

Visually lossless compression of digitized radiographs based on contrast sensitivity and visual masking

Damon M. Chandler^a, Nathan L. Dykes^b, and Sheila S. Hemami^a

^aVisual Communications Lab, School of Electrical and Computer Engineering;

^bDept. of Clinical Sciences, Section of Veterinary Imaging, College of Veterinary Medicine;
Cornell University, Ithaca, NY 14853

ABSTRACT

A visually lossless compression algorithm for digitized radiographs, which predicts the maximum contrast that wavelet subband quantization distortions can exhibit in the reconstructed image such that the distortions are visually undetectable, is presented. Via a psychophysical experiment, contrast thresholds were measured for the detection of 1.15–18.4 cycles/degree wavelet subband quantization distortions in five digitized radiographs; results indicate that digitized radiographs impose image- and frequency-selective effects on detection. A quantization algorithm is presented which predicts the thresholds for individual images based on a model of visual masking. When incorporated into JPEG-2000 and applied to a suite of images, results indicate that digitized radiographs can be compressed in a visually lossless manner at an average compression ratio of 6.25:1, with some images requiring visually lossless ratios as low as 4:1 and as great as 9:1. The proposed algorithm thus yields images that require the minimum bit-rate such that the reconstructed images are visually indistinguishable from the original images. The primary utility of the proposed algorithm is its ability to provide image-adaptive visually lossless compression, thereby avoiding overly conservative or overly aggressive compression.

Keywords: Image compression, wavelet, contrast threshold, visual masking

1. INTRODUCTION

Modern image compression algorithms exploit the fact that the human visual system is an imperfect sensor. In this paradigm, an exact bit-for-bit reconstruction of the original image is unnecessary; rather, the data can be coded in a non-invertible or *lossy* fashion. Lossy compression algorithms can be classified as either *visually lossless* or *visually lossy*, depending on whether or not the distortions are visible in the reconstructed image. In particular, visually lossless compression algorithms aim to encode digital images at the minimum bit-rate such that the original and reconstructed images are indistinguishable when viewed by a human.

Of primary concern in the design of such an algorithm is the determination of a *visually lossless threshold*, which specifies the maximum amount of compression that can be applied before the resulting image appears distorted. This criterion is particularly important for compression of medical imagery wherein the presence of visible compression artifacts may give rise to misinterpretation and thus misdiagnosis. For medical images, visually lossless thresholds have traditionally been reported in terms of average compression ratios, which are measured for specific entropy coders and specific imaging modalities (e.g., 6:1 compression for digital coronary angiograms using JPEG¹). However, because the results of such an approach represent an average over a sample of images, and because some images are inherently more compressible than others, the use of average threshold compression ratios may result in overly aggressive compression of certain images and overly conservative compression of other images. Furthermore, such an approach does not fully exploit the frequency-selective sensitivity of human vision, nor does it account for image-specific effects such as luminance and contrast masking. These properties can be used to better fine-tune the quantization stage of compression.

This paper presents the results of a psychophysical experiment and an associated quantization strategy for visually lossless compression of digitized radiographs. Here, we advocate the use of *contrast detection thresholds* assessed specifically for compression-induced distortions as a measure of the visually lossless threshold. Using a three-alternative forced-choice paradigm, contrast detection thresholds were measured for 1.15–18.4 cycles/degree distortions induced via quantization of single wavelet subbands in five digitized radiographs. Results indicate that radiographs impose both frequency-selective and image-selective effects on detection thresholds: On average,

D.M.C.: E-mail: dmc27@cornell.edu; N.L.D.: E-mail: nld2@cornell.edu; S.S.H.: E-mail: hemami@ece.cornell.edu

greater detection thresholds were found for 1.15 and 18.4 cycles/degree distortions, and lower thresholds were found for 2.3 and 4.6 cycles/degree distortions. However, these thresholds also demonstrated significant variations from image to image, a finding which precludes the use of average detection thresholds for visually lossless compression; rather, these data suggest that detection thresholds must be predicted on a per-image basis.

To this end, an algorithm which predicts contrast thresholds for detection of wavelet subband quantization distortions in individual images based on a model of visual masking is presented. The predicted contrast detection thresholds are used as part of a contrast-based quantization scheme in which the wavelet subbands of an image are quantized such that the resulting distortions exhibit the threshold contrasts in the reconstructed image. The proposed quantization strategy was incorporated into JPEG-2000^{2,3} and the modified coder was used to compress 20 digitized radiographs. The compression results revealed an average threshold compression ratio of 6.25:1, with some images exhibiting threshold compression ratios as low as 4:1 and as great as 9:1. The primary utility of the proposed algorithm is its ability to provide *image-adaptive* visually lossless compression, thereby avoiding overly conservative compression of certain images and overly aggressive compression of others.

This paper is organized as follows. Section 2 provides a brief review of previous work and an overview of wavelet-based image compression. Section 3 describes the psychophysical experiment performed to assess masked detection thresholds for wavelet subband quantization distortions; the results and analyses of these experiments are presented in Section 4. Section 5 describes the visual masking model, quantization algorithm, and compression results. General conclusions are presented in Section 6.

2. BACKGROUND

2.1. Previous work

Studies which have assessed the effects of compression on medical images can generally be divided into two types: (1) those which evaluate the effects of compression of diagnostic quality; and (2) those which seek the maximum allowable compression such that the compressed images* are visually indistinguishable from the original images. In this paper, we focus solely on the later type, under the assumption that an image which is visually lossless is also diagnostically lossless.⁴ Visually lossless thresholds have traditionally been assessed by first generating images compressed to various bit-rates, and then conducting psychophysical experiments in which the compressed images are visually compared to the originals. The psychophysical experiments typically involve either visual-discrimination tasks or quality ratings. The results of these experiments are then reported as the maximum compression ratio (minimum bit-rate) such that the compressed image is visually indistinguishable from the original.

For example, using JPEG compression and a quality-rating paradigm, Brennecke *et al.*¹ reported that digital coronary angiograms can be compressed up to a compression ratio of 6:1 while still yielding compressed images that are rated visually equivalent to the originals. Similarly, Slone *et al.*⁴ tested the effects of both JPEG compression and a wavelet-based compression algorithm on CRT-displayed postanterior digital chest radiographs. Relative to 16-bits/pixel originals, the compression ratio at the visually lossless threshold occurred between 8:1 and 16:1 for JPEG and between 8:1 and 11:1 for WTCQ. In a later study⁵ using both CRT- and film-based displays, Slone *et al.* found between 10:1 and 20:1 JPEG compression to yield visually lossless images, depending on the viewing distance. Foos *et al.*⁶ evaluated the performance of JPEG-2000 on a variety of digitized and natively digital image-types. For most image-types, the compressed images were rated equivalent to the originals at bit-rates above 2.0 bits/pixel (compression ratios < 7:1, relative to 15-bits/pixel originals).

The results of these studies suggest that the average compression ratio at the visually lossless threshold lies somewhere between 6:1 and 20:1, depending on the compression algorithm and the image's content/type. Unfortunately, these dependencies severely confound the utility of the results. For example, decreasing the codeblock size in JPEG-2000 significantly decreases the performance of the entropy encoder. In this case, compressing an image to a compression ratio of 6:1 might yield an image which is as distorted as that produced using a much larger compression ratio had the compression been applied using JPEG-2000's default codeblock size. Moreover, results assessed in terms of average compression ratios do not take into account each image's

*We use the term *compressed image* or *reconstructed image* to refer to the image reconstructed from the compressed data.

