Visual Quality Assessment of Watermarked Medical Images

Jason A. Dowling [a], Birgit M. Planitz [a] and Anthony J. Maeder [a], Jiang Du [b], Binh Pham * [b] and Colin Boyd*[b], Shaokang Chen [c], Andrew P. Bradley* [c] and Stuart Crozier* [c]

[a] e-Health Research Centre / CSIRO ICT Centre. 20/300 Adelaide St, Brisbane, QLD 4001, Australia (+61 7 3024 1600) {jason.dowling; birgit.planitz; anthony.maeder}@csiro.au
[b] Faculty of Information Technology, Queensland University of Technology. GPO Box 2434, Brisbane QLD 4001, Australia (+61 7 3864 2111) {j.du@isrc.; b.pham@;c.boyd@}qut.edu.au
[c] School of Information Technology & Electrical Engineering, University of Queensland. Brisbane QLD 4072, Australia (+61 7 3365 1111) {shoakang;bradley;stuart}@itee.uq.edu.au

*Affiliated with National ICT Australia

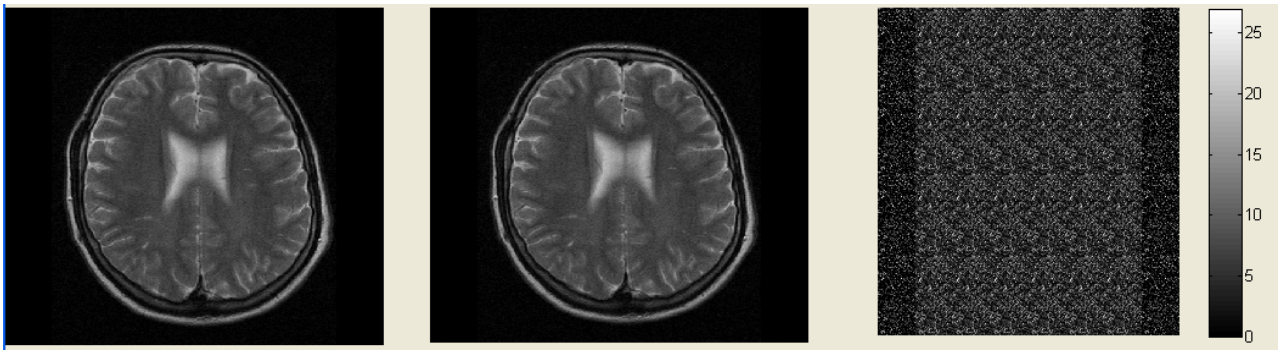
ABSTRACT

Increasing transmission of medical images across multiple user systems raises concerns for image security. Hiding watermark information in medical image data files is one solution for enhancing security and privacy protection of data. Medical image watermarking however is not a widely studied area, due partially to speculations on loss in viewer performance caused by degradation of image information. Such concerns are addressed if the amount of information lost due to watermarking can be kept at minimal levels and below visual perception thresholds. This paper describes experiments where three alternative visual quality metrics were used to assess the degradation caused by watermarking medical images. Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) medical images were watermarked using different methods: Block based Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) with various embedding strengths. The visual degradation of each watermarking parameter setting was assessed using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Measure (SSIM) and Steerable Visual Difference Predictor (SVDP) numerical metrics. The suitability of each of the three numerical metrics for medical image watermarking visual quality assessment is noted. In addition, subjective test results from human observers are used to suggest visual degradation thresholds.

Keywords: digital image watermarking, information hiding, perceptual factors, human observers, medical image modalities

1. INTRODUCTION

Digital image watermarking involves placing a hidden message (the *payload*) within the body of an image. When this message is extracted by a receiver it can be used to prove ownership, to identify if an image has been altered, and the location of any alterations[1]. As medical images are increasingly captured, transmitted and stored in a digital format, a watermarking method that can prove authenticity and identify tampering (or *attacks*) will be valuable. For example, software already exists to insert imperceptibly a lesion into a digital medical image[7]. In a digital image watermarking application an imaging specialist could open this tampered image and receive a warning message that a section of the image (containing the inserted lesion) has been altered.



(a) (b) (c)

Fig. 1. Sample watermarked image. The original image (a), has been watermarked to include a 32 bit message (image b). The pixel value difference between the original and watermarked images is shown in (c).

A crucial requirement for any medical image watermarking scheme is that the method used to embed the hidden message does not visibly degrade image quality[1]. An additional consideration for medical images is that the most commonly used image format (DICOM) also includes text header containing image information and often patient details. Therefore tampering of a DICOM format image can include both the image body and header information.

To increase the *robustness* of watermark insertion, the payload is multiplied by a global scaling factor (called the *embedding strength*) before insertion. The *Bit Error Ratio (BER)* is a commonly used measure of watermark extraction errors. BER represents the ratio of bits incorrectly extracted to the total number of bits extracted[2]. As the embedding strength of watermarking algorithm is increased, there will usually be a corresponding decrease in the BER when the watermark is extracted[9]. However there is also an inverse relationship between the embedding strength of a watermark and visual quality. Therefore, a watermarking method which is more capable of being robust to tampering will usually degrade image quality more than a method with a lower embedding strength. Figure 1 shows a sample Magnetic Resonance Imaging (MRI) image (a) in which a 32 bit payload has been embedded (b) with an embedding strength of 10. Although Figures (a) and (b) appear identical, the difference between the two images is shown in figure 1 (c).

