Definition of Masks Related to Psychovisual Features for Video Quality Assessment

López, J.P.; Rodrigo, J.A.; Jiménez, D. and Menéndez, J.M.

Abstract— **Video Quality Assessment needs to correspond to human perception. Pixel-based metrics (PSNR or MSE) fail in many circumstances for not taking into account the spatio-temporal property of human's visual perception. In this paper we propose a new pixel-weighted method to improve video quality metrics for artifacts evaluation. The method applies a psychovisual model based on motion, level of detail, pixel location and the appearance of human faces, which approximate the quality to the human eye's response. Subjective tests were developed to adjust the psychovisual model for demonstrating the noticeable improvement of an algorithm when weighting the pixels according to the factors analyzed instead of treating them equally. The analysis developed demonstrates the necessity of models adapted to the specific visualization of contents and the model presents an advance in quality to be applied over sequences when a determined artifact is analyzed.**

Keywords*— *subjective assessment, video quality, psychovisual model, mask, motion

I. INTRODUCTION

VIDEO quality measurement plays an important role for most image processing applications, but it is still a difficult task because of the weaknesses of the error sensitivity based framework and the difficulty of finding models which adjust perfectly to viewer's vision [1]. Pixel-based metrics, such as Root Mean Square Error (RMSE) or Peak Signal-to-Noise Ratio (PSNR) are dominant in practice. However, these metrics do not take into account the spatio-temporal properties of human's visual perception, which is the reason why it fails under many circumstances [2]. Assuming that human visual perception is highly adapted for extracting structural information from a scene [3], quality metrics should offer a closer response to visual perception.

Techniques analyzed include detection of motion and isolation of edges to create visual attention and psychovisual models. Zhai proposed a psychovisual metric based on the free-energy principle [4]. Saliency diagrams and sharpness evaluation on edges are helpful to create visual models that focus on user attention, as demonstrated on static images [5]. Content-weighted video quality assessment has been developed by other researchers [6], based on texture, edge and smooth regions, exportable to video sequences. Picture region division with pondered weights studied in [7]. Motion is a key factor on visual attention, because it calls human eye's attention compared to static object [8].

In this paper, we propose the inclusion of a psychovisual model for improving objective video quality metrics, a novel statement and a key factor for assuring that the final spectator obtains a signal quality correlated to human visual system.

II. OBJECTIVES AND PLATFORM DESCRIPTION

For defining the coefficients of influence for each video characteristic, three sequences containing different content and artificially impaired were mainly used. Impairment is produced over motion, level of detail, pixel location and the appearance of human faces. In [9] the characteristics of this sequences are collected for each of the degradation when applying the distorted image only to one on the six regions of interest for each phenomena (details or spatial entropy; motion or temporal entropy; center, lateral or corner position and faces presence), conserving the rest of the picture sharped. For each of these six phenomena the inverse version was also developed, i.e. sharp pixels where the phenomena is located and the rest of the picture distorted. Results from subjective evaluation for each of these sequences measured in MOS (Mean Opinion Score), full-reference metrics like PSNR or MSE and the conventional blurring metric are collected from the former study. This information is used to create pixel masks for pixel-based metrics in order to assign different weights to each pixel depending on subjective relevance.

III. DEFINITION OF MASKS

For each of the phenomena, a mask is implemented. The level of importance of the mask is indicated in grey level images. Black pixels are not taken into account, while white pixels are primordial for the final algorithm. Motion, detail and facial masks have binary values. Position masks offer a shade of grey to determine the level of importance of the lateral areas.

The four masks are previously combined for the final psychovisual model in order of influence to the algorithm. Each frame of the sequence possesses a different and unique psychovisual model mask. Fig. 1 shows the final composite mask for a given frame.

After analyzing the subjective results in [9], coefficients were assigned to the four different masks related to each of the image characteristic selected. The values were defined ordering the level of importance derived from observers' opinions. The most highlighted effect in video sequences is face detection. Sequences with a lower PSNR obtained better scores than sequences with higher PSNR when low quality was concentrated in pixels contained in faces ROI's.

The same conclusions can be extracted from the region analysis. Sequences got better results when distortion was applied to the center rather than to the rest of the image.

Following the same pattern it has been decided to consider motion distortion to affect more the final result than detail distortion, both considered to be more important than lateral or corner distortions.

TABLE I. COEFFICIENTS OF INFLUENCE TO THE PSYCHOVISUAL MODEL DEPENDING ON THE ROI BELONGING

Characteristics		Pixels ROI Belonging	Partial coef.	Norm coef.	Influence (%)
Faces		Pixels ∈ (ROI Faces)	1	0,29	28,6%
Pos.	Center	Pixels ∈ (ROI Center)	0,9	0,26	25,7%
	Lateral	Pixels ∈ (ROI Lateral)	0,4	0,11	11,4%
	Corner	Pixels ∈ (ROI Corner)	0,1	0,03	2,9%
Motion		Pixels ∈ (ROI Motion)	0,6	0,17	17,1%
Details		Pixels ∈ (ROI Detail)	0,5	0,14	14,3%

The following chain of influence has been considered: Faces > Central > Motion > Detail > Lateral > Corner.