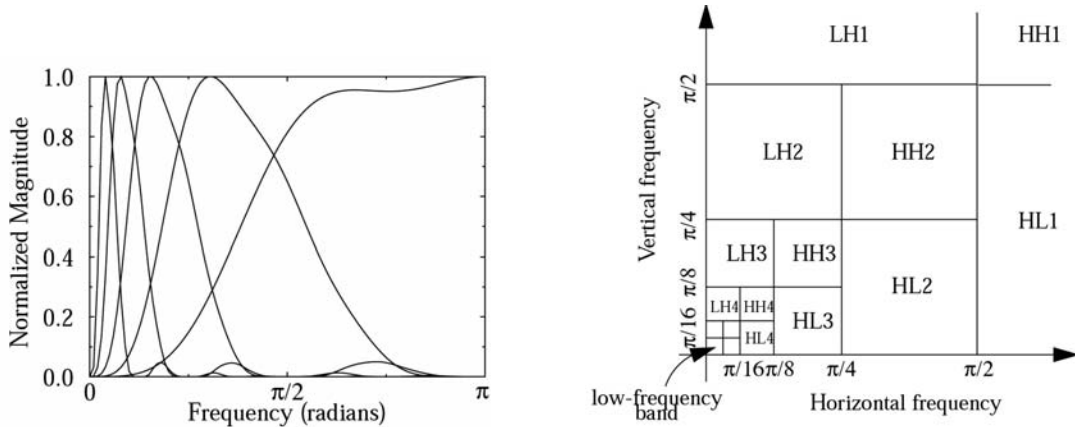


Figure 1. *Left:* One-dimensional frequency response of a five-level hierarchical 9/7 biorthogonal synthesis filter bank. *Right:* Tiling of the two-dimensional frequency plane by a five-level hierarchical wavelet decomposition. Only the upper right quadrant is shown, and the fifth-level bands are not labeled. (See Section 2.2 of the text for details.)

inherent compressibility. Because images of greater predictability are also more compressible, compression ratios will necessarily demonstrate a degree of variability across images, regardless of whether the data are encoded in a lossless or lossy fashion. In the latter case, compressing all images to a fixed compression ratio results in lower distortion for images which are easily compressed and greater distortion for those which are more difficult to compress.

Rather than evaluating visually lossless thresholds based on bit-rates or compression ratios, this paper advocates the use of *contrast detection threshold*, which is defined as the minimum contrast needed to detect a visual target. By investigating and reporting results in terms of the contrast that distortions can exhibit in the reconstructed image such that the distortions are visually undetectable, our data provide a bit-rate-independent measure of visually losslessness. Furthermore, because contrast detection thresholds play a crucial role in our current understanding of human vision⁷⁻⁹ the proposed approach facilitates the use of human-visual-system characteristics to further fine-tune the quantization stage of lossy compression.

2.2. Wavelet-based image compression

State-of-the-art image compression algorithms employ a discrete wavelet transform (DWT) front-end which separates an image into spatial-frequency and orientation components. This process, which is most often performed via a filter-bank/lifting¹⁰ implementation, results in a tiling of the spatial-frequency plane whereupon the image is represented as a series of spatial-frequency bands called *subbands*. Figure 1 depicts the frequency responses of the (one-dimensional) filters and the corresponding tiling of the spatial frequency plane that results from a five-level DWT using the 9/7 biorthogonal filters¹¹⁻¹³ (also used in JPEG-2000).

As shown in Figure 1, an n -level DWT will yield $3n + 1$ subbands; each level contains an LH band, an HL band, and an HH band. The LH (HL) subbands are low-pass (high-pass) filtered horizontally and high-pass (low-pass) filtered vertically and hence contain horizontal (vertical) edge information with frequencies in the range given by the one-dimensional frequency response. Their corresponding spatial-frequency content can be described by their orientation and by their frequencies in cycles per degree of visual angle.¹² The HH bands are high-pass filtered in both directions and thus contain both 45° and 135° edge information. The three subbands in a level are collectively referred to as a *scale*, with finer scales containing higher-frequency information and coarser scales containing lower-frequency information. The coarsest scale contains four subbands, in which the additional band is the low-frequency LL band (represented by the lower left of the tiling shown in Figure 1).

After an image is transformed into its spatial-frequency representation, the coefficients within each subband are quantized. Let $c(s)$ denote a coefficient of subband s . Quantization, which constitutes the lossy stage of image compression algorithms, is a non-invertible process which maps the value of each coefficient $c(s)$ to an output

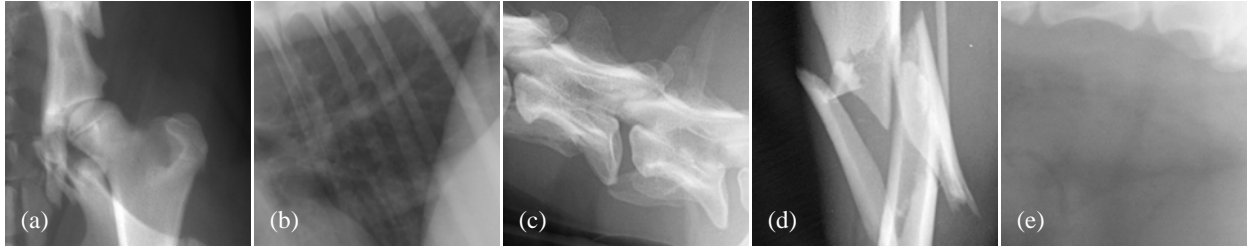


Figure 2. Five 512×512 segments from digitized radiographs which served as masks in the experiment: (a) *hip*; (b) *chest*; (c) *neck*; (d) *leg*; (e) *abdomen*.

level $\hat{c}(s)$ from a discrete set via $\hat{c}(s) = \Delta \times \left\lfloor \frac{c(s)}{\Delta} + \frac{1}{2} \right\rfloor$. The parameter Δ denotes the quantization step size; the larger the step size, the larger the quantization-induced error. The error is given by $e(s) = \hat{c}(s) - c(s)$, which manifests itself in the reconstructed image as a wavelet basis function (distortion) with amplitude proportional to $e(s)$. Quantization of all coefficients of subband s results in distortions which constitute a superposition of scaled wavelet basis functions (see Figure 3).

During compression, the quantization step size (Δ) for each subband is typically selected based on rate-distortion criteria using mean-squared error (MSE) as the distortion metric; this is the approach used in baseline JPEG-2000. However, in order to maximally compress a medical image in a *visually lossless* manner, the step sizes must be selected such that the resulting distortions are just below the threshold of *visual detection*. Attaining this goal requires an understanding of both the visual detectability of the distortions and the masking effects imposed upon this detectability when the distortions are viewed within a medical image. The following section describes a psychophysical experiment designed to assess the effects of medical images on the detectability of wavelet subband quantization distortions.

3. PSYCHOPHYSICAL EXPERIMENT

A psychophysical experiment was performed to assess visual detection thresholds for targets consisting of wavelet subband quantization distortions presented against backgrounds (masks) consisting of five digitized radiographs.

3.1. Apparatus

Stimuli were displayed on an NEC FP2141SB 22-inch monitor at a display resolution of 33.7 pixels/cm and a frame rate of 85 Hz. The digital-pixel-value-to-luminance response of the display was mapped via a digital lookup table to conform to the DICOM grayscale standard display function,¹⁴ with minimum, maximum, and mean luminances of 0.7, 105.7, and 36.9 cd/m^2 , respectively. Stimuli were viewed binocularly through natural pupils in a darkened room at a distance of approximately 58 cm.

3.2. Stimuli

Stimuli consisted of 512×512 -pixel luminance modulations which subtended 3.5×3.5 degrees of visual angle. Each stimulus was composed of a target (wavelet subband quantization distortions) and a mask (digitized radiograph).