Unfortunately there is no standard method for automatically evaluating the amount of visible degradation of watermarked images. Peak Signal to Noise Ratio (PSNR) is frequently used in the literature; however it does not consider characteristics of the human visual system[8]. Therefore there are two main questions addressed in the current research. Firstly, *is there a reliable alternate image quality measure which can be used to assess watermarked image quality degradation*? Secondly, *based on the results of human subjective studies, is there a threshold of perceptibility which can be used to calibrate an automated quality measure*.

2. METHODOLOGY

To address the first research question, a range of medical images were watermarked using two different methods at three levels of embedding strength. These images were then assessed using three numerical quality measures discussed below. Following this, watermarked versions of these images were altered with three common image manipulation methods, and watermark robustness/fragility were assessed. All program and analysis code was written in Matlab.

2.1 Images

A total of 60 medical images were used in this study, sourced from the University of Queensland (*UQ*), and the CSIRO ICT Centre BioMedIA Lab (*BML*). To investigate the effects of image modality Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) images were used. The size of these images was either 256x256 or 512x512 pixels. In

most images the bit range of was [0,12], although some images with a range of [0,9], [0,10], and [0,12] were also included.

2.2 Watermarking methods

The encoding and decoding procedures used by both frequency domain methods, DWT and DCT, are presented below.

Encoding:

1. Divide original image into bxb blocks
2. Generate payload of length 32 bits from image DICOM header
3. Generate watermark from payload using PN sequences
4. Compute DWT/DCT of each bxb image block
5. Embed watermark in each bxb image block
6. Compute IDWT/IDCT of each bxb image block

Decoding:

1. Divide watermarked image into bxb blocks
2. Generate PN sequences for '0' and '1' bits
3. Compute DWT/DCT of each bxb watermarked image block
4. Correlate pre-specified sections of DWT/DCT block with '0' and '1' PN sequences
5. Select bit ('0' or '1') where PN sequence corresponds to highest correlation as current bit value
6. Return extracted 32-bit payload message for each image block

2.3 Tamper detection method

To investigate the effect of image changes on watermark extraction and visual quality three common types of image manipulation (or tampering) were applied to watermarked images at three different levels of degradation (low, medium and high). The embedding strengths used to generate each of the 60 watermarked images were selected based on the initial DWT and DCT watermarking results (summarized in Figures 1 and 2) to produce the lowest extraction errors, while retaining reasonable visual quality results (PSNR~50dB).

The tampering methods used (with the parameter values for low, medium and high levels of degradation in brackets) were: *edge enhancement* (Sobel with α=0.1; 0.5; 0.9) , *histogram stretching* (window/level method with threshold = background mean; background mean+signal variation; background mean + 2 x signal variation) and *JPEG Compression* (Quality factors: 100; 75; 50).

Tampering was detected by computing the BER of each manipulated, watermarked block and locating image blocks where BER>0. If BER>0 for an image block, there is a high probability that the block has been manipulated. Note that BER >0 for non-degraded images, when watermark extraction errors occur.

The objective of the image manipulation tests was to assess the *fragility* of watermarking when using selected parameter settings. Note that DWT and DCT are not directly comparable at these parameter settings, but the BER results (when applying image manipulations) quantify watermark fragility at the preset visual quality level.

2.4 Visual degradation assessment

PSNR, the Structural Similarity Measure (SSIM), and the Steerable Visual Difference Predictor (SVDP) were used to measure the amount of visual quality degradation between original and watermarked medical images. These three

metrics were chosen as they range from placing a low (PSNR) to high (SVDP) emphasis on the Human Visual System (HVS).

The first measure, PSNR, is a common numerical model that is used to determine visual differences between original and watermarked images. PSNR is expressed as

$$PSNR = 10\log_{10}\frac{M \times N \times Max(x_{ij}{}^2)}{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(x_{ij} - y_{ij})^2}$$

where x is the original image, y is the modified image, and xij and yij are intensity values for the pixels on row i and column j on images x and y respectively. M is the number of rows, N is the number columns, and Max(xij) is the greatest pixel value in the original image x. PSNR does not always successfully mimic the response of the HVS, but is commonly used in watermarking. It is poor at comparing different watermarking methods, but provides a simple indicator for quantifying the similarity between original and watermarked images[8].

