Spatial Information and Disparity-based Weighting Applied to Stereoscopic Video Quality Assessment

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

***Abstract*— This paper presents new results on the objective evaluation of stereoscopic video quality. A balance between the scores provided by classical objective algorithms and the disparity information of the reference video is presented, along with a comparison between the approximations used to compute the local spatial information. The performance of the proposed technique was verified using statistical metrics (correlation coefficients and root mean square error) that compare the scores provided by the objective algorithms and the subjective results provided by the NAMA3DS1–COSPAD1 stereoscopic video quality database. The obtained results suggest a significant improvement on the performance of the objective algorithms for the proposed technique.**

***Keywords*— Stereoscopic Video, Objective Evaluation Techniques, Spatial Perceptual Information, Disparity**

I. Introduction

Objective evaluation is a fast and a low cost alternative to time-consuming subjective evaluations. Successfully objective algorithms were developed to evaluate the 2D video quality [1], but the 3D video has a new component that needs to be considered in the design of algorithms: the depth. Recently, the disparity has been used as a depth estimation [2].

Objective algorithms are computational models, which use statistical characteristics of the video combined with features of the Human Visual System (HVS), to estimate the quality score, classified according to the availability of the original signals as: *full reference*, in which the original video is compared with the video under test (degraded); *reduced reference*, whenever only characteristics of the original video are available for comparison with the video under test, and *no reference*, in which only the video under test is used for quality assessment.

This paper presents an investigation on the *full reference* objective algorithms for 3D video quality measurement when they are combined with the disparity and spatial informations present in the reference video. The focus of this research is to verify the impact of the disparity and spatial information on the visual attention of the human visual system and its consequence to the perceived 3D video quality.

The visual attention is a cognitive ability that involves search, selection and focus of relevant stimuli [3]. Experiments indicate that the human visual attention is not equally distributed throughout the image environment, but concentrates in a few regions [4]. Several approaches [5], [6], [7], [8], [9] indicate that the inclusion of methods to identify the visual attention of a scene, i.e., assign a weight to the visual importance of regions on the image, enhances the evaluation provided by the objective metrics.

Therefore, the proposed approach combines a weighting, that is a function of the disparity of the 3D original video, with the score provided by objective algorithms. The paper presents a comparative analysis of proposals to identify the spatial information, and improve the assessment of the 3D video quality. The following statistical metrics were used to compare the performance of the proposed approach with the classical algorithms: Pearson Linear Correlation Coefficient (PLCC), Spearman Rank-Order Correlation Coefficient (SROCC), Kendall Rank-Order Correlation Coefficient (KROCC) and Root Mean Square Error (RMSE). The Confidence Intervals for the PLCC are also presented.

The remaining of this paper is organized as follows. Section II describes some objective algorithms for 2D video quality used to predict the 3D video quality. Section III presents the approach used to introduce the disparity information contained in the reference video into the objective algorithm. Section IV describes the approches to estimate the rate of change of luminance and the local spatial information and how it was introduced into the objective metrics. The subjective evaluation with 3D video sequences, performed by NAMA3DS1-COSPAD1 [10] is reviewed in Section V. A discussion on the performance of the proposed algorithms is presented in Section VI. Section VII presents the conclusions.

II. Objective Approaches

Let $V = [v_L(x, y, n), v_R(x, y, n)]$ be a 3D video sample, in which the scalar functions v_L and v_R correspond to the left and right views, respectively; (x, y) represents the rectangular spatial coordinates and n represents the frame number. Let yet F and H be 3D reference video and the video under test, respectively. A full reference objective algorithm for 3D video quality assessment is a function G, such that its image $(\mathrm{G}(F, H))$ represents the quality of H with respect to F.

The following convention is used in this paper

$$\mathrm{G}(F, H) = \frac{\mathrm{G}(f_L, h_L) + \mathrm{G}(f_R, h_R)}{2}, \tag{1}$$

since the importance of the left and right views is the same.

