# Analysis of the Window Size Used by Objective Algorithms to 3D 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*— The authors describe the results about the effect of the window size in objective algorithms, based on the Structural Similarity Index (SSIM), applied to estimate the 3D video quality. The evaluated 3D video samples, provided by NAMA3DS1-COSPAD1 project, contain spatial impairments due to the H.264 coding. The correlation between the sets of objective measurements and subjective scores are presented.**

***Keywords*— 3D Video Quality, Objective Evaluation Techniques, Structural Similarity**

## 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, 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 [1].

Several objective algorithms use the sliding window technique, with a static size, to subdivide the video signal, performing a local quality measurement [2]. The size of the window is related with the spatial resolution of the video signals and it is used by objective algorithms to indicate the amount of the visual information necessary to compute a local quality score. Nevertheless, there are many video services and systems that use video signals with various spatial resolutions such that a static window size is inappropriate, so that a evaluation of the window size used by objective algorithms is necessary.

This paper presents an investigation on the effect of the window size, in the high spatial resolution scenario, used by *full reference* objective algorithms, for 3D video quality measurement of the H.264/AVC coded videos, when they are combined with the disparity and spatial informations present in the reference video.

## II. Objective Algorithms

### A. Strucutural Similality Index

The Structural SIMilarity (SSIM) [2] 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.

Let $f(x, y, n)$ and $h(x, y, n)$ be scalar functions that represent the original and under test 2D videos, respectively, in which $(x, y)$ represents the rectangular spatial coordinates and $n$ represents the frame number. 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 $s(f, h)$

$$l = \frac{2\mu_f\mu_h + C_1}{\mu_f^2 + \mu_h^2 + C_1}, c = \frac{2\sigma_f\sigma_h + C_2}{\sigma_f^2 + \sigma_h^2 + C_2}, s = \frac{\sigma_{fh} + C_3}{\sigma_f\sigma_h + C_3} \quad (1)$$

in which $\mu$ is the sample average, $\sigma$ is the sample standard deviation, $\sigma_{fh}$ 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

$$\text{SSIM}(f, h) = [l(f, h)]^\alpha \cdot [c(f, h)]^\beta \cdot [s(f, h)]^\gamma, \quad (2)$$

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

In practice the SSIM is computed for an $8\times8$ sliding squared window. For two videos which are subdivided into $J$ windows, the SSIM is computed as

$$\text{SSIM}(f, h) = \frac{1}{J}\sum_{j=1}^{J}\text{SSIM}(f_j, h_j). \quad (3)$$

### B. Perceptual Weighted Structural Similarity Index

Regis *et al.* [3] 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.

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 windows of $8 \times 8$ pixels, and the local SI in each window is computed as

$$\text{SI}(f_j) = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}(\mu_j \ - \ |\nabla f_j(k)|)^2}, \quad (4)$$

in which, $\mu_j$ represents the sample average of the perceptual map in a $j$-th window and $K$ is the number of gradient vectors in the $j$-th window. For the case that the frames are partitioned uniformly in squares $8\times8$, $K = 64$. The Perceptual Weighted Structural Similarity Index (PW–SSIM) is computed as

$$\text{PW–SSIM}(f, h) = \frac{\sum_{j=1}^{J}\text{SSIM}(f_j, h_j)\cdot\text{SI}(f_j)}{\sum_{j=1}^{J}\text{SI}(f_j)}. \quad (5)$$

## III. Disparity Weighting Technique

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.

The disparity that is present in a 3D video sample is an information related to the stereo perception [1]. 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.

Let $F$ and $H$ be reference and under test 3D video samples, respectively. The disparity map, $\mathrm{D}(F)$, is computed as

$$\mathrm{D}(F(x,y,n)) = |f_L(x,y,n) - f_R(x,y,n)|, \;\; \forall \;\; (x,y,n). \quad (6)$$

The introduction of the disparity information in the 2D objective metrics uses the weighted average of the objective measurements with the disparity map [4]. This approach was implemented in two objective metrics, SSIM and PW–SSIM, producing the DSSIM and DPW–SSIM.