The second measure used was SSIM, which is a region-based numerical metric that places more emphasis on the HVS than PSNR. Mathematically, it is expressed as

$$SSIM(Rx, Ry) = LC(Rx, Ry)^{\alpha} \cdot CC(Rx, Ry)^{\beta} \cdot SC(Rx, Ry)^{\lambda}$$

SSIM measures the similarity in luminance (LC), contrast (CC), and structure (SC) of image regions[10] α,β, and λ are ≥ 1 and are used to weight the importance of each of the three components. Luminance comparison is given by

$$LC(Rx, Ry) = \frac{2\mu_{Rx}\mu_{Ry} + C_1}{\mu_{Rx}^2 + \mu_{Ry}^2 + C_1}$$

where μRx and μRy are the means of image regions Rx and Ry respectively, and C1 is a constant. Contrast comparison is expressed as

$$CC(Rx, Ry) = \frac{2\sigma_{Rx}\sigma_{Ry} + C_2}{\sigma_{Rx}^2 + \sigma_{Ry}^2 + C_2}$$

where σRx and σRy are the standard deviations and σRx2 and σRy2 are the variances of image regions Rx and Ry respectively, and C2 is a constant. Finally, structure comparison is given by

$$SC(Rx, Ry) = \frac{\sigma_{RxRy} + C_3}{\sigma_{Rx}\sigma_{Ry} + C_3}$$

where σRxRy is the correlation coefficient between regions Rx and Ry, and C3 is a constant. SSIM is a region based metric, so the effects of watermarking certain image regions, rather than the degradation of the whole image, are computed. This is important for readers of medical images, as it is undesirable to have visible changes to regions of interest on medical images.

The final measure used was the SVDP which is used to measure Just Noticeable Differences (JND) between two images. JND is a psycho-visual model based on the HVS. The HVS has a limited sensitivity so that it does not react to small stimuli and cannot distinguish between signals with infinite precision. This limitation is dependant on characteristics of signals, such as luminance, frequency, orientation, and contrast. Thus, there exist visibility thresholds for HVS, which can be considered as the minimal perceptible differences[3,4]. This metric compares the original image and watermarked image in different sub-bands and different orientations, and calculates the number of JNDs between them. Generally, the SVDP model is composed of four operations: luminance normalisation, computing the contrast sensitivity function, applying the steerable transformation and applying a visual difference predictor. Currently, most medical imaging display devices, such as monitors, automatically normalise luminance before display. For example, the monitor proposed for visual quality testing, the SIEMENS SMD 21500 LCD monitor, complies with the CIE/DICOM greyscale response. That is, it follows the DICOM/GSDF standard so that the perceived brightness is linearly proportional to the discrete pixel values in medical images. The GSDF is defined as a function of the JND index as the following:

$$\log_{10} L(j) = \frac{a + c \cdot Ln(j) + e \cdot (Ln(j))^2 + g \cdot (Ln(j))^3 + m \cdot (Ln(J))^4}{1 + b \cdot Ln(j) + d \cdot (Ln(j))^2 + f \cdot (Ln(j))^3 + h \cdot (Ln(j))^4 + k \cdot (Ln(j))^5},$$

where $a = -1.3011877$, $b = -2.5840191 \times 10^{-2}$, $c = -8.024636 \times 10^{-2}$, $d = -1.0320229 \times 10^{-1}$, $e = 1.3646699 \times 10^{-1}$, $f = 2.8745620 \times 10^{-2}$, $g = -2.5468404 \times 10^{-2}$, $h = -3.1978977 \times 10^{-3}$, $k = 1.2992634 \times 10^{-4}$, $m = 1.3635334 \times 10^{-3}$.

In being compliant with the GSDF, it was assumed that the monitor normalises the luminance and that normalisation did not need to be performed for visual quality testing.

In the second step a 2D CSF was applied to filter the image as if viewed from a certain distance. The closest distance from the observer to the image was set as one times the image height, which is closer than is generally used by readers of medical images.

The steerable decomposition applied was developed from the Matlab Pyrtools package (Laboratory for Computer Vision, New York University[5]), which decomposes the image into four different levels.

Finally, the visual difference predictor that was used to describe the possibility of detection of the differences for each pixel and the image quality is measured by two parameters: $P_{\max}$ and P_{sum}:

$$P_{\max} = \max_{i,j \in I} (P(i,j)),$$

$$P_{sum} = \left(\sum_{i,j \in I} P(i,j)^4 \right)^{0.25}.$$

2.5 Subjective Quality assessment

To address the second question considered in this research subjective testing was also conducted so that visual thresholds for the PSNR and SVDP numerical metrics could be determined. Visual thresholds are defined such that human observers would not be able to detect significant differences between an original image and images which had been degraded with Gaussian noise. Watermarked images were not used in this experiment as an unbiased method of calibrating the visual model was required.

The visual quality experiment was conducted with 20 volunteers from the CSIRO ICT e-Health Centre and National ICT Australia. Each experiment took approximately 20 minutes to complete. After reading an instruction sheet and signing a consent form, the visual acuity of each participant was assessed using a Snellen chart. Those without normal or corrected-to-normal visual acuity (n=2) were excluded.

Participants who passed the acuity screening (n=20) were led into a dimly lit lab approximately 3.5 x 3 metres in size. The only sources of lighting in this room were from a Siemens 21" gray scale flat screen panel display (SMD 21500), and a 20 Watt halogen desk lamp positioned 20cm behind the monitor. The halogen light faced away from the monitor and was angled downwards at 45◦ to provide a diffuse light source.

To assist participants determine the difference between the original and degraded images used in the experiment, they were required to initially complete a training task. During this task eight pairs of medical images were displayed on the Seimens monitor. Each image-pair consisted of an original image and one of four degraded versions which had been modified using the Matlab imnoise function with mean=0 and variance values of 1*(10^-7), 5*(10^-8), 1*(10^-8) and 3*(10^-9) . Participants were able to move between the original and degraded images by pressing "1" or "2" on the numeric keypad and hit "enter" to continue to the next image.