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

A. Peak Signal-to-Noise Ratio

Let $f(x,y,n)$ and $h(x,y,n)$ be scalar functions that represent 2D video sequences. The Mean Square Error (MSE) between the signals is computed as

$$\text{MSE}(f,h) = \frac{1}{N \cdot X \cdot Y} \sum_{n=1}^{N} \sum_{x=1}^{X} \sum_{y=1}^{Y} [f(x,y,n) - h(x,y,n)]^2 . \tag{2}$$

The Peak Signal-to-Noise Ratio (PSNR) is computed as

$$\text{PSNR}(f,h) = 20 \cdot \log_{10} \left[\frac{\text{MAX}}{\sqrt{\text{MSE}(f,h)}} \right] \text{dB}, \tag{3}$$

in which MAX is the maximum value of the gray scale and $\text{MSE}(f,h)$ is the Mean Square Error between f and h.

B. Structural Similarity Index

The Structural SIMilarity (SSIM) [11] is a full-reference approach to image and video quality assessment based on the assumption that the 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 human visual system.

The $\text{SSIM}(f,h)$ 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 $s(f,h)$:

$$l(f,h) = \frac{2\mu_f \mu_h + C_1}{\mu_f^2 + \mu_h^2 + C_1}, \tag{4}$$

$$c(f,h) = \frac{2\sigma_f \sigma_h + C_2}{\sigma_f^2 + \sigma_h^2 + C_2}, \tag{5}$$

$$s(f,h) = \frac{\sigma_{fh} + C_3}{\sigma_f \sigma_h + C_3}, \tag{6}$$

in which 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

$$\text{SSIM}(f,h) = [l(f,h)]^\alpha \cdot [c(f,h)]^\beta \cdot [s(f,h)]^\gamma, \tag{7}$$

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

$$\text{SSIM}(f,h) = \frac{1}{J} \sum_{j=1}^{J} \text{SSIM}(f_j, h_j). \tag{8}$$

C. Perceptual Weighted Structural Similarity Index

Regis *et al.* [12] 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 [8]. 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

$$\text{SI}(f_j) = \sqrt{\frac{1}{K-1} \sum_{k=1}^{K} (\mu_j \; - \; |\nabla f_j(k)|)^2}, \tag{9}$$

in which, μ_j represents the sample average of the perceptual map in a j-block and K is the number of gradient vectors in the 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

$$\text{PW–SSIM}(f,h) = \frac{\sum_{j=1}^{J} \text{SSIM}(f_j,h_j) \cdot \text{SI}(f_j)}{\sum_{j=1}^{J} \text{SI}(f_j)}. \tag{10}$$

III. Disparity Weighting Technique

The disparity that is present in a 3D video sample is an information related to the stereo perception [2]. This information is computed as the difference between two corresponding pixels in the left and right views. Indeed, as is well know, the disparity should be considered in the development of objective algorithms, to improve the correlation between the objective prediction and the subjective scores.

The disparity map, $\text{D}(F)$, is computed as

$$\text{D}(F(x,y,n)) = |f_L(x,y,n) - f_R(x,y,n)|, \;\; \forall \;\; (x,y,n). \tag{11}$$

The introduction of the disparity information in the 2D objective metrics uses the weighted average of the objective measurements with the disparity map [13]. This approach was implemented in two objective metrics, PSNR and SSIM, producing the DPSNR and DSSIM.

The DPSNR_L, i.e., DPSNR for the left view, is computed as

$$\text{DPSNR}_L(F,H) = 20 \cdot \log_{10} \left[\frac{\text{MAX}}{\sqrt{\text{DMSE}_L(F,H)}} \right] \text{dB}. \tag{12}$$

The DMSE and the DPSNR for the right view (DMSE_R and DPSNR_R) are computed in the same manner. Then the overall DPSNR is the average of the DPSNR_L and DPSNR_R.

The DSSIM is computed as

$$\text{DSSIM}(F,H) = \frac{\sum_{j=1}^{J} \text{SSIM}(F_j,H_j) \cdot \text{D}(F_j)}{\sum_{j=1}^{J} \text{D}(F_j)}, \tag{14}$$

$$\text{DMSE}_L(F,H) = \frac{\sum_{n=1}^{N}\sum_{x=1}^{X}\sum_{y=1}^{Y}[f(x,y,n)-h(x,y,n)]^2 \cdot \text{D}(F(x,y,n))}{\sum_{n=1}^{N}\sum_{x=1}^{X}\sum_{y=1}^{Y}\text{D}(F(x,y,n))} \tag{13}$$

in which $\text{D}(F_j)$ is the average disparity contained in block j.

