PERCEPTUAL QUANTITATIVE QUALITY ASSESSMENT OF JPEG2000 COMPRESSED CT IMAGES WITH VARIOUS SLICE THICKNESSES

Jean-François Pambrun and Rita Noumeir

École de Technologie Supérieure, University of Quebec
Department of Electrical Engineering
1100 Notre-Dame West, Montreal, Quebec, Canada, H3C 1K3
rita.noumeir@etsmtl.ca

ABSTRACT

Modern medical equipments produce huge amounts of data that need to be archived for long periods and efficiently transferred over networks. Data compression plays an essential role in reducing the amount of medical imaging data. Medical images can usually be compressed by a factor of three before any degradation appears. Higher compression levels are desirable but can only be achieved with lossy compression, thus scarifying image quality. The diagnosis value of compressed medical images has been studied and recommendations about maximum acceptable compression ratios have been provided based on qualitative visual analysis. It has been suggested, without further investigation, that CT images, with thicknesses below five mm, cannot undergo lossy compression if diagnostic value needed to be preserved. In this paper, we present an objective quantitative quality assessment of compressed CT images using Visual Signal to Noise Ratio. Our results show that visual fidelity can be significantly affected by two factors, slice thickness and exposure time, for images compressed using the same compression ratio.

Index Terms— Medical Imaging, Compression, JPEG2000, Human Visual System (HVS), Quality Assessment, Computerized Tomography (CT), Slice Thickness, Exposure Time

1. INTRODUCTION

The amount of data produced by modern medical imaging devices is enormous [1, 2].This, with exhaustive archiving requirements [3] and growing expectations for instant availability, put tremendous strain on health care IT infrastructures. Part of this burden can be eliminated with the use of data compression to significantly reduce storage and improve network efficiency. Image compression can usually reduce image file size by a factor of three [4] without any information loss. Achieving compression beyond this point involves using irreversible lossy compression that compromises reconstruction

accuracy leading to images whose quality is affected. However, using irreversible compression in the medical domain is challenging because diagnostic accuracy must be preserved while multiple factors affect the level of achievable compression without a significant loss in quality.

Digital Imaging and Communications in Medicine (DICOM), which is the most widely used standard in medical imaging, allows image compression with JPEG for lossy image compression as well as JPEG-LS for lossless operations. However, recent advancement in wavelet based signal processing led to development of new compression algorithms and DICOM was supplemented in 2002 [4] to support JPEG2000. This modern codec provides new and improved features that are especially valuable in the medical domain[5, 6], including: efficient reversible and irreversible compression within a single codec; region of interest support; multiple bit depth support; extensive progressive rendering and transmission support and efficient compression on a wide range of image types. Like most compression algorithms, JPEG2000's compression level is set using one of two options: compression ratio (CR) or quality factor (QF). The former specifies the ratio between compressed and original image file sizes while the second specifies a target mean squared error (MSE). In an effort to promote medical image compression in Canada, the Canadian Association of Radiologists (CAR) produced, in 2008, a set of guidelines [7] based on a pan-Canadian study involving hundreds of radiologists where both diagnostic accuracy and perceptual evaluation of image quality were analyzed to define diagnostically acceptable compression ratios for different imaging modalities. Other studies [8] with similar approaches have been criticized [9] because compression ratios correlates poorly with image degradation. It has already been suggested [9, 10] that both of these metrics are unreliable at predicting the image quality perceived by the human visual system (HVS). In this paper we present a quantitative quality assessment of JPEG2000 compressed computerized tomography (CT) images using Visual Signal to Noise Ratio (VSNR)[11] which is a perceptual quality metric based on the HVS. Furthermore, we study the

effect of slice thickness and exposure time on compressibility. The scope of this paper is not to provide optimal compression ratios based on the theses parameters as it would require a panel of experienced radiologists. Instead, we are suggesting that specifying CRs for each imaging modality isn't ideal.

