B-SSIM: Structural Similarity Index for Blurred Videos

José V. de M. Cardoso, Carlos D. M. Regis and Marcelo S. Alencar

***Abstract*— This paper presents the B-SSIM, a new metric to video quality assessment based on the structural similarity index and the spatial perceptual information. It shows a better correlation with the quality perceived by the human visual system, than the metrics PSNR, SSIM and MultiScale-SSIM (MS-SSIM).**

***Keywords*— Video Quality Assessment, Structural Similarity, Spatial Perceptual Information, Human Visual System.**

I. Introduction

Visual quality is an important factor in video communication systems and services. Especially due to an increased demand and the variety of video services distributed over the Internet and mobile networks. Compression and transmission introduce a variety of artifacts and distortions in the digital video, such as, blurring, blocking, Gaussian noise and salt & pepper noise, causing loss of visual quality [1].

The most accurate way to determine the quality of a video is by measuring it using psychophysical experiments with human subjects, called subjective video quality assessment [2], in which evaluators watch video samples and assign a score to the quality according to individual criteria of judgment. However, the implementation of this approach is complex, has a high cost and demands human resources.

An alternative is to resort to methods of objective video quality assessment, that use algorithms to measure the visual quality. This model is faster and has lower cost than subjective solutions and its results allow continuous monitoring of the quality of videos or define optimal compression parameters. Nevertheless, popular objective metrics, such as PSNR (Peak Sinal-to-Noise Ratio) and MSE (Mean Squared Error) do not usually show good correlation with subjective scores, compromising the reliability of this approach [2].

Currently, one of the most studied objective metrics is the Structural SIMilarity Index (SSIM), based on the concept that the Human Visual System (HVS) is highly adapted to recognize structural information in the visual environment. Its results show substantial progress on image and video quality assessment. Nevertheless, this metric does not have the same success in the case of videos that present blurring distortion [3].

A new approach to assess the quality of digital video objectively in the presence of blurring distortion is proposed, called B-SSIM, based on SSIM and the Spatial Perceptual Information (SI) of the videos.

II. SSIM: Structural SIMmilarity Index

The Structural SIMilarity Index (SSIM) is a metric proposed by Wang et al [4], based on the hypothesis that the structural information of the image is a highly sensitive measure to the HVS and that the change in this structural information can provide a good approximation to the visual quality. Let $f = \{f_i \mid i = 1, 2, 3, \ldots, P\}$ be the original video signal and $h = \{h_i \mid i = 1, 2, 3, \ldots, P\}$ be the distorted video signal, computed as the set of three measures over the pixel luminance plane: luminance comparison $l(f, h)$, constrast comparison $c(f, h)$ and structural comparison $s(f, h)$,

$$l(f,h) = \frac{2\mu_f\mu_h + C_1}{\mu_f^2 + \mu_h^2 + C_1}, \quad c(f,h) = \frac{2\sigma_f\sigma_h + C_2}{\sigma_f^2 + \sigma_h^2 + C_2}, \tag{1}$$

$$s(f,h) = \frac{\sigma_{fh} + C_3}{\sigma_f\sigma_h + C_3}, \tag{2}$$

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

The structural similarity index is described as

$$\text{SSIM}(f,h) = [l(f,h)]^\alpha \cdot [c(f,h)]^\beta \cdot [s(f,h)]^\gamma, \tag{3}$$

in which usually $\alpha = \beta = \gamma = 1$ [4].

In pratice 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 D blocks, the SSIM is computed as

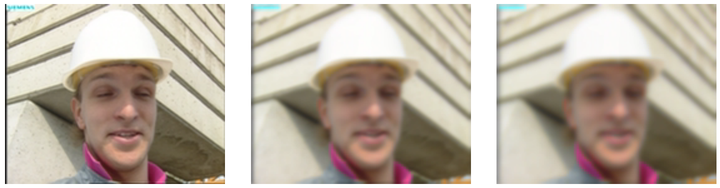
$$\text{SSIM}(f,h) = \frac{1}{D}\sum_{i=1}^{D} \text{SSIM}(f_i, h_i). \tag{4}$$

III. B-SSIM: Structural SIMilarity Index for Blurred Videos

The parameter that measures the complexity of the spatial details of the video samples is called Spatial Perceptual Information (SI). It is higher for more spatially complex samples [5]. The SI is computed taking into account the gradients in the vertical and horizontal directions using the Sobel filters in the n-th video frame ($\text{Sobel}(\text{F}_n)$), then the standard deviation of the magnitude of the gradients ($\text{std}[\text{Sobel}(\text{F}_n)]$) is calculated for each video frame. The highest value among the standard deviations represents the SI of the video sample. This process is

$$\text{SI} = \max\{\text{std}[\text{Sobel}(\text{F}_n)]\}. \tag{5}$$

The blurring distortion is presented as a reduction of edge sharpness and a loss of spatial detail. In real applications, this degradation is due to the exclusion of high frequency



(a) Original frame (b) 2-Blurred frame (c) 4-Blurred frame

Fig. 1: Blurring degradation in the "Foreman" video frame.

coefficients in the quantization process [1]. Fig. 1a shows an example of this distortion, in which the "2-blurred" means two applications of the mean filter 3×3 on the video and "4-blurred" means four applications.

