



Perceptually lossless medical image coding

Wu, David; Tan, Damian; Baird, Marilyn; DeCampo, J; White, Christopher; Wu, Henry

https://researchrepository.rmit.edu.au/discovery/delivery/61RMIT_INST:ResearchRepository/12247856540001341?#13248375860001341

Wu, Tan, D., Baird, M., DeCampo, J., White, C., & Wu, H. (2006). Perceptually lossless medical image coding. *IEEE Transactions on Medical Imaging*, 25(3), 335–344. <https://doi.org/10.1109/TMI.2006.870483>

Published Version: <https://doi.org/10.1109/TMI.2006.870483>

Repository homepage: <https://researchrepository.rmit.edu.au>

© 2006 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Downloaded On 2022/08/10 05:16:03 +1000

Perceptually Lossless Medical Image Coding

David Wu*, Damian M. Tan, Marilyn Baird, John DeCampo, Chris White, and Hong Ren Wu

Abstract—A novel perceptually lossless coder is presented for the compression of medical images. Built on the JPEG 2000 coding framework, the heart of the proposed coder is a visual pruning function, embedded with an advanced human vision model to identify and to remove visually insignificant/irrelevant information. The proposed coder offers the advantages of simplicity and modularity with bit-stream compliance. Current results have shown superior compression ratio gains over that of its information lossless counterparts without any visible distortion. In addition, a case study consisting of 31 medical experts has shown that no perceivable difference of statistical significance exists between the original images and the images compressed by the proposed coder.

Index Terms—Biomedical imaging, double blind testing, image coding, just-not-noticeable- difference, medical image coding, perceptually lossless image coding, 2-staged forced choice, vision model.

I. INTRODUCTION

ADVANCED medical imaging technologies, such as computed tomography (CT), magnetic resonance imaging (MRI) [1], [2], and traditional radiography performed using computed radiography (CR) [3] and digital radiography (DR) [1], [3] are fundamental tools in providing more efficient and effective healthcare systems and services. The key to the proliferation of these technologies is the digital representation of images. Digital medical images have potential benefits in terms of durability and portability. In addition, it offers versatility, enabling or expanding its applications in medical imaging. Durability permits a digital image to be stored indefinitely without any degradation in image fidelity and portability allows a digital image to be transmitted to any desired destination over communication networks with relative ease. Problems involving storage space and network bandwidth requirements arise when large volumes of images are to be stored or transmitted, as is the case with medical images [4]. From the diagnostic imaging

point of view, the challenge is how to deliver clinically critical information in the shortest time possible. A solution to this problem is through image compression.

Generally, image compression schemes can be classified into two distinct categories, reversible (information lossless) and irreversible (information lossy) compression [4]. Reversible compression schemes are highly desired as information integrity is maintained throughout the whole encoding and decoding process. However, at best, the existing state-of-the-art technology achieves compression ratios between 2:1 to 4:1 and, thus, has been the limiting factor in its proliferation [5]. On the other hand, irreversible compression schemes provide greater compression gains at the expense of information integrity. Any deterioration in image integrity may lead to visible distortions, if the degradation is not monitored and controlled properly. Nevertheless, it is the degree of loss of diagnostic information, which must be ascertained to determine the acceptance of a medical image compression strategy. As such, studies have shown evidence that irreversible image compression with compression ratios ranging from 10:1 up to 20:1 are achievable for medical images without significantly impairing their diagnostic value [5]–[8]. A possible approach to circumventing the limitations of both reversible and irreversible image compression is through perceptually (or visually) lossless image coding (PLIC) [9]. PLIC provides greater compression gain than reversible techniques while yielding compressed images without any degradation in visual quality. The focus of this paper centers on a novel perceptually lossless image coding technique — presented as an alternative for the compression of medical images. It is important to note that an objective of this paper is to determine if images compressed by the proposed coder (PC) are perceptually lossless to the original images. Thus, perceptual image enhancement as well as feature extraction are not considered here.

Based on the JPEG2000 coding framework [10], the heart of the PC is the implementation of an advanced visual pruning function combined with a human vision model [11], [12] to identify and to remove visually insignificant/irrelevant information as well as to offer the benefits of simplicity and modularity. Furthermore, the visual pruning function can be embedded into any discrete Wavelet transform-based coder while maintaining bitstream compliance. This has been demonstrated previously in [13], [14] based on the Set Partitioning of Hierarchical Trees (SPIHT) coding framework [15]. Current results have shown improved coding performance over the JPEG-LS LOCO-I algorithm [16] without any perceivable visual distortions. In terms of subjective performance, based on a subjective assessment with 31 medical expert viewers, no perceivable differences of statistical significance exists between the original images and the images compressed by the PC.

Manuscript received December 10, 2004; revised December 2, 2005. *Asterisk indicates corresponding author.*

*D. Wu was with the School of Computer Science and Software Engineering, Monash University, Clayton Campus, VIC 3800, Melbourne, Australia. He is now with the School of Electrical and Computer Engineering, Royal Melbourne Institute of Technology, VIC 3001, Melbourne, Australia (e-mail: dwu8@op-tusnet.com.au).

D. M. Tan and H. R. Wu were with the School of Computer Science and Software Engineering, Monash University, Clayton Campus, VIC 3800, Melbourne, Australia. They are now with the School of Electrical and Computer Engineering, Royal Melbourne Institute of Technology, VIC 3001, Melbourne, Australia (e-mail: henry.wu@rmit.edu.au).

M. Baird is with the Department of Medical Imaging & Radiation Sciences, Monash University, Clayton Campus, VIC 3800, Melbourne, Australia.

J. DeCampo was with the Department of Diagnostic Imaging, Southern Health Monash Medical Centre, Clayton, VIC 3168, Melbourne, Australia. He is now with the Royal Perth Hospital, WA 6000, Perth, Australia.

C. White is with the School of Business Systems, Monash University, Clayton Campus, VIC 3800, Melbourne, Australia.

Digital Object Identifier 10.1109/TMI.2006.870483

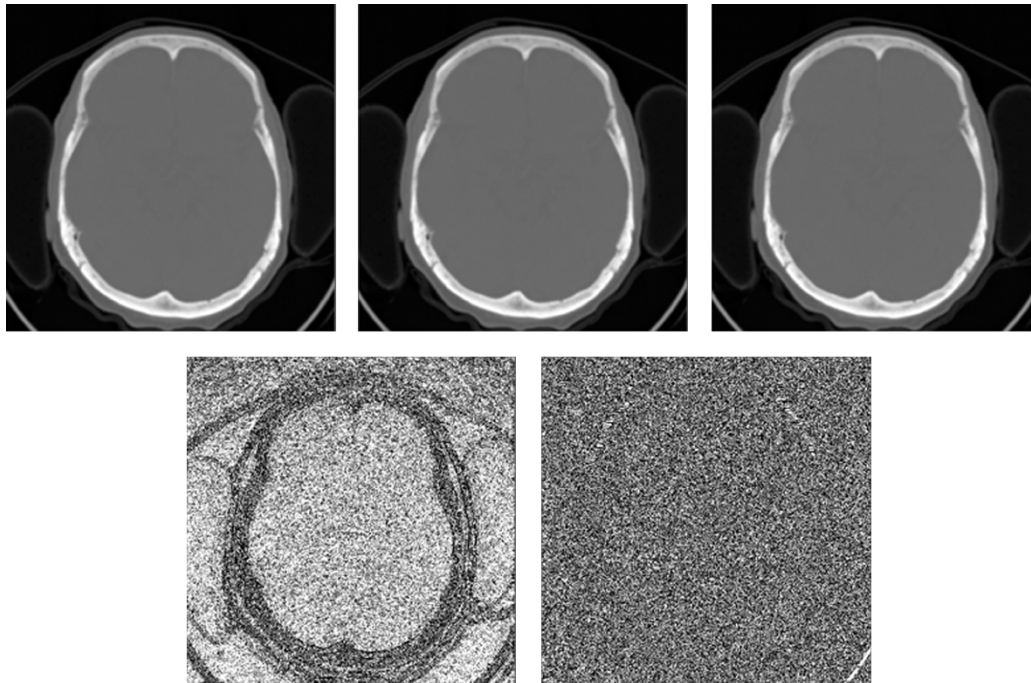


Fig. 1. Top (left to right): Original; proposed coder; NLOCO ($d = 2$) [16] (see Section IV); Bottom (left to right): Difference of PC and original; Difference of NLOCO ($d = 2$) and original. As an example, this image shows the distribution of the pixel differences. Both error images have been normalized to the largest absolute pixel difference between the two images, which was 12. Here, NLOCO only has a maximum pixel difference of 2. The errors in the PC difference image are concentrated only in areas where visual masking occurs (see Section III), while errors in the NLOCO difference image are distributed across the entire image. White pixels in the error images represent no difference while black pixels represent that a difference exists.