The DSSIM is computed as

$$\mathrm{DSSIM}(F,H) = \frac{\sum_{j=1}^{J} \mathrm{SSIM}(F_j,H_j) \cdot \mathrm{D}(F_j)}{\sum_{j=1}^{J} \mathrm{D}(F_j)}, \quad (7)$$

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

The DPW–SSIM is computed as

$$\text{DPW–SSIM}(F,H) = \frac{\sum_{j=1}^{J} \mathrm{SSIM}(F_j,H_j) \cdot \mathrm{SI}(F_j) \cdot \mathrm{D}(F_j)}{\sum_{j=1}^{J} [\mathrm{SI}(F_j) \cdot \mathrm{D}(F_j)]}. \quad (8)$$

## IV. Experiments, Simulation Results and Discussion

The NAMA3DS1-COSPAD1 stereoscopic video quality database [5] provides subjective results for the tests, using the Absolute Category Rating with Hidden Reference (ACR-HR) method, for coding and spatial degradations scenarios, which include H.264/AVC coding. The Quantization Parameter (QP = 32, QP = 38 and QP = 44) was used to generate different levels of spatial degradation.

The effect of the window used in SSIM based objective algorithms was evaluated for the following sizes: $8 \times 8$ (standard), $12 \times 12$, $20 \times 20$ and $30 \times 30$. The Pearson Linear Correlation Coefficient (PLCC), Spearman Rank-Order Correlation Coefficient (SROCC), Kendall Rank-Order Correlation Coefficient (KROCC) and Root Mean Square Error (RMSE) were used to assess the accuracy, monotonicity and consistency of the objective model prediction with respect to human subjective scores.

Table I presents the performance of the objective algorithms for each window size considered. The boldface values represent the better performance for the objective algorithms in function of the window size. In general, the results suggest that the increase of the window size reduces the performance of the objective algorithms such as DPW-SSIM and PW-SSIM. On the other hand, the best results for the DSSIM and SSIM were obtained with window size of $12 \times 12$ and $20 \times 20$, respectively. For future works the authors intend to evalute the impact of the window size in the objective algorithms using others stereoscopic video quality database.

TABLE I: Performance measures of the objective algorithms.

(a) Window Size $8 \times 8$

| Algorithm | PLCC | SROCC | KROCC | RMSE |
|---|---|---|---|---|
| **SSIM** | 0.730523 | 0.716222 | 0.555117 | 0.744770 |
| **PW–SSIM** | **0.915983** | **0.906776** | **0.756978** | **0.437573** |
| **DSSIM** | 0.901635 | 0.892266 | 0.746354 | 0.471688 |
| **DPW–SSIM** | **0.954403** | **0.937166** | **0.815412** | **0.325572** |

(b) Window Size $12 \times 12$

| Algorithm | PLCC | SROCC | KROCC | RMSE |
|---|---|---|---|---|
| **SSIM** | 0.771426 | 0.749624 | 0.586990 | 0.693981 |
| **PW–SSIM** | 0.909449 | 0.887885 | 0.730417 | 0.453495 |
| **DSSIM** | **0.912704** | **0.907871** | **0.762290** | **0.445650** |
| **DPW–SSIM** | 0.943981 | 0.919370 | 0.788851 | 0.359902 |

(c) Window Size $20 \times 20$

| Algorithm | PLCC | SROCC | KROCC | RMSE |
|---|---|---|---|---|
| **SSIM** | 0.801577 | **0.779740** | **0.613551** | 0.652070 |
| **PW–SSIM** | 0.893660 | 0.882136 | 0.735730 | 0.489408 |
| **DSSIM** | 0.905383 | 0.899932 | 0.746354 | 0.463072 |
| **DPW–SSIM** | 0.923606 | 0.899932 | 0.751666 | 0.418080 |

(d) Window Size $24 \times 24$

| Algorithm | PLCC | SROCC | KROCC | RMSE |
|---|---|---|---|---|
| **SSIM** | 0.805835 | 0.773169 | 0.602926 | 0.645798 |
| **PW–SSIM** | 0.884212 | 0.861054 | 0.703857 | 0.509412 |
| **DSSIM** | 0.895905 | 0.882136 | 0.725105 | 0.484502 |
| **DPW–SSIM** | 0.912356 | 0.883231 | 0.725105 | 0.446495 |

(e) Window Size $30 \times 30$

| Algorithm | PLCC | SROCC | KROCC | RMSE |
|---|---|---|---|---|
| **SSIM** | **0.809333** | 0.775633 | 0.592302 | **0.640573** |
| **PW–SSIM** | 0.871708 | 0.842711 | 0.682608 | 0.534431 |
| **DSSIM** | 0.881318 | 0.854210 | 0.693233 | 0.515343 |
| **DPW–SSIM** | 0.895858 | 0.862423 | 0.693233 | 0.484606 |

## Referências

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

[2] 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.

[3] 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.

[4] 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.

[5] 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 *4th International Workshop on Quality of Multimedia Experience (QoMEX)*, July 2012.

## Acknowledgements

The authors sincerely thank to UFCG/PIBIC, Iecom, IFPB and CNPq.