WAVELET-BASED FOVEATED IMAGE QUALITY MEASUREMENT FOR REGION OF INTEREST IMAGE CODING

Zhou Wang[1], *Alan C. Bovik*[1], *and Ligang Lu*[2]

[1]Laboratory for Image and Video Engineering (LIVE), Dept. of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084, USA
[2]Video and Image Systems, IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, USA
(zwang@ece.utexas.edu, bovik@ece.utexas.edu, lul@us.ibm.com)

ABSTRACT

Region of interest (ROI) image and video compression techniques have been widely used in visual communication applications in an effort to deliver good quality images and videos at limited bandwidths. Most image quality metrics have been developed for uniform resolution images. These metrics are not appropriate for the assessment of ROI coded images, where space-variant resolution is necessary. The spatial resolution of the human visual system (HVS) is highest around the point of fixation and decreases rapidly with increasing eccentricity. Since the ROIs are usually the regions "fixated" by human eyes, the foveation property of the HVS supplies a natural approach for guiding the design of ROI image quality measurement algorithms. We have developed an objective quality metric for ROI coded images in the wavelet transform domain. This metric can serve to mediate the compression and enhancement of ROI coded images and videos. We show its effectiveness by applying it to an embedded foveated image coding system.

1. INTRODUCTION

Region of interest (ROI) image coding allows the assignment of more bits to the ROIs than other parts of the image. It is a useful tool for visual communication applications where the available bandwidth is limited. While there has been a large amount of work in uniform resolution image quality measurement, little has been done in the assessment of ROI coded images. Quality assessment method plays an important role in ROI image coding, because image coding is essentially an optimization procedure that maximizes the image quality with a limited number of bits, where the quality metric serves as a guide for bit assignment. The development of ROI image quality metrics is also very important for the postprocessing or quality enhancement of ROI coded images. However, uniform resolution image quality measurement approaches such as peak signal-to-noise ratio (PSNR) are still inappropriately used for the evaluation of ROI image coding and postprocessing [1, 2].

The motivation of this work is that the human visual system (HVS) is highly space-variant in sampling, coding, processing and understanding. The spatial resolution of the HVS is the highest around the point of fixation (foveation point) and decreases rapidly with increasing eccentricity. This feature delivers a natural way to define an image quality measure for the case that the human eyes are fixating at a given point in the image. For example, a foveated PSNR (F-PSNR) metric was proposed in [3] for foveated video compression. By thinking of the ROIs as collections of pixels that are possibly "fixated" by human eyes, a natural ROI image quality metric can be designed that utilizes a foveation model of the HVS.

In this paper, we develop a foveation-based HVS model in the discrete wavelet transform (DWT) domain because wavelet analysis supplies a convenient way to simultaneously examine localized spatial as well as frequency information. A new image quality metric called the foveated wavelet image quality index (FWQI) is then defined for ROI coded images.

2. FOVEATED WAVELET IMAGE QUALITY MEASUREMENT

The photoreceptors (cones and rods) and ganglion cells are non-uniformly distributed in the retina in the human eye [4]. The density and sensitivity of cone receptors and ganglion cells play important roles in determining the ability of our eyes to resolve what we see. The resolution is the highest around the foveation point and decreases dramatically with increasing eccentricity. Psychological experiments had been conducted to measure the contrast sensitivity as a function of retinal eccentricity [5-7]. In [5], a model that fits the experimental data was given by

This research is supported in part by IBM Corporation, Texas Instrument, Inc., and Texas Advanced Technology Program.

$$CT(f,e) = CT_0 \exp\left(\alpha f \frac{e+e_2}{e_2}\right), \quad (1)$$

where f is the spatial frequency (cycles/degree), e is the retinal eccentricity (degrees), CT_0 is a constant minimal contrast threshold, α is the spatial frequency decay constant, e_2 is the half-resolution eccentricity, and $CT(f, e)$ is the visible contrast threshold as a function of f and e. The best fitting parameter values given in [5] are α = 0.106, e_2 = 2.3, and CT_0 = 1/64. It was also reported in [5] that the same values of α and e_2 provide a good fit to the data in [6] with CT_0 = 1/75, and an adequate fit to the data in [7] with CT_0 = 1/76, respectively. We use the parameter selections as in [5]. The contrast sensitivity is defined as:

$$CS(f,e) = \frac{1}{CT(f,e)}. \quad (2)$$

For a given e, equation (1) can be used to find its critical frequency or so called cutoff frequency f_c by setting CT to 1.0 (the maximum possible contrast) and solving for e

$$f_c = \frac{e_2 \ln(1/CT_0)}{(e+e_2)\alpha} \text{ (cycles/degree).} \quad (3)$$