This paper is organized with Section II providing a brief overview of medical image coding and an introduction to PLIC. Section III presents the PC, the underlying vision model and its adaptation into the JPEG 2000 coding framework [17]. Section IV evaluates the performance of the PC through coding and subjective analyzes. Finally, a conclusion is drawn in Section V.

II. MEDICAL IMAGE CODING

A. An Overview

The coding of medical images differs from the coding of standard natural images in that it is imperative that the integrity of the diagnostic information in medical images are maintained while providing a reduction in storage space and network transmission bandwidth requirements. Inevitably, the ultimate solution is through reversible compression. However, at present, the existing state-of-the-art reversible technologies cannot achieve a significant reduction in bit-rate deemed adequate for the current practical applications in biomedical imaging [5]. A survey of reversible coding techniques for medical images can be found in [18]. Other approaches such as progressive image transmission [19], [20] and irreversible image coding schemes [17], [21]–[23] have been investigated and applied to alleviate this problem.

Irreversible coding techniques generally can be subclassed as rate-driven, quality-driven, error-driven and hybridized. Rate-driven coders encode images at a given bit-rate and, thus, a quality level cannot be guaranteed. Examples of rate-driven coders include JPEG Baseline [24], standard JPEG2000 [10], vector quantization [22], and fractal coding [23]. In contrast, quality-driven coders encode images at a given quality level while attempting to achieve a best possible minimum

bit-rate. Error-driven techniques encode an image such that the maximum absolute pixel error/difference is no more than a specified error value, d . Error-driven techniques do not fall into the categories of rate-driven or quality-driven since they do not guarantee to achieve a fixed bit-rate or perceptual quality criterion. Near-lossless coders such as JPEG-LS near-lossless [16] is an example of an error-driven coder.

Finally, hybridized coding techniques [25], [26] encode areas of an image [region of interest (ROI)] with a reversible coding technique and the remaining areas with an irreversible technique. The selection of the ROIs can be done manually or automatically. Although automated ROIs selection is a practical solution, the recognition and segmentation of the ROIs is a complex issue [27].

B. Perceptually Lossless Image Coding

PLIC falls into the quality-driven category, that is, to encode an image at the best possible minimum rate such that it is indistinguishable from the original [9]. A common misconception is that perceptually lossless can be achieved with any rate-driven coder through “tuning” the bit-rate to a point where no loss of detail can be seen. The key issue here is that image quality is dependant on image content and rate-driven coders, such as baseline JPEG [24], cannot guarantee reliable identification and removal of visually insignificant/irrelevant information (Fig. 1 and Table I). Table I demonstrates that there is no fixed error or bit-rate during the encoding phase of PLIC and, thus, picture quality is image content dependant. These errors were obtained by subtracting the PC compressed images from their respective original images.

TABLE I
THE MAXIMUM AND MINIMUM PIXEL DIFFERENCES BETWEEN THE ORIGINAL
IMAGE AND THE IMAGES COMPRESSED BY THE PC

Image Name	PC Maximum Pixel Difference	PC Minimum Pixel Difference
Knee	18	-18
SideBrain	14	-18
Brain2	12	-12

The theoretical significance of PLIC is that by employing an advanced model of human vision it is possible to identify visually irrelevant/insignificant information and thereby to remove only psycho-visual redundancy [4]. Thus, the framework of this problem can be posed as what is the minimum threshold of human vision such that it is just below or equal to the just-not-noticeable-difference (JNND) threshold, that is

$$d(\mathbf{x}, \hat{\mathbf{x}}) \leq \text{JNND} \quad (1)$$

where d is a distortion function computing the visually significant difference between an original image \mathbf{x} and a reconstructed image $\hat{\mathbf{x}}$, and JNND is the JNND level. A solution to this problem is best described by models of the Human Visual System (HVS). Although the concept of PLIC has appeared previously in [9], its application to medical imaging is still in its infancy as seen in the limited treatment of the subject in medical imaging literature [28], [29]. The contributions of this paper include an embedded advanced human vision model to identify and to remove visually insignificant/irrelevant information while maintaining bitstream compliance with the JPEG 2000 coding framework [10] and subsequently retaining compliance with the Digital Imaging and Communications in Medicine (DICOM) standard [30]. In addition, a subjective assessment of the coder performance with 31 medical expert viewers using 16-bit medical (CT, MRI, and CR) images is presented.

III. PROPOSED CODER

A. Vision Modeling

Traditional metrics, such as the mean squared error (MSE), the peak-signal-to-noise ratio (PSNR) and its variations [4] have served as the basic means of quantifying visual distortions and quality. These metrics are commonly classified as objective raw mathematical measurements and offer the advantage of simplicity in computation, requiring only a processed image and the original image. However, it is well known that these metrics do not correlate well with what is perceived by a human observer [31]. A solution is to utilize metrics which incorporate the perceptual characteristics of the HVS. This approach has demonstrated its effectiveness in picture quality/impairment assessments [32]–[35].

The HVS can be described in three parts [36]. The first part describes the optical characteristics of the human eye with respect to its sensitivity relative to background luminance levels and varying spatio-temporal frequencies. This sensitivity is termed “contrast sensitivity” [37], which is functionally described as the contrast sensitivity function (CSF). The second part is the visual pathway and this provides a link between the eye and the visual cortex. Finally, the third part describes

the formation of images within the visual cortex. Neuron interactions in the visual cortex leads to the visual masking phenomenon [38] Visual masking affects a visual signal by diminishing its visibility when it is within the presence of another visual signal. This occurs between neurons from similar (intra) and different (inter) frequency, orientation and color channels. It is these interactions that are modeled to describe the visual masking effect.

The contrast gain control (CGC) (Fig. 2) coined by Watson and Solomon [33] serves as a vision model template implemented here. This vision model template is a unification of other earlier vision models by Teo and Heeger [34] and by Watson and Solomon [33]. The CGC consists of a linear transform, a masking response and a pooling and detection phase. The CGC takes two inputs, that is, a reference (original) image and a processed image.

1) *Linear Transform*: A linear transform (T) takes into account of the frequency and orientation selectivity of the HVS

$$\mathbf{X} = T(\mathbf{x}) \quad (2)$$

where, \mathbf{X} and \mathbf{x} are the neural and pixel domain images, respectively. Immediately after a linear transform, a set of CSF frequency sensitive weights (Table II) are applied to modulate the neural image to the sensitivity levels of the human eye. One issue in vision modeling is the selection of filters for linear transform. Over-complete (redundant) transforms, such as the steerable-pyramid transform (SPT), are known to provide a more accurate account of the visual mechanisms of the HVS since they are free of “aliasing” and are shift invariant [33], [34], [38], [39]. However, a drawback of using an over-complete transform is that it requires additional resources to code. To counter this problem, critically sampled (nonredundant) transforms can be employed. The vision model here uses the discrete wavelet transform (DWT) with the Daubechies 9/7 (D97) filter set, which are used for both vision modeling and coding. An alternative is to employ separate transform filters, one for modeling and one for coding. This leads to a significantly higher computational complexity. The 5/3 filter set by Le Gall and Tabatabai [40], adopted in the current JPEG 2000 standard for reversible coding [10], was not considered for the current investigation due to its short filter length. Although computationally less demanding, short filter lengths may cause dramatic ringing distortions when coefficients are quantized.

Although there are issues, such as aliasing and shift variance, associated with using the DWT with the D97 filter set, it nevertheless provides several practical advantages for the current application. One such advantage is that it is linear and complete. Therefore, this model can be embedded into any Wavelet based coding framework, such as the JPEG2000 [10], while maintaining bit-stream compliance and, thus, would not require a specialized decoder. Subsequently, the second advantage is that bit-stream compliance with the JPEG2000 coding leads to compliance with the DICOM standard [30]. Due to the sensitivity¹ of the vision model, only a 5-level Mallat [41] Wavelet transform was employed.