The DPW–SSIM is computed as

$$\text{DPW–SSIM}(F,H) = \frac{\sum_{j=1}^{J}\text{SSIM}(F_j,H_j)\cdot\text{SI}(F_j)\cdot\text{D}(F_j)}{\sum_{j=1}^{J}[\text{SI}(F_j)\cdot\text{D}(F_j)]}. \tag{15}$$

IV. Local Spatial Perceptual Information

The Spatial Perceptual Information (SI) quantifies the complexity of the spatial details present in a video sequence, and it increases with the spatial complexity of the samples [14]. The SI is computed by means of gradient vectors, which in turn, are computed using the Sobel approximations for the derivative in the n-th video frame ($\text{Sobel}(F_n)$). The standard deviation of the magnitude of the gradient vectors ($\text{std}[\text{Sobel}(F_n)]$) is calculated for each video frame. The highest value among the standard deviations represents the SI of the video sample. This process is mathematically represented as

$$\text{SI} = \max\{\text{std}[\text{Sobel}(V_n)]\}. \tag{16}$$

The gradient vectors ($\vec{\nabla}V$) estimate the rate of change of luminance values of the pixels along the horizontal and vertical directions, their maginitude is computed as

$$|\vec{\nabla}V| = \left[\left(\frac{\partial V}{\partial x}\right)^2 + \left(\frac{\partial V}{\partial y}\right)^2\right]^{1/2}. \tag{17}$$

In the discrete domain, the derivative operation is computed by approximations based in finite differences as

$$\frac{\partial V}{\partial x} \approx \sum_{y=1}^{Y}\sum_{x=1}^{X} v(x,y)\cdot e_x(x,y), \tag{18}$$

in which $v(x,y)$ is the luminance and $e_x(x,y)$ is the element of the approximation matrix.

The approaches to approximation considered in this work are: Sobel (19), Prewitt (20) and Robert-cross (21) Their approximation matrices, for orthogonal directions, are:

$$\begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}, \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix}; \tag{19}$$

$$\begin{bmatrix} -1 & 0 & +1 \\ -1 & 0 & +1 \\ -1 & 0 & +1 \end{bmatrix}, \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ +1 & +1 & +1 \end{bmatrix}; \tag{20}$$

$$\begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & +1 \\ -1 & 0 \end{bmatrix}. \tag{21}$$

For the Sobel and Prewitt approaches, the rate of change of luminance is computed in a coordinate system with 0 and $\frac{\pi}{2}$ rad as orthogonal directions. For the Robert-cross operator, the angle of the coordinate system is rotated by $\frac{\pi}{4}$ rad [15].

The Laplacian is defined as

$$\nabla^2 V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2}. \tag{22}$$

In image processing, the Laplacian is an isotropic operator, i.e., its value is independent of the direction of the edge. The Laplacian operation is approximated in a manner similar to Equation 18, and its matrix is defined as

$$\begin{bmatrix} 0 & +1 & 0 \\ +1 & -4 & +1 \\ 0 & +1 & 0 \end{bmatrix}. \tag{23}$$

The local spatial information, computed by Equation 9, was modified to verify the performance of the objective algorithms, based in the PW (DPW) technique, when combined with the information extracted by the differential operators.

The general equation for an objective algorithm, based on the PW technique, is

$$\text{M–PW–SSIM} = \frac{\sum_{j=1}^{J}\text{SSIM}(F_j,H_j)\cdot\text{SI}(F_j)}{\sum_{j=1}^{J}\text{SI}(F_j)}, \tag{24}$$

in which M = {R, P, L} denotes the differential approximation used to compute the magnitude of the gradient vector, or the Laplacian, in Equation 9 as: R (Roberts), P (Prewitt) and L (Laplace). The standard PW-SSIM means that the Sobel approximation was used.