2. RELATED WORK

In [12], the authors concluded that image degradation could not be controlled only with compression ratios because deterioration levels are also dependent on image content. Likewise,they suggested that quality factors are more appropriate but the lack of standardization in their definition makes objective comparisons impossible. Most QFs are based on Peak Signal to Noise Ratio(PSNR), which is derived from MSE. This technique is also criticized in other studies, including [10], for being inaccurate. Possible solutions could involve taking advantage of our knowledge of the human visual system (HVS) or natural scene statistics in the design of quality metrics. Several approaches have been explored and multiples algorithms have been presented: Wavelet Visible Difference Predictor (WVDP)[13], High- Dynamic Range Visual Difference Predictor(HDR-VDP)[14], Structural Similarity (SSIM) [15], Discrete wavelet SSIM (DWT-SSIM) [16], Complex Wavelet SSIM (CW-SSIM) [17], Visual Information Fidelity (VIF)[18] and (VSNR)[11]. Most of these HVS metrics rely on two principles: contrast sensitivity function (CSF) and visual masking. Contrast sensitivity function refers to the decrease of contrast sensitivity as special frequency is increased. The CSF can be used to weigh compression errors before estimating the perceived image deterioration. It is defined in terms of cycles per degrees; therefore viewing conditions, pixels per inches and viewing distance must be taken into account. On the other hand, visual masking is a phenomenon where signals are locally masked by a background texture and compression artifacts can be hidden to some extent.

A comparative statistical analysis of multiples HVS metrics was performed in [19] where 29 images were degraded by JPEG and JPEG2000 compression, white noise, Gaussian blur and JPEG2000 transmission trough a noisy wireless channel. In total 779 images were produced. Each image was subjectively evaluated by an average of 23 observers rating them from bad to excellent quality. Because the original image was also evaluated, a difference mean opinion score was calculated. This score was then compared to the different image quality assessment algorithms. CR was not taken into account for JPEG and JPEG2000 based deterioration, however among ten algorithms tested PSNR performed the worst while VIF performed best. But perhaps the most interesting thing about this work is that the authors made their database [20] available for future research as the LIVE quality assessment database. This database is now widely used as a standard comparative tool to evaluate HVS based quality metrics.

Based on this tool, VSNR, used in this paper, performed well especially with JPEG2000 compression.

HVS metrics have already been applied in the medical imaging domain. In [21, 22, 23] a group of researchers used the High Dynamic Range Visual Difference Predictor(HDR-VDP) algorithm on abdomen and body images compressed with ratio ranging from 4:1 to 15:1. PSNR, HDR-VDP, Multi-Scale Structural Similarity and five radiologist's pooled assessment of artifacts were compared. The images were transformed using a single windows/level with a center of 20 Hounsfield unit(HU) and range of 400 HU. Readers could switch between the original and compressed images and were asked to establish if both images were distinguishable. They concluded that HDR-VDP was far more accurate than PSNR or MS-SSIM at predicting the observer's response. However, HVS metrics are sensitive to viewing condition and needs to be calibrated accordingly. DICOM specifies appropriate display parameters. Later, in [24], the same authors used half of the radiologist assessment of distinguishable pairs to fine tune the algorithms calibration while using the other half for validation. Using the derived calibration parameters, they were able to improve prediction accuracy. These advancements are very interesting but suffer from some limitations. Most notably, HVS metrics were only used to predict distinguishable and indistinguishable pairs. This is useful to achieve visually lossless compression but it is possible to achieve higher compression ratio and still retain diagnostic accuracy.

Before higher compression can be achieved, better knowledge of the factors that can impair image compressibility is required. It has been suggested in [25] that noise present in the original image is one of these factors as it increases the observer's ability to perceive a difference when quickly switching between the original and compressed images. These differences were mainly due to attenuated noise components in the compressed images but these were still acceptable for diagnostic purposes. They also noted that the denoised images resulting from compression were preferred by some observers. In [8] and [3], they also noted that acceptable compression ratios are highly dependent on noise level and that, in the case of CT scans, this level is affected by the radiation dose as well as acquisition time. In [21], [22] and [23], the authors used two set of images of different slice thickness generated from the same raw projection data ; one group was reconstructed with slice thickness of 0.67mm while the other at 5mm. They showed, with radiologist observations, that thick slices (5mm) could be slightly more compressed that thin slices (0.67mm) to achieve the visually lossless threshold. Similar results were observed in two studies, [26] and [27], which used 3D JPEG200 compression in an effort to further increase compression ratios of CT scans.
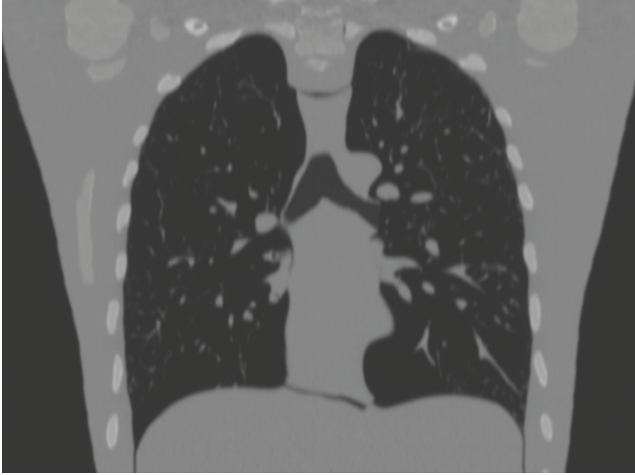
Fig. 1: A coronal scout view that shows 30 equally spaced axial images from 48 series considered for a total of 1440 images