Finally, participants completed the main testing task. This task was similar to the training exercise above, however 15 base images were each processed with four levels of Gaussian noise to generate a set of 60 image pairs. Participants were asked to review each pair of images and select the original image. Once they had decided which displayed image was the original, participants pressed the enter key to record their choice and move to the next image. Participants were instructed to select images randomly if they were unable to detect any difference between the two image versions.

3. RESULTS

3.1 Watermarking embedding strength

BER results for the images watermarked with the DWT and DCT methods are shown in Figures 2 and 3 respectively. These results do not include image tampering. The PSNR, SSIM, and SVDP Max results are shown in Figures 4 (DWT) and 5 (DCT). In these graphs the embedding strength is shown to scale on x-axes of plots where the mean (-o-) and standard deviation (-+-) are shown. The minimum and maximum values are represented with the symbol -x-. As the DWT method results in a lower amount of visual degradation for higher embedding strengths, the DWT embeddings strength values in figures 2 to 4 are higher than the DCT method.

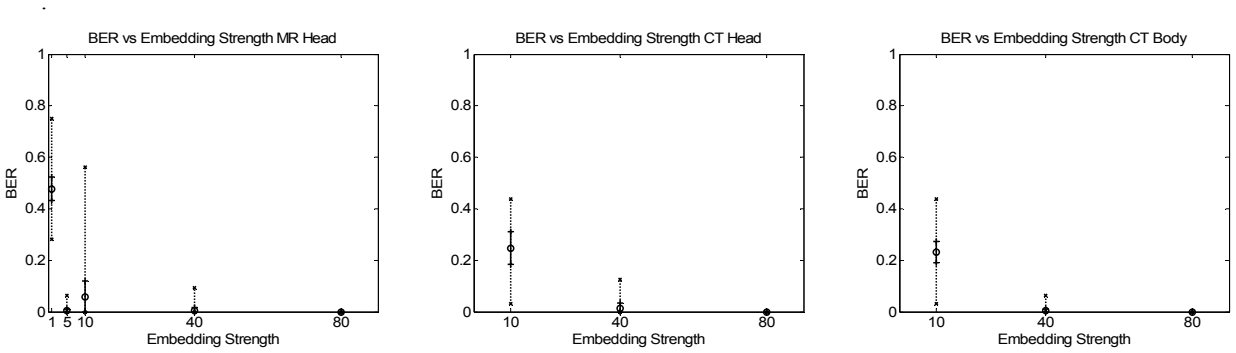


Fig. 2. DWT watermarking method: BER Results for MR Head, CT Head and CT Body at different embedding strengths.

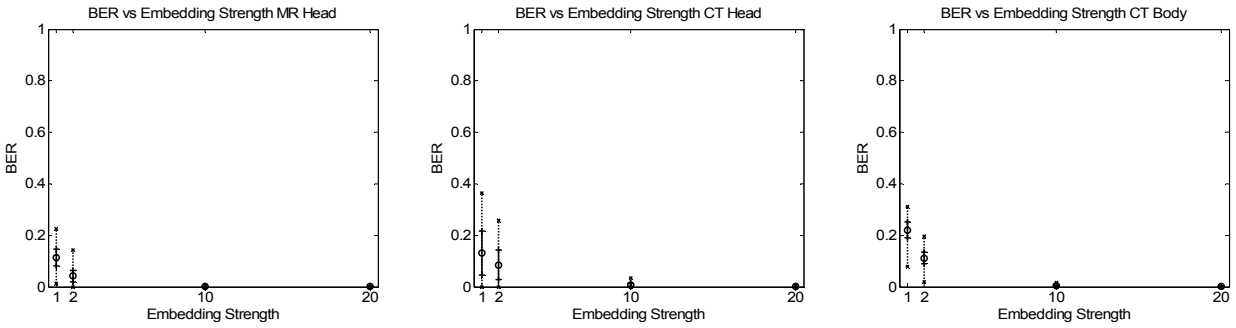


Fig. 3. DCT watermarking method: BER Results for MR Head, CT Head and CT Body at different embedding strengths

