World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:1, No:8, 2007

A New Image Psychovisual Coding Quality Measurement based Region of Interest

M. Nahid, A. Bajit, A. Tamtaoui, and E. H. Bouyakhf

***Abstract*—**To model the human visual system (HVS) in the region of interest, we propose a new objective metric evaluation adapted to wavelet foveation-based image compression quality measurement, which exploits a foveation setup filter implementation technique in the DWT domain, based especially on the point and region of fixation of the human eye. This model is then used to predict the visible divergences between an original and compressed image with respect to this region field and yields an adapted and local measure error by removing all peripheral errors. The technique, which we call foveation wavelet visible difference prediction (FWVDP), is demonstrated on a number of noisy images all of which have the same local peak signal to noise ratio (PSNR), but visibly different errors.

We show that the FWVDP reliably predicts the fixation areas of interest where error is masked, due to high image contrast, and the areas where the error is visible, due to low image contrast. The paper also suggests ways in which the FWVDP can be used to determine a visually optimal quantization strategy for foveation-based wavelet coefficients and to produce a quantitative local measure of image quality.

***Keywords*—**Human Visual System, Image Quality, Image Compression, foveation wavelet, region of interest ROI.

I. INTRODUCTION

THE psychovisual experiments demonstrates that spatially, the resolution, or sampling density, has the highest value at the point of the fovea and drops rapidly away from that point as a function of eccentricity. As a result, when a human observer gazes at a point in a real-world image, the region around the point of fixation is projected into the fovea, sampled with the highest density and perceived with the highest contrast sensitivity. In conclusion the sampling density and contrast sensitivity decrease dramatically with increasing the viewing angle namely called eccentricity with respect of that point of fixation.

The motivation behind *foveation image compression scheme* is that there exists considerable high-frequency information redundancy in the peripheral regions, so much more efficient representation of images can be obtained by removing or reducing such information redundancy, based on the foveation point(s) and the viewing distances [1-3]. The first aim of that scheme is *foveation filtering*, which foveate a uniform resolution image, such that when the human eyes gaze at the point of fixation, they cannot distinguish between the original and the foveated versions of that image. If attention is focused at the central foveation point, both of images have the same appearance (depending on the viewing distance). In order to evaluate or compare image compression techniques we need to reliably measure the quality of compressed images by taking into account the famous observer mean opinion score (MOS). Many mathematical measures are often used such as mean squared error (MSE) and peak signal to noise ratio (PSNR). However, these measures often have a poor correlation with MOS and functions, that take advantage of properties of the human visual system (HVS), are often incorporated to improve their performance [4]. Recently, techniques based on multiple channel models of the HVS have been shown to improve correlation to MOS [5]. From these HVS models it is possible to predict, on a pixel by pixel basis, if the noise introduced in the compressed image will be visible to a human observer. The VDP [6] map inspired on HVS criteria provides an indication of the degree of visual error as a function of image location. The wavelet transform is one of the most powerful techniques for image compression [7], because of its similarities to the multiple channel models of the HVS. The DWT [8], [9] decomposes the image into a limited number of spatial frequency channels, with respect to the cortical decomposition. Despite this limitation the quality measure still a goal of the wavelet visible difference predictor (WVDP) [10] to visually optimize image compression scheme.

In this paper, we propose a new objective metric adapted to foveation based image coding quality measurement, which exploit a foveation setup filter implementation based on the point of fixation and region of interest ROI. This metric exploits the human visual system quality criteria (HVS) to construct a foveation matrix mask used weighting original and degraded images. The algorithm aims to determine the regions of interest in the wavelet domain. First it decomposes the images into DWT domain and yields a probability error detection using a Minkowski probability summation. Then to focus the quality evaluation in these regions, it masks the wavelet coefficients using the foveation mask which eliminates all peripheral errors and keeps those concentrated in that regions. Finally it maps the probability detection map and displays the foveation probability scale FPS used for wavelet foveation based image coding quality measurement.

LMRIAF, UFR ACSYS: University Med V and INPT Institute, Rabat Morocco.

