An Improved Visual Pruning Algorithm for Perceptually Lossless Medical Image Coding

David Wu, Damian M. Tan and Hong Ren Wu
School of Electrical and Computer Engineering,
Royal Melbourne Institute of Technology,
Melbourne, Victoria, Australia, 3001
Email: Henry.Wu@rmit.edu.au

Abstract—An improved algorithm for perceptually lossless coding of medical images is presented in this paper. Built on the JPEG 2000 coding framework, the proposed coder combines an improved Visual Pruning algorithm with an advanced model of the Human Visual System to identify and to remove visually insignificant/irrelevant information. Current results have shown superior compression ratio gains over that of its lossless counterparts without any loss in visual fidelity.

I. INTRODUCTION

Advanced medical imaging systems play an important role in providing a non-invasive, fast and accurate diagnosis. The key to this is through the digitisation of medical images. Digital medical images can be stored indefinitely without any deterioration in quality and can be transmitted to any geographical location with relative ease. One type of medical application that utilizes these features is telemedicine [1]. However, as the demand increases, so does the requirements for storage space and network transmission bandwidth. Thus, the challenge is how to deliver clinically critical information in a smaller package. A solution to this problem is through image compression. In general, there are two broad categories for encoding digital images; they are reversible and irreversible coding. Reversible coding schemes offer the advantage of having no loss of information. Although desirable, the current state-of-the-art reversible image coding schemes do not provide adequate compression ratio gains for applications in medical imaging [2]. Counter to this is irreversible coding schemes, which provide greater compression ratio gains at the expense of information integrity. However, a distinction of what information is lost must be ascertained. Past publications have shown that compression ratios up to 1:20 are achievable without any loss in diagnostic information [2]–[6].

An alternative to these schemes is perceptually lossless image coding (PLIC). PLIC provides the best of both worlds, that is, greater compression ratio gain than reversible coding schemes, while producing images without any visible loss. The effectiveness of this concept was demonstrated in [7], with 27 radiologists and 4 radiographers. This paper presents a novel algorithm for visual pruning (VP). Built on the JPEG 2000 coding framework [8], the algorithm is combined with an advanced model of the Human Visual System (HVS) [9] so that only visually irrelevant/insignificant information is removed. Thus having the ability to encode images at a perceptually lossless quality. Other key features of this algorithm are its simplicity and modularity. These features enable the algorithm to be implemented into any other Wavelet transform based coding framework without disrupting bit-stream compliance. The effectiveness of modularity has been demonstrated in [10], [11] with a different algorithm using the Set Partitioning in Hierarchical Trees (SPIHT) coding framework [12].

This paper is presented with the following sections. Section II describes the vision model employed in the improved VP algorithm. Section III, describes the VP algorithm as well as its adaptation into the JPEG 2000 coding framework. Section IV evaluates the improved VP algorithm and finally a conclusion in Section V.

II. THE PERCEPTUAL DISTORTION METRIC

Traditional raw mathematical objective metrics such as Mean squared error (MSE) and Peak-signal-to-noise ratio (PSNR) has served as one of the basic means for quantifying visual distortions/quality [13]. However, the main drawback of these traditional metrics is that they do not correlate with what is perceived by a human observer [14]. Thus, the need for metrics that incorporate the perceptual characteristics of the HVS, that is, a perceptual distortion metric. The effectiveness of these HVS based metrics have been demonstrated in past publications [15]–[18]. This section describes a perceptual distortion metric embedded with a vision model, however, development and fundamental ideas are left the references [9], [16], [18], [19].