The similar approach is used for the DPW technique, in which the general equation is

$$\text{M–DPW–SSIM}(F,H) = \frac{\sum_{j=1}^{J}\text{SSIM}(F_j,H_j)\cdot\text{SI}(F_j)\cdot\text{D}(F_j)}{\sum_{j=1}^{J}[\text{SI}(F_j)\cdot\text{D}(F_j)]}. \tag{25}$$

V. Subjective Experiment: NAMA3DS1-COSPAD1

The NAMA3DS1-COSPAD1 stereoscopic video quality database provides subjective results [16], [10] for the tests, using the Absolute Category Rating with Hidden Reference (ACR-HR) method [17], for coding and spatial degradations scenarios, which include H.264 coding. The Quantization Parameter (QP = 32, QP = 38 and QP = 44) was used to generate different levels of spatial degradation.

The diversity of the spatial and temporal features of the video samples available in the NAMA3DS1-COSPAD1 were quantified using the Spatial Perceptual Information and Temporal Perceptual Information (TI). The ITU-T Recommendation P.910 [14] suggests that, to form a set of video samples for a subjective evaluation some video parameters are taken into account, in particular the SI and TI, to prevent that the subjective evaluation becomes tiring and boring for the observers. It is important that the chosen videos present a variety of values of SI and TI.

Fig. 1 indicates the heterogeneity of sources available in the NAMA3DS1-COSPAD1, in which case each point represents a reference stereoscopic video.

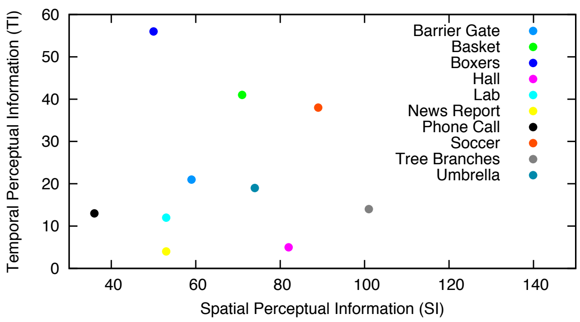


Fig. 1: Spatial and temporal heterogeneity found in the video sequences available in NAMA3DS1-COSPAD1.

VI. Simulation Results

The performance of the algorithms was measured using the following estimators: Pearson Linear Correlation Coefficient (PLCC), Spearman Rank-Order Correlation Coefficient (SROCC), Kendall Rank-Order Correlation Coefficient (KROCC) and Root Mean Square Error (RMSE). These estimators were used to assess the accuracy, monotonicity and consistency of the objective model prediction with respect to human subjective scores available from NAMA3DS1-COSPAD1, allowing the comparison and validation of the performance of the proposed algorithms with state of the art algorithms.

The PLCC, SROCC, KROCC and RMSE were computed after performing a non-linear regression on the objective video quality assessment algorithmic measures, using a four parameter monotonic cubic polynomial function to fit the objective prediction to the subjective quality scores. The function is the following [18],

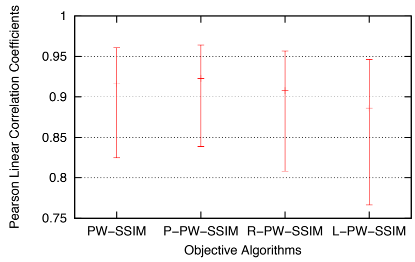
$$\text{DMOS}_t^{(p)} = \beta_1 + \beta_2 \cdot Q_t + \beta_3 \cdot Q_t^2 + \beta_4 \cdot Q_t^3, \tag{26}$$

in which Q_t represents the quality that a video quality assessment algorithm predicts for the t-th video in the NAMA3DS1-COSPAD1 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_t) and the fitted objective scores ($\text{DMOS}_t^{(p)}$). The MATLAB® function `nlpredci` was used to obtain the DMOS predicted scores, after the least squares optimization. The results of the statistical measures are presented in Table I, the best results are shown in boldface.

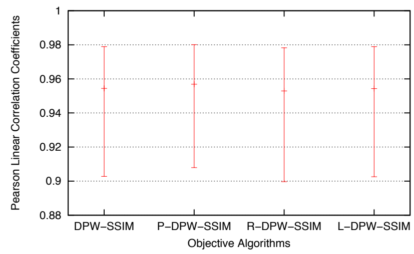
The results indicate that the Prewitt approximation is the best choice for the PW and DPW objective algorithms, providing high correlation coefficients and low RMSE. Nevertheless, the performance of the Laplacian operator is very interisting, because it needs fewer operations to compute the local perceptual information [5].