II. Wavelet Visible Difference Predictor

The first part, in the algorithm of our system foveation-based quality measure scheme shown in Fig. 1, is composed of 4 stage functions, respectively wavelet decomposition, Error wavelet decomposition, Detection Probability and a Watson Error Sensitivity as a final step. The original and noisy images are first transformed to the wavelet domain using a 5 level decomposition. The second part, aims to adapt the measure to images compressed under foveation aspect, is foveation setup and localized region of interest measure to finally map the errors using the minkowski summation algorithm.

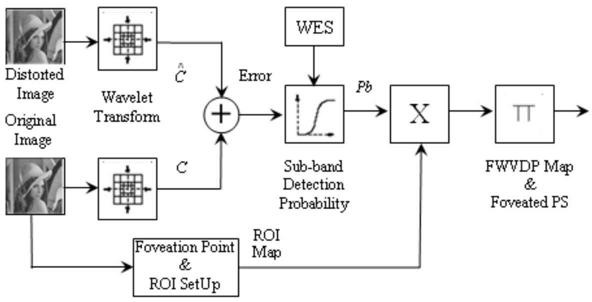


Fig. 1 Foveation Based Image Compression Quality Measure Scheme

As shown in Fig. 3 the differences between them are tested against a masking, or threshold elevation function. The masking function has two parts; the first is the minimum threshold when there is little or no image contrast (no masking). The second is an increasing function of image contrast that defines the effect of contrast masking.

The complete masking function is determined from psychovisual threshold visibility of wavelet coefficients noise added to wavelet coefficients with respect to level and orientation decomposition introduced by Watson model in [11], and from the masking effect due to the contrast in the sub-band. As shown in Fig. 2, thresholds increase rapidly with wavelet spatial frequency, and with orientation from lowpass to horizontal/vertical to diagonal.

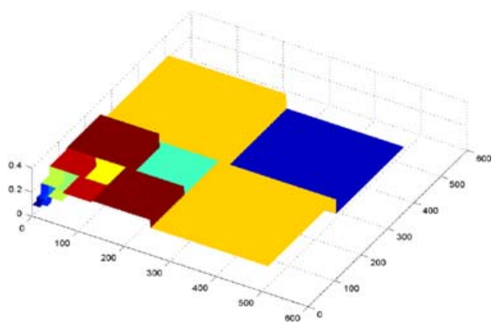


Fig. 2 Foveation Filter error sensitivity mask for viewing distance V = 3.58

This means that the minimum threshold is a function of frequency level and orientation only, while the masking effect is also a function of the actual value of the wavelet coefficients. The amount of masking that occurs in a compressed image depends both on the original image content, and also to the compressed image content due to compression effect. For example, if the original image contained a highly textured (busy) area, one might assume that we could expect a significant masking effect to occur in that area. However, if the compression scheme effectively smoothes that busy area, then this assumption would be incorrect and little, or no, visual masking would occur. In addition, if a smooth area in the original is made highly textured by the compression scheme, then again no visual masking should occur. This effect is called mutual masking and is usually accounted for by taking the minimum in each band for both the original and noisy thresholds. In this way only areas that are highly textured in both the original and compressed images produce a significant masking effect. Once the minimum threshold elevation has been calculated, it is used in conjunction with the coefficient differences to calculate a detection probability for each coefficient in each sub-band. A psychometric function then converts these differences, as a ratio of the threshold elevation, to sub-band detection probabilities Fig. 3.

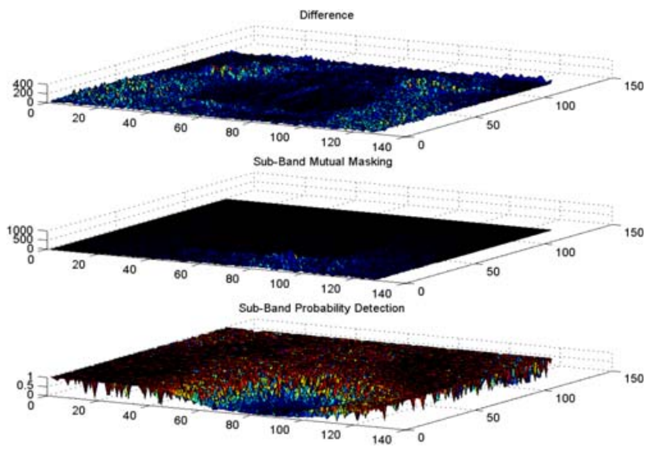


Fig. 3 Probability detection, mutual masking and difference at level 2

The psychometric function calculates the probability of detecting a visible difference in each sub-band, for each coefficient in the DWT. The final output of the WVDP is a probability map, i.e., the detection probability at each pixel in the image. Therefore, the probability of detection in each sub-band (channel) must be combined for every spatial location in the image, using a minkowski probability summation of all channel's detection probabilities.