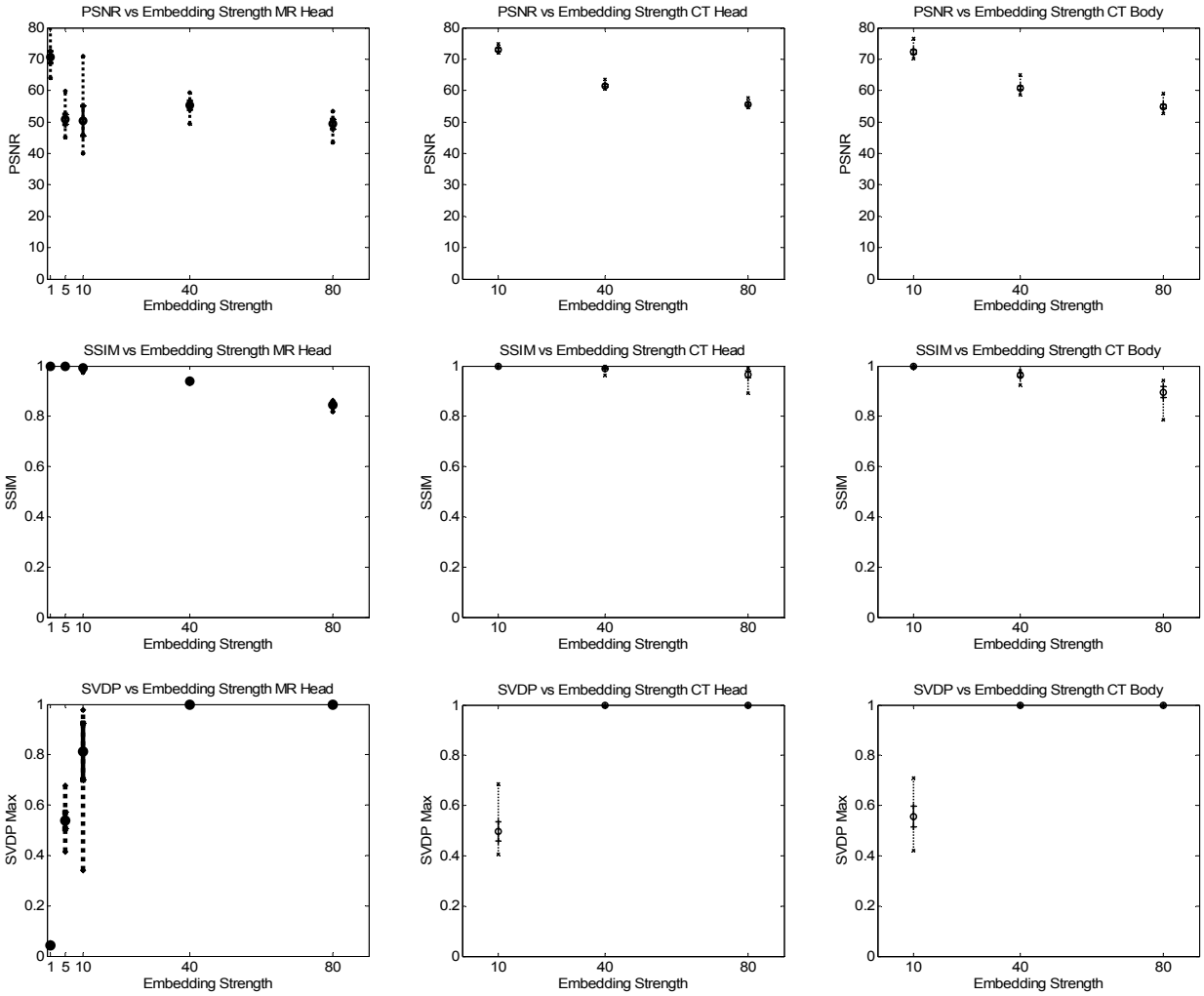
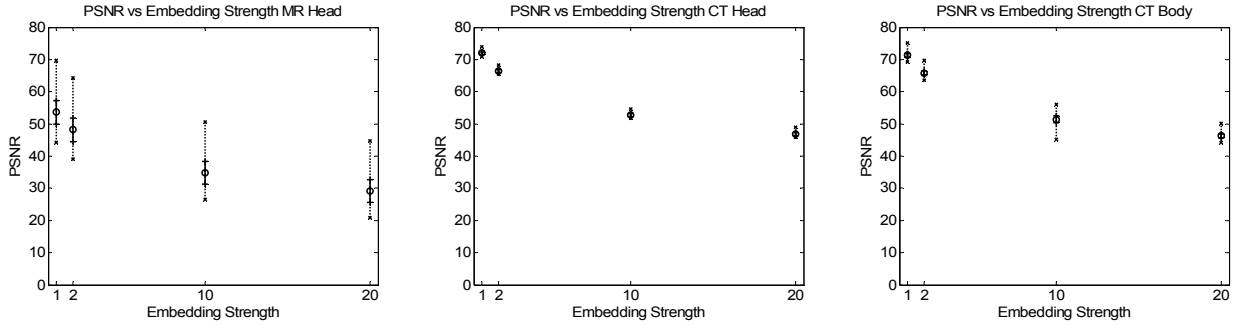


Fig. 4. DWT watermarking method: Visual Quality Results for MR Head, CT Head and CT Body at different embedding strengths.