TABLE I: Performance measures of the objective algorithms.

Algorithm	PLCC	SROCC	KROCC	RMSE
PSNR	0.774946	0.721424	0.533869	0.689299
SSIM	0.730523	0.716222	0.555117	0.744770
DPSNR	0.863640	0.838604	0.640111	0.549789
DSSIM	0.901635	0.892266	0.746354	0.471688
PW–SSIM	**0.915983**	**0.906776**	**0.756978**	**0.437573**
L–PW–SSIM	0.886193	0.872006	0.709169	0.505301
R–PW–SSIM	0.907625	0.879398	0.730417	0.457821
P–PW–SSIM	**0.922928**	**0.908693**	**0.767602**	**0.419859**
DPW–SSIM	**0.954403**	**0.937166**	**0.815412**	**0.325572**
L–DPW–SSIM	0.954302	0.936619	0.810099	0.325923
R–DPW–SSIM	0.952903	0.934429	0.810099	0.330758
P–DPW–SSIM	**0.956889**	**0.945380**	**0.820724**	**0.316774**



(a) PW Technique



(b) DPW Technique

Fig. 2: The 95% confidence intervals for the PLCC.

Fig. 2 presents the PLCC with a 95% confidence interval for the PW and DPW techniques. The Fisher Z transformation was applied to the PLCC (ρ) to produce the confidence interval. The Fisher Z transformation is defined as

$$Z = \frac{1}{2}\log_e\left(\frac{1+\rho}{1-\rho}\right), \tag{27}$$

in which Z follows the Normal distribution with variance

$$\sigma_Z^2 = \frac{1}{\mathcal{N}_s - 3}, \tag{28}$$

and $\mathcal{N}_s$ is the total number of samples. The confidence interval for this variable is defined as,

$$\text{IC}(z, 1-\alpha) = (Z - z_{1-\alpha} \cdot \sigma_Z, Z + z_{1-\alpha} \cdot \sigma_Z). \tag{29}$$

For $\alpha = 0.05$, i.e., an interval with 95% of confidence, $z_{0.95} = 1.96$, and Equation 29 is rewritten as

$$\text{IC}(z, 0.95) = (Z - 1.96 \cdot \sigma_Z, Z + 1.96 \cdot \sigma_Z). \tag{30}$$

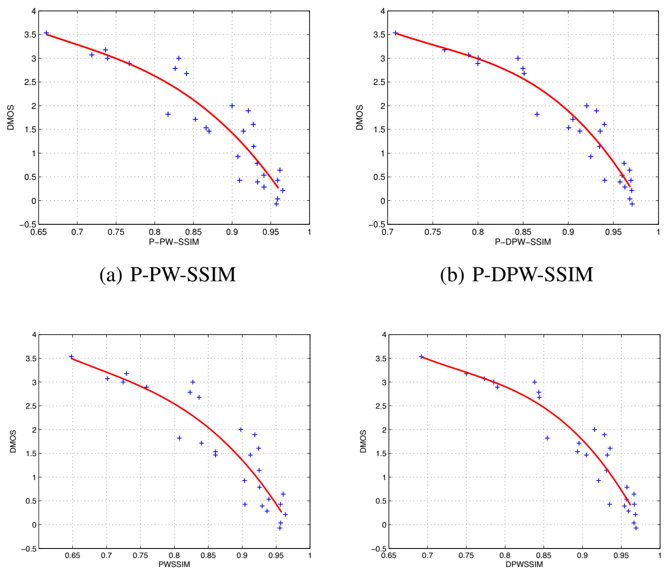
The inverse of the Fisher Z transformation is

$$\rho = \frac{e^{2z} - 1}{e^{2z} + 1}, \tag{31}$$

and the confidence interval in terms of ρ is defined as,

$$\mathrm{IC}(\rho, 0.95) = \left(\frac{e^{2\cdot(Z-1.96\cdot\sigma_Z)} - 1}{e^{2\cdot(Z-1.96\cdot\sigma_Z)} + 1}, \frac{e^{2\cdot(Z+1.96\cdot\sigma_Z)} - 1}{e^{2\cdot(Z+1.96\cdot\sigma_Z)} + 1} \right). \quad (32)$$

Fig. 2 suggests that, besides an increase in the correlation coefficient, there was a reduction in the length of the PLCC confidence interval for the algorithms that use the disparity weighting technique and the local spatial information computed by the Prewitt approximation.



