χSET: Image Coder based on Contrast Band-Pass Filtering

Jaime Moreno†‡, Oswaldo Morales†, and Ricardo Tejeida†
†Superior School of Mechanical and Electrical Engineers, National Polytechnic Institute of Mexico, IPN Avenue, Lindavista, Mexico City, 07738, Mexico.
‡Signal, Image and Communications Department, University of Poitiers, Poitiers, 30179, France. e-mail:jmorenoe@ipn.mx

Abstract—Noise is fatal to image compression performance because it can be both annoying for the observer and consumes excessive bandwidth when the imagery is transmitted [1], [2]. Some noise, in addition to some numerical redundancy, is removed during the quantization process, but in some circumstances the removed information is easily perceived by the observer, leading to annoying visual artifacts. Perceptual quantization reduces unperceivable details and thus improve both visual impression and transmission properties. In this work, we apply perceptual criteria in order to define a perceptual forward and inverse quantizer. It is based on the CBPF, a low-level computational model that reproduces color perception in the Human Visual System. Our approach consists in performing a local quantization of wavelet transform coefficients using some of the human visual system behavior properties. It is performed applying a local weight for every coefficient. The CBPF allows to recover these weights from the quantized data, which avoids the storing and transmission of these weights. We apply this perceptual quantizer to the Hi-SET coder. The comparison between JPEG2000 coder and the combination of Hi-SET with the proposed perceptual quantizer (χSET) is shown. The latter produces images with lower PSNR than the former, but they have the same or even better visual quality when measured with well-known image quality metrics such as MSSIM, UQI or VIF, for instance. Hence, χSET obtain more compressed (i.e. lower bit-rate) images at the same perceptual image quality than JPEG2000. Keywords: Human Visual System, Contrast Sensitivity Function, Perceived Images, Wavelet Transform, Peak Signal-to-Noise Ratio, No-Reference Image Quality Assessment, JPEG2000.

1. Introduction

Digital image compression has been a research topic for many years and a number of image compression standards has been created for different applications. The JPEG2000 is intended to provide rate-distortion and subjective image quality performance superior to existing standards, as well as to supply functionality [3]. However, JPEG2000 does not provide the most relevant characteristics of the human visual system, since for removing information in order to compress the image mainly information theory criteria are applied. This information removal introduces artifacts to the image that are visible at high compression rates, because of many pixels with high perceptual significance have been discarded. Hence, it is necessary an advanced model that removes information according to perceptual criteria, preserving the pixels with high perceptual relevance regardless of the numerical information. The Chromatic Induction Wavelet Model presents some perceptual concepts that can be suitable for it. Both CBPF and JPEG2000 use wavelet transform. CBPF uses it in order to generate an approximation to how every pixel is perceived from a certain distance taking into account the value of its neighboring pixels. By contrast, JPEG2000 applies a perceptual criteria for all coefficients in a certain spatial frequency independently of the values of its surrounding ones. In other words, JPEG2000 performs a global transformation of wavelet coefficients, while CBPF performs a local one. CBPF attenuates the details that the human visual system is not able to perceive, enhances those that are perceptually relevant and produces an approximation of the image that the brain visual cortex perceives. At long distances, the lack of information does not produce the well-known compression artifacts, rather it is presented as a softened version, where the details with high perceptual value remain (for example, some edges).

2. Chromatic Induction Wavelet Model: Brief description.

The Chromatic Induction Wavelet Model (CBPF) [4] is a low-level perceptual model of the HVS. It estimates the image perceived by an observer at a distance $d$ just by modeling the perceptual chromatic induction processes of the HVS. That is, given an image $I$ and an observation distance $d$, CBPF obtains an estimation of the perceptual image $I_p$ that the observer perceives when observing $I$ at distance $d$. CBPF is based on just three important stimulus properties: spatial frequency, spatial orientation and surround contrast. This three properties allow to unify the chromatic assimilation and contrast phenomena, as well as some other perceptual processes such as saliency perceptual processes.