Five digitized radiographs obtained from the Cornell College of Veterinary Medicine served as masks in the experiment. The original 2048×2560 images were acquired using a Lumisys 75 laser film scanner (8-bits/pixel grayscale, 146 dots/inch), and then a 512×512 region containing diagnostically relevant information was selected from each image. The 512×512 images contained digital pixel values in the range 0–255 and mean physical luminances of 14.2 cd/m^2 (*hip*), 13.2 cd/m^2 (*chest*), 28.0 cd/m^2 (*neck*), 27.4 cd/m^2 (*leg*), and 25.0 cd/m^2 (*abdomen*); scaled representations of these images are shown in Figure 2.

A discrete wavelet transform (DWT) was applied to each 512×512 image using the 9/7 biorthogonal DWT filters^{11–13} and five levels of decomposition. Targets (distortions) centered at spatial frequencies of 18.4, 9.2,

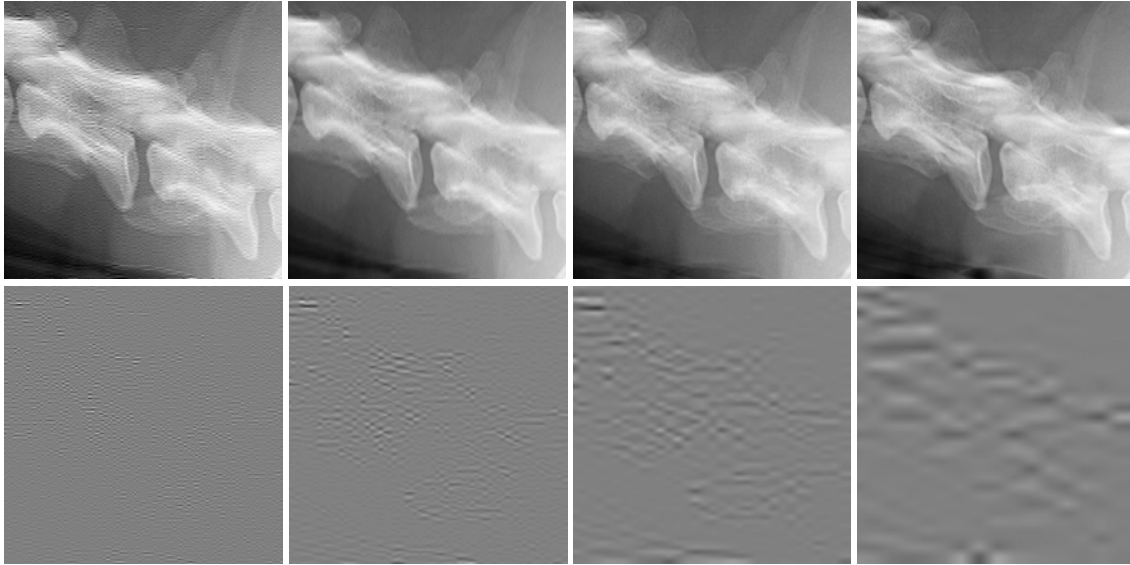


Figure 3. *Top row:* Image *neck* containing distortions generated via quantization of the LH subband at decomposition levels 2, 3, 4, and 5 (left to right); stimuli generated via quantization of the LH level-1 subband are not shown. *Bottom row:* The distortions contained in each image of the top row offset by solid gray.

4.6, 2.3, and 1.15 cycles/degree and one orientation (horizontal) were generated by uniformly quantizing all coefficients of the LH subband at the first through fifth decomposition levels, respectively; the quantization step size was adaptively chosen such that the root-mean-squared (RMS) contrast of the resulting distortions [see Equation (1)] was as specified by the QUEST staircase procedure (described in the following section). Following quantization of the subband, an inverse DWT was applied to generate the distorted image. The top row of Figure 3 depicts image *neck* reconstructed following coarse quantization of individual subbands. To facilitate visibility of the distortions, the bottom row of Figure 3 depicts the distortions presented against a solid gray background.

3.3. Procedures

Thresholds were measured by using a spatial three-alternative forced-choice procedure. On each trial, observers concurrently viewed three adjacent images placed upon a uniform 26.2 cd/m^2 background. Two of the images contained the mask alone and the other image additionally contained one of the previously described targets (distortions); the image to which the target was added was randomly selected at the beginning of each trial. Observers indicated via keyboard input which of the three images contained the target (i.e., which image was distorted). Target RMS contrasts were guided via an adaptive QUEST staircase procedure¹⁵ using software derived from the Psychophysics Toolbox.^{16,17} Contrast detection threshold was defined as the 75%-correct point on a Weibull function, which was fitted to the data following each series of 32-trial tracks.

Each experimental session began with five minutes of adaptation to a blank 26.2 cd/m^2 display. Before each series of trials, observers were briefly shown a high-contrast, spatially randomized version of the distortions to facilitate a signal-known-exactly condition^{18,19} (i.e., to minimize subjects' uncertainty in the orientation and frequency of the target). During each trial, an auditory tone indicated stimulus onset, and auditory feedback was provided to denote correct and incorrect responses. Response time was limited to within 20 seconds of stimulus onset during which time all three images remained visible; observers were instructed to examine all three images before submitting a response.

3.4. Observers

Three observers, ND, DC, and YS, participated in the study. ND was a board-certified radiologist with 30 years of primary diagnostic experience, but with little exposure to compression-induced distortions. DC and YS were engineering graduate students with extensive experience in evaluating and detecting compression-induced

distortions in natural images, but with little exposure to radiographs. All observers had normal or corrected-to-normal visual acuity.

4. RESULTS

Thresholds are reported in terms of the RMS contrast²⁰ of the target (distortions). Let t denote a target presented against an image mask m ; the RMS contrast of t , $C(t)$, is defined as

$$C(t) = \frac{1}{\mu_{L(m)}} \left(\frac{1}{N} \sum_{i=0}^N [L(t_i) - \mu_{L(t)}]^2 \right)^{1/2} \quad (1)$$

where $\mu_{L(m)}$ denotes the average luminance of the mask, $\mu_{L(t)}$ denotes the average luminance of the target, $L(t_i)$ denotes the luminance of the i^{th} pixel of t , and N represents the total number of pixels.

The luminance values in Equation (1), which are defined in standard units of cd/m^2 , are given by²¹

$$L = (\epsilon + kI)^\gamma \quad (2)$$

where L and I represent luminance and (8-bit) digital pixel value, respectively. The parameters ϵ , k , and γ represent, respectively, the black-level offset, the pixel-value-to-voltage scaling factor, and the gamma of the display monitor. When the response characteristics of the display monitor used in this experiment were adjusted to match the DICOM grayscale standard display function, these parameters were $\epsilon = 0.922$, $k = 0.008$, and $\gamma = 4.425$.

4.1. Results and Analysis

For the five images tested, contrast detection thresholds for 1.15, 2.3, 4.6, 9.2, and 18.4 cycles/degree distortions averaged over all images and all observers were, respectively, 0.019 ± 0.017 (1 standard deviation), 0.010 ± 0.005 , 0.010 ± 0.003 , 0.013 ± 0.002 , and 0.031 ± 0.007 . On average, observers were least sensitive to (i.e., exhibited the largest thresholds for) 1.15 and 18.4 cycles/degree distortions, and were most sensitive to (lowest thresholds for) 2.3–4.6 cycles/degree distortions. Both a one-way ANOVA and a Kruskal-Wallis test (non-parametric one-way ANOVA) showed a significant effect of spatial frequency on detection thresholds ($df = 4$, $F = 7.22$, $p < 0.001$; and $df = 4$, $X^2 = 12.44$, $p < 0.02$; respectively).