(a) P-PW-SSIM

(b) P-DPW-SSIM

(c) PW-SSIM

(d) DPW-SSIM

Fig. 3: Scatter plots of subjective scores (DMOS) versus model prediction. Each sample point represents a 3D test video sample.

Fig. 3 presents the trend between the set of subjective scores and the set of objective results. The observation of the scatter plots indicates a lower dispersion around the prediction curve for algorithms that use the disparity weighting.

VII. Conclusions

This paper presents new contributions for the objective stereoscopic video quality assessment. The proposed algorithms use a disparity weighting strategy to take into account the stereoscopic information and different approaches to compute the local spatial information, to improve the evaluation. A reliable assessment is important for systems and services that use video processing techniques.

The simulation results suggest a significant increase on the evaluation capacity of the objective algorithms that use the proposed techniques, in relation to the classical algorithms. The Prewitt approximation present the best results. The Laplacian operator appears as a good alternative because of the low computational cost in relation to the first order derivative operators. For future work, the authors intend to consider temporal information for the objective evaluation and, compare with recent published algorithms for stereoscopic video quality assessment.

Acknowledgments

This research was supported by the UFCG/PIBIC, National Council for Scientific and Technological Development (CNPq), IFPB Campus João Pessoa and Iecom.

Referências

[1] 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.

[2] A. Benoit, P. L. Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP J. Image and Video Processing*, vol. 2008, 2008.

[3] M. Corbetta, "Frontoparietal cortical networks for directing attention and the eye to visual locations: identical, independent, or overlapping neural systems?" *Proc Natl Acad Sci U S A*, vol. 95, no. 3, pp. 831–8, 1998. [Online]. Available: http://www.biomedsearch.com/nih/Frontoparietal-cortical-networks-directing-attention/9448248.html

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

[5] 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. New York, NY, USA: ACM, 2012, pp. 75–78. [Online]. Available: http://doi.acm.org/10.1145/2382636.2382654

[6] 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 the 18th Brazilian symposium on Multimedia and the Web*, ser. WebMedia '12. New York, NY, USA: ACM, 2012, pp. 71–74. [Online]. Available: http://doi.acm.org/10.1145/2382636.2382653

[7] W. Y. L. Akamine and M. C. Q. Farias, "Incorporating visual attention models into image quality metrics," in *Proceedings of the Sixth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, 2012.

[8] 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.

[9] 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, 2011.

[10] M. Urvoy, M. Barkowsky, R. Cousseau, Y. Koudota, V. Ricorde, P. Le Callet, J. Gutierrez, and N. Garcia, "Nama3ds1-cospad1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3d stereoscopic sequences," in *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, july 2012, pp. 109 –114.

[11] 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.

[12] 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 30th Brazilian Symposium of Telecommunications (SBrT'12)*, 2012.

[13] C. D. M. Regis, J. V. M. Cardoso, I. P. Oliveira, and M. S. de Alencar, "Objective estimation of 3d video quality: A disparity-based weighting strategy," in *Proceedings of the IEEE International Symposium On Broadband Multimedia Systems and Broadcasting*, 2013.

[14] International Telecommunication Union, "Recommendation P.910: Subjective video quality assessment methods for multimedia applications," ITU-T, Tech. Rep., April 2008.

[15] M. Juneja and P. S. Sandhu, "Performance evaluation of edge detection techniques for images in spatial domain," *International Journal of Computer Theory and Engineering*, no. 5, Dec. 2009.

[16] IRCCyN-IVC, "Nantes-madrid 3d stereoscopic database," Julho 2012. [Online]. Available: http://www.irccyn.ec-nantes.fr/spip.php?article1052

[17] International Telecommunication Union, "Recommendation BT.500-11: Methodology for the subjective assessment of the quality of television pictures," ITU-T, Tech. Rep., 2002.

[18] ——, "Objective perceptual assessment of video quality: Full reference television," ITU-T, Tech. Rep., 2004.