The vision model [9] in this paper is based on the unified vision model template, the contrast gain control (CGC) (Figure 1), by Watson and Solomon [18]. The CGC consists of three parts and takes in two inputs, a processed image and a reference image, both of which are subjected to the three part process. The three parts of the CGC are linear transform, visual masking response and detection and pooling. A linear transform (Equation 1) takes into account of the frequency and orientation selectivity of the HVS [18], [20]. In general, a linear transform can be expressed as

\[ X = T(x) \] (1)

where, X and x represents the neural and pixel domain images, respectively. Immediately following the linear transform, a set of contrast sensitive weights are applied to modulate the
neural image to the sensitivity levels of the human eye. The selection of a transform is an issue. Generally over-complete transforms, such as the Steerable-Pyramid [21] (SPT), are used since they can closely represent the mechanics of the HVS [20]. However, although over-complete transforms, like the SPT, are ‘alias-free’, they require additional resources to code. Thus to counter this problem, a Mallat [22] Wavelet transform with the Daubechies 9/7 filter set [23] (D97) is used. Only 5 levels of decomposition was used due to the sensitivity of the vision model. The drawback of using the Wavelet transform with the D97 filters is contrary to using over-complete transforms [20]. Despite these issues, using the 5 level Mallat [22] Wavelet transform opens up practical advantages such as bitstream compliance with JPEG2000 [24], which subsequently conforms with the Digital Imaging and Communications in Medicine (DICOM) standard [25].

Visual masking causes a visual signal to be hidden or diminished within the presence of another visual signal. This occurs between neurons from similar (intra-) and different (inter-) frequency, orientation and colour channels. It is these interactions that are modelled (Equation 2).

$$R_{Z,l,\theta}[m,n] = k_Z \cdot \frac{E_{Z,l,\theta}[m,n]}{I_{Z,l,\theta}[m,n] + \gamma_Z},$$

(2)

where $m$ and $n$ are the spatial frequency coordinate of a coefficient, $E_{Z,l,\theta}[m,n]$ and $I_{Z,l,\theta}[m,n]$ are excitation and inhibition functions, $k_Z$ and $\gamma_Z$ are the scaling and saturation constants, $Z \in \{\Theta, \Upsilon\}$, with $\Theta$ and $\Upsilon$ specifying the inter-orientation and intra-frequency masking domains, respectively. $l = \{1, 2, 3, 4, 5\}$ and $\theta = \{1, 2, 3\}$ represent the frequency levels and the orientation bands, respectively. The excitation and inhibition functions for each domain are defined as follows:

$$E_{\Theta,l,\theta}[m,n] = X_{l,\theta}[m,n]^{\varphi_\Theta}$$

(3)

$$E_{\Upsilon,l,\theta}[m,n] = X_{l,\theta}[m,n]^{\varphi_\Upsilon}$$

(4)

1Sensitivity here refers to the vision model being calibrated to only a 5-level Wavelet transform.

2Colour masking is not considered. This paper focuses on grayscale images.

3Inter-frequency masking was omitted to simplify the model.
frequency level would have the largest neighbouring. This approach attempts to equalize the uneven spatial coverage between images of different frequency levels inherent in multi-resolution representations. The neighbourhood variance \( \sigma_{m,n} = \frac{1}{M} \sum_{w=-M}^{M} \sum_{v=-M}^{M} (X_l(a[w, v]) - \mu)^2 \), with \( \mu \) representing the mean, has been added to the inhibition process to account for texture masking [26]. Exponents \( \rho \) and \( q \) are governed by the condition \( \rho > q > 0 \) according to [18]. Currently, \( q \) is set to 2.

The final component of the model detects the perceptually significant difference between two images. A squared-error (l2 norm) function defines the distortion within each masking channel. The total distortion is the sum of the distortions over all masking channels, given as

\[
D_T = \sum_{l=1}^{L_{max}} \sum_{m=1}^{M_l} \sum_{n=1}^{N_l} g_Z \cdot |R_{Z,X}[m,n] - R_{Z,D}[m,n]|^2
\]  