The problem with the first part of the algorithm is its global measure. However, the foveation-based compression removes all peripheral frequencies from the region of eye's fixation; therefore this compression requires a localized quality measurement centered at the point of fixation. This approach makes the principle goal of the second part of the algorithm and will be detailed in the following paragraph.

III. Foveation Setup & Quality Measure

The motivation behind foveation image compression [12-14] scheme is that there exists considerable high-frequency information redundancy in the peripheral regions, so much more efficient representation of images can be obtained by removing or reducing such information redundancy, based on the foveation point(s) and the viewing distances. The first aim of that scheme is *foveation filtering*, which foveate a uniform resolution image, such that when the human eyes gaze at the point of fixation, they cannot distinguish between the original and the foveated versions of that image. The fixation point determines the foveation sensitivity mask to weight the decomposed image; as a result all frequencies around the region of interest will be either reduced or removed from the image spectrum. This operation is obtained using the foveation filter mask. One feature of the foveation filter is its modification of the spectrum occupation depending on the viewing observation distance as shown in Fig. 4.

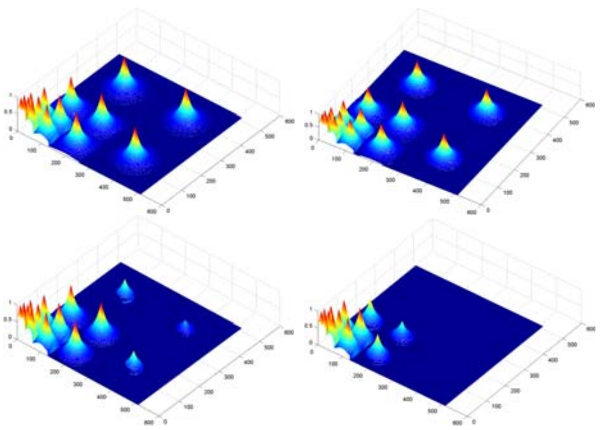


Fig. 4 Foveation Filter error sensitivity mask in the DWT domain. The top-left, top-right, bottom-left, and bottom-right figures are for viewing distance V = 1, 3, 6 and 10

This foveation filter mask depends on many essential parameters like the display Nyquist and the cut-off Frequencies. The first one express the visible frequencies towards the fixation region of interest, the second one show limit of visible frequencies without display aliasing in human visual cortex. The minimum of them determinates final visible frequency spectrum in the area of interest. These frequencies are weighted using the contrast sensitivity function to form the foveation filter. This shape progressively eliminates higher frequencies with an increasing observation distance. In other words its shape leaves high frequency region and move lower one. As a result, the human observer is progressively unable to detect high frequencies in the image when distance increases.

Finally, in the second part of our quality metric; we use the foveation filter mask to setup the region of interest depending on the point of fixation in the world digital image. The masking operation effect removes all peripheral regions errors in DWT domain made by the first part of the scheme. As a result only the focused error areas will be taken into account in our quality measure processing.

One illustrative demonstration is given in Fig. 5 where we show the original errors in left, the region of interest in the right and the localized probabilities detection in the bottom.

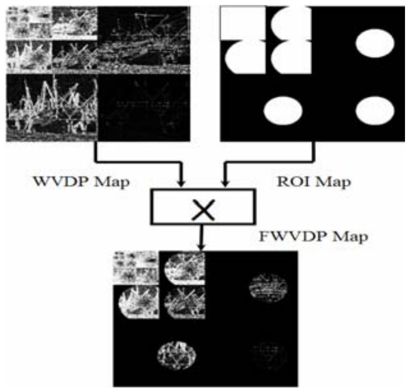


Fig. 5 Errors probabilities detection (left top), ROI (right top) and localized errors probabilities detection (bottom)

The channel probability maps are combined in the limited memory probability summation stage. In this stage, the channel probability maps from higher levels of the pyramid are upsampled to match the resolution of the lowest pyramid level and then lowpass filtered to remove sharp edges. The probability maps are then combined using a method we call limited memory probability summation. The method uses probability summation to combine the N largest visual channel responses at each spatial location. At a given spatial location, the N largest visual channel responses are combined according to the largest probability of detecting a difference in a single visual channel. In Fig. 6, we show errors map.