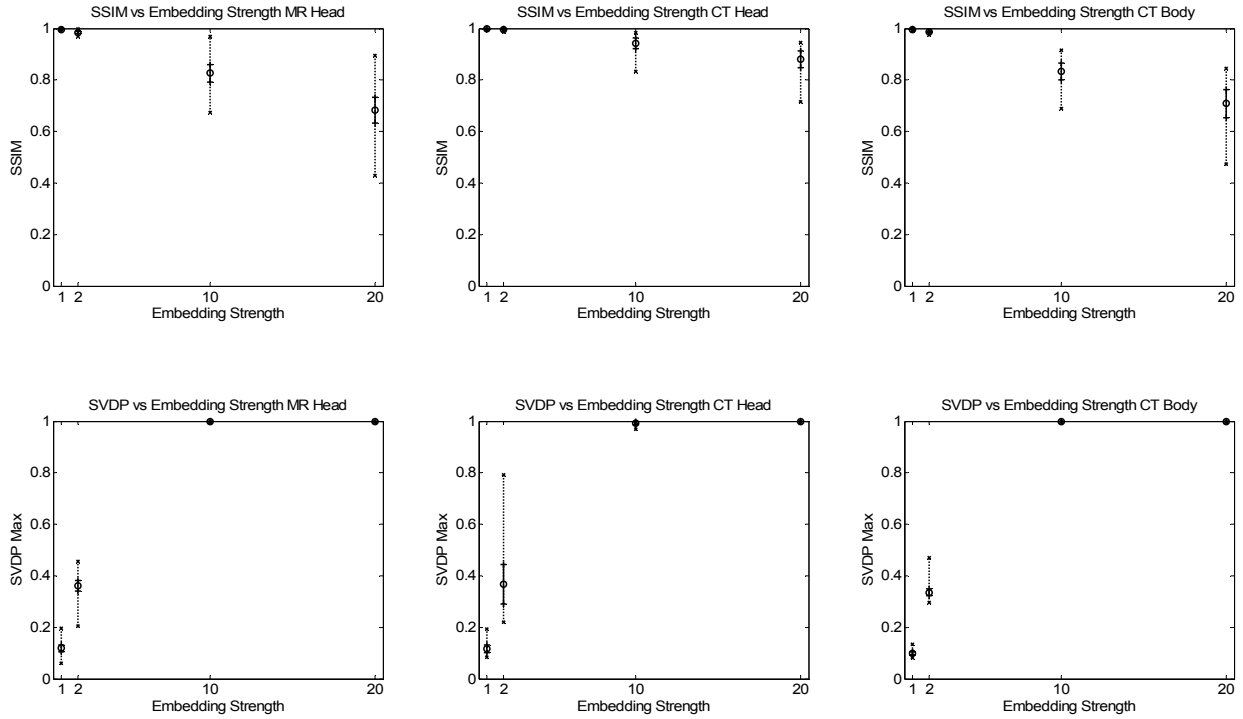


Fig. 5. DCT watermarking method: BER Results for MR Head, CT Head and CT Body at different embedding strengths.

3.2 Image tampering

The results of different levels of image modification on DWT and DCT watermark extraction are shown in tables 1-3. Figures 6-8 show image blocks and associated BERs for DCT watermark extraction at the three levels of degradation.

Table 1- Means and (standard deviations) of edge enhancement BER results.

Watermarking Method	DWT			DCT		
Degradation Level	Low	Medium	High	Low	Medium	High
UQ Head MRI	0.0000 (0.0000)	0.0328 (0.0343)	0.0500 (0.0398)	0.0045 (0.0032)	0.0338 (0.0120)	0.0958 (0.0229)
BML Head MRI	0.0108 (0.0187)	0.0644 (0.0504)	0.0816 (0.0521)	0.0608 (0.0457)	0.1278 (0.0626)	0.2165 (0.0582)
UQ Head CT	0.0120 (0.0272)	0.0529 (0.0411)	0.0625 (0.0460)	0.0080 (0.0120)	0.0325 (0.0453)	0.0688 (0.0850)
BML Body CT	0.0141 (0.0201)	0.0570 (0.0418)	0.0723 (0.0617)	0.0058 (0.0045)	0.0355 (0.0148)	0.0795 (0.0258)

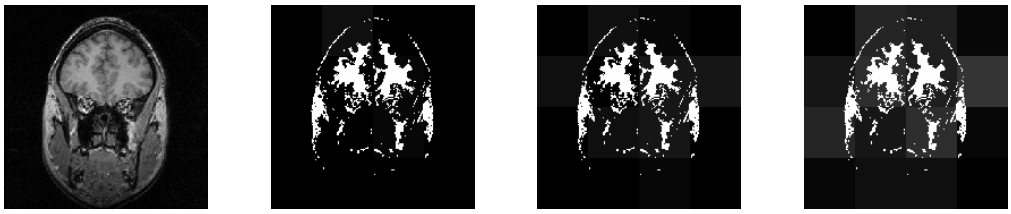


Fig. 2. DCT watermarked MR Head Image Block BERs at different edge enhancement levels. Block shades range from black (BER=0) to white (BER=1). Binary version of watermarked image is overlaid on blocks for reference.

Table 2 - Means and (standard deviations) of histogram stretching BER results.

Watermarking Method	DWT			DCT		
Degradation Level	Low	Medium	High	Low	Medium	High
UQ Head MRI	0.0000 (0.0000)	0.0375 (0.0360)	0.0297 (0.0277)	0.0055 (0.0037)	0.2293 (0.0640)	0.3114 (0.0734)
BML Head MRI	0.0230 (0.0336)	0.0517 (0.0411)	0.0606 (0.0479)	0.1349 (0.1110)	0.3463 (0.0696)	0.3993 (0.0734)
UQ Head CT	0.0144 (0.0243)	0.0361 (0.0281)	0.0553 (0.0701)	0.0094 (0.0140)	0.2454 (0.1198)	0.2929 (0.2232)
BML Body CT	0.0184 (0.0209)	0.0594 (0.0481)	0.0600 (0.0488)	0.0395 (0.0097)	0.3995 (0.0672)	0.4240 (0.0756)