(7)

where \( R_{Z,X}[m,n] \) and \( R_{Z,D}[m,n] \) are the masking responses of the two images, \( a \) and \( b \), respectively, and \( g_Z \) being the channel gain. \( L_{max} \) is the maximum number of decompositions and \( Z \in \{ \Theta, \Upsilon \} \), where \( \Theta \) and \( \Upsilon \) are the inter-orientation and intra-frequency masking domains. In this paper, \( L_{max} \) is set to 5. Given that the dimensions of the image are \( M \times N \), \( M_l \) and \( N_l \) are defined as \( M_l = M \times 2^{l-L} \) and \( N_l = N \times 2^{l-L} \), respectively, where \( L = L_{max} + 1 \). The pooling equation, Equation (7), pools all coefficients spanning all frequencies and orientations, and provides an overall perceptual distortion, \( D_T \), between the two images.

III. VISUAL PRUNING

A. The Proposed Algorithm

By employing the perceptual distortion metric described in Section II, the VP algorithm (Figure 3) can identify and remove visually irrelevant/insignificant information. For a given frequency level, \( l \), orientation band, \( \theta \), a spatial frequency location \((m,n)\) and \( B \in \{0,1,...,P\} \), the VP algorithm can be described in two stages. The first stage computes a set of distortion measures (Equation 7), \( D_T(i,\theta,m,n) = \{D_T(i,\theta,m,n) | i \in B\} \), and a set of percentage responses, \( R_P(i,\theta,m,n) = \{R_P(i,\theta,m,n) | i \in B\} \), from a reference image and a set of processed images, \( V_{i,\theta,m,n} \). \( P \) is positive non-zero value specifying the precision of truncation. Larger values of \( P \) will result in finer truncated coefficients and hence provide a more accurate account of the visual distortions. The processed images are generated by percentage coefficient truncation (PCT)

\[
\tilde{X}_{i,\theta}[m,n] = \begin{cases} X_{i,\theta}[m,n] \times \left( \frac{P-i}{P} \right) & i \in B \\ \end{cases}
\]  

(8)

where, \( \tilde{X}_{i,\theta}[m,n] \) and \( X_{i,\theta}[m,n] \) are the processed and original coefficients, respectively. Large values of \( P \) are desirable at this stage since visual distortions can be accurately modeled. \( P \) is set to 100 in this paper. Equation 8 is applied to each coefficient separately at each frequency band except the LL band. The percentage response, \( R_{P,i}(\theta,m,n) \), for a given reference coefficient and a distorted coefficient, is defined as

\[
R_{P,\{i,\theta\}}(m,n) = \sum_{Z} R_{\{i,\theta\}}(m,n) \sum_{Z} R_{\{i,\theta\}}(m,n)
\]  

(9)

where \( R_{\{i,\theta\}}(m,n) \) and \( R_{\{i,\theta\}}(m,n) \) are, respectively, the masking response, for a referenced and a distorted coefficient, taken from Equation (2). \( \Theta \in \{ \Theta, \Upsilon \} \), denotes the orientation and local responses, respectively. Equation (9) provides a measurement of the depreciation of the response energy over both the intra-frequency and inter-orientation channels.

The last stage gathers the set of distortion measures, \( D_T(i,\theta,m,n) \), the set of percentage responses, \( R_P(i,\theta,m,n) \) and performs visually adaptive coefficient pruning. By comparing \( D_T(i,\theta,m,n) \) and \( R_P(i,\theta,m,n) \) to a set of pre-determined JNND thresholds, \( T_D \) and \( T_P \), respectively, a coefficient is truncated (Equation 8) to a perceptually optimal bit-plane level, \( i_{opt} \), only when a distortion measure from \( D_T(i,\theta,m,n) \) is less than or equal to a JNND threshold, \( T_D(i,\theta) \) and when a percentage response from \( R_P(i,\theta,m,n) \) is less than or equal to a percentage response threshold \( T_P(i,\theta) \). Thus,

\[
\tilde{X}_{i_{opt},\theta}[m,n] = X_{i,\theta}[m,n] \times \left( \frac{P-i_{opt}}{P} \right)
\]  