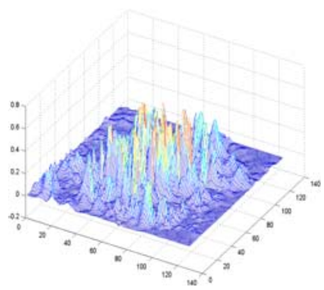


Fig. 6 Local errors map in the Region Of Interest ROI (Lena image)

Summing probability errors of each DWT perceptual channel produce the same probability score factor PS. There are two different approaches; the first is global summation of all DWT errors which yields the factor PS, the second is regions of interest summation of only fixation point region errors yielding FPS factor. As a basic remark the global measure is inefficient as the probability score PS nearest 1, which makes the assessment really difficult. The local probability score gives much more significant judgment with respect to viewing conditions.

IV. Results and Discussions

Hard studies have been done in the DWT domain using our new quality metric based region of interest, with respect to varying observation distances and bite rates, most useful test images, and a special fixation points (Smooth and textured regions) parameters. For this purpose our new quality metric based region of interest is used to predict the visual differences between the original and foveated-based images as illustrated in figure Fig. 7 (fixation point is at the Boat name).

Fig. 7 Boat original (left) and Boat foveated (right) images

The comparison of compression quality performances between the original and foveation-based images is done by evaluating their foveation probability score factor FPS. This factor is obtained by summing all probability errors in the DWT after eliminating all non interesting regions by multiplying the error map by the DWT regions of interest map plotted in figure Fig. 8.

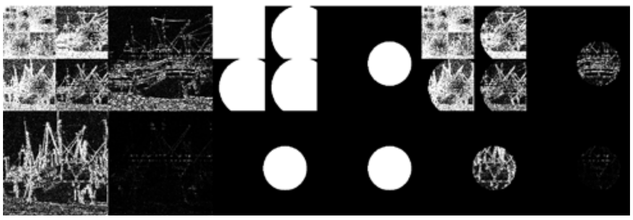


Fig. 8 Errors probabilities detection (left), ROI (middle) and localized errors probabilities detection (right)

A Minkowski summation is then run to process the probability error detection map from the regions of interest map. The last map is then used to provide the probability scores PS (global quality measure) and FPS (local quality measure). The global and local probability detection maps are plotted in Fig. 9.

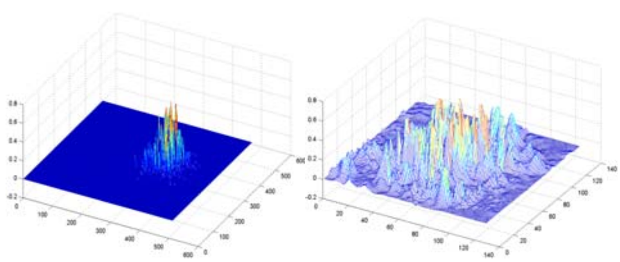
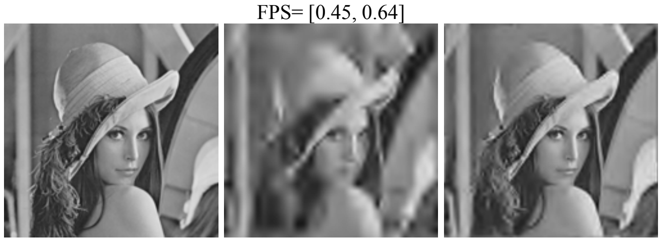
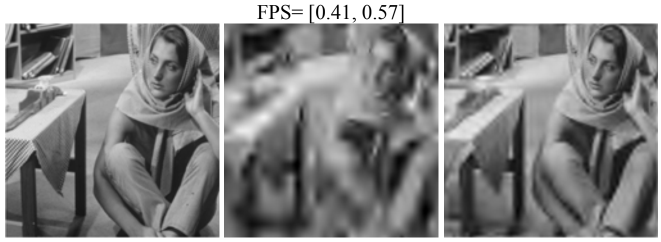