Figure 4(a) depicts contrast thresholds averaged over all images and all observers. Figures 4(b)–4(f) depicts the mean detection thresholds (for all observers) measured for each of the five images. In each graph, the horizontal axis denotes the center spatial frequency of the distortions and the vertical axis denotes the RMS contrast of the distortions. Solid black circles indicate contrast detection thresholds; error bars denote ± 1 standard deviation of the mean. For comparison, light-gray lines/open circles represent previously measured average thresholds for wavelet subband quantization distortions in the unmasked condition.²² Black lines represent fits of the masking model described in Section 5.1.

4.2. Discussion

Previous masking experiments have traditionally revealed greatest elevations in contrast thresholds when the spatial frequency of the mask is nearby that of the target.^{22–26} In particular, when the mask is a natural image, which is composed predominantly of low spatial frequencies, the greatest elevation in threshold is found to occur for low spatial frequency distortions, whereas distortions of high spatial frequencies typically incur only a minor elevation in threshold.²²

For the five images tested, the results of this experiment indicate that digitized radiographs also impose frequency-selective elevations of detection thresholds. Whereas detection thresholds for wavelet subband quantization distortions presented in the absence of a mask are typically greatest for high-frequency distortions and lowest for low-frequency distortions, the data of Figure 4(a) show that, on average, thresholds measured in the presence of a radiograph are most elevated for 1.15 cycles/degree distortions (by a factor of 11.2) and least elevated for 18.4 cycles/degree distortions (factor of 1.2).

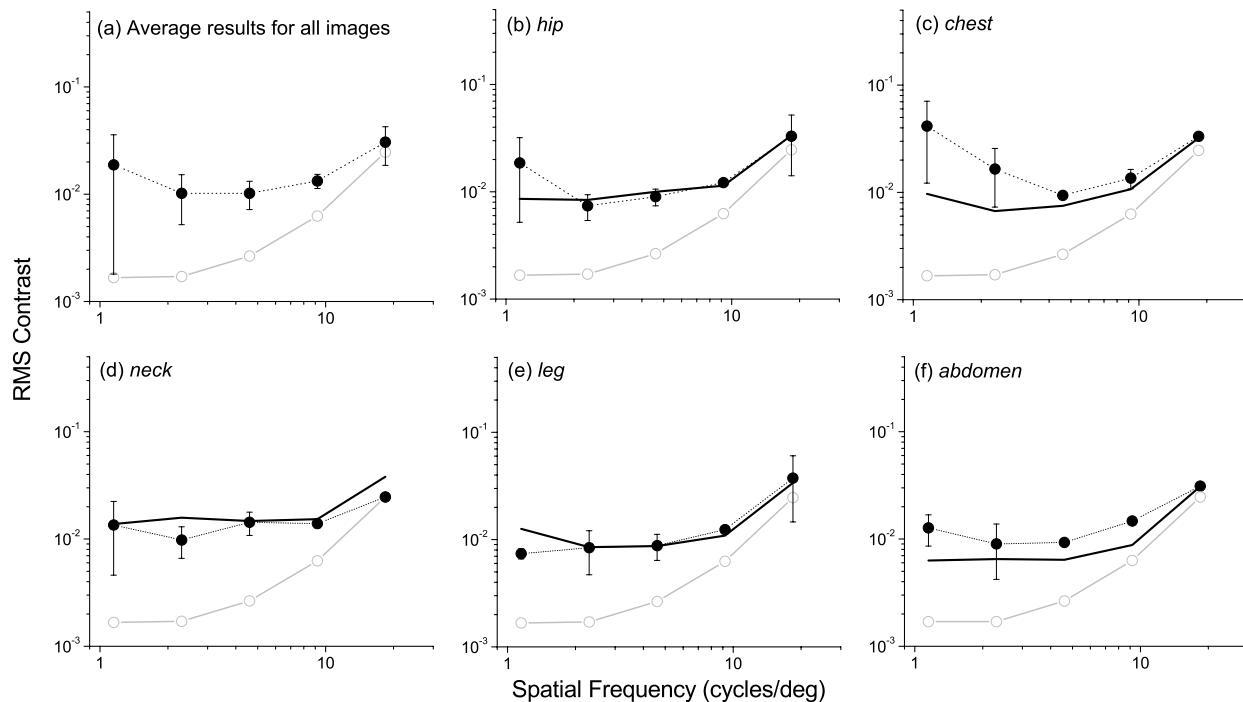


Figure 4. *Black circles:* Contrast detection thresholds for wavelet subband quantization distortions presented against various digitized radiographs; (a) average thresholds for all images; (b)-(f) thresholds for individual images; error bars denote ± 1 standard deviation of the means. *Open circles/light-gray lines:* Contrast detection thresholds measured in a previous study for wavelet subband quantization distortions presented against a solid gray background.²² *Solid black lines:* Detection thresholds predicted by the masking model described in Section 5.1.

However, as noted by the relatively large standard deviations in Figure 4(a) and by the results for individual images in Figures 4(b)-4(f), detection thresholds show a marked variation from image to image. These results suggest that digitized radiographs exhibit both frequency-selective *and* image-selective effects on an observer’s ability to detect wavelet subband quantization distortions and therefore advocate the use of image-specific results during compression. The following section describes a method for predicting detection thresholds on a per-image basis and an algorithm for quantizing each subband such that the compression-induced distortions exhibit the predicted RMS contrasts in the reconstructed image.

5. APPLICATION TO COMPRESSION

Given a digitized radiograph and parameters (ϵ , k , and γ) of the display monitor on which the image is to be viewed, detection thresholds for 1.15–18.4 cycles/degree distortions are first predicted based on a model of visual masking (Section 5.1). These data are then adjusted to account for visual summation (Section 5.2), and then the DWT subbands of the image are quantized such that the resulting distortions are at the adjusted threshold contrasts (Section 5.3). Finally, the quantized coefficients are entropy encoded to generate the compressed stream. The following sections provide the details of the algorithm; compression results are presented in Section 5.4.

5.1. Predicting detection thresholds

To predict detection thresholds on a per-image basis, a descriptive model of visual masking is employed. In this model, the psychophysical response $R(t|m)$ of an observer in the task of detecting target t in the presence of mask m is given by

$$R(t|m) = \frac{[g_t C(t)]^p}{b^q + [g_m C(m)]^q} \quad (3)$$

where $C(t)$ and $C(m)$ denote the contrast of the target and the contrast of the mask, respectively; g_t and g_m represent the contrast gain of the detecting and inhibitory mechanisms, respectively; p and q model the output nonlinearities of these respective mechanisms; and b represents a semi-saturation constant designed to prevent an infinite response in the unmasked condition (i.e., when $C(m) = 0$).

At the threshold of masked detection, $R(t|m) = 1$ and $C(t) = CT(t|m)$, where $CT(t|m)$ denotes the masked contrast threshold of the target; Equation (3) can then be rearranged as

$$CT(t|m) = \frac{(b^q + [g_m C(m)]^q)^{1/p}}{g_t} \quad (4)$$

to yield $CT(t|m)$ given the contrast of the mask $C(m)$ and given the model parameters g_t , g_m , p , q , and b . For simplicity, we set $p = q = 2$ and $b = 0.01$. The contrast gains of the detecting mechanisms can be determined by using thresholds assessed in the unmasked condition,²² in which case $C(m) = 0$, $CT(t|m) = CT(t)$, and then Equation (4) is used to obtain the contrast gains via $g_t = b/CT(t)$. The only remaining parameters, the contrast gains of the inhibitory mechanisms $\{g_m\}$, were fitted to masked thresholds for distortions measured using natural-texture backgrounds.²⁷ The $\{g_t\} = \{5.99, 5.84, 5.11, 2.46, 0.35\}$ and $\{g_m\} = \{0.84, 0.83, 0.74, 0.42, 0.16\}$ were used for 1.15, 2.3, 4.6, 9.2, and 18.4 cycles/degree mechanisms, respectively.