Given a pixel $\boldsymbol{x}$ in an N pixels wide image, its distance from the foveation point $\boldsymbol{x}_f$ is $d(\boldsymbol{x}) = \left\|\boldsymbol{x} - \boldsymbol{x}^f\right\|_2$ (pixels) and its eccentricity is given by $e(v, \boldsymbol{x}) = \tan^{-1}\left(d(\boldsymbol{x})/Nv\right)$, where v is the viewing distance in image width. In Fig. 1, we show the normalized contrast sensitivity as a function of pixel position for N = 512 and v = 1 and 6, respectively. f_c as a function of pixel position is also given. The CS is normalized so that the highest value is 1.0 at 0 eccentricity. The maximum perceived resolution is also limited by the display resolution $r \approx \pi Nv/180$ (pixels/degree). The Nyquist display frequency is given by $f_d = r/2$ (cycles/degree). Combining this with (3), the cutoff frequency for $\boldsymbol{x}$ is $f_m(\boldsymbol{x}) = \min(f_c(d(\boldsymbol{x})), f_d)$. We define the foveation-based error sensitivity as:

$$S_f(v,f,\boldsymbol{x}) = \begin{cases} \dfrac{CS(f,e(v,\boldsymbol{x}))}{CS(f,0)} & for \quad f \le f_m(\boldsymbol{x}) \\ 0 & for \quad f > f_m(\boldsymbol{x}) \end{cases}. \quad (4)$$

Now let us consider the wavelet transforms. In the 1-D DWT, the input discrete signal $\boldsymbol{s}$ is convolved with highpass and lowpass analysis filters and downsampled by two, resulting in transformed signals $\boldsymbol{s}_H$ and $\boldsymbol{s}_L$. The signal $\boldsymbol{s}_L$ can be further decomposed and the process may be repeated multiple times. The number of repetitions defines the wavelet decomposition level λ. For image processing, the horizontal and vertical wavelet decompositions are applied alternatively, yielding LL, HL, LH and HH subbands. The LL subband may be further decomposed and the process repeated multiple times. Let (λ,θ) represent the subband of level λ and orientation θ, where θ is an index representing the LL, LH, HH or HL subband. The wavelet coefficients at different subbands supply information of variable perceptual importance. In [8], psychovisual measurement results were given for the visual sensitivity in wavelet decompositions. A model that fits the experimental data is $\log Y = \log a + k(\log f - \log g_\theta f_0)^2$ [8], where Y is the visually detectable noise threshold and $f = r2^{-\lambda}$ [8] is the spatial frequency. The visual sensitivity in subband (λ,θ) is given by:

$$S_W(\lambda,\theta) = \frac{A_{\lambda,\theta}}{Y_{\lambda,\theta}} = \frac{A_{\lambda,\theta}}{A_{\lambda,\theta} a 10^{k(\log(2^\lambda f_0 g_\theta / r)^2}}, \quad (5)$$

where $A_{\lambda,\theta}$ is the basis function amplitude given in [8]. Let $\boldsymbol{B}_{\lambda,\theta}$ denote the set of wavelet coefficient positions residing in subband (λ,θ). For each $\boldsymbol{B}_{\lambda,\theta}$, we calculate the corresponding foveation point $\boldsymbol{x}^f_{\lambda,\theta}$ in it. Given a wavelet coefficient at $\boldsymbol{x} \in \boldsymbol{B}_{\lambda,\theta}$, its equivalent distance from the foveation point in the spatial domain is given by $d_{\lambda,\theta}(\boldsymbol{x}) = 2^\lambda \left\|\boldsymbol{x} - \boldsymbol{x}^f_{\lambda,\theta}\right\|_2$. With this distance, we have

$$S_f(v,f,\boldsymbol{x}) = S_f(v, r2^{-\lambda}, d_{\lambda,\theta}(\boldsymbol{x})) \; for \; \boldsymbol{x} \in \boldsymbol{B}_{\lambda,\theta}. \quad (6)$$

A foveation-based visual sensitivity model in the DWT domain is obtained by combining (5) and (6):

$$S(v,\boldsymbol{x}) = \left[S_w(\lambda,\theta)\right]^{\beta_1} \cdot \left[S_f(v, r2^{-\lambda}, d_{\lambda,\theta}(\boldsymbol{x}))\right]^{\beta_2} \quad for \; \boldsymbol{x} \in \boldsymbol{B}_{\lambda,\theta}, \quad (7)$$