¹Sensitivity refers to model parameters tuned to a specific level of decomposition.

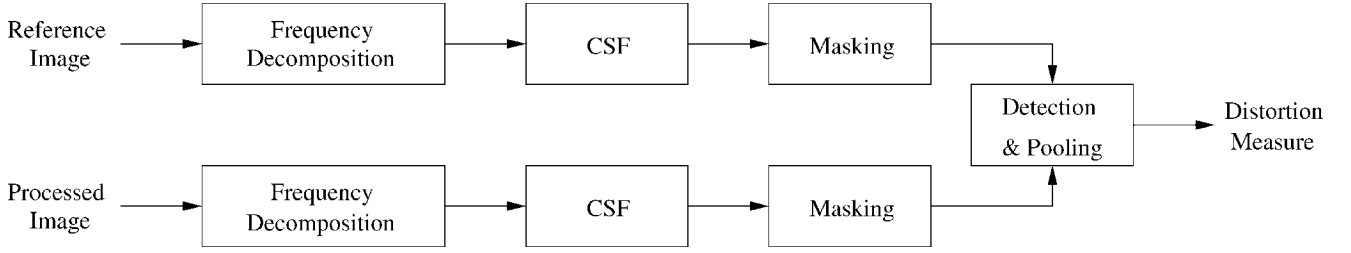


Fig. 2. CGC model. The example given here only models the primary visual cortex. The CSF can be applied prior to or after the frequency decomposition. The difference between the two approaches is the domain in which the CSF operates in – time or frequency. Here, the CSF operates in the frequency domain.

TABLE II
THE VISION MODEL PARAMETERS DERIVED FROM SUBJECTIVE EXPERIMENTS
IN [11]. THE CSF – LL WEIGHT IS NOT USED

Parameters	
$CSF - LL$	1.4800
$CSF - 1$	1.5500
$CSF - 2$	1.7700
$CSF - 3$	1.6800
$CSF - 4$	1.2900
$CSF - 5$	0.8050
k_{Υ}	1.0880
k_{Θ}	0.9876
γ_{Υ}	5.5550
γ_{Θ}	7.6800
p_{Υ}	2.5800
p_{Θ}	2.3950
q	2.0000
g_{Υ}	0.7588
g_{Θ}	0.4834

2) *Masking Response*: The second stage of the CGC is the masking response, which itself is encompassed into a single multichannel response function, of the form [12]

$$R_{\mathcal{Z},l,\theta}[m,n] = k_{\mathcal{Z}} \cdot \frac{E_{\mathcal{Z},l,\theta}[m,n]}{I_{\mathcal{Z},l,\theta}[m,n] + \gamma_{\mathcal{Z}}^q} \quad (3)$$

where m and n are the spatial frequency coordinate of a coefficient, $E_{\mathcal{Z},l,\theta}[m,n]$ and $I_{\mathcal{Z},l,\theta}[m,n]$ are excitation and inhibition functions, $k_{\mathcal{Z}}$ and $\gamma_{\mathcal{Z}}^q$ are the scaling and saturation constants (Table II), $\mathcal{Z} \in \{\Theta, \Upsilon\}$, with Θ and Υ representing the interorientation and intrafrequency masking domains, respectively². $l = \{1, 2, 3, 4, 5\}$ 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:

$$E_{\Theta,l,\theta}[m,n] = \mathbf{X}_{l,\theta}[m,n]^{p_{\Theta}} \quad (4)$$

$$E_{\Upsilon,l,\theta}[m,n] = \mathbf{X}_{l,\theta}[m,n]^{p_{\Upsilon}} \quad (5)$$

$$I_{\Theta,l,\theta}[m,n] = \mathbf{X}_{l,\theta}[m,n]^q + \sum_{\alpha=1, \alpha \neq \theta}^3 \mathbf{X}_{l,\alpha}[m,n]^q \quad (6)$$

$$I_{\Upsilon,l,\theta}[m,n] = \frac{8}{A_l} \sum_{u=m-l}^{m+l} \sum_{v=n-l}^{n+l} \mathbf{X}_{l,\theta}[u,v] + \sigma_l[m,n]^q \quad (7)$$

²Interfrequency masking was omitted to simplify the model.

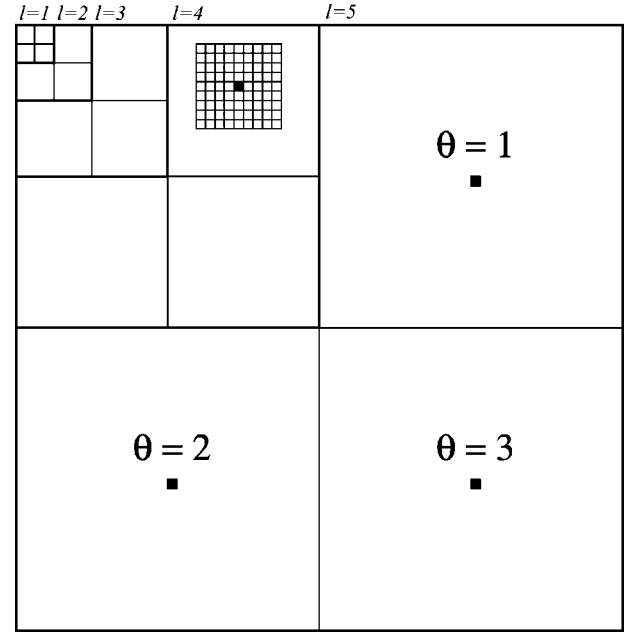


Fig. 3. Orientation and spatial frequency locations of the hierarchical (Mallat wavelet) decomposition. Each frequency level has three orientated bands, $\theta = \{1, 2, 3\}$, except for the lowest frequency level. At the lowest frequency, there is an additional isotropic band (LL) (top left corner). At frequency level 4, the center (shaded) coefficient represents $X_{4,1}[m,n]$ and the surrounding coefficients are $X_{4,1}[u,v]$ with $u = \{m-4, m-3, \dots, m+3, m+4\}$ and $v = \{n-4, n-3, \dots, n+3, n+4\}$.

where $\mathbf{X}_{l,\theta}[m,n]$ is the transform coefficient at orientation θ , spatial frequency location $[m,n]$ and frequency level l (see Fig. 3). $I_{\Theta,l,\theta}[m,n]$ is the sum of transformed coefficients spanning all orientations. $I_{\Upsilon,l,\theta}[m,n]$ is the sum of neighboring coefficients about $\mathbf{X}_{l,\theta}[m,n]$ (Fig. 3). The neighborhood, $A_l = (2l+1)^2$, is a square area surrounding $\mathbf{X}_{l,\theta}[m,n]$, whose size is dependant on the frequency level of $\mathbf{X}_{l,\theta}[m,n]$. Thus, coefficients from the highest frequency level would have the largest neighborhood. This approach attempts to equalize the uneven spatial coverage inherent in multiresolution representations. The neighborhood variance $\sigma_l[m,n] = (1/A_l) \sum_{u=m-l}^{m+l} \sum_{v=n-l}^{n+l} (\mathbf{X}_{l,\theta}[u,v] - \mu_l[m,n])^2$, with $\mu_l[m,n]$ representing the neighborhood mean, has been added to the inhibition process to account for texture masking [42]. Exponents $p_{\mathcal{Z}}$ and q are governed by the condition $p_{\mathcal{Z}} > q > 0$ according to [33]. Currently, q is set to 2.