Thresholds measured using small, relatively homogeneous masks are reasonably well-fitted by Equation (4); however, this equation does not generally perform well at predicting thresholds for targets presented against non-homogeneous, full-sized images. To overcome this limitation, the original image is divided into overlapping blocks, and then Equation (4) is applied to each block to compute *local contrast thresholds*; these data are then used to estimate contrast thresholds for detection of distortions in the full-sized image.

5.1.1. Step 1: Compute local contrast thresholds

To compute local contrast thresholds, the original image first is divided into overlapping $M \times M$ blocks, where the block size $M = 8, 16, 32, 64,$ or 128 for 18.4, 9.2, 4.6, 2.3, or 1.15 cycles/degree distortions, respectively. These block sizes were selected based on the spatial extent of the wavelets, and the total number of blocks is set to either the number of coefficients in the corresponding subband or 64×64 , whichever is greater. The local RMS contrast of the image in each block, $C(m_{blk})$, is then computed via Equation (1) using $t = m$ and $N = M \times M$; this process yields a $C(m_{blk})$ for each $M \times M$ block.[†] The $\{C(m_{blk})\}$ are then used with Equation (4) to compute a corresponding set of local contrast thresholds $\{CT(t_{blk}|m_{blk})\}$. Figure 5(a) depicts local contrast thresholds predicted for detection of 4.6 cycles/degree distortions in image *leg*; each point corresponds to a $CT(t_{blk}|m_{blk})$ predicted for the $M \times M$ block centered at the corresponding spatial location in the image; brighter points correspond to greater thresholds.

5.1.2. Step 2: Generate the baseline distortions

For each block, $CT(t_{blk}|m_{blk})$ obtained in Step 1 specifies the contrast threshold for detecting target t_{blk} in the corresponding $M \times M$ region of the image (m_{blk}). However, these data do not take into account the spatial location of the target, i.e., *where* in the image the target may appear; in particular, targets consisting of quantization distortions exhibit positive spatial correlations with the image. As a result, the $\{CT(t_{blk}|m_{blk})\}$ alone are insufficient at predicting thresholds for detection of quantization distortions in the full-sized image.

To account for the spatial correlations, baseline distortions are generated via actual quantization of the target's corresponding subband[‡] followed by an inverse DWT. This process yields zero-mean distortions, t_0 , that are spatially correlated with the original image (cf the bottom row of Figure 3). As described next, the distorted image, \hat{m} , is generated by scaling and then adding t_0 to the original image (i.e., $\hat{m} = m + \alpha t_0$, for some scaling factor α).

[†]To delineate edges from textures, each $M \times M$ block is recursively subdivided and then $C(m_{blk})$ is set to the minimum contrast of all sub-blocks. The depths of recursion are respectively, 0, 0, 1, 2, 2 for $M = 8, 16, 32, 64, 128$.

[‡]Quantization is performed using step size $\Delta = 200\sigma/2^n$, where σ denotes the standard deviation of the subband and n denotes its DWT decomposition level.

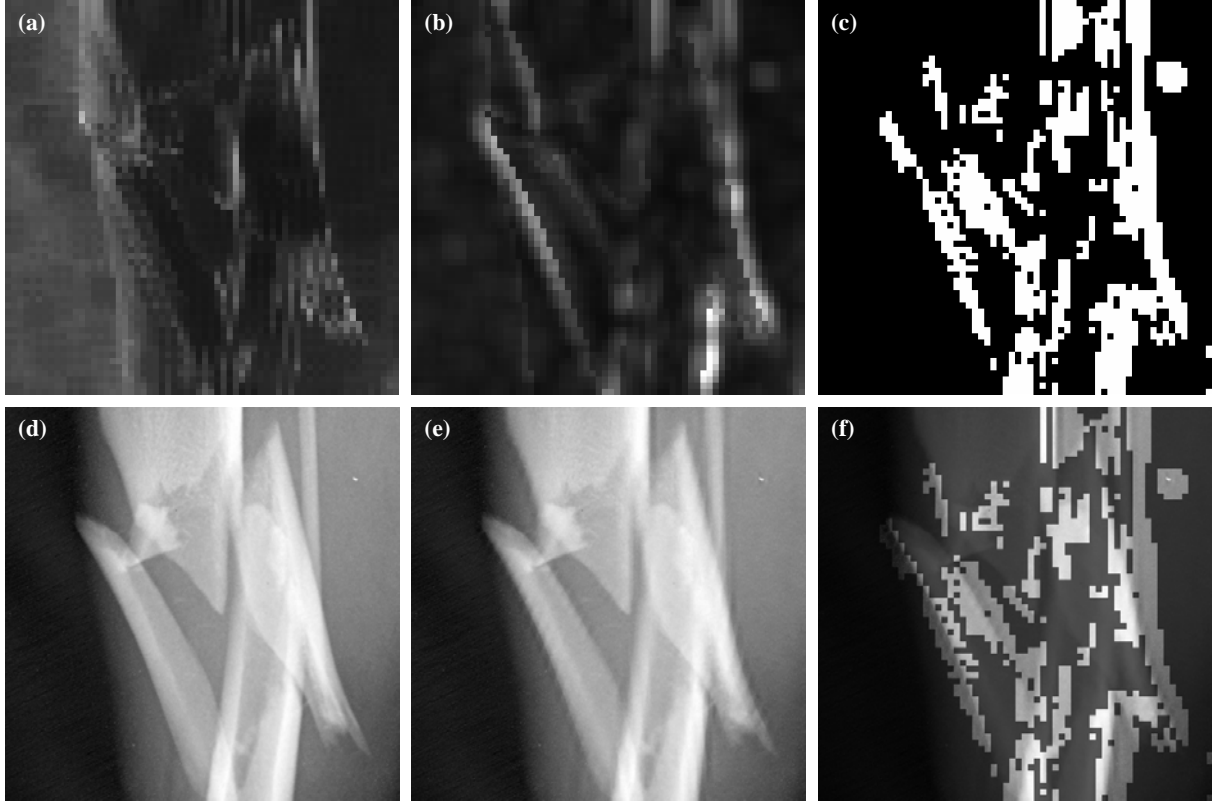


Figure 5. (a) Local contrast thresholds predicted in the initial stage of the masking model (Section 5.1.1); brighter points denote greater thresholds. (b) Local distortion contrasts computed from the distortions in (e). (c) A binary map indicating blocks in which the local contrast of the distortions is greater than the corresponding local contrast threshold. (d) The original image *leg*. (e) Image *leg* containing 4.6 cycles/degree HL distortions. (f) Translucent overlay of (c) on (e).

5.1.3. Step 3: Iterate to find contrast threshold

Let $C(t_{blk})$ denote the RMS contrast of the distortions in an $M \times M$ block of the distorted image. Using the $\{CT(t_{blk}|m_{blk})\}$ obtained in Step 1, and using t_0 generated in Step 2, a bisection search is performed in which t_0 is scaled and then added to the original image in an iterative fashion until 25%[§] of the blocks contain visible distortions, i.e., $C(t_{blk}) > CT(t_{blk}|m_{blk})$ in 25% of the blocks. The following pseudo-code outlines this iterative procedure:

1. Set $\alpha_{lo} = 0.05$ and $\alpha_{hi} = 10$
2. Repeat
 3. Set $\alpha = \frac{1}{2}(\alpha_{lo} + \alpha_{hi})$
 4. Generate the distorted image via $\hat{m} = m + \alpha t_0$
 5. Compute the local contrast of the distortions, $C(t_{blk})$, in each $M \times M$ block of the distorted image
 6. Set $p =$ percentage of blocks in which $C(t_{blk}) > CT(t_{blk}|m_{blk})$
 7. If $p \approx 25\%$ Exit
 8. Else adjust α_{lo} or α_{hi} accordingly
9. While ($\alpha_{lo} < \alpha_{hi}$)

[§]We have found 25% to provide a reasonably accurate fit to thresholds measured for both natural and medical images.