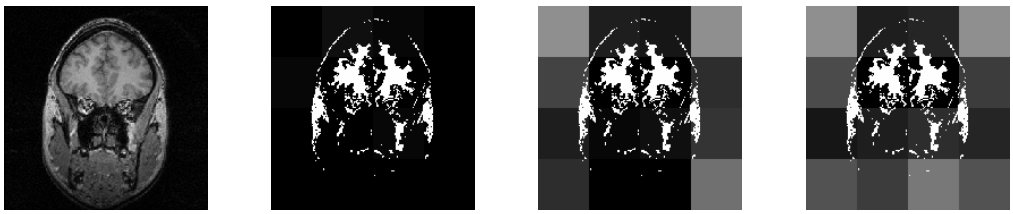


Fig. 3. DCT watermarked MR Head Image Block BERs at different histogram stretch levels. Block shades range from black (BER=0) to white (BER=1). Binary version of watermarked image is overlaid on blocks for reference.

Table 3 - Means and (standard deviations) of JPEG BER results.

Watermarking Method	DWT			DCT		
Degradation Level	Low	Medium	High	Low	Medium	High
UQ Head MRI	0.0000 (0.0000)	0.0250 (0.0192)	0.0547 (0.0285)	0.0043 (0.0037)	0.3904 (0.0216)	0.4503 (0.0242)
BML Head MRI	0.0064 (0.0142)	0.0344 (0.0384)	0.0478 (0.0511)	0.0570 (0.0448)	0.3838 (0.0206)	0.4516 (0.0200)
UQ Head CT	0.0048 (0.0173)	0.0096 (0.0197)	0.0192 (0.0203)	0.0072 (0.0112)	0.0190 (0.0109)	0.1698 (0.0069)
BML Body CT	0.0037 (0.0119)	0.0086 (0.0154)	0.0116 (0.0176)	0.0038 (0.0031)	0.0408 (0.0085)	0.2252 (0.0209)

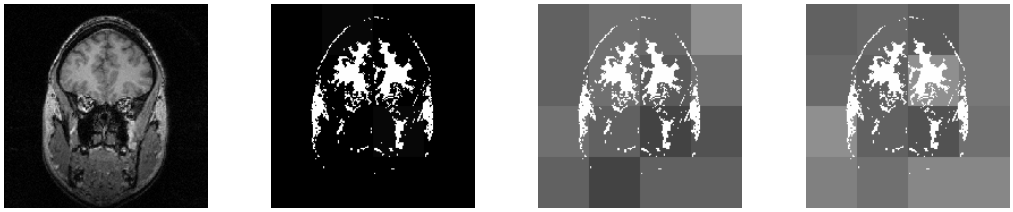


Fig. 4. DCT watermarked MR Head Image Block BERs at different JPEG compression levels. Block shades range from black (BER=0) to white (BER=1). Binary version of watermarked image is overlaid on blocks for reference.

3.3 Subjective quality assessment

Mean subjective quality results for the 20 subjects are shown in Figure 9. These figures indicate that the medical image perceptual threshold values should be approximately SVDP Sum = 0.5, SVDP Max = 0.13 and PSNR = 57.

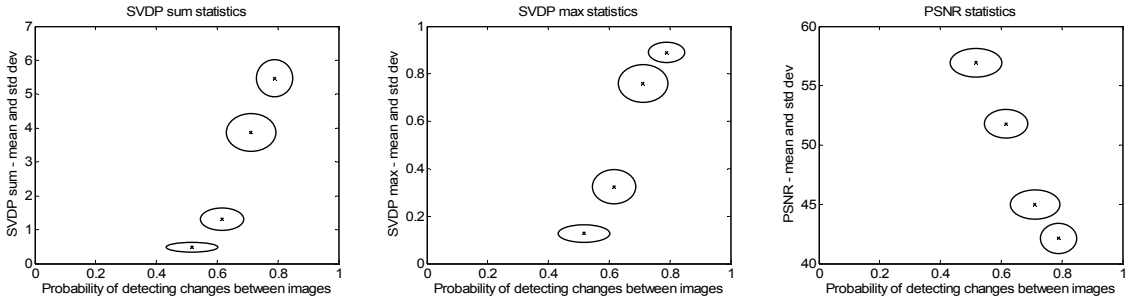


Fig. 5. Mean (x) and standard deviation (o) subjective quality assessment results for SVDP Sum, SVDP Max and PSNR.

4. DISCUSSION

4.1 Watermarking method and tamper detection

The results in Figures 2 and 3 show differences in BER at different embedding strengths for each image modality. The watermarking method applied has an effect on the BER, as can seen in the DCT CT Body results. There do not appear to be significant visual quality differences between medical image types. Images of different modalities generate similar PSNR, SSIM and SVDP outcomes. PSNR and SSIM results for MRI data are slightly lower than for CT data due to the lower bit range of selected MRI data.

When comparing the three numerical quality metrics, PSNR and SSIM results demonstrated comparable behaviour across all image types, watermarking methods and watermarking parameter settings. However, SVDP outcomes did not always follow the same result patterns as PSNR and SSIM. Initial subjective testing indicates that there are correlations between human observer results and SVDP results. Therefore, a conservative initial conclusion is that SVDP is suitable for medical image watermark assessment.