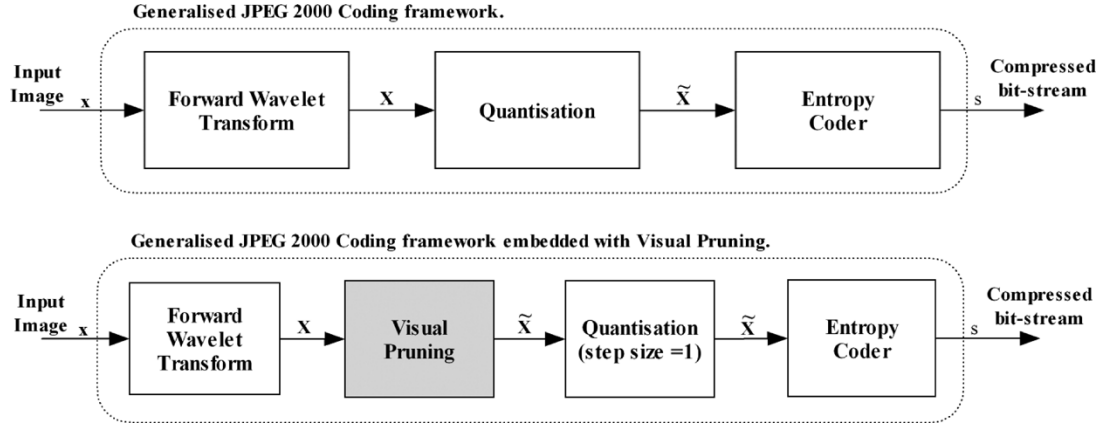


Fig. 4. Generalized JPEG 2000 (hierarchical Mallat wavelet transform [41]) coder embedded with the visual pruning function. Modularity and simplicity is achieved without disrupting the bit-stream flow. Thus; a specialized decoder is not required.

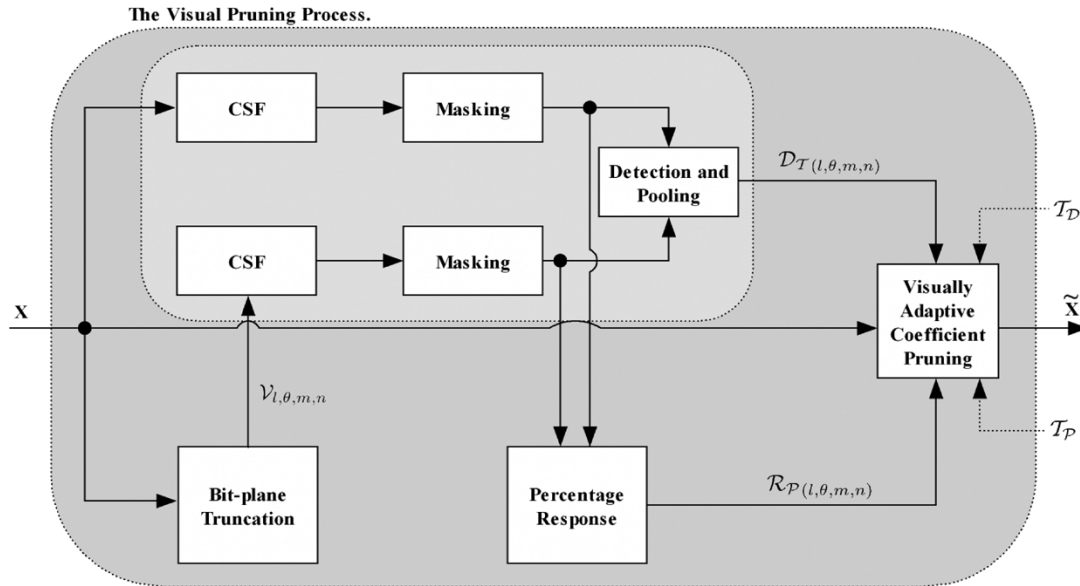


Fig. 5. Visual pruning function.

3) *Detection and Pooling*: The final component of the model detects the perceptually significant difference between two images. A squared-error (l_2 norm) function defines the distortion within each masking channel. The total distortion is the sum of the distortions over all masking channels, given as [12]

$$D_{T_i, \theta}[m, n] = \sum_{\mathcal{Z}} g_{\mathcal{Z}} \cdot |R_{a_{\mathcal{Z}, i, \theta}}[m, n] - R_{b_{\mathcal{Z}, i, \theta}}[m, n]|^2 \quad (8)$$

where $R_{a_{\mathcal{Z}, i, \theta}}$ and $R_{b_{\mathcal{Z}, i, \theta}}$ are the masking responses of the two images, a and b , respectively, and $g_{\mathcal{Z}}$ being the channel gain (Table II) with $\mathcal{Z} \in \{\Theta, \Upsilon\}$, where Θ and Υ are the interorientation and intrafrequency masking domains. The pooling equation, (8), pools the spatial and orientation masking responses for each individual coefficient and provides an overall perceptual distortion, $D_{T_i, \theta}[m, n]$, between the two images.

B. Coder Adaptation

Whilst it is possible to embed the vision model into the post-compression rate-distortion optimization stage [10] and to re-

place the MSE distortion metric as done so previously [11], the approach taken here embeds the vision model into a visual pruning (VP) function. This modular approach enables the VP function to be easily adapted into other Wavelet based coding frameworks while maintaining bit-stream compliance [13], [14] (Fig. 4). The VP function (Fig. 5) consists of two stages. For each frequency level, l , at each orientation, θ , and at a particular location (m, n) , the first stage takes in a reference coefficient, $\mathbf{X}_{l, \theta}[m, n]$ and generates a set of distorted coefficients, $\mathcal{V}_{l, \theta, m, n}$. These distorted coefficients are generated through progressive bit-plane truncation from the least significant bit (lsb), upwards. That is, given $\mathcal{B} \in \{0, 1, 2, \dots, \beta\}$

$$\tilde{\mathbf{X}}_{i, \theta}[m, n] = \left\{ \left\lfloor \frac{\mathbf{X}_{l, \theta}[m, n]}{2^i} \right\rfloor \times 2^i \mid i \in \mathcal{B} \right\} \quad (9)$$

where $\lfloor \cdot \rfloor$ is a truncation function, $\tilde{\mathbf{X}}_{i, \theta}[m, n]$ is a coefficient from a reference image truncated to the i^{th} bit-plane, where β specifies the maximum number of bit-planes for the largest coefficient in the transformed image, \mathbf{X} . Thus, $\mathcal{V}_{l, \theta, m, n} = \{\tilde{\mathbf{X}}_{i, \theta}[m, n] \mid i \in \mathcal{B}\}$. Immediately, each distorted

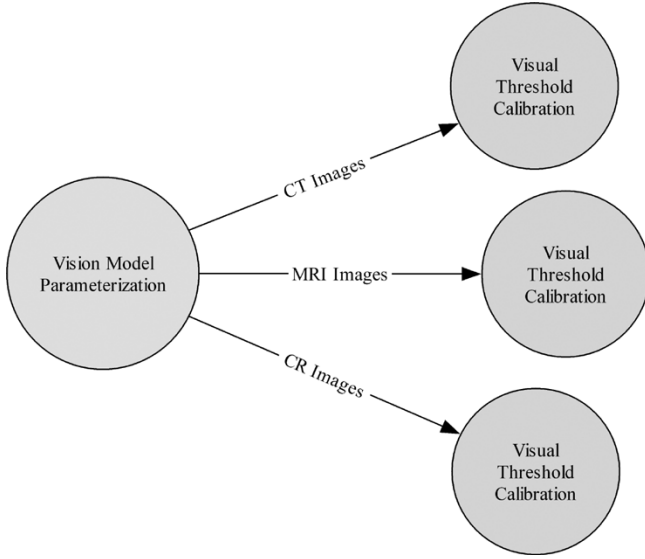


Fig. 6. The parameterization process. The first stage is the vision model parameterization. The second stage is the calibration of the visual thresholds. In this figure, the vision model parameters are subjectively parameterized once to capture the visual nature of the images governed by the visual mechanics of the observers. Visual thresholds are then calibrated to the JNND level for each type (modality) of medical images for optimal coding performance.

coefficient from the set, $\mathcal{V}_{l,\theta,m,n}$ is compared with the reference coefficient using the vision model described in the previous section. This generates a set of perceptual distortion measures, $\mathcal{D}_{\mathcal{T}}(l,\theta,m,n) = \{D_{\mathcal{T}_{i(l,\theta)}}[m,n] \mid i \in \mathcal{B}\}$, and a set of percentage responses, $\mathcal{R}_{\mathcal{P}}(l,\theta,m,n) = \{R_{\mathcal{P}_i(l,\theta,m,n)} \mid i \in \mathcal{B}\}$. The percentage response, $R_{\mathcal{P}_i(l,\theta,m,n)}$, for a given reference coefficient and a distorted coefficient, is defined as

$$R_{\mathcal{P}_i(l,\theta,m,n)} = \left\{ \frac{\sum_{\mathcal{Z}} R_{i_{(\mathcal{Z},l,\theta)}}[m,n]}{\sum_{\mathcal{Z}} R_{o_{(\mathcal{Z},l,\theta)}}[m,n]} \mid i \in \mathcal{B} \right\} \quad (10)$$

where $R_{o_{(\mathcal{Z},l,\theta)}}[m,n]$ and $R_{i_{(\mathcal{Z},l,\theta)}}[m,n]$ are, respectively, the masking response, for a reference and a distorted coefficient, taken from (3). $\mathcal{Z} \in \{\Theta, \Upsilon\}$, denotes the orientation and local responses, respectively. Equation (10) provides a measurement of the depreciation of the response energy over both the intrafrequency and interorientation channels.