Fig. 2: Steps of our evaluation scheme

(a) Abdomen	(b) Spine	(c) Lung

Fig. 3: One slice transformed using the Abdomen (a), Spine(b), and Lung(c) window/levels. The dashed square shows the area that was analyzed with VSNR

3. METHODOLOGY

3.1. Subject and acquisition parameters

The subject used in this experiment is an anthropomorphic thoracic phantom (Kyotokagaku Incorporated, Tokyo, Japan) scanned multiples times using different acquisition parameters in an effort [28] to provide a database for lung nodule size estimation. These images are available through the National Biomedical Imaging Archive (NBIA) [29] created by National Cancer Institute (NCI) and the National Institute of Biomedical Imaging and Bioengineering (NIBIB). The phantom is made of urethane and epoxy to mimic soft tissues and bones but does not contain lung parenchyma and some structures are filled with air [28]. Image series were scanned using a Philips 16-row scanner (Mx8000 IDT, Philips Healthcare, Andover, MA) with various acquisition parameters including multiple exposures (20-25 mAs, 50 mAs, 100 mAs, 200 mAs), slice collimations (16x0.75 mm, 16x1.5 mm), pitches (0.9, 1.2) and slice thicknesses (.8 mm, 1.5 mm, 3 mm with a collimation of 16x0.75 mm and 2 mm, 3 mm, 5 mm with a collimation of 16x1.5 mm). In total, 48 series were used. The number of images per series varied from 119 for 5mm slices to 799 for .8 mm slices. Thirty images, equally spaced, were selected from each serie as depicted in Figure 1 for a total of 1440 slices.

Each image is compressed using the open source and freely available OpenJpeg JPEG2000 codec [30] at 20 different compression ratios ranging from 4:1 to 60:1. This library is capable of accurately handling medical images composed of 16 signed bits per sample and it is used in widely deployed projects such as Grassroots DICOM (GDCM) [31] and Insight Segmentation and Registration Toolkit (ITK) [32]. Compression ratios are calculated based on file size and are equivalent to the 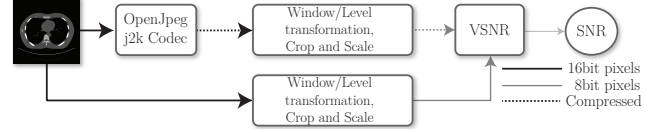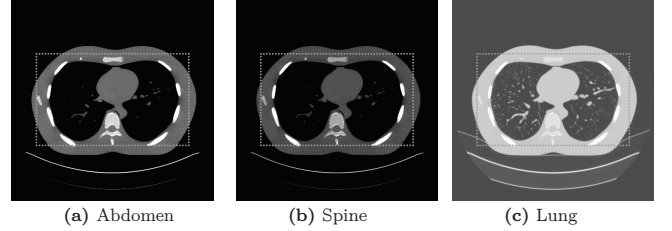number of bits in the compressed streams compared to number of bits allocated in the original files. Some codec vendors use the more mathematically rigorous definition of bits in the compressed stream compared to stored bits in the original image. Typically, each pixel of a CT image in DICOM native format have 12 stored bits in a 16 bits allocated space.