The investigation that resulted in the proposed method started from the observation that the SI of the video samples is closely related to the blurring distortion. Fig. 2 shows that an increase in the amount of the blurring distortion causes a reduction in the SI.

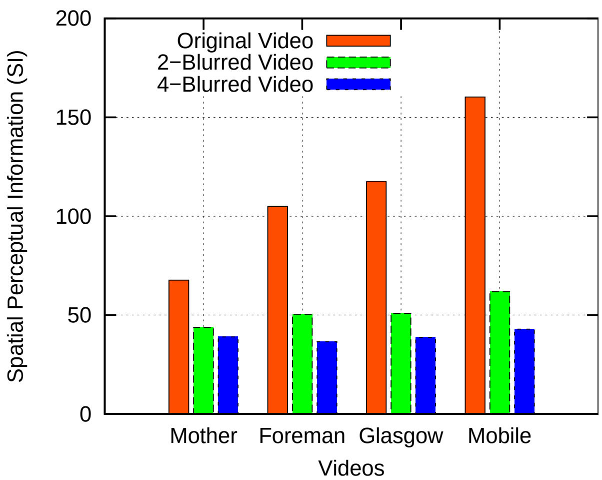


Fig. 2: Relationship between SI and the blurring distortion.

A comparison between the SI of the original and processed videos provides a good approximation to the quality perceived by the HVS regarding the blurring degradation.

A modification is proposed for the SSIM, taking into account the SI to evaluate the quality of videos that present this type of degradation. It is called Blur-inspired Structural Similarity Index (B-SSIM).

Initially, define the comparison function of the spatial perceptual informations $b(f, h)$ as

$$b(f, h) = \frac{2\mathrm{SI}_f \mathrm{SI}_h}{\mathrm{SI}_f^2 + \mathrm{SI}_h^2}, \quad (6)$$

in which SI_f and SI_h are the spatial perceptual informations of the original and processed video, respectively. The B-SSIM is described as,

$$\text{B-SSIM}(f, h) = b(f, h) \cdot \frac{1}{D} \sum_{i=1}^{D} \mathrm{SSIM}(f_i, h_i). \quad (7)$$

IV. Simulation and Results

For subjective evaluation, the selected videos were: "Foreman", "Glasgow", "Mobile & Calendar" and "Mother and Daughter". They display two intensity levels of blurring distortion, as seen in Fig. 1. These videos were analyzed by 40 evaluators, using the Absolute Category Rating (ACR) method [5], and their scores were assigned according as a scale of five discrete values.

The efficiency of the B-SSIM metric was assessed by the correlation between its results and the values of the subjective evaluation (Mean Opinion Score − MOS). For the videos presented in Fig. 1b and Fig. 1c, B-SSIM $= 0.6434$ and SSIM $= 0.8241$, B-SSIM $= 0.4526$ and SSIM $= 0.7276$, respectively.

These results indicate that the effect of the SI on the objective video quality assessment, provided by the B-SSIM measure, is more sensitive to the blurring artifact, resulting in an improvement that better represents the quality perceived by the HVS. This improvement is shown in Table I, considering the Pearson Correlation Coefficient (PCC) and the Spearman Rank-order Correlation Coeficient (SROCC) between the MOS and the results obtained using objective metrics, suggesting that the proposed metric has a better ability to predict the visual quality perceived by HVS.

TABLE I: Pearson and Spearman Rank-order Correlation Coefficients

Model	**PCC**	**SROCC**
PSNR	0.677	0.738
SSIM	0.806	0.738
MS-SSIM	0.817	0.786
B-SSIM	**0.905**	**0.881**

Acknowledgments

The authors would like to thank CNPq/PIBITI, UFCG/COPELE, IFPB and Iecom for supporting research.

References

[1] S. Winkler, *Digital Video Quality: Vision Models and Metrics*, WileyEditors, Ed. Wiley, 2005.

[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, jan. 2009.

[3] J. V. de Miranda Cardoso, A. C. S. Mariano, C. D. M. Regis, and M. S. Alencar, "Comparação das métricas objetivas baseadas na similaridade estrutural e na sensibilidade ao erro," *Revista de Tecnologia da Informação e Comunicação (RTIC)*, no. 2, pp. 33–40, 2012.

[4] Z. Wang, L. Lu, and A. Bovik, "Video quality assessment using structural distortion measurement," in *International Conference on Image Processing.*, vol. 3, 2002, pp. III–65 – III–68 vol.3.

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