(10)

where

\[
i_{opt} = \max \left\{ i \in B \mid \left(D_T(i,\theta,m,n) \leq T_D(i,\theta) \right) \land \left(R_P(i,\theta,m,n) \leq T_P(i,\theta) \right) \right\}
\]

B. Determining \( T_D(i,\theta) \) and \( T_P(i,\theta) \)

Both \( T_D(i,\theta) \) and \( T_P(i,\theta) \) were derived from subjective experiments. For each orientation \( \theta \) and each frequency level \( l \), there is a set of pre-determined thresholds \( T_D \) and \( T_P \), for \( \theta = \{1,2,3\} \) and \( l = \{1,2,3,4,5\} \). These thresholds have been obtained through the testing of approximately 2560 (32x32 pixels) 16-bit medical greyscale sub-images, for each modality. These sub-images originated from a particular base image (512x512 pixels), which was distorted in 10 different ways through bit-plane filtering. Each of these 10 distorted images was then partitioned into 256 (32x32 pixels) individual pieces. Sub-image testing is preferred in this case over the complete image testing because it is able to quantify the...
different local thresholds levels in different regions within images, i.e., the segmented test is better equipped to capture the localised variation in image quality. The threshold level for the experiment is set at the JNND level for visually lossless quality encoding. Once the JNND level of test materials have been mapped, the thresholds $T_D$ and $T_P$ can be determined by soliciting the responses (2) and (7) of the sub-images in the JNND map. In other words, only sub-images at the JNND level will be used to determined the thresholds $T_D(l,\theta)$ and $T_P(l,\theta)$. Finally, fine tuning was performed through a stringent test of flipping back and forth the encoded image with the original image. This employs the temporal sensitivities of the HVS to ensure that ‘distortion flickers’ between the two images are not perceivable.

C. JPEG 2000 Adaptation

The VP algorithm is independent and does not require a specialised decoder. Thus it can be implemented in any Wavelet based coding framework (Figure 4). Here, the VP algorithm is applied to the wavelet coefficients immediately after a forward Wavelet transform. The alternative is to apply the algorithm on a per code block basis, however, this method can significantly impede the accuracy of the distortion metric thus leading to lower coding performance [27].

IV. PERFORMANCE EVALUATION

The proposed coder (PC-PCT) in this paper will be evaluated against four benchmark coders. The first is the LOCO lossless coder [28], the second is the NLOCO near-lossless coder [28], the third is the JPEG2000 lossless mode (J2KL) [24] and the fourth is a variant of the (PC-PCT), (PC-BCT) [7], which employs a bit-plane coefficient truncation (BCT) approach. A parameter, $d$, for NLOCO is set to 2. Here, $d$ represents the maximum pixel difference between the original image and the compressed image. Current results (Table I & II), in terms of bits-per-pixel(bpp), show that the coding performance of the (PC-PCT) outperforms the LOCO and J2KL lossless coders in all instances, averaging 70% more compression. On the other hand, there was on average, a 5% and 14% compression gain over the NLOCO near-lossless coder and the PC-BCT coder, respectively. A PCT approach provides a more accurate account of the visual distortions since each coefficient is pruned at finer fractional steps depending on the precision $P$, whereas, in the BCT approach each coefficient is pruned at fixed step sizes in powers of $2$ ($1,2,4, ..., 2^3$). A drawback of the PCT approach is that it has a higher computational complexity over the BCT when $P$ is greater than 32, that is, the greater the precision the greater the computational complexity. The impact of $P$ on the coding performance is dependant on the threshold values $T_D(l,\theta)$ and $T_P(l,\theta)$. Generally, it is desirable to use larger values of $P$ when determining the aforementioned JNND visual thresholds. Nevertheless and more importantly, no distortions were perceivable in the images compressed by PC-PCT (Figure 5).