For comparison with the $\{CT(t_{blk}|m_{blk})\}$ depicted in Figure 5(a), the $\{C(t_{blk})\}$ computed in Line 5 are shown in Figure 5(b) for a particular (suprathreshold) value of α . Locations in which the intensity of a point in Figure 5(b) is greater than the intensity of the co-located point in Figure 5(a) correspond to blocks in which the distortions are visible, i.e., blocks in which $C(t_{blk}) > CT(t_{blk}|m_{blk})$; these blocks are shown as a binary map in Figure 5(c). For reference, the original image (m), the distorted image (\hat{m} generated in Line 4), and a translucent overlay of the binary map upon this distorted image, are shown in Figures 5(d), 5(e), and 5(f), respectively. Notice how the masking model successfully identifies the regions which contain suprathreshold distortions.

The iterative procedure typically converges within 4–5 iterations, whereupon the contrast threshold $CT(t|m)$ for detection of distortions (t) in the entire image (m) is estimated as the RMS contrast of the distortions in \hat{m} . Steps 1, 2, and 3 are repeated for each possible target t (i.e., for LH, HL, and HH, 1.15–18.4 cycles/degree distortions) to yield a set of contrast thresholds $\{CT(t|m)\}$. Thresholds predicted via this approach for horizontally oriented (LH) distortions using the five images described in Section 3 are plotted as solid black lines in Figure 4.

5.2. Adjustments for visual summation

The masking model described in the previous section yields contrast detection thresholds for distortions generated via quantization of individual DWT subbands; however, in a compression application, all $L = 3n + 1$ subbands of an n -level DWT are quantized simultaneously. In this case, the resulting distortions constitute a *compound* target containing components of multiple orientations and multiple spatial frequencies. If the compound target is more easily detected than are any of its components, the visual responses to those components are said to have *summed*.⁸ Accordingly, the individual contrast thresholds need to be lowered in order to ensure that the compound target is at the threshold of detection.

Let \mathbf{t} denote a compound target composed of distortions generated via quantization of all L subbands; and let t_s denote its s^{th} component containing distortions generated via quantization of a single subband s . Visual summation is modeled by using a Minkowski metric in which the psychophysical response of the observer to the compound target is given by^{8, 28}

$$R(\mathbf{t}) = \left(\sum_{s=1}^L R(t_s)^\beta \right)^{1/\beta} \quad (5)$$

where the parameter β denotes the amount of summation; e.g, $\beta = \infty$ corresponds to no summation and $\beta = 1$ corresponds to complete or linear summation. Under this model, to ensure that the compound target is at the threshold of detection, the contrast threshold for each component is scaled by the factor $L^{-1/\beta}$.

Whereas summation experiments performed in the unmasked paradigm have traditionally yielded $\beta \in [3.5, 5]$ (termed “probability summation”), summation of responses to targets consisting of wavelet subband quantization distortions presented against natural-image masks was previously found to be much greater, yielding $\beta \in [1.1, 1.8]$.²² Although summation in the presence of medical-image masks remains an open question, here we have elected to err on the side of moderation and assume $\beta = 1.8$. For a 5-level DWT, ignoring the contribution of the LL subband, $L = 15$ and thus the summation-adjusted contrast threshold for t_s , $CT(t_s|m, \mathbf{t})$, is given by $CT(t_s|m, \mathbf{t}) = L^{-1/1.8} CT(t_s|m) \approx 0.22 CT(t_s|m)$, where $CT(t_s|m)$ denotes the contrast threshold for t_s as measured in the previous section.

5.3. Contrast-based quantization

Let $C(t_s)$ denote the RMS contrast of the distortions in the reconstructed image induced via quantization of subband s . To generate a visually lossless image given a set of summation-adjusted contrast thresholds, $\{CT(t_s|m, \mathbf{t})\}$, each subband must be quantized such that $C(t_s) = CT(t_s|m, \mathbf{t})$. This step is accomplished by first computing for each $C(t_s)$ a corresponding $D(s)$, where $D(s)$ represents the mean-squared error between the original and quantized subband coefficients. $D(s)$ is computed from $C(t_s)$ via²⁹

$$D(s) \approx 2^{2n_s} C^2(t_s) \zeta^2 \quad (6)$$

where the accuracy of the approximation depends only on the extent to which the DWT filters deviate from orthogonality.³⁰ Here, n_s denotes the DWT decomposition level of s and the quantity ζ , which accounts for image and monitor characteristics, is defined as

$$\zeta \equiv \frac{\mu_{L(m)}}{k\gamma}(\epsilon + k\mu_m)^{1-\gamma} \quad (7)$$

where μ_m and $\mu_{L(m)}$ represent the image’s average digital pixel value and luminance, respectively. [Recall that ϵ , k , and γ are display parameters which govern the relationship between pixel value and luminance; see Equation (2).]

For each subband s , Equation (6) is used to compute a $D(s)$ from the corresponding threshold contrast $C(t_s) = CT(t_s|m, \mathbf{t})$. Quantization of s is then performed by using standard techniques such that MSE in s is as specified by $D(s)$. For example, using JPEG-2000, $D(s)$ can be effected by truncating the bit-planes generated by the embedded coder.³¹ Alternatively, subband s can be quantized in an iterative fashion until $D(s)$ is met to the desired accuracy. (See also the techniques presented in Refs. 32 and 33.)

5.4. Compression results and discussion

We have used the visual masking and quantization strategies presented in the previous sections to supplement the current JPEG-2000 algorithm³ (using quantization via bit-plane truncation).^{††} Figures 6(a) and 6(d) depict original 512×512 images of pelvic and thoracic limbs. The compressed versions of these images are shown in Figures 6(b) and 6(e), respectively. The compressed streams require 1.47 bits/pixel (*pelvic limb*) and 1.96 bits/pixel (*thoracic limb*), resulting in compression ratios of, respectively, 5.45:1 and 4.08:1 (relative to the original 8-bits/pixel representation). The peak signal-to-noise ratios of the compressed images are 43.2 dB and 42.3 dB, respectively. The total RMS contrast of the distortions is approximately 0.02 for both images (using the display parameters specified in Section 4). The distortions contained in the compressed images are presented against a gray background in Figure 6(c) and 6(f). Notice that the distortions, which are visible Figures 6(c) and 6(f), are not detectable in the compressed images as predicted by the masking model. To facilitate viewing, these images are also available online.³⁴

5.4.1. Average compression ratio

For comparison with previous studies, the proposed compression algorithm was applied to a suite of 20 512×512 images containing diagnostically relevant subject matter; these images were cropped from 2048×2560 8-bits/pixel original digitized radiographs obtained in the same manner described in Section 3.2. The 20 images were chosen based on anatomical structure: four images contained shots of thorax, four of abdomen, four of neck/back/tail, four of pelvic limb, and four of thoracic limb. The resulting compression ratios are shown in Figure 7(a). These data yield an average compression ratio of 6.25:1, with a minimum of 4.08:1 and a maximum of 9.21:1; the corresponding bit-rates are, respectively, 1.28, 0.87, and 1.96 bits/pixel. For reference, compression ratios obtained using lossless JPEG-2000 are also plotted in Figure 7(a) as solid gray circles.