where β_1 and β_2 are parameters used to control the magnitudes of S_w and S_f, respectively. We use β_1 = 1 and β_2 = 2.5. Fig. 2 shows $S(v,\boldsymbol{x})$ for v = 1, 3, 6 and 10, respectively. For the evaluation of image quality, instead of using the traditional error summation methods, we designed a new quality index [9] by modeling any signal distortion as a combination of three factors: loss of correlation, mean distortion and variance distortion. For any 2-D signal, the measurement results are a 2-D quality map as well as an overall quality index. Readers can refer to [9] and *http://anchovy.ece.utexas.edu/~zwang/research/quality_index/demo.html* for more details and demonstrative images of the new quality index. In this paper, we adapt the index into the DWT domain and define a foveated wavelet image quality index (FWQI) as:

$$FWQI = \frac{\sum_{n=1}^{M} S(v,\boldsymbol{x}_n) \cdot |c(\boldsymbol{x}_n)| Q(\boldsymbol{x}_n)}{\sum_{n=1}^{M} S(v,\boldsymbol{x}_n) \cdot |c(\boldsymbol{x}_n)|}, \quad (8)$$

where M is the number of the wavelet coefficients, $c(\boldsymbol{x}_n)$ is the wavelet coefficient of the original image at location $\boldsymbol{x}_n$, and $Q(\boldsymbol{x}_n)$ is the quality value at location $\boldsymbol{x}_n$ in the quality index map. Since $S(v,\boldsymbol{x}_n)$ varies with v, FWQI of an test image is a function of v, instead of a single value.

The above model is developed for the case of a single

foveation point. We consider the ROIs as the groups of possibly fixated pixels. This corresponds to the case of multiple foveation points. Our model can easily adapt to this case. Suppose that there are P foveation points in the image, with $S_i(v, \boldsymbol{x})$ for $i = 1,2,\cdots,P$, then the overall error sensitivity should be given by the maximum value of them: $S(v, \boldsymbol{x}) = \max_{i=1\cdots P}(S_i(v, \boldsymbol{x}))$.

3. IMAGE CODING USING THE FOVEATED QUALITY METRIC

SPIHT [10] is a very efficient progressive wavelet image coding algorithm. We designed a modified SPIHT algorithm and tuned it using the above FWQI model to optimize the foveated visual quality at any given bit rate. We call the new coding algorithm the embedded wavelet image coding (EFIC) algorithm [11]. The encoded bitstream can be truncated at arbitrary places to create reconstructed images with different quality and depth of foveation. Fig. 3 gives the FWQI comparison of the EFIC and SPIHT compressed 8bits/pixel (bpp) "Zelda" images at 0.015265, 0.0625 and 0.25bpp, respectively. FWQI for each image is given as a function of the viewing distance. It can be observed that significant quality gain is achieved throughout the whole range of the viewing distances. Fig. 4 shows the SPIHT and EFIC decoded images. Compared with SPIHT, EFIC provides better foveated visual quality. When sufficient bit rate is available, the EFIC coded image approaches uniform resolution.

4. CONCLUSIONS

We propose a foveation-based sophisticated ROI image quality metric in the wavelet transform domain. This metric can serve as a very useful tool for foveated ROI image coding and quality enhancement.

REFERENCES

[1] E. Atsumi, and N. Farvardin, "Lossy/lossless region-of-interest image coding based on set partitioning in hierarchical trees," *IEEE International Conference on Image Processing*, vol. 1, pp. 87-91, 1998.

[2] J. Jung, S. Joung, Y. Jang, and J. Paik, "Enhancement of region-of-interest coded images by using adaptive regularization," *IEEE International Conference on Consumer Electronics*, pp. 62-63, 2000.

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

[4] W. S. Geisler, *Vision Notes*, The University of Texas at Austin, 1999.

[5] W. S. Geisler, and J. S. Perry, "A real-time foveated multiresolution system for low-bandwidth video communication," *Proc. of SPIE*, vol. 3299, 1998.

[6] T. L. Arnow, and W. S. Geisler, "Visual detection following retinal damage: predictions of an inhomogeneous retino-cortical model," *Proceedings of SPIE*, vol. 2674, pp. 119-130, 1996.

[7] M. S. Banks, A. B. Sekuler, and S. J. Anderson, "Peripheral spatial vision: limits imposed by optics, photoreceptors, and receptor pooling," *Journal of the Optic. Soc. of America*, vol. 8, pp. 1775-1787, 1991.

[8] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Trans. Image Processing*, vol. 6, no. 8, pp. 1164-1175, Aug. 1997.

[9] Z. Wang and A. C. Bovik, "A universal image quality index," submitted to *IEEE Signal Proc. Letters*, 2001.

[10] A. Said, and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits & Systems for Video Tech.*, vol. 6, no. 3, pp. 243-250, June 1996.

[11] Z. Wang and A. C. Bovik, "Embedded foveation image coding," submitted to *IEEE Trans. Image Processing*, 2001.