The perceptual image $I_p$ is recovered by weighting these $\omega_{n, o}$ wavelet coefficients using the extended Contrast Sensitivity Function (CSF) of the human visual system. The CSF is a non-linear function that models the contrast sensitivity of the human visual system at different spatial frequencies and orientations. The weights $\omega_{n, o}$ are calculated using a set of functions that depend on the specific application and the desired perceptual quality.

The CBPF model uses these weights to perform a local quantization of the wavelet coefficients, which reduces the number of coefficients that are transmitted. This process is repeated for each level of the wavelet decomposition, and the resulting coefficients are used to reconstruct the perceptual image $I_p$. The reconstruction process is performed using a set of inverse wavelet transformations, which take into account the weights $\omega_{n, o}$ and the original coefficients.

The CBPF model has been shown to be effective in reducing the visual artifacts produced by traditional compression methods, while preserving the visual quality of the original image. This makes it a promising candidate for use in applications where high-quality images are required, such as medical imaging and surveillance.

Overall, the CBPF model represents an important contribution to the field of image compression, as it offers a solution to the problem of perceptual quality in compressed images. Its ability to adapt to the specific requirements of different applications makes it a versatile and powerful tool for image compression.
sitivity Function (e-CSF). The e-CSF is an extension of the psychophysical CSF [5] considering spatial surround information (denoted by $r$), visual frequency (denoted by $\nu$), which is related to spatial frequency by observation distance) and observation distance ($d$). Perceptual image $I_p$ can be obtained by

$$I_p = \sum_{s=1}^{n} \sum_{o=v,h,dgl} \alpha(\nu, r) \omega_{s,o} + c_n, \quad (1)$$

where $\alpha(\nu, r)$ is the e-CSF weighting function that tries to reproduce some perceptual properties of the HVS. The term $\alpha(\nu, r) \omega_{s,o}$ is considered the perceptual wavelet coefficients of image $I$ when observed at distance $d$ and is written as:

### 3. JPEG2000 Global Visual Frequency Weighting

In JPEG2000, only one set of weights is chosen and applied to wavelet coefficients according to a particular viewing condition (100, 200 or 400 dpi’s) with fixed visual weighting [3, Annex J.8]. This viewing condition may be truncated depending on the stages of embedding, in other words at low bit rates, the quality of the compressed image is poor and the detailed features of the image are not available since at a relatively large distance the low frequencies are perceptually more important.

The table 1 specifies a set of weights which was designed for the luminance component based on the CSF value at the mid-frequency of each spatial frequency. The viewing distance is supposed to be 4000 pixels, corresponding to 10 inches for 400 dpi print or display. The weight for $LL$ is not included in the table, because it is always 1. Levels 1, 2, ..., 5 denote the spatial frequency levels in low to high frequency order with three spatial orientations, horizontal, vertical and diagonal.

#### Table 1

<table>
<thead>
<tr>
<th>$s$</th>
<th>horizontal</th>
<th>vertical</th>
<th>diagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0.731 668</td>
</tr>
<tr>
<td>3</td>
<td>0.564 344</td>
<td>0.564 344</td>
<td>0.285 968</td>
</tr>
<tr>
<td>4</td>
<td>0.179 609</td>
<td>0.179 609</td>
<td>0.043 903</td>
</tr>
<tr>
<td>5</td>
<td>0.014 774</td>
<td>0.014 774</td>
<td>0.000 573</td>
</tr>
</tbody>
</table>

### 4. Perceptual Forward Quantization

#### 4.1 Forward

Quantization is the only cause that introduces distortion into a compression process. Since each transform sample at the perceptual image $I_p$ (1) is mapped independently to a corresponding step size either $\Delta_s$ or $\Delta_n$, thus $I_p$ is associated with a specific interval on the real line. Then, the perceptually quantized coefficients $Q$, from a known viewing distance $d$, are calculated as follows:

$$Q = \sum_{s=1}^{n} \sum_{o=v,h,dgl} \text{sign}(\omega_{s,o}) \left( \frac{\alpha(\nu, r) \omega_{s,o}}{\Delta_s} \right) + \left\lfloor \frac{c_n \Delta_n}{\Delta_s} \right\rfloor \quad (2)$$