![Image](image.png)

**Figure 4.** A generalised Wavelet-based image coder embedded with the Visual Pruning algorithm. The Visual Pruning algorithm is applied immediately after a forward transform. Note that the Quantiser was disabled (step size set to 1).

The VP algorithm is independent and does not require a specialised decoder. Thus it can be implemented in any Wavelet based coding framework (Figure 4). Here, the VP algorithm is applied to the wavelet coefficients immediately after a forward Wavelet transform. The alternative is to apply the algorithm on a per code block basis, however, this method can significantly impede the accuracy of the distortion metric thus leading to lower coding performance [27].

IV. PERFORMANCE EVALUATION

The proposed coder (PC-PCT) in this paper will be evaluated against four benchmark coders. The first is the LOCO lossless coder [28], the second is the NLOCO near-lossless coder [28], the third is the JPEG2000 lossless mode (J2KL) [24] and the fourth is a variant of the (PC-PCT), (PC-BCT) [7], which employs a bit-plane coefficient truncation (BCT) approach. A parameter, $d$, for NLOCO is set to 2. Here, $d$ represents the maximum pixel difference between the original image and the compressed image. Current results (Table I & II), in terms of bits-per-pixel(bpp), show that the coding performance of the (PC-PCT) outperforms the LOCO and J2KL lossless coders in all instances, averaging 70% more compression. On the other hand, there was on average, a 5% and 14% compression gain over the NLOCO near-lossless coder and the PC-BCT coder, respectively. A PCT approach provides a more accurate account of the visual distortions since each coefficient is pruned at finer fractional steps depending on the precision $P$, whereas, in the BCT approach each coefficient is pruned at fixed step sizes in powers of $2$ ($1,2,4, ..., 2^3$). A drawback of the PCT approach is that it has a higher computational complexity over the BCT when $P$ is greater than 32, that is, the greater the precision the greater the computational complexity. The impact of $P$ on the coding performance is dependant on the threshold values $T_D(l,\theta)$ and $T_P(l,\theta)$. Generally, it is desirable to use larger values of $P$ when determining the aforementioned JNND visual thresholds. Nevertheless and more importantly, no distortions were perceivable in the images compressed by PC-PCT (Figure 5).

