2	Wireless	All Data
PSNR	0.5492	0.4645	0.3891	0.6690	0.5621
SSIM	0.6656	0.5119	0.5491	0.5401	0.5444
MS-SSIM	0.6919	0.7219	0.6604	0.7170	0.7441
S-MOVIE	0.7252	0.7378	0.6587	0.7883	0.7451
T-MOVIE	0.7920	0.7383	**0.8252**	0.8371	**0.8217**
MOVIE	0.7902	**0.7622**	0.7595	0.8386	0.8116
S-ViMSSIM	0.7834	0.7503	0.7515	0.7837	0.7796
T-ViMSSIM	**0.8810**	0.6890	0.7909	0.8219	0.8122
ViMSSIM	0.8117	0.7322	0.7978	0.8327	**0.8260**
PW-SSIM	0.7274	0.6363	0.6694	0.6412	0.6423
TP-VQI	0.7141	0.7569	0.7785	**0.8519**	0.7461
VAA-PW-SSIM	0.7298	0.6100	0.6569	0.7313	0.6559
BD-PW-SSIM	0.7260	0.6134	0.6861	0.7290	0.6558
BD-TPW-SSIM	**0.8328**	**0.7874**	**0.7998**	**0.8554**	0.7684

TABLE II: SROCC considering the scenarios presented in LIVE [19].

Algorithm	H.264	IP	MPEG-2	Wireless	All Data
PSNR	0.4296	0.3206	0.3588	0.4334	0.3684
SSIM	0.6514	0.4550	0.5545	0.5233	0.5257
MS-SSIM	0.7051	0.6534	0.6617	0.7285	0.7361
S-MOVIE	0.7066	0.7046	0.6911	0.7927	0.7270
T-MOVIE	0.7797	0.7192	**0.8170**	0.8114	**0.8055**
MOVIE	0.7664	0.7157	**0.7733**	0.8109	0.7890
S-ViMSSIM	0.7713	0.6521	0.7694	0.7340	0.7690
T-ViMSSIM	**0.8580**	0.6650	0.7499	0.7951	0.7984
ViMSSIM	**0.8559**	0.6774	0.7630	0.8111	**0.8211**
PW-SSIM	0.7289	0.5626	0.6482	0.6167	0.6336
TP-VQI	0.7938	**0.7571**	0.7421	**0.8231**	0.7394
VAA-PW-SSIM	0.6728	0.5706	0.5942	0.7304	0.6461
BD-PW-SSIM	0.6786	0.5822	0.7012	0.7227	0.6454
BD-TPW-SSIM	0.7996	**0.7268**	0.7494	**0.8486**	0.7624

the edge detection provides better correlation to video quality assessment when compared to SSIM.

The scatter plots presented in Fig. 2 illustrate the non-linear behavior of the measurements of the proposed algorithm with respect to the concept of quality of the observers on the subjective experiments performed with the LIVE Video Quality Database. In which, the line represents the curve of the logistic function (discussed in [23]), and the dots are the measured values, in axis of BD-TPW-SSIM is seen the results of avaliation of the algorithm, and in axis Difference Mean Opinion Scores (DMOS), its the result of subjective evaluation. How much closer is the dots of the line, better is the correlation.

VII. Conclusions

The authors proposed an algorithm to VQA that subdivided the quality computation in spatial analysis and temporal analysis. The overall quality assessment is an average of these two analysis. A classification algorithm based on egde detection that consider the arithmetic mean of amplitude of vector edge considering only blocks of size 8×8 that highest average between all blocks, different of others algorithm, that considerated all blocks in the frame. To predict the temporal perceptual quality was used a simple technique in which the PW-SSIM index is computed between the frames that contain the differences among the pixels in the same spatial position and in subsequent frames. The proposed algorithm was validated using the LIVE Video Quality Database. It showed satisfactory correlations and is an good alternative to VQA.

Acknowledgements

This research was supported by the Federal Institute of Education, Science and Technology of Paraíba (IFPB - Campus Campina Grande), by the National Council for Scientific and Technological Development (CNPq/PIBIC), by the Federal University of Campina Grande (UFCG), and by the Institute for Advanced Studies in Communications (Iecom).

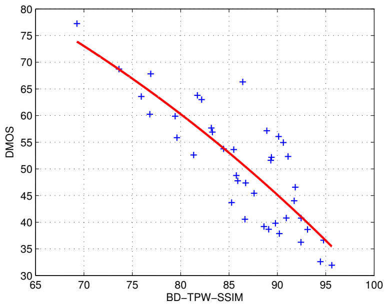
Referências

[1] Z. Wang, H. R. Sheikh, and A. C. Bovik, "Objective video quality assessment," in *In the Handbook of Video Databases: Design and Applications*. CRC Press, 2003, pp. 1041–1078.