Fig. 9 Global errors map (Left), and ROI errors map (Right)

FPS= [0.45, 0.64]

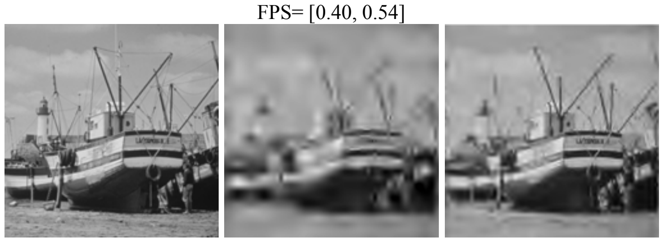


FPS= [0.41, 0.57]

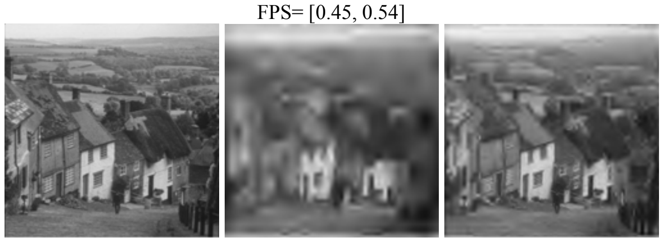


FPS= [0.59, 0.72]

FPS= [0.40, 0.54]



FPS= [0.45, 0.54]



FPS= [0.42, 0.49]

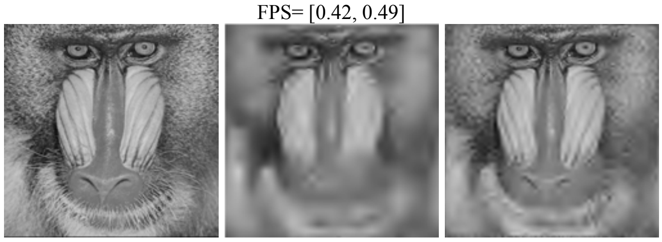


Fig. 10 Foveation Probability Score FPS Evaluation for Lena, Barbara, Zelda, Boat, Goldhill and Mandrill test images for a viewing distance of V = 4. The FPS factors quality are given for each test image (left column: original images) coded at a bit rate of 0.03125 bpp (middle column: low bit rate) and 0.125 bpp (right column: high bit rate)

Our new metric is compared to a number of more conventional measures of image quality such as mean squared error MSE and peak signal to noise ratio PSNR for many Test Images. These comparisons have shown that our objective quality measure correlates well with the mean opinion score MOS which is the average human quality decision and image coding performance judgment.

Traditional metrics assess the quality over the whole image witch is not compatible to foveated images. As well known and explained before, the foveation filtering removes or reduces all high peripheral frequencies that are out of the region of interest ROI. So as the SPIHT quality is spread at the whole image, then the comparison is not useful in peripheral region of the fixation point. As a basic remark the global measure is inefficient as the quality factor is always nearest 1 for both compression techniques, which makes the assessment really difficult. The local probability score gives much more significant assessment with respect to viewing conditions. The concluded idea is that the global image compression quality measure based-foveation is inefficient.

This urges us to focus locally the image quality assessment around the gazed point. The model evaluation is Foveation Quality Wavelet Index will thereafter be called FQWI metric. The latter yields a foveation probability score named FPS that varies from 0 to 1 corresponding to a variation scale factor from bad to high image coding quality.

The high this factor is the high the quality is and the threshold is pointed around a value of 0.35 under which the distortion is perceptible to human eyes.

Figure 10 illustrates the foveation image coding for different images with respect to a given bit rates 0.03125bpp and 0.125bpp and a viewing condition of V=4. In Fig. 10 the left column corresponds to the original, the medium column corresponds to the degraded images at 0.03125bpp coding bit rate and the right column corresponds to degraded images at 0.125bpp coding bit rate.