<table>
<thead>
<tr>
<th>Image Name</th>
<th>Dimensions (pixels)</th>
<th>Image Type</th>
<th>Bitrate (bpp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ankle 1</td>
<td>2572 x 2040</td>
<td>CR</td>
<td>5.852</td>
</tr>
<tr>
<td>Ankle 2</td>
<td>1516 x 2044</td>
<td>CR</td>
<td>5.695</td>
</tr>
<tr>
<td>Body 1</td>
<td>512 x 512</td>
<td>CT</td>
<td>4.191</td>
</tr>
<tr>
<td>Body 2</td>
<td>512 x 512</td>
<td>CT</td>
<td>5.933</td>
</tr>
<tr>
<td>Body 3</td>
<td>512 x 512</td>
<td>CT</td>
<td>6.853</td>
</tr>
<tr>
<td>Brain 1</td>
<td>512 x 512</td>
<td>CT</td>
<td>3.380</td>
</tr>
<tr>
<td>Brain 2</td>
<td>512 x 512</td>
<td>CT</td>
<td>3.841</td>
</tr>
<tr>
<td>Brain 3</td>
<td>512 x 512</td>
<td>MR</td>
<td>4.687</td>
</tr>
<tr>
<td>Brain 4</td>
<td>208 x 256</td>
<td>MR</td>
<td>3.989</td>
</tr>
<tr>
<td>Brain 5</td>
<td>512 x 512</td>
<td>CT</td>
<td>3.681</td>
</tr>
<tr>
<td>Brain 6</td>
<td>512 x 512</td>
<td>CT</td>
<td>3.597</td>
</tr>
<tr>
<td>Brain 7</td>
<td>512 x 512</td>
<td>CT</td>
<td>6.175</td>
</tr>
<tr>
<td>Brain 8</td>
<td>512 x 512</td>
<td>CT</td>
<td>3.962</td>
</tr>
<tr>
<td>Brain 9</td>
<td>512 x 512</td>
<td>CT</td>
<td>3.190</td>
</tr>
<tr>
<td>Chest 1</td>
<td>2496 x 2048</td>
<td>CR</td>
<td>6.736</td>
</tr>
<tr>
<td>Chest 2</td>
<td>2496 x 2048</td>
<td>CR</td>
<td>6.821</td>
</tr>
<tr>
<td>Chest 3</td>
<td>2496 x 2048</td>
<td>CR</td>
<td>6.259</td>
</tr>
<tr>
<td>Chest 4</td>
<td>2496 x 2048</td>
<td>CR</td>
<td>6.198</td>
</tr>
<tr>
<td>Chest 5</td>
<td>2496 x 2048</td>
<td>CR</td>
<td>6.020</td>
</tr>
<tr>
<td>Chest 6</td>
<td>1516 x 2044</td>
<td>MR</td>
<td>5.912</td>
</tr>
<tr>
<td>Elbow</td>
<td>2044 x 1514</td>
<td>CR</td>
<td>6.177</td>
</tr>
<tr>
<td>Knee</td>
<td>512 x 512</td>
<td>MR</td>
<td>7.143</td>
</tr>
<tr>
<td>Leg</td>
<td>2040 x 2570</td>
<td>CR</td>
<td>3.729</td>
</tr>
<tr>
<td>Liver</td>
<td>512 x 512</td>
<td>CT</td>
<td>5.029</td>
</tr>
<tr>
<td>Neck</td>
<td>2040 x 2570</td>
<td>CR</td>
<td>4.674</td>
</tr>
<tr>
<td>Pelvis</td>
<td>3732 x 3062</td>
<td>CR</td>
<td>5.915</td>
</tr>
<tr>
<td>Saddle</td>
<td>256 x 256</td>
<td>MR</td>
<td>5.972</td>
</tr>
<tr>
<td>Spine 1</td>
<td>2040 x 2570</td>
<td>CR</td>
<td>6.437</td>
</tr>
<tr>
<td>Spine 2</td>
<td>512 x 512</td>
<td>CT</td>
<td>3.526</td>
</tr>
<tr>
<td>Spine 3</td>
<td>512 x 512</td>
<td>CT</td>
<td>6.106</td>
</tr>
<tr>
<td><strong>AVERAGE</strong></td>
<td></td>
<td></td>
<td><strong>5.322</strong></td>
</tr>
</tbody>
</table>

**Table I**

**SHOWS THE CODING PERFORMANCE, IN TERMS OF BITRATE, BETWEEN THE PC-PCT AND THE LOSSLESS CODERS, LOCO AND J2KL. EACH IMAGE HAS A MAXIMUM BIT-DEPTH OF 16 BITS PER PIXEL.

V. CONCLUSION

This paper presents an improved visual pruning algorithm for perceptually lossless image coding. Built on the JPEG2000 coding framework [24] and embedded with an advanced Human vision model, the VP algorithm can identify and remove visually insignificant/irrelevant information. In terms of coding performance, the PC outperforms its lossless counterparts in all instances. The key features of the VP algorithm is its simplicity and modularity. Hence, it does not require a specialised decoder and can be implemented into any wavelet based image coder while maintaining bit-stream compliance. More importantly, there was no perceivable loss in fidelity.
The authors would like to thank Southern Health Monash Medical Centre and the patients for permission to use their images.

ACKNOWLEDGMENT

The authors would like to thank Southern Health Monash Medical Centre and the patients for permission to use their images.

REFERENCES


Fig. 5.  *Left:* Original.  *Right:* Proposed Codet.  *Top down:* Knee; SideBrain; Chest6(cropped).