Using a wavelet-based trellis-coded-quantization compression algorithm (WTCQ), Slone *et al.*⁴ reported average compression ratios for visually lossless compression to lie somewhere between 8:1 and 11:1 relative to 16-bits/pixel originals; these data correspond to bit-rates of 2.0 and 1.45 bits/pixel, respectively. However, as mentioned in Ref. 4, the pixel values of the 16-bits/pixel originals were reduced to 12 bits/pixel prior to compression as required by the WTCQ coder. When compression ratios are computed relative to 12-bits/pixel originals, the results of Ref. 4 reveal a compression-ratio range of 6:1 to 8.25:1. Our result of 6:25:1 is consistent with these data. However, we stress that the primary utility of the proposed algorithm is its ability to provide *image-adaptive* compression. Had our 20 images been compressed to a fixed compression ratio of, e.g., 7:1, according to our masking model, six of the images would have been compressed in an overly conservative fashion, 13 of the images would have been compressed in an overly aggressive fashion, and only one would have been on par.

^{††}Because a single contrast $C(t_s)$ is selected for each subband, the contrast-based approach presented here operates within the bounds of Part I of the JPEG-2000 standard and therefore generates a fully compliant JPEG-2000 codestream; see Annex E in ISO/IEC FDIS15444-1:2000.²

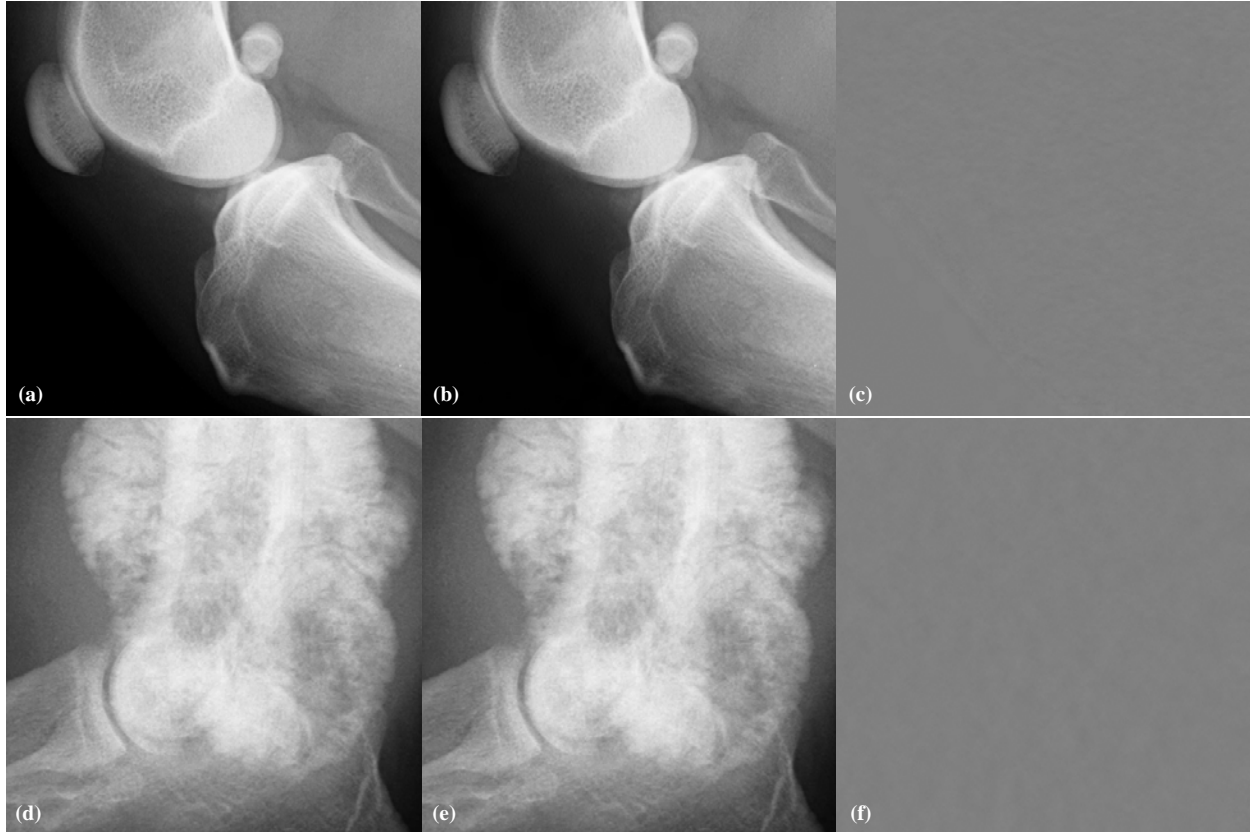


Figure 6. (a) Original image *pelvic limb*. (b) Image *pelvic limb* compressed to the visually lossless threshold using the proposed algorithm (1.47 bits/pixel for this image). (c) The distortions contained in (b) presented against a solid gray background. (d) Original image *thoracic limb*. (e) Image *thoracic limb* compressed to the visually lossless threshold using the proposed algorithm (1.96 bits/pixel for this image). (f) The distortions contained in (e) presented against a solid gray background. These images were designed to be viewed from approximately two picture heights; to facilitate viewing, these images are also available online.³⁴

5.4.2. Verification of visual losslessness

To verify that images compressed using our algorithm are indeed visually lossless and therefore assumed suitable for primary diagnostic interpretation, an additional psychophysical experiment was performed on subject ND (board-certified radiologist; see Section 3.4) using 15 512×512 images randomly selected from the 20 images described in the previous section (three from each anatomical group). The experimental apparatus was identical to that described in Section 3.1. As before, a spatial three-alternative forced-choice procedure was used in which the subject concurrently viewed three adjacent images placed upon a uniform 26.2 cd/m^2 background. Two of the images contained the original and the third image contained the compressed image; the position of the compressed image was randomly selected at the beginning of each trial. Via keyboard input, the subject indicated which of the three images contained the compressed image. This procedure was conducted for 30 trials using each of the 15 compressed images.

The probability of randomly selecting the correct image out of the three possibilities is $1/3$. Therefore, if the compressed images are indeed visually lossless, ND is expected to perform no better than chance and should accordingly submit a correct response (i.e., correctly identify the compressed image) on 33.3% of the trials. The results of this experiment, which are plotted for individual images in Figure 7(b), revealed that ND was correct on 35.6% of the trials, with a standard error of 2.4%. This result is well within the 95% confidence interval of 28.9–37.4% given by the limited number of trials. Although additional testing is certainly required, these preliminary results suggest that images compressed with the proposed algorithm are visually indistinguishable

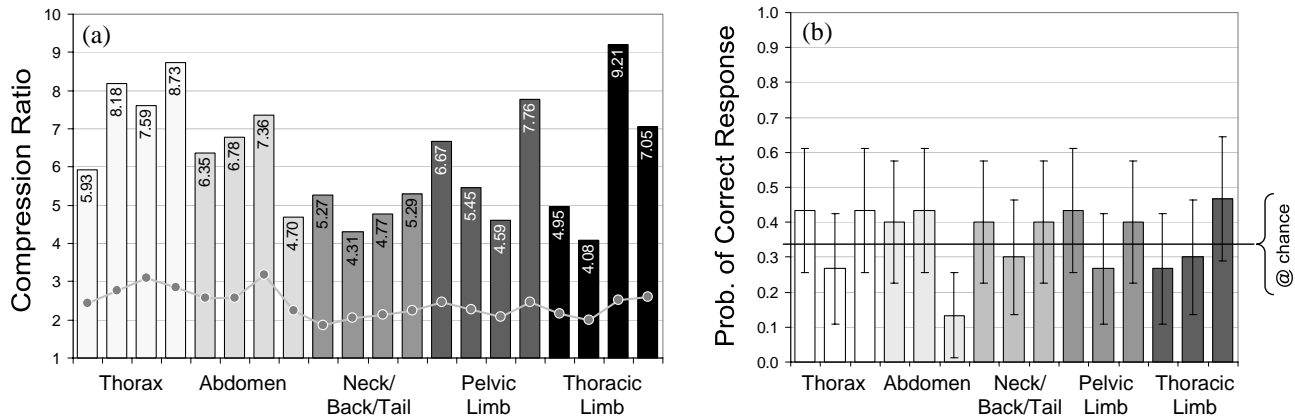


Figure 7. (a) Compression ratios for the proposed algorithm applied to a suite of 20 512×512 segments containing diagnostically relevant subject matter cropped from 2048×2560 8-bits/pixel original digitized radiographs. (b) Performance of subject ND on correctly identifying the compressed image out of three possibilities (two original, one compressed); the probability of randomly selecting the correct image was $1/3$; error bars denote 95% confidence intervals.

from the originals and may therefore be suitable for primary diagnostic interpretation.