3.2. Objective quality assessment

3.3. Image Compression

In this experiment, a full reference HVS based metric, VSNR[11], was used to compare images compressed at different CRs with the original ones. When presented with a base and distorted image, this algorithm provides an objective measure of visual fidelity, in decibels, ranging from zero to infinity. The infinity symbol is used to express deterioration below the threshold of perception. Viewing conditions including screen resolution, distance and display pixel-value-to-luminance are taken into account. Pixel-value-to-luminance is modelled with $L(P) = (b + kP)^\gamma$ and the following parameters are adjustable: spatial resolution (r) of the display, viewing distance (d), black level offset (b), pixel-value-to-voltage ratio (k) and gamma (γ). The parameters used in this experiment are presented in Table 1. VSNR uses several

Table 1: VSNR configuration parameters

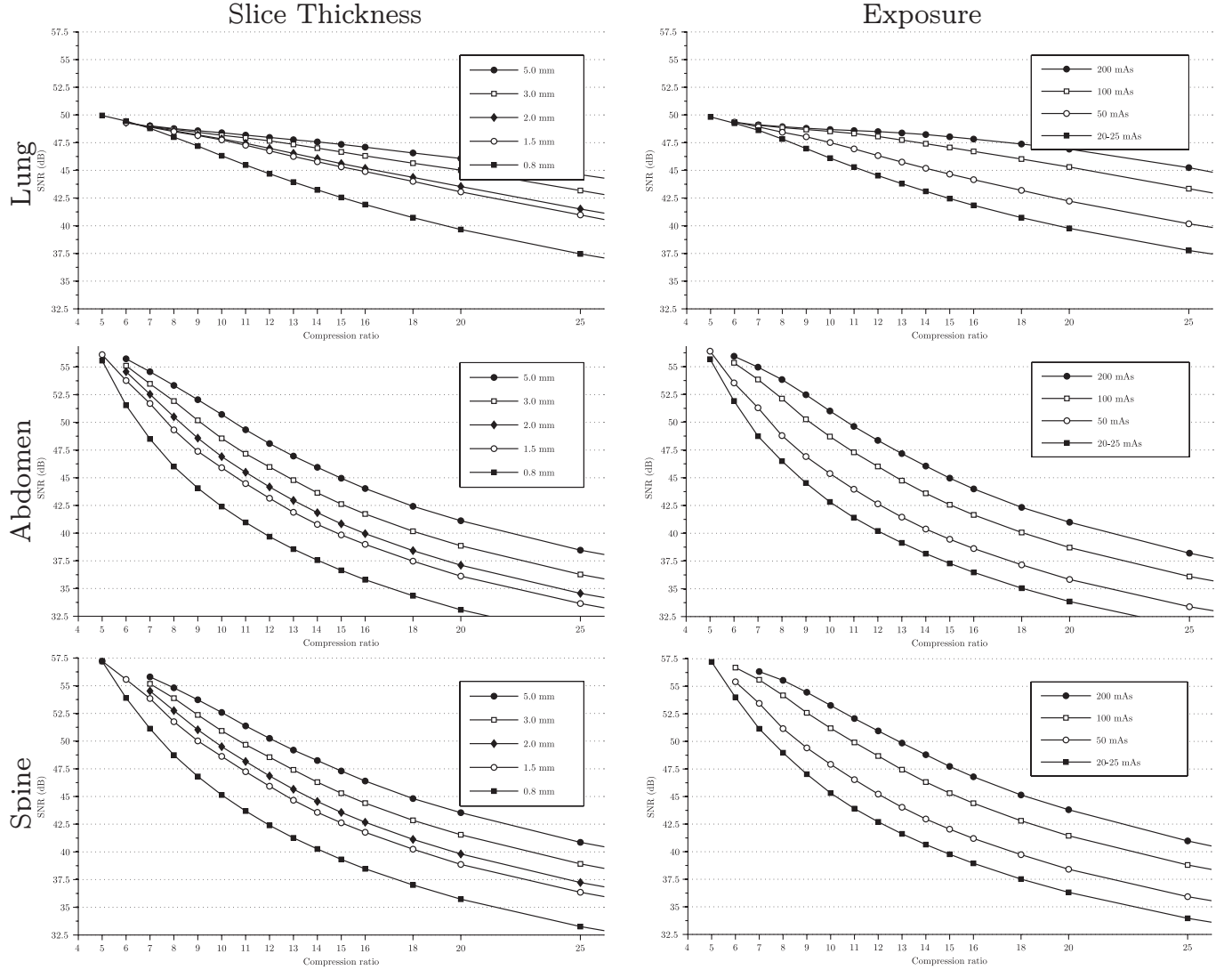| Parameter	Symbol	Value
Spatial resolution	r	122 (dpi)
viewing distance	d	16 (po)
black-level offset	b	0
pixel-value-to-voltage ratio	k	0.02874
gamma	γ	2.2