As a start point, the greatest influence (faces) has been assigned with a weight value of 1 and the following elements have been assigned with decreasing values. These values have been normalized to obtain the final distortion coefficients.

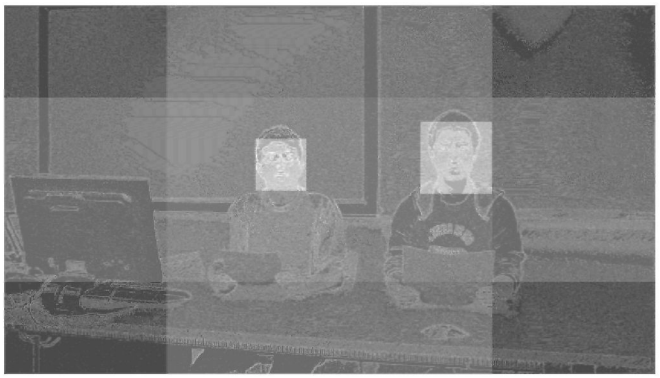

Fig. 1. Example of final mask in sequence News Report

The coefficients are justified with the results obtained from subjective tests. The focus is on the cases where contradictions exist between the MOS scores and classical metrics such as PSNR or MSE, and psychovisual models are especially developed for introducing visual attention factors to correct the values of distortion metrics. An example of the final mask for a frame in sequence News Report can be seen in Fig. 1.

IV. CONCLUSIONS

Most of video quality metrics are based on objective parameters, such as PSNR, MSE or SNR, which compare the original image to the degraded one, establishing a relationship between them. These metrics are often not corresponded to the human visual system because they are not adapted to what human eye most probably sees. The visual attention systems and saliency diagram based on the factors analyzed in the state of the art, improves the response of metrics, as in the case of blurring.

Characterization of video sequences to concrete artifacts and visual attention models is a key factor when developing a useful database for assessing quality. In this paper we have presented a way to weight these metrics in order to adapt their results to a more realistic situation. Based on previous studies a chain of influence has been stablished in order to create pixel-weight masks, taking into account the appearance of faces in the frame, pixel location, motion and frame detail.

The metrics with correction derived by spatial and temporal entropy regions obtain better results than without taking into account the image characteristics. This means than human eye is more sensitive to artifacts in the central areas of the picture. Also, motion and level of detail detected in the video frame are an important factor to generate an attention model which helps to improve the result of a metric as blurring.

ACKNOWLEDGMENT

The work developed by Universidad Politécnica de Madrid was performed in the framework of project TEC2012-38402-C04-01 HORFI, which is partially funded by the Spanish Ministry of Science and Innovation.

REFERENCES

[1] Z.Wang, A.C.Bovik, and L.Lu, "Why is image quality assessment so difficult?", IEEE International Conference on Acoustics, Speech, & Signal Processing, vol, 4, pp. 3313-3316, May 2002.

[2] Xiao, Feng. "DCT-based video quality evaluation." *Final Project for EE392J* 769 (2000).

[3] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. Image Processing, IEEE Transactions on, 13(4), 600-612.

[4] Zhai, G., Wu, X., Yang, X., Lin, W., & Zhang, W. (2012). A psychovisual quality metric in free-energy principle. *Image Processing, IEEE Transactions on*, *21*(1), 41-52.

[5] Wooding, D. S. (2002, March). Fixation maps: quantifying eye-movement traces. In Proceedings of the 2002 symposium on Eye tracking research & applications (pp. 31-36). ACM.

[6] Li, C.; Yuan, W.; Bovik, A.C.; Wu, X.; , "No-reference blur index using blur comparisons," Electronics Letters , vol.47, no.17, pp.962-963, August 18 2011

[7] Nojiri, Y., Yamanoue, H., Ide, S., Yano, S., & Okana, F. (2006). Parallax distribution and visual comfort on stereoscopic HDTV. In Proceedings of IBC (No. 3, pp. 373-380).

[8] Seshadrinathan, K., & Bovik, A. C. (2010). Motion tuned spatio-temporal quality assessment of natural videos. Image Processing, IEEE Transactions on, 19(2), 335-350.

[9] J. P. López, J. A. Rodrigo, D. Jiménez y J. M. Menéndez. (2014). "Insertion of Impairments in Test Video Sequences for Quality Assessment Based on Psychovisual Characteristics". In Artificial Intelligence, Modelling and Simulation (AIMS), 2014 Second International Conference on. IEEE.