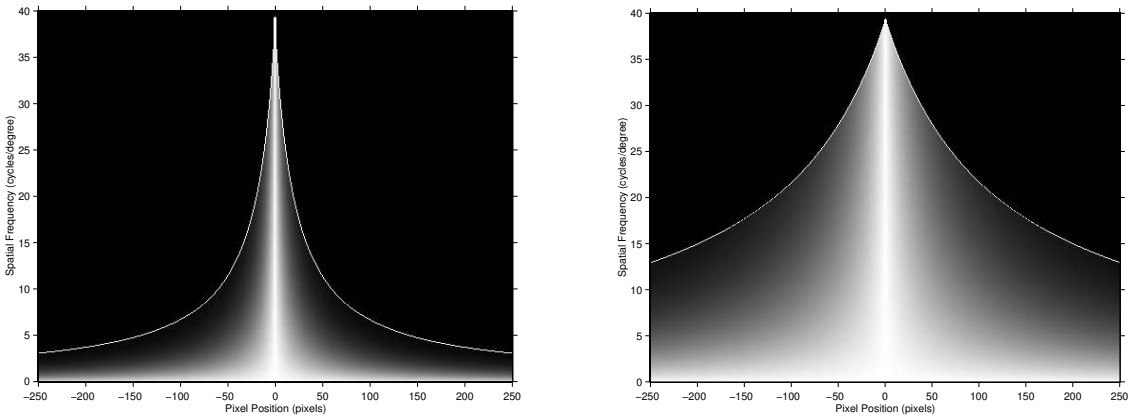

Fig. 1 Normalized contrast sensitivity (Brightness indicates the strength of contrast sensitivity) for $N = 512$ and $v = 1$ (Left) and $v = 6$ (Right) times of the image width, respectively. The white curves show the cutoff frequencies.

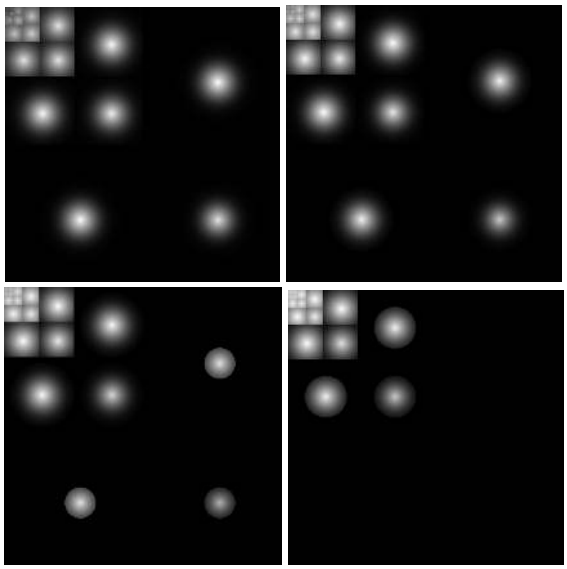

Fig. 2 Foveation-based error sensitivity mask in the DWT domain. The top-left, top-right, bottom-left, and bottom-right are for v = 1, 3, 6 and 10 times of the image width, respectively. (Brightness logarithmically enhanced for display purpose)

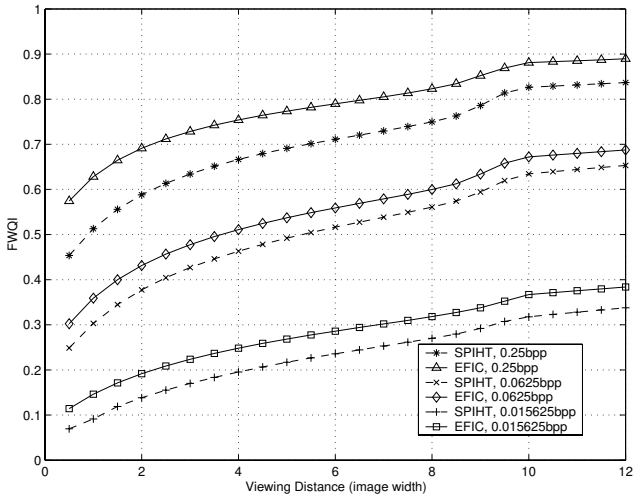

Fig. 3 FWQI comparison of EFIC and SPIHT compressed "Zelda" image (512×512, 8bpp) at 0.015625bpp, 0.0625bpp and 0.25bpp.

Fig. 4 "Zelda" image (512×512, 8bpp) compression result comparison at 0.015625 (CR=512:1), 0.0625 (CR=128:1) and 0.25bpp (CR=32:1). Upper: SPIHT coded images; Bottom: EFIC coded images.