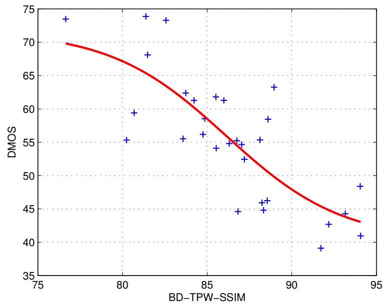
[2] Z. Wang and A. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *Signal Processing Magazine, IEEE*, vol. 26, no. 1, pp. 98–117, 2009.

[3] H. U. Keval, "Effective, design, configuration, and use of digital cctv," Ph.D. dissertation, University College London, 2009.

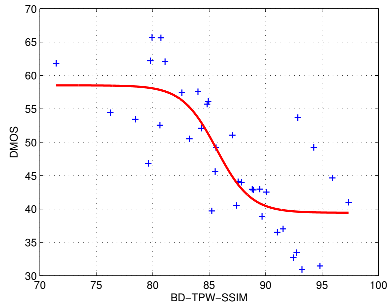
[4] R. Lopes, C. Regis, W. Lopes, and M. de Alencar, "Adaptvod - an adaptive video-on-demand platform for mobile devices," in *Multimedia and Ubiquitous Engineering (MUE), 2011 5th FTRA International Conference on*, 2011, pp. 257–262.



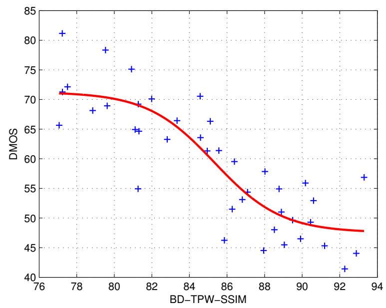
(a) H.264



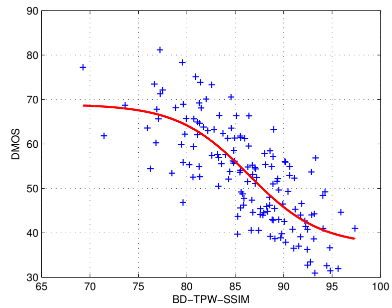
(b) IP



(c) MPEG-2



(d) Wireless



(e) All Data

Fig. 2: The behavior between the measurements of the proposed algorithm and the scores of the subjective experiments performed in LIVE Video Quality Database.

[5] C. D. M. Regis, "Avaliação de técnicas de redução da resolução espacial de vídeos para dispositivos móveis," Master's thesis, Universidade Federal de Campina Grande, 2009.

[6] C. R. D. Estrada, "Avaliação automática de qualidade de videoconferência de alta definição," Master's thesis, Universidade Federal do Rio de Janeiro, 2011.

[7] J. V. M. Cardoso, A. C. S. Mariano, C. D. M. Regis, and M. S. Alencar, "Comparação das métricas objetivas de qualidade de vídeos baseadas na similaridade estrutural e na sensibilidade ao erro," *Revista de Tecnologia da Informação e Comunicação*, pp. 33–40, 2012.

[8] 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.

[9] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, Mar 2001.

[10] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600 –612, april 2004.

[11] 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.

[12] H. Liu and I. Heynderickx, "Studying the added value of visual attention in objective image quality metrics based on eye movement data," in *16th IEEE International Conference on Image Processing*, nov. 2009, pp. 3097 –3100.

[13] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001. [Online]. Available: http://papers.klab.caltech.edu/83/1/391.pdf

[14] J. You, A. Perkis, and M. Gabbouj, "Improving image quality assessment with modeling visual attention," in *2nd European Workshop on Visual Information Processing (EUVIP)*, july 2010, pp. 177 –182.

[15] 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 the 18th Brazilian symposium on Multimedia and the web*, ser. WebMedia '12. São Paulo, SP, Brasil: ACM, 2012, pp. 75–78. [Online]. Available: http://doi.acm.org/10.1145/2382636.2382654

[16] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Addison-Wesley Publishing Company, 2000.

[17] A. S. University, "Yuv video sequences," in *teste*. http://trace.eas.asu.edu/yuv/, 2013, aNI-0133525 and ANI-0136774.

[18] ITU-T, "ITU-T recommendation P.910, subjective video quality assessment methods for multimedia applications," September 1999.

[19] 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

[20] 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.

[21] C. D. M. Regis, J. V. M. Cardoso, Ítalo Pontes Oliveira, and M. S. Alencar, 2013.

[22] 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.

[23] K. Seshadrinathan, R. Soundararajan, A. Bovik, and L. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1427 –1441, june 2010.

[24] P. Vu, C. Vu, and D. Chandler, "A spatiotemporal most-apparent-distortion model for video quality assessment," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, 2011, pp. 2505–2508.