Fig. 4: Mean SNR versus compression ratio, for different slice thicknesses and exposure times, for images transformed using Lung, Abdomen and Spine window/levels

properties of the human visual system including contrast sensitivity, visual masking and global precedence. The images were scaled by a factor of 1.5 using bi-linear interpolation to emulate a two by two display mode on a three megapixels screen. The computation was only applied in a region of interest shown in Figure 3. This was done to reduce computational requirements and ensure that the vast empty black space around the phantom does not affect the results because it is almost always indistinguishable. To further mimic typical viewing conditions, each image pair is transformed using three different window/level configurations before visual fidelity is evaluated. The windows/levels that were used are lung (center: -600, width: 2000), spine (center: 100, width: 500) and abdomen (center: 40, width: 400). Figure 2 shows the various steps of our evaluation approach.

4. RESULTS

Over 28 thousand image pairs were compared for each window/level configuration. That is, 30 images from each of the 48 series of varying pitch (1.2, 0.9), exposure (20-25mAs, 50mAs, 100mAs, 200mAs) and 5 slice thicknesses (0.8mm, 1.5mm, 2mm, 3mm(repeated twice with different slice collimation), 5mm) compressed at 20 compression ratios ranging from 4:1 to 60:1. Figure 4 shows mean SNR versus CR curves for varying slice thicknesses and exposures for the lung , spine and abdomen window/level. Curves for the different slice thicknesses are averaged over more than 200 images while those from the exposure graph are averaged from 360 samples. These graphs clearly show that for each decrease in slice thickness or exposure, visual fidelity is decreased for the same compression ratio. Figure 5 exposes both extremes,
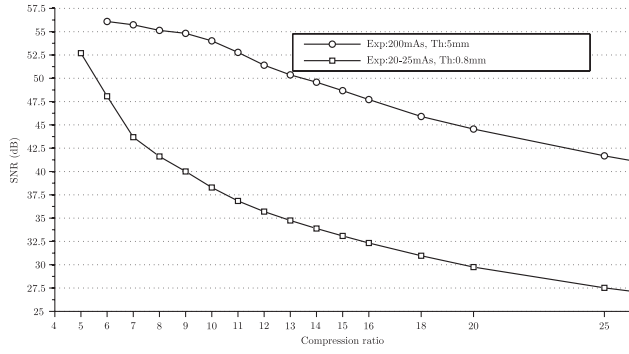
Fig. 5: Mean SNR versus compression ratio for best and worst possible cases
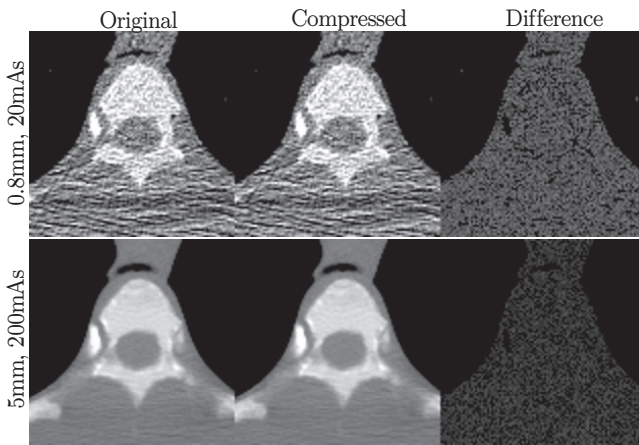

Fig. 6: Example of the best possible case (Top) against the worst possible case (Bottom)

it presents the best possible case for image compressibility, 5mm slice acquired with an exposure of 200mAs, and the worst possible case, 0.8mm acquired with an exposure of 20-25mAs. Both are transformed using abdomen window/level. It shows widely different image fidelity for the same compression ratios. Each curve from this figure was built by averaging 60 images. To illustrate the difference between both cases Figure 6 shows image pairs compressed at 10:1 from each one along with the corresponding difference image.