The amount and location of watermark degradation, in a block-based watermarked image, depends on the type and level of image manipulation applied. Three different manipulations were tested on DWT and DCT block-based watermarked images: edge enhancement results showed that BER levels vary according to the image and watermarking method tested, however no distinct behaviour emerges for any particular case. It was also shown that higher BERs result in areas where greater edge enhancement has been applied to the watermarked image. Histogram stretching resulted in different outcomes for DWT and DCT watermarking: DWT was more robust and DCT more fragile. Being the more fragile method, DCT provided better evidence/location of image tampering. The results for this manipulation showed that the selection of watermarking method depends on the fragility/robustness requirements of a system.

JPEG compression outcomes also demonstrated that the selection of watermarking method depends on the fragility/robustness requirements of the system. DCT again showed to be the more fragile system, particularly for MR data, which resulted in significant differences in BER levels for low-medium manipulation degradation levels.

From the summary given, it can be seen that one of the factors that will influence the choice of watermark type for a system is the level of robustness/fragility that is required by the user of the system.

4.2 Subjective Quality assessment

The results in Figure 9 suggest threshold values of approximately SVDP Sum = 0.5, SVDP Max = 0.13 and PSNR = 57 for p = 0.5 (equal probability of detecting changes/no changes between images). However the probability of detecting differences between the noisy and original images varied greatly; even within the same Gaussian noise level. Therefore further research (greater sample size and further calibration) is required to ensure that these threshold values reflect the perceptual threshold where a watermarked image is indistinguishable from its original image.

4.3 Conclusion and future work

In summary, initial analyses have shown that SVDP is suitable for assessing visual degradation that is caused by medical image watermarking. This is expected as SVDP focuses on the HVS for visual quality assessment. In a watermarking application based on the DCT and DWT methods, it would be appropriate for a user to make a choice between DWT and DCT depending on the requirements of the watermarking system. If greater robustness is desired, DWT is more appropriate. If more fragility is required, DCT is more appropriate, particularly for MR data.

The next phase of the project will include investigating different medical image watermarking scenarios and determining the best watermarking method and parameter settings for each scenario. For example, a user wishes to embed a watermark into Head MRI images that results in minimal visual degradation and is relatively robust against histogram stretching type image manipulations. DWT watermarking would be a suitable choice for this requirement, because it is

more robust than DCT against histogram degradation levels. Various scenarios, such as this example, will be investigated and are currently being implemented in an interactive demonstration system.

In addition, the current watermarking implementation requires a conversion from integer to double point precision numbers during the conversion between the frequency and spatial domains. This conversion can result in rounding errors, requiring a higher embedding strength. Integer-based DWT and DCT embedding algorithms would ensure zero extraction errors (BER = 0) for non-degraded watermarked images and will be investigated in future work.

Acknowledgements: National ICT Australia is funded by the Australian Government's Department of Communications, Information Technology, and the Arts and the Australian Research Council through Backing Australia's Ability and the ICT Research Centre of Excellence programs and the Queensland Government.

The authors would like to acknowledge Dr Sebastien Ourselin (ICT Centre, CSIRO) and Dr Andrew P. Bradley (University of Queensland) for providing medical images to test the watermarking systems.

REFERENCES

1. Coatrieux, G., Main, H., Sankur, B., Rolland, Y., and Collorec, R., "Relevance of Watermarking in Medical Imaging", IEEE-embs Information Technology Applications in Biomedicine, pp. 250-255, 2000.
2. Cox, I.J., Miller, M.L., Bloom, J.A., Digital Watermarking. Morgan Kaufmann . 2002.
3. Daly, S., A visual model for optimizing the design of image processing algorithms, In Proc. of IEEE International Conference on Image Processing, Austin, Texas, USA, pp. 16-20, 1994.
4. Daly, S., The Visible Differences Predictor, In Digital Images and Human Vision, Anonymous MIT Press, 1993.
5. Laboratory for Computational Vision, 2005. Publicly Available Software Packages. Website: http://www.cns.nyu.edu/~lcv/software.html (date accessed: 5/2006).
6. Lin, T., Podilchuk, C. I., and Delp, E. J., "Detection of image alterations using semi-fragile watermarks", Proc. of the SPIE Int. Conf. on Security and Watermarking of Multimedia Contents II, vol. 3971, pp. 152-163, 2000.
7. Madsen M. T., Berbaum, K. S., Ellingson A., Thompson, B. H. and Mullan, B. F. "Lesion removal and lesion addition algorithms in lung volumetric data sets for perception studies", SPIE Conf. on Medical Imaging, vol. 6146, pp. 61460T-1 – 10, February 2006.
8. Petitcolas, F.A. Watermarking schemes evaluation. *IEEE. Signal Processing*, vol. 17, no. 5, pp. 58–64, 2000.
9. Planitz, B. and Maeder, A., "Perceptually-limited modality-adaptive medical image watermarking", SPIE Conf. on Medical Imaging, vol. 6146, pp. 61460V-1 - 10, February 2006.
10. Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., Image Quality Assessment: From Error Visibility to Structural Similarity, *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600-612, April 2004.