Unlike the classical techniques of Visual Frequency Weighting (VFW) on JPEG2000, which apply one CSF weight per sub-band [3, Annex J.8], Perceptual Quantization using CBPF ($\rho$SQ) applies one CSF weight per coefficient over all wavelet planes $\omega_{s,o}$. In this section we only explain Forward Perceptual Quantization using CBPF (F-$\rho$SQ). Thus, (2) introduces the perceptual criteria of the Perceptual Images (1) to each quantized coefficient of the Dead-zone Scalar Quantizer [3, Annex J.8]. A normalized quantization step size $\Delta = 1/128$ is used, namely the range between the minimal and maximal values at $I_p$ is divided into 128 intervals. Finally, the perceptually quantized coefficients are entropy coded, before forming the output code stream or bitstream.

#### 4.2 Inverse

The proposed Perceptual Quantization is a generalized method, which can be applied to wavelet-transform-based image compression algorithms such as EZW, SPIHT, SPECK or JPEG2000. In this work, we introduce both forward (F-$\rho$SQ) and inverse perceptual quantization (I-$\rho$SQ) into the Hi-SET coder [6], [7], [8]. An advantage of introducing $\rho$SQ is to maintain the embedded features not only of Hi-SET algorithm but also of any wavelet-based image coder. Thus, we call Perceptual Quantization + Hi-SET = PH$i$SET or $\chi$SET.

Both JPEG2000 and $\chi$SET choose their VFWs according to a final viewing condition. When JPEG2000 modifies the quantization step size with a certain visual weight, it needs to explicitly specify the quantizer, which is not very suitable for embedded coding. By contrast, $\chi$SET neither needs to store the visual weights nor to necessarily specify a quantizer in order to keep its embedded coding properties.

The main challenge underlies in to recover not only a good approximation of coefficients $Q$ but also the visual weight $\alpha(\nu, r)$ (Eq. 2) that weighted them. A recovered approximation $\hat{Q}$ with a certain distortion $\Delta$ is decoded from the bitstream by the entropy decoding process. The VFWs were not encoded during the entropy encoding process, since it would increase the amount of stored data. A possible solution is to embed these weights $\alpha(\nu, r)$ into $\hat{Q}$. Thus, our goal is to recover the $\alpha(\nu, r)$ weights only using the information from the bitstream, namely from the Forward quantized coefficients $\hat{Q}$.

The reduction of the dynamic range is uniformly made by the perceptual quantizer, thus the statistical properties of $I$ are maintained in $\hat{Q}$. 
Therefore, our hypothesis is that an approximation $\hat{\alpha}(\nu, r)$ of $\alpha(\nu, r)$ can be recovered applying CBPF to $\hat{Q}$, with the same viewing conditions used in $I$. That is, $\hat{\alpha}(\nu, r)$ is the recovered e-CSF. Thus, the perceptual inverse quantizer or the recovered $\hat{\alpha}(\nu, r)$ introduces perceptual criteria to the Inverse Scalar Quantizer [3, Annex J.8] and is given by:

$$\hat{\alpha}(\nu, r) = \begin{cases} \sum_{x=1}^{n} \sum_{r \in S, h, d} \text{sign}(\omega_{x,r}) \Delta_{x} \left( \frac{|\omega_{x,r}| + \delta}{\hat{\alpha}(\nu, r)} \right), & |\omega_{x,r}| > 0 \\ + (|\hat{\alpha}| + \delta) \cdot \Delta_{\hat{\alpha}}, & |\omega_{x,r}| = 0 \end{cases}$$

For the sake of showing that the encoded VFWs are approximately equal to the decoded ones, that is $\alpha(\nu, r) \approx \hat{\alpha}(\nu, r)$, we perform the following experiment:

We employ the process shown in Fig. 1(a) for all the images of the CMU [9], CSIQ [10] and IVC [11] Image Databases. We chose for evaluating these assessments the implementation provided in [12], since it is based on the parameters proposed by the author of each indicator. In order to obtain $\hat{\alpha}(\nu, r)$, we measure the lineal correlation between the original $\alpha(\nu, r)$ applied during the F-$\rho$SQ process and the recovered $\hat{\alpha}(\nu, r)$. Table 2 shows that there is a high similarity between the applied VFW and the recovered one, since their correlation is 0.9849, for gray-scale images, and 0.9840, for color images.