V. Conclusion

In this paper, we propose an optimized foveation-based image quality measure algorithm, which exploit an adapted foveation filter implementation based on the point and region of fixation, by exploiting the human visual system quality criteria (HVS), to evaluate a quality. This quality evaluation is based especially on a Foveation probability scale named FPS, which provides a good tool by opposite to Peak Signal Noise Ratio PSNR which still approximately invariant for many compressed foveated images with respect to different points of fixation.

To achieve this paper, note that our compression and quality evaluation systems makes a great part of a great project concerning the real time video coding and quality assessing in a wireless GSM networks infrastructure. These systems will be incorporated to provide the final scheme.

References

[1] E.-C. Chang and C. Yap, "A wavelet approach to foveating images" in *Proc. 13th ACM Symp. Computational Geometry*, 1997, pp. 397–399.

[2] E.-C. Chang, "Foveation techniques and scheduling issues in thin wire visualization," Ph.D. dissertation, New York Univ., 1998.

[3] E.-C. Chang, S. Mallat, and C. Yap. (1999) Wavelet Foveation. [Online]. Available: http://www.cs.nyu.edu/visual/

[4] S. J. P. Westen, R. L. Lagendijk and J. Biemond, "Perceptual Image Quality based on a Multiple Channel HVS Model," Proceedings of ICASSP, pp. 2351-2354, 1995

[5] C. Zetzsche and G. Hauske, "Multiple Channel Model Prediction of Subjective Image Quality," SPIE, Human Vision, Visual Processing, and Display, 1077, pp. 209-215, 1989.

[6] S. Daly, "The visible differences predictor: An algorithm for the assessment of image fidelity," in Digital Images and Human Vision (A. B. Watson, ed.), pp. 179-205, Cambridge, MA: MIT Press, 1993W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[7] D. J. Heeger and P. C. Teo, "A model of perceptual image fidelity," Proc. Of IEEE Int'l Conf. on Image Proc., Oct. 23-26 1995, Washington, D.C., USA, pp. 343-345.

[8] S. Mallat "A wavelet Tour of Signal and Image Processing". Edition 2. 2002.

[9] Marc Antonini, Michel Barlaud, Pierre Mathieu, and Ingrid Daubechies, "Image Coding Using Wavelet Transform," IEEE Transaction on Image Processing, vol. 1, no. 2, pp. 205–220, April 1992.

[10] Andrew P. Bradley "A Wavelet Visible Difference Predictor" Member IEEE Transactions on Image Processing Vol. 8. No. 5. May 1999.

[11] Andrew B. Watson, Gloria Y. Yang, Joshua A. Solomon, and John Villasenor "Visibility of Wavelet Quantization Noise" Member IEEE Transactions on Image Processing VOL. 6, NO. 8, AUGUST 1997.

[12] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video quality assessment," *IEEE Trans. Multimedia*, vol. 3, 2001.

[13] "Foveated video compression with optimal rate control," *IEEE Trans. Image Processing*, vol. 10, pp. 977–992, July 2001.

[14] "Rate control for foveated MPEG/H.263 video," in *IEEE ICIP*, vol. 2, 1998.

M. Nahid Studied electronics at ENSET (Ecole Normale Supérieure de l'Enseignement Technique) Institute of Mohammedia, Morocco in 1994, received the Aggregation diploma of Electronics Engineering, from ENSET Institute of Rabat, Morocco in 2000.
He received the DESA (Diplôme des Etudes Supérieures Approfondies) diploma in Computer Engineering, Telecoms and Multimedia from Med V University of Rabat in 2004, and conducted doctoral research at Images and Videos Coding and Quality Assessment under Psychovisual Quality criteria since 2004.

A. Bajit Studied electronics at ENSET (Ecole Normale Supérieure de l'Enseignement Technique) Institute of Rabat, Morocco in 1993, received the Aggregation diploma of Electronics Engineering, from ENSET Institute of Rabat, Morocco in 2000.
He received the DESA (Diplôme des Etudes Supérieures Approfondies) diploma in Computer Engineering, Telecoms and Multimedia from Med V University of Rabat in 2004, and conducted doctoral research at Images and Videos Coding and Quality Assessment under Psychovisual Quality criteria since 2004.