The last stage gathers the set of distortion measures, $\mathcal{D}_{\mathcal{T}}(l,\theta,m,n)$, the set of percentage responses, $\mathcal{R}_{\mathcal{P}}(l,\theta,m,n)$ and performs visually adaptive coefficient pruning. By comparing $\mathcal{D}_{\mathcal{T}}(l,\theta,m,n)$ and $\mathcal{R}_{\mathcal{P}}(l,\theta,m,n)$ to a set of predetermined JNND thresholds, $\mathcal{T}_{\mathcal{D}}$ and $\mathcal{T}_{\mathcal{P}}$, respectively, a coefficient is truncated [see (9)] to a perceptually optimal bit-plane level, i_{opt} , only when a distortion measure from $\mathcal{D}_{\mathcal{T}}(l,\theta,m,n)$ is less than or equal to a JNND threshold, $\mathcal{T}_{\mathcal{D}}(l,\theta)$ and when a percentage response from $\mathcal{R}_{\mathcal{P}}(l,\theta,m,n)$ is less than or equal to a percentage response threshold $\mathcal{T}_{\mathcal{P}}(l,\theta)$. Thus

$$\tilde{\mathbf{X}}_{i_{\text{opt}},l,\theta}[m,n] = \left\lfloor \frac{\mathbf{X}_{l,\theta}[m,n]}{2^{i_{\text{opt}}}} \right\rfloor \times 2^{i_{\text{opt}}} \quad (11)$$

where

$$i_{\text{opt}} = \max \left\{ i \in \mathcal{B} \mid (\mathcal{D}_{\mathcal{T}_{i(l,\theta,m,n)}} \leq \mathcal{T}_{\mathcal{D}}(l,\theta)) \right. \\ \text{and} \\ \left. (\mathcal{R}_{\mathcal{P}_{i(l,\theta,m,n)}} \leq \mathcal{T}_{\mathcal{P}}(l,\theta)) \right\}.$$

All transform coefficients are subjected to this perceptual filtering operation except for those in the isotropic lowpass band (LL). The values in both $\mathcal{T}_{\mathcal{D}}$ and $\mathcal{T}_{\mathcal{P}}$ were derived from subjective experiments. For each orientation, θ , and each frequency level, l , there are unique pairs of predetermined thresholds $\mathcal{T}_{\mathcal{D}}(l,\theta)$ and $\mathcal{T}_{\mathcal{P}}(l,\theta)$, for $\theta = \{1, 2, 3\}$ and $l = \{1, 2, 3, 4, 5\}$.

C. Parameterization

There are two stages to the parameterization process. The first concerns the vision model parameters (Table II), which are subjectively determined [12] by capturing the visual nature of the images governed by the visual mechanics of the observers. The second is a set of visual thresholds $\mathcal{T}_{\mathcal{D}}$ and $\mathcal{T}_{\mathcal{P}}$, which are mapped to the JNND level for perceptually lossless encoding (Fig. 6). Vision model parameters affect the accuracy of the visual distortion measure and the thresholds capture the visually sensitive nature of the images. Therefore, a change in vision model parameters would require a re-calibration of the thresholds. Hence, suboptimal vision model parameters may impede the compression ratio gain once the visual thresholds have been mapped to the JNND level for a particular type (modality) of medical images.

Nevertheless, while a direct importation of vision model parameters (Table II) from an 8-bit natural image coder [11], [12] to a 16-bit medical image coder may be less than adequate for the desired application, these imported parameters provide a rough indication of the performance capability of the PC. The set of visual thresholds $\mathcal{T}_{\mathcal{D}}$ and $\mathcal{T}_{\mathcal{P}}$ were obtained through the testing of approximately 5120 (32×32 pixels) 16-bit medical greyscale subimages. These subimages originated from a particular base image (512×512 pixels), which was distorted in 20 different ways through bit-plane filtering. These 20 distorted images were then partitioned into 256 (32×32 pixels) individual pieces. Subimage testing is preferred in this case over the complete image testing because it is able to quantify the different local threshold levels in different regions within images, i.e., the segmented test is better equipped to capture the localized variation in image quality. Due to the varying nature of medical images (CT, CR and so forth), each set of medical images of a particular type required their own set of $\mathcal{T}_{\mathcal{D}}$ and $\mathcal{T}_{\mathcal{P}}$ thresholds for optimal coding performance, thus, the above procedure was subsequently repeated. Once the JNND level of test materials have been mapped, the thresholds $\mathcal{T}_{\mathcal{D}}(l,\theta)$ and $\mathcal{T}_{\mathcal{P}}(l,\theta)$ can be determined by soliciting the responses (3) and (8) of the subimages in the JNND map. In other words, only subimages at the JNND level will be used to determine the thresholds $\mathcal{T}_{\mathcal{D}}(l,\theta)$ and $\mathcal{T}_{\mathcal{P}}(l,\theta)$. A stringent test of flipping back and forth the encoded image with the original image was used. This employs the temporal sensitivities of the HVS to ensure that “distortion flickers” between the two images are not visible.

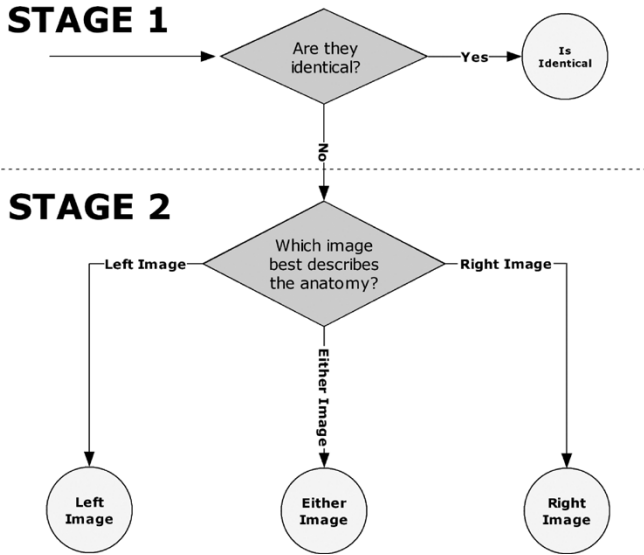


Fig. 7. Double blind 2SFC experiment flow. The first stage asks the viewer if the two images are identical (“Yes” or “No”). If “No”, then the second stages asks the viewer which of the image in the pair best describes the anatomy (“Left,” “Right,” “Either”).

IV. CASE STUDY

A. Methodology

There are two aims to this study. The first is to ascertain if differences can be perceived between original images and images compressed by the PC. The secondary purpose is to determine if the PC retains clinical information. To answer these questions, a double blind two staged forced choice (2SFC) (Fig. 7) comparative experiment with two benchmark coders was conducted. These coders were the JPEG-LS LOCO-I lossless coder [16] and the JPEG-LS NLOCO near lossless coder [16] with $d = 2$. The error, d , specifies the maximum pixel difference between the original image and the NLOCO compressed image. Only 3 coders were chosen to simplify the analysis of the results.

A 2SFC is favored over a standard dichotomous forced choice (DFC) [43] experiment, since it minimizes systematic errors in the experiment [44]. The experiment was conducted using 2 calibrated Barco 10-bit 20.8-inch medical grade LCDs, which were placed next to one another. Each screen has a maximum resolution of 1536×2048 pixels (3 megapixels). The source material for the test consisted of 30 medical images from 30 different patients, comprising CT, CR and MRI images. Each medical image has a bit-depth of 16 bpp and a spatial resolution ranging from 208×256 pixels up to 3732×3062 pixels. 30 source images produced a total of 90 test images from 3 coders, each with 30 images used in 9 permutations, e.g., LOCO versus LOCO, PC versus LOCO, LOCO versus PC, and so forth. Each viewer had to evaluate 30 different pairs of images.