<table>
<thead>
<tr>
<th>Image Database</th>
<th>8 bpp gray-scale</th>
<th>24 bpp color</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMU</td>
<td>0.9840</td>
<td>0.9857</td>
</tr>
<tr>
<td>CSIQ</td>
<td>0.9857</td>
<td>0.9851</td>
</tr>
<tr>
<td>IVC</td>
<td>0.9840</td>
<td>0.9840</td>
</tr>
<tr>
<td>Overall</td>
<td><strong>0.9849</strong></td>
<td><strong>0.9844</strong></td>
</tr>
</tbody>
</table>

Fig. 1(b) depicts the PSNR difference (dB) of each color image of the CMU database, that is, the gain in dB of image quality after applying $\hat{\alpha}(\nu, r)$ at $d = 2000$ centimeters to the $\hat{Q}$ images. On average, this gain is about 15 dB. Visual examples of these results are shown by Fig. 2, where the left images are the original images, central images are perceptual quantized images after applying $\alpha(\nu, r)$ and right images are recovered images after applying $\hat{\alpha}(\nu, r)$.

After applying $\hat{\alpha}(\nu, r)$, a visual inspection of these sixteen recovered images show a perceptually lossless quality. We perform the same experiment for gray-scale and color images with $d = 20, 40, 60, 80, 100, 200, 400, 800, 1000$ and $2000$ centimeters, in addition to test their objective and subjective image quality by means of the PSNR and MSSIM metrics, respectively, across the CMU(Fig. 3) Image Database.

In Figure 3, green functions denoted as F-$\rho$SQ are the quality metrics of perceptual quantized images after applying...
Thus, either for gray-scale or color images, both PSNR and MSSIM estimations of the quantized image \( Q \) decrease regarding \( d \), the longer \( d \) the greater the image quality decline. When the image decoder recovers \( Q \) and it is perceptually inverse quantized, the quality barely varies and is close to perceptually lossless, no matter the distance.

5. Experiments and Results

For the sake of comparing the performance between the JPEG2000 [13] and \( \chi \)SET coders, both algorithms are tested according to the process depicted in Fig. 4. First a \( \chi \)SET compression with certain viewing conditions is performed, which gives a compressed image with a particular bit-rate (bpp). Then, a JPEG2000 compression is performed with the same bit-rate. Once both algorithms recover their distorted images, they are compared with some numerical image quality estimators such as: MSSIM [14], PSNR [15], SSIM [16], VIF [17], UQI [18] and WSNR [19].

This experiment is performed across the CMU [9] and IVC [11] Image Databases. Image quality estimations are assessed by the six metrics mentioned before.

Figs. 5 and 6 show the perceptual quality, estimated by 5(a) MSSIM, 5(c) SSIM, 6(a) UQI, 6(b)VIF and 6(c) WSNR, in addition to the objective quality 5(b) PSNR, of the recovered color images both for JPEG2000(Blue function) and \( \chi \)SET(Green function) as a function of their compression rate. For this experiment, we employ the CMU Image Database and the Kakadu implementation for JPEG2000 compression [20]. On the average, a color image compressed at 1.0 bpp (1:24 ratio, stored in 32 KBytes) by JPEG2000 coder has MSSIM=0.9424, SSIM=0.8149, UQI=0.5141, VIF=0.2823 and WSNR=29.2 of perceptual image quality, and PSNR=30.11 of objective image quality, while by \( \chi \)SET has MSSIM=0.9780, SSIM=0.8758, UQI=0.6249, VIF=0.4387, WSNR=35.41 and PSNR=31.84.

In Figure 7, we can see these differences when images (a-b)Lenna, (c-d)Girl2 and (e-f)Tiffany are compressed at 0.92 bpp, 0.54 bpp and 0.93 bpp, respectively, by JPEG2000 and \( \chi \)SET. For example for these three images, the average difference of MSSIM is 0.0321 in favor of \( \chi \)SET. Therefore, for this image database, \( \chi \)SET has clearly improvement of visual quality than JPEG2000.