A quadratic regression was performed for each of the three window/level configurations (Lung, Abdomen and Spine) with respect to a constant term, compression ratio, slice thickness and exposure, respectively. P-values were ze-

Table 2: Spearman's rank correlation coefficients

	Ratio	Thickness	Exposure
Lung	-0.8740	0.1815	0.2733
Abdomen	-0.8872	0.2114	0.2242
Spine	-0.8796	0.1965	0.2269

ros in all cases and t-statistic was high (> 47) showing statistical significance for all variables. R-square was respectively 0.89, 0.86 and 0.85 for each window/level. As it can be observed in the graphics, linear coefficients show that image fidelity is less affected when using lung (ratio:-0.55, thickness:2.19 and exposure:0.06) than both abdomen (ratio:-1.48, thickness:3.32 and exposure:0.09)and spine (ratio:-1.36, thickness:3.28 and exposure:0.09) windows/levels. Spearman's correlation coefficients are presented for each window/level in Table 2 showing significant correlation between visual fidelity and compression ratio but also with slice thickness and exposure.

5. CONCLUSION

Results from objective visual fidelity assessment with VSNR indicated that at least two factors can impair compressibility of CT scans: slice thickness and exposure time. This may be explained by increasing noise as thickness and exposure are reduced. Moreover, high frequencies are attenuated as more raw projection data are averaged to produce thicker slices, which may also explain higher compressibility. The consequence of these factors is that CT images displayed using the abdomen window/level, with thickness of 0.8mm, compressed at 8:1, appear as deteriorated as images with thickness of 5mm, compressed at 14:1. Similarly, images acquired at 200mAs and compressed at 8:1 are comparable to images acquired at 20-25mAs and compressed at 14:1, for the same window/level. In the most extreme comparison, images with thickness of 0.8mm, acquired at 20-25mAs, and compressed at 8:1 are similar to images with thickness of 5mm, acquired at 200mAs and compressed at 25:1. The window/level settings used to evaluate image fidelity also had an impact on results. These showed that visual fidelity for the lung configuration was less affected than spine and abdomen. This is expected due to the width of the window which are respectively 2000, 500 and 400. Since typical computer monitors can only display 256 gray levels, with a window width on 2000, 8 HU level are converted into a single gray value. Image deterioration below 4 HU would not be noticeable in the transformed image. The same deterioration could produce an error equal to two pixel values when transformed with the abdomen window. Consequently, Windows/level transformations should be selected with care when designing this kind of study.

In Canada, the CAR based their image compression recommendations on a pan-Canadian study of diagnostic accuracy that only used CRs as the quality metric. This study did however voice concerns about the effect of slice thickness and did not include thin slices below 2.5mm. Because of this, the Canadian Association of Radiologists does not recommend lossy compression of CT scans with slices thinner than 2.5mm. Our results suggest that specifying a unique CR for CT scans is not an optimal quality control scheme because CRs are not closely correlated to image fidelity. This is fur-

ther supported by the lack of standards in the way CRs are calculated; some vendors use bit allocated while others uses bit stored as a reference resulting in a 30% difference between implementations. To improve accuracy, CR should be standardized and recommendations could be defined with respect to multiple parameters including modality, slice thickness, exposure time, etc.

6. REFERENCES

[1] K H Lee, H J Lee, J H Kim, H S Kang, K W Lee, H Hong, H J Chin, and K S Ha, "Managing the CT data explosion: initial experiences of archiving volumetric datasets in a mini-PACS.," *Journal of digital imaging*, vol. 18, no. 3, pp. 188–95, Sept. 2005.

[2] G D Rubin, "Data explosion: the challenge of multidetector-row CT.," *European journal of radiology*, vol. 36, no. 2, pp. 74–80, Nov. 2000.

[3] D Koff, P Bak, P Brownrigg, D Hosseinzadeh, A Khademi, A Kiss, L Lepanto, T Michalak, H Shulman, and A Volkening, "Pan-Canadian evaluation of irreversible compression ratios ("lossy" compression) for development of national guidelines.," *Journal of digital imaging*, vol. 22, no. 6, pp. 569–78, 2009.

[4] K H Lee, Y H Kim, B H Kim, K J Kim, T J Kim, H J Kim, and S Hahn, "Irreversible JPEG 2000