In order to ascertain the effectiveness of the PC, viewers were allowed to change the windows and levels [30] as they would in their examinations. The testing software changed the windows and levels simultaneously on both screens. This was applied through linear contrast stretching [4] [see (12)] as specified in the DICOM standard [30]

$$\tilde{x}[i, j] = C(m \cdot (x[i, j] - l_x)) \quad (12)$$

TABLE III
MODEL 1 (ARTEFACT ANALYSIS) ANALYSES WHEN TWO TRULY IDENTICAL IMAGES ARE SHOWN. THIS TABLE SHOWS THE COMPUTED AVERAGE RESPONSE (I_{avg}) AS IDENTICAL, WHICH SERVES AS A “BASELINE” FOR “IDENTICAL”

	A vs A	B vs B	C vs C
I	0.69	0.79	0.77
\bar{I}	0.31	0.21	0.23
$I + \bar{I}$	1.00	1.00	1.00
I_{avg}	0.75		

where $x[i, j]$ and $\tilde{x}[i, j]$ are the original and transformed pixels respectively at location (i, j) , m , the gradient, is defined as $m = 1023/(u_x - l_x)$ with $u_x = w_c + w_w/2$ and $l_x = w_c - w_w/2$, and C is a clamping function that maintains pixel values within an unsigned 10-bit range. Here w_c and w_w are the window center (brightness) and the window width (contrast), respectively [30]. Differences were readily perceivable between 8-bit and 10-bit screens, with the 10-bit screens having better contrast. Nevertheless, the images were still perceptually lossless in both instances.

The experiment was conducted at Monash Medical Centre in a standard radiological reporting room, simulating a typical evaluation environment. However, only one person and an experiment supervisor were allowed in the room during the experiment to prevent contamination of results. The 2SFC approach poses the following questions to each viewer: 1) Are the image pairs identical (“Yes” or “No”) and 2) If not, which of the image in the pair best displays the anatomy (“Left,” “Right,” or “Either,” if no preference was held). It is important to note that each viewer was asked not to determine the pathology or to perform a diagnosis, so as to avoid inducing “emotional” biases into the results. A total of 31 medical expert viewers completed the experiment, amongst whom 27 were radiologists and 4 were radiographers. Of the 27 radiologists, 15 were consultants and 12 were trainee imaging consultants. An interesting observation was that the viewing behavior of each expert was not a simple evaluation of picture quality but also a natural reaction to search for pathological and physical traits in the medical image.

B. Analysis and Results

Four data analysis models have been employed to interpret the raw results of the experiment. The first two models aid the first aim of the experiment, while the last two models aid the last aim of the experiment, as discussed in the previous subsection. The first model determines a baseline for “identicalness” (the average success rate of identification). This baseline is ascertained by gathering the results for the first question of the experiment when two truly identical images are shown. An image is shown on both the left and the right screens and the viewers are asked if they are identical.

To simplify the representation of the results, the lossless LOCO coder is labeled as “A,” the PC is labeled as “B” and the NLOCO coder is labeled as “C.” The results of the first model are shown in Table III. In Table III, I is the response in each category perceived as identical, \bar{I} is the response in each category perceived as not identical and I_{avg} is the average of

TABLE IV
MODEL 2 DETERMINES IF THE PROPOSED CODER PRODUCES IMAGES THAT ARE PERCEIVED AS BEING IDENTICAL TO THE ORIGINAL

	A vs B	A vs C	C vs B
Response as Identical (x)	0.74	0.70	0.71
Mean (\bar{x}) (Baseline for identical)	0.75		
Standard ($\sigma_{\bar{x}}$) Deviation	0.01		
Z-Score (z)	-0.78	-3.57	-2.85
Outcome	Not Significant (Is Identical)	Is Significant (Not Identical)	Is Significant (Not Identical)

TABLE V
MODEL 3 (ARTEFACT ANALYSIS) ANALYSES WHEN TWO TRULY IDENTICAL IMAGES ARE SHOWN. THIS TABLE SHOWS THE COMPUTED AVERAGE RESPONSE AS IDENTICAL AND EITHER (IE_{avg}), WHICH SERVES AS A "BASELINE" FOR "CLINICALLY RETAINING"

	A vs A	B vs B	C vs C
IE	0.72	0.81	0.79
\bar{IE}	0.28	0.19	0.21
$IE + \bar{IE}$	1.00	1.00	1.00
IE_{avg}	0.78		

the responses in I , across all three categories. In the event of when an image compressed by A is shown against the same image compressed by A (A versus A), 69.39% (rounded up) in proportion have correctly determined (a success) that the two images are identical. In the event of B versus B, 78.79% have correctly determined that the two images are identical and in the case for C versus C, 76.64% have correctly responded that the two images are identical. Thus, the baseline for "identicalness" is formed by taking the average of these successes, across each category, which leads to an average success rate of correct identification of 74.94%.

The second model takes into account of the baseline determined in the first model. Here, the aim is to determine if the results for A versus B, A versus C, and C versus B are, in each event, considered as identical, that is, if they significantly far from the baseline of "identicalness" (74.94%). This is done by computing the z-score [see (13)] of the two tail 95% confidence interval test [44], [45]

$$z = \frac{x - \bar{x}}{\sigma_{\bar{x}}} \quad (13)$$

where z is the computed z-score, x and \bar{x} represent the sample observation and sample mean, respectively, and $\sigma_{\bar{x}} = \sqrt{b_s \times b_f}$ is the sample standard deviation, where b_s and b_f represent the binomial success and binomial failure, respectively [44], [45].³ Hence, when $|z| \leq 1.96$, there is no significant variation [45] and, thus, there is no perceivable difference of statistical significance. The results for this model can be found in Table IV. Therefore, in the case for A versus B, 73.81% of selections considered both images as identical. By taking the average success rate of identification (74.94%) found in the previous model and

³Normal distribution is approximated with binomial data.

TABLE VI
MODEL 4 DETERMINES IF THE IMAGES BY THE ORIGINAL (LOSSLESS) CONTAIN THE SAME DIAGNOSTIC VALUE AS THE IMAGES COMPRESSED WITH THE PROPOSED CODER

	A vs B	A vs C
Response as Identical & Either (x)	0.77	0.73
Mean (\bar{x}) (Baseline for Identical & Either)	0.78	
Standard ($\sigma_{\bar{x}}$) Deviation	0.01	
Z-Score (z)	-0.46	-3.38
Outcome	Not Significant (Is Clinically Retaining)	Is Significant (Not Clinically Retaining)

TABLE VII
CODING PERFORMANCE, IN TERMS OF BIT-RATE, OF LOCO, PC AND NLOCO $d = 2$. EACH IMAGE HAS A MAXIMUM BIT-DEPTH OF 16 BITS PER PIXEL

Image Name	Dimensions (pixels)	Image Type	Bitrate (bpp)		
			LOCO	PC	NLOCO _{d2}
Ankle 1	2572 × 2040	CR	5.852	3.945	3.557
Ankle 2	1516 × 2044	CR	5.695	3.758	3.395
Body 1	512 × 512	CT	4.191	2.650	2.494
Body 2	512 × 512	CT	5.933	4.475	4.116
Body 3	512 × 512	CT	6.851	4.675	4.517
Brain 1	512 × 512	CT	3.380	2.040	1.963
Brain 2	512 × 512	CT	3.841	1.595	2.058
Brain 3	512 × 512	MR	4.687	3.981	2.959
Brain 4	208 × 256	MR	3.898	2.500	1.624
Brain 5	512 × 512	CT	3.681	2.163	2.165
Brain 6	512 × 512	CT	3.597	1.427	1.893
Brain 7	512 × 512	CT	6.175	4.641	4.339
Brain 8	512 × 512	CT	3.962	1.640	2.124
Brain 9	512 × 512	CT	3.190	1.213	1.616
Chest 1	2496 × 2048	CR	6.786	4.803	4.468
Chest 2	2496 × 2048	CR	6.821	4.818	4.506
Chest 3	2496 × 2048	CR	6.259	4.320	3.948
Chest 4	2496 × 2048	CR	6.198	4.288	3.894
Chest 5	2496 × 2048	CR	6.020	4.061	3.737
Chest 6	1516 × 2044	CR	5.912	4.015	3.598
Elbow	2044 × 1514	CR	6.177	4.253	3.868
Knee	512 × 512	MR	7.143	5.897	4.838
Leg	2040 × 2570	CR	5.759	3.789	3.422
Liver	512 × 512	CT	5.029	2.690	2.783
Neck	2040 × 2570	CR	4.674	3.042	2.779
Pelvis	3732 × 3062	CR	5.915	4.093	3.683
SideBrain	256 × 256	MR	5.972	4.996	3.745
Spine 1	2040 × 2570	CR	6.437	4.498	4.173
Spine 2	512 × 512	CT	3.526	2.620	1.888
Spine 3	512 × 512	CT	6.106	4.675	4.255
AVERAGE			5.322	3.585	3.280

the standard deviation (0.014594), the z-score computed was -0.775240 . By taking a 95% two tail test, $0.775240 \leq 1.96$, the results show that 73.81% is not significantly far from 74.94% and therefore, in this experiment, images compressed by B are seen as identical without any perceivable difference of statistical significance to the images compressed by A. This process is repeated for cases, A versus C and C versus B, both of which were not seen as identical.