6. Conclusions

In this work we defined both forward (F-\( \rho \)SQ) and inverse (I-\( \rho \)SQ) perceptual quantizer using CBPF. We incorporated it to Hi-SET, proposing the new perceptual image compression system \( \chi \)SET. In order to measure the effectiveness of the perceptual quantization, a performance analysis is done using six image quality assessments such as MSSIM, PSNR, SSIM, UQI, VIF and WSNR, which measured the image quality between reconstructed and original images. The experimental results show that the solely usage of the Forward Perceptual Quantization improves the JPEG2000 compression and image perceptual quality. In addition, when both Forward and Inverse Quantization are applied into Hi-SET, it significatively improves the results regarding the JPEG2000 compression.
PROCESS FOR COMPARING JPEG2000 AND $\chi$SET. GIVEN SOME VIEWING CONDITIONS A $\chi$SET COMPRESSION IS PERFORMED OBTAINING A PARTICULAR BIT-RATE. THUS, A JPEG2000 COMPRESSION IS PERFORMED WITH SUCH A BIT-RATE.

Therefore, we developed a perceptual quantizer algorithm that, in contrast to the JPEG2000 global Frequency weighting, performs a local quantization, that is coefficient-by-coefficient. Similarly to JPEG2000, it is not necessary to store the applied weighting for inverse quantizing is because CBPF properties permits to predict perceptual weighting $a$ posteriori.

Furthermore, when $\chi$SET is compared with stated-of-the-art perceptual image coders it obtains good results both for objective and subjective images quality.

Acknowledgment

This work is supported by National Polytechnic Institute of Mexico by means of Project No. 20131312 granted by the Academic Secretary and the Committee of Operation and Promotion of Academic Activities (COFAA), National Council of Science and Technology of Mexico by means of Project No. 204151/2013, LABEX $\Sigma$-LIM France, Coimbra Group Scholarship Programme granted by University of Poitiers and Region of Poitou-Charentes, France.
Fig. 7  
EXAMPLE OF RECOVERED COLOR IMAGES Lenna, Girl2 AND Tiffany OF
THE CMU IMAGE DATABASE COMPRESSED AT (A AND B) 0.92 BPP, ((B
AND C) 0.54 BPP AND (C AND D) 0.93 BPP, RESPECTIVELY. JPEG2000
IMAGES ARE COMPRESSED USING TABLE 1 AND s = 3.

References

compression by combining a conventional wavelet-coder with texture  
1294, Nov. 2002.
[3] M. Boliek, C. Christopoulos, and E. Majani, Information Technology:  
JPEG2000 Image Coding System, JPEG 2000 Part I final committee  
draft version 1.0 ed., ISO/IEC JTC1/SC29 WG1, JPEG 2000, April  
2000.
and blue-yellow chromatic gratings,” Journal of Physiology, vol. 359,  
Embedded quadTrees,” IEEE Data Compression Conference, p. 470,  
March 2011.
[7] ——,” Image coder based on Hilbert Scanning of Embedded quadTrees,”  
June 2011, under review in Digital Signal Processing.
[8] ——,” Image coder based on Hilbert Scanning of Embedded quadTrees:  
An introduction of Hi-SET coder,” IEEE International Conference on  
Multimedia and Expo, July 2011.
USC-SIPI image database. Signal and Image Processing Institute of  
the University of Southern California. [Online]. Available:  
http://sipi.usc.edu/database/
strategy for full-reference image quality assessment,” in Proc. SPIE,  
vol. 742, 2009.
/jfoulard.ece.cornell.edu/gaubatz/metrixmux/. Cornell  
University Visual Communications Laboratory.
Fundamentals, Standards and Practice, ser. ISBN: 0-7923-7519-X.  
for image quality assessment,” in Conference Record of the Thirty-
Seven Asilomar Conference on Signals, Systems and Computers.,  
image/video quality assessment,” Electronics Letters, vol. 44, no. 13,  
2006.
assessment: from error visibility to structural similarity;” IEEE  
for formulation of quality measures incorporated in halftoning algo-

rithms,” IEEE International Conference on Acoustics, Speech and Signal  
http://www.kakadusoftware.com/