6. CONCLUSION

A visually lossless wavelet-based algorithm for compression of digitized radiographs was proposed. Via a psychophysical experiment, contrast thresholds were measured for detection of 1.15-18.4 cycles/degree wavelet sub-band quantization distortions presented against five digitized radiographs. An algorithm which predicts these contrast thresholds based on a model of visual masking is then used as part of a contrast-based quantization strategy. Compression results using a suite of 512×512 radiographs indicate that digitized radiographs can be compressed with JPEG-2000 in a visually lossless manner at an average compression ratio of 6.25:1. This average compression ratio is expected to increase when the algorithm is applied to full-sized images which also contain non-diagnostically relevant regions. We are in the process of designing additional tests to verify that the visually lossless images are also diagnostically lossless.

REFERENCES

1. R. Brennecke, U. Burgel, R. Simon, G. Rippin, H. P. Fritsch, T. Becker, and S. E. Nissen, "American college of cardiology/european society of cardiology international study of angiographic data compression phase iii: measurement of image quality differences at varying levels of data compression," *J. Am Coll Cardiol* **35**(5), pp. 1388–1397, 2000.
2. "Information technology–JPEG 2000 image coding system: Core coding system," Tech. Rep. ISO/IEC FDIS15444-1: 2000, International Organization for Standardization, Geneva, Switzerland, August 2000.
3. I. F.-. ISO/IEC JTC 1/SC 29/WG 1, "Information technology-coding of still pictures-part 5: Reference software," tech. rep., 2001.
4. 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* **215**(2), pp. 543–553, 2000.
5. R. M. Slone, E. Muka, and T. K. Pilgram, "Irreversible JPEG compression of digital chest radiographs for primary interpretation: assessment of visually lossless threshold," *Radiology* **228**(2), pp. 425–429, 2003.
6. D. H. Foos, E. Muka, R. M. Slone, B. J. Erickson, M. J. Flynn, D. A. Clunie, L. Hildebrand, K. Kohm, and S. Young, "JPEG 2000 compression of medical imagery," in *Proc. SPIE Medical Imaging 2000*, G. Blaine and E. Siegel, eds., **3980**, pp. 85–96, February 2000.
7. R. L. DeValois and K. K. DeValois, *Spatial Vision*, Oxford University Press, New York, 1990.
8. N. Graham, *Visual Pattern Analyzers*, Oxford University Press, New York, 1989.

9. D. Regan, *Human Perception of Objects: Early Visual Processing of Spatial Form Defined by Luminance, Color, Texture, Motion, and Binocular Disparity*, Sinauer Associates, Sunderland, MA, 2000.
10. I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.* **4**, pp. 247–269, 1998.
11. J. Villasenor, B. Belzer, and J. Liao, "Wavelet filter evaluation for image compression," *IEEE Trans. Image Process.* **4**, pp. 1053–1060, 1995.
12. A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Trans. Image Process.* **6**, pp. 1164–1175, 1997.
13. M. G. Ramos and S. S. Hemami, "Suprathreshold wavelet coefficient quantization in complex stimuli: psychophysical evaluation and analysis," *J. Opt. Soc. Am. A* **18**, pp. 2385–2397, 2001.
14. "Digital imaging and communications in medicine (DICOM) part 14: Grayscale standard display function," Tech. Rep. PS3.14-2003, NEMA, Rosslyn, VA, 2003.
15. A. B. Watson and D. G. Pelli, "Quest: A Bayesian adaptive psychometric method," *Perception and Psychophysics* **33**, pp. 113–120, 1983.
16. D. H. Brainard, "The psychophysics toolbox," *Spatial Vision* **10**, pp. 433–436, 1997.
17. D. G. Pelli, "The videotoolbox software for visual psychophysics: Transforming numbers into movies," *Spatial Vision* **10**, pp. 437–442, 1997.
18. R. A. Smith and D. J. Swift, "Spatial-frequency masking and Birdsall's theorem," *J. Opt. Soc. Am. A* , pp. 1593–1599, 1985.
19. Y. Zhang, B. Pham, and M. P. Eckstein, "JPEG 2000 encoder options on model observer performance in signal known exactly but variable tasks (SKEV)," in *Medical Imaging 2003: Image Perception, Observer Performance, and Technology Assessment*, E. A. K. D. P. Chakraborty, ed., *Proc. SPIE* **5034**, pp. 371–382, May 2003.
20. B. Moulden, F. A. A. Kingdom, and L. F. Gatley, "The standard deviation of luminance as a metric for contrast in random-dot images," *Perception* **19**, pp. 79–101, 1990.
21. C. Poynton, "The rehabilitation of gamma," in *Proc. SPIE Human Vision and Electronic Imaging III*, B. E. Rogowitz and T. N. Pappas, eds., pp. 232–249, (San Jose, CA), 1998.
22. D. M. Chandler and S. S. Hemami, "Effects of natural images on the detectability of simple and compound wavelet subband quantization distortions," *J. Opt. Soc. Am. A* **20**, July 2003.
23. G. E. Legge and J. M. Foley, "Contrast masking in human vision," *J. of Opt. Soc. Am.* **70**, pp. 1458–1470, 1980.
24. J. Nachmias, "Masked detection of gratings: The standard model revisited," *Vis. Res.* **33**, pp. 1359–1365, 1993.
25. J. M. Foley, "Human luminance pattern mechanisms: masking experiments require a new model," *J. Opt. Soc. Am. A* **11**, pp. 1710–1719, 1994.
26. J. M. Foley and C. C. Chen, "Pattern detection in the presence of maskers that differ in spatial phase and temporal offset: threshold measurements and a model," *Vision Res.* **39**, pp. 3855–3872, 1999.
27. M. Masry, D. Chandler, and S. Hemami, "Digital watermarking using local contrast-based texture maskings," in *Proceedings of the 37th Asilomar Conference on Signals, Systems and Computers*, November 2003.
28. R. F. Quick, "A vector magnitude model of contrast detection," *Kybernetik* **16**, pp. 65–67, 1974.
29. D. M. Chandler and S. S. Hemami, "Dynamic contrast-based quantization for lossy wavelet image compression," *IEEE Trans. Image Process.* , March 2005. in press.
30. M. Rabbani and R. Joshi, "An overview of the JPEG 2000 still image compression standard," *Signal Processing: Image Communication* **17**, pp. 3–48, 2002.
31. M. W. Marcellin, M. A. Lepley, A. Bilgin, T. J. Flohr, T. T. Chinen, and J. H. Kasner, "An overview of quantization in JPEG-2000," *Signal Processing: Image Communication* **17**, pp. 73–84, 2002.
32. D. M. Chandler and S. S. Hemami, "Contrast-based quantization and rate control for wavelet-coded images," *Proc. IEEE Int. Conf. on Image Processing* , 2002.
33. J. Minguillon and J. Pujol, "Uniform quantization error for Laplacian sources with applications to JPEG standard," in *Proc. SPIE Mathematics of Data/Image Coding, Compression, and Encryption*, M. S. Schmalz, ed., **3456**, pp. 77–88, 1998.
34. <http://foulard.ece.cornell.edu/dmc27/med.images/MI2005.htm>.