In the third and fourth models, the aim is to determine if the same level of clinical information is retained statistically between the original and a compressed image by either LOCO

TABLE VIII
CODING PERFORMANCE, IN TERMS OF BIT-RATE, OF LOCO, JPEG 2000 (J2KL) AND NLOCO $d = 1$. EACH IMAGE HAS A MAXIMUM BIT-DEPTH OF 16 BITS PER PIXEL

Image Name	Dimensions (pixels)	Image Type	Bitrate (bpp)		
			LOCO	J2KL	NLOCO $_{d1}$
Ankle 1	2572 × 2040	CR	5.852	5.892	4.260
Ankle 2	1516 × 2044	CR	5.695	5.707	4.095
Body 1	512 × 512	CT	4.191	4.327	2.993
Body 2	512 × 512	CT	5.933	6.235	4.696
Body 3	512 × 512	CT	6.851	6.987	5.246
Brain 1	512 × 512	CT	3.380	3.558	2.321
Brain 2	512 × 512	CT	3.841	3.500	2.493
Brain 3	512 × 512	MR	4.687	4.650	3.501
Brain 4	208 × 256	MR	3.898	1.523	2.378
Brain 5	512 × 512	CT	3.681	3.789	2.575
Brain 6	512 × 512	CT	3.597	3.230	2.336
Brain 7	512 × 512	CT	6.175	6.459	4.920
Brain 8	512 × 512	CT	3.962	3.627	2.581
Brain 9	512 × 512	CT	3.190	2.815	2.046
Chest 1	2496 × 2048	CR	6.786	6.847	5.199
Chest 2	2496 × 2048	CR	6.821	6.874	5.235
Chest 3	2496 × 2048	CR	6.259	6.312	4.683
Chest 4	2496 × 2048	CR	6.198	6.288	4.662
Chest 5	2496 × 2048	CR	6.020	6.051	4.472
Chest 6	1516 × 2044	CR	5.912	5.965	4.324
Elbow	2044 × 1514	CR	6.177	6.200	4.608
Knee	512 × 512	MR	7.143	6.903	5.563
Leg	2040 × 2570	CR	5.759	5.740	4.144
Liver	512 × 512	CT	5.029	4.942	3.461
Neck	2040 × 2570	CR	4.674	4.598	3.343
Pelvis	3732 × 3062	CR	5.915	5.993	4.414
SideBrain	256 × 256	MR	5.972	5.927	4.445
Spine 1	2040 × 2570	CR	6.437	6.482	4.863
Spine 2	512 × 512	CT	3.526	3.458	2.381
Spine 3	512 × 512	CT	6.106	6.375	4.842
AVERAGE			5.322	5.316	3.903

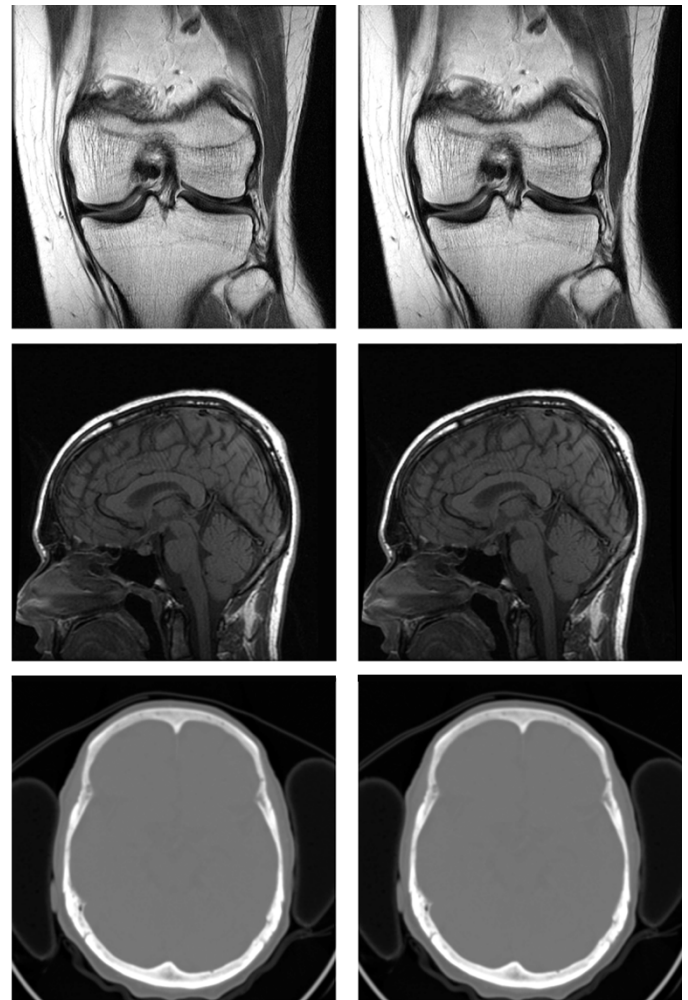


Fig. 8. Left: Original. right: Proposed coder. Top down: knee; sidebrain; brain2.

near-lossless or PC. Both the third and fourth models are similar to the first and second models, respectively. The difference here is that the “either” selections from the second question are taken into consideration as well as the first question of the experiment. The baseline for clinical retention was 77.57%, which can be found in Table V. In Table V, IE is the response in each category perceived as identical and chosen as either. \overline{IE} is the response in each category perceived as not identical and chosen as not either. IE_{avg} is the average of the responses in IE , across all three categories.

In the case for A versus B, images compressed by B had the same level of clinical information as those compressed by A with no perceivable difference of statistical significance (Table VI). Similarly for A versus C, statistically, images compressed by C did not have the same level of clinical information as that of the original images. It is important to re-iterate that this paper focuses on the question of if the images compressed by the PC were perceptually lossless to the original images. Therefore, enhancement effects are not considered.

In terms of coding performance the proposed coder has on average a compression gain of 48% compared with the LOCO lossless coder and a 9% compression loss against the NLOCO near lossless coder (Table VII). However, the statistical analysis in Table IV, has shown that a significant difference between the original images and the images compressed by NLOCO can be

perceived. For completeness, Table VIII includes the JPEG2000 [10] lossless mode and NLOCO $d = 1$ bit-rate. On average, the PC has a compression gain of 9% over NLOCO $d = 1$. Nevertheless, there were no visible distortions in the images compressed by the PC (Fig. 8).

V. CONCLUSION

A perceptually lossless coder for medical images based on the JPEG 2000 coding framework is proposed. The proposed coder outperforms the LOCO coder [16] while preserving the visual fidelity of the image. A double blind comparative experiment with 31 medical experts has shown that no perceivable difference of statistical significance exists between the original images and the corresponding images compressed by the proposed coder, while offering the same level of clinical information. Furthermore, the heart of the proposed coder is the implementation of a visual pruning function combined with a vision model [11], [12] to identify and to remove visually insignificant information, achieving simplicity and modularity. The visual pruning function can be embedded into any Wavelet based coding framework while maintaining bit-stream compliance.

ACKNOWLEDGMENT

The authors would like to thank GE Medical Systems Australia Pty. Ltd. for their generous support in providing the equipment and calibration software, and the 31 radiologists/radiographers at Southern Health Monash Medical Centre and the Faculty of Medicine at Monash University for their active participation in this research and the patients for permission to use their images. They would like to offer special thanks to Prof. W. Anderson of the School of Biomedical Sciences at Monash University, for his foresight and support to this work. Finally, they would like to thank the reviewers for their valuable and constructive comments, which helped improve the quality of this paper. The first author is a recipient of an Australian Postgraduate Award.

REFERENCES

- [1] P. V. Peck, "New medical imaging technology," in *Proc. Special Symp. Maturing Technologies and Emerging Horizons in Biomedical Engineering*, Nov. 1988, pp. 113–114.
- [2] R. Archarya, R. Wasserman, J. Stevens, and C. Hinojosa, "Biomedical imaging modalities: A tutorial," *Comput. Med. Imag. Graphics*, vol. 19, no. 1, pp. 3–25, 1995.
- [3] X. Cao and H. K. Huang, "Current status and advances of digital radiography and PACS," *IEEE Eng. Med. Biol.*, vol. 19, pp. 80–88, Sep.-Oct. 2000.
- [4] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
- [5] B. J. Erickson, *Irreversible Compression of Medical Images*. Great Falls, VA: Society for Computer Applications in Radiology, Nov. 2000.
- [6] T. Ishigaki, S. Sakuma, M. Ikeda, Y. Itoh, M. Suzuki, and S. Iwai, "Clinical evaluation of irreversible image compression: Analysis of chest imaging with computed radiography," *J. Radiol.*, vol. 175, pp. 739–743, 1990.
- [7] R. M. Slone, D. H. Foos, B. R. Whiting, E. Muka, D. A. Rubin, T. K. Pilgram, K. S. Kohm, S. S. Young, P. Ho, and D. D. Hendrickson, "Assessment of visually lossless irreversible image compression: Comparison of three methods by using an image-comparison workstation," *Radiology-Computer Applications*, vol. 215, no. 2, pp. 543–553, 2000.
- [8] A. J. Maeder and M. Deriche, "Establishing perceptual limits for medical image compression," *Proc. SPIE (Medical Imaging 2001: Image Perception and Performance)*, vol. 4324, pp. 204–210, 2001.
- [9] A. B. Watson, "Receptive fields and visual representations," *Proc. SPIE*, vol. 1077, pp. 190–197, 1989.
- [10] D. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standard and Practice*, 1st ed. Norwell, MA: Kluwer Academic, 2002.
- [11] D. M. Tan, H. R. Wu, and Z. Yu, "Perceptual coding of digital monochrome images," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 239–242, Feb. 2004.
- [12] D. M. Tan and H. R. Wu, *Perceptual Image Coding*, H. R. Wu and K. R. Rao, Eds. Boca Raton, FL: CRC, 2005, Digital Video Image Quality and Perceptual Coding.
- [13] D. Wu, D. M. Tan, and H. R. Wu, "Visually lossless adaptive compression of medical images," in *Proc. 4th Int. Conf. Information, Communications & Signal Processing and 4th Pacific Rim Conf. Multimedia*, 2003, pp. 458–463.
- [14] —, "Vision model based approach to medical image compression," in *Proc. Int. Symp. Consumer Electronics*, 2003, Paper ISCE3058.
- [15] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.
- [16] M. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1309–1324, Aug. 2000.
- [17] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1151–1170, Jul. 2000.
- [18] D. A. Clunie, "Lossless compression of greyscale medical images – Effectiveness of traditional and state of the art approaches," in *Proc. SPIE (Med. Imag.)*, vol. 3980, 2000, pp. 74–84.
- [19] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: An overview," *IEEE Trans. Consumer Electron.*, vol. 46, no. 4, pp. 1103–1127, Nov. 2000.
- [20] W. J. Hwang, C. F. Chine, and K. J. Li, "Scalable medical data compression and transmission using wavelet transform for telemedicine applications," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 1, pp. 54–63, Mar. 2003.
- [21] Y. Q. Zhang, M. H. Loew, and R. L. Pickholtz, "A combined-transform coding (CTC) scheme for medical images," *IEEE Trans. Med. Imag.*, vol. 11, no. 2, pp. 196–202, Jun. 1992.
- [22] G. Poggi and R. A. Olshen, "Pruned tree-structured vector quantization of medical images with segmentation and improved prediction," *IEEE Trans. Image Process.*, vol. 4, pp. 734–742, Jun. 1995.
- [23] M. H. Loew and D. Li, "Medical image compression using a fractal model with condensation," in *Proc. 16th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society*, vol. 1, Nov. 1994, pp. 714–715.
- [24] W. B. Pennebaker and J. L. Mitchell, *JPEG: Still Image Data Compression Standard*. New York: Van Nostrand Reinhold, 1992.
- [25] M. J. Askelof, L. Carlander, and C. Christopoulos, "Region of interest coding in JPEG2000," *Signal Process. Image Commun.*, vol. 17, no. 1, pp. 105–111, Jan. 2002.
- [26] M. Penedo, W. A. Pearlman, P. G. Tahoces, M. Souto, and J. J. Vidal, "Region-based wavelet coding methods for digital mammography," *IEEE Trans. Med. Imag.*, vol. 22, no. 10, pp. 1288–1296, Oct. 2003.
- [27] L. F. Verheij, J. A. K. Blokland, A. M. Vossepoel, R. Valkema, J. A. J. Camps, S. E. Papapoulos, O. L. M. Bijvoet, and E. K. J. Pauwels, "Automatic region of interest determination in dual photon absorptiometry of the lumbar spine," *IEEE Trans. Med. Imag.*, vol. 10, no. 2, pp. 200–206, Jun. 1991.
- [28] Y. Chee and K. Park, "Medical image compression using the characteristics of human visual system," in *Proc. 16th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society*, vol. 1, Engineering Advances: New Opportunities for Biomedical Engineers, Nov. 1994, pp. 618–619.
- [29] N. Lin, T. Yu, and A. K. Chan, "Perceptually lossless wavelet-based compression for medical images," *Proc. SPIE (Medical Imaging 1997: Image Display)*, vol. 3031, pp. 763–770, 1997.
- [30] (2004) DICOM Standard. ACR-NEMA. [Online]. Available: <http://medical.nema.org/>
- [31] B. Girod, *What's Wrong with Mean-Squared Error?*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, Digital Images and Human Vision.
- [32] H. R. Wu and K. R. Rao, Eds., *Digital Video Image Quality and Perceptual Coding*. Boca Raton, FL: CRC, 2005, Signal Processing and Communications.
- [33] A. B. Watson and J. A. Solomon, "A model of visual contrast gain control and pattern masking," *J. Opt. Soc. Am. A*, pp. 2379–2391, 1997.
- [34] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, Nov. 1994, pp. 982–986.
- [35] Z. Yu, H. R. Wu, S. Winkler, and T. Chen, "Vision model based impairment metric to evaluate blocking artifacts in digital video," *Proc. IEEE*, vol. 90, no. 1, pp. 154–169, Jan. 2002.
- [36] B. Wandell, *Foundations of Vision*: Sinauer Associates, Inc., 1995.
- [37] F. L. Van Nes and M. A. Bouman, "Spatial modulation transfer function in the human eye," *J. Opt. Soc. Am.*, vol. 57, no. 3, pp. 401–406, 1967.
- [38] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transform," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 587–607, Feb. 1992.
- [39] S. Mallat, "Wavelets for a vision," *Proc. IEEE*, vol. 84, pp. 604–614, Apr. 1996.
- [40] D. Le Gall and A. Tabatabai, "Sub-band coding of digital images using symmetric short kernel filters and arithmetic coding techniques," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 1988, pp. 761–765.
- [41] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
- [42] R. Safranek and J. Johnston, "A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression," in *Proc. IEEE ICASSP*, 1989, pp. 1945–1948.
- [43] R. M. Kaplan and D. P. Saccuzzo, *Psychological Testing: Principles Applications and Issues*, 5th ed. London, U.K.: Thomson Learning, 2001.
- [44] C. White, H. R. Wu, D. Tan, R. L. Martin, and D. Wu, *Experimental Design for Digital Image Quality Assessment*, 1st ed. Zetland, Sydney, Australia: Epsilon Publishing, 2004.
- [45] D. R. Anderson, N. J. Harrison, D. Sweeney, J. A. Rickard, and T. A. Williams, *Statistics for Business and Economics*, 1st ed. New York: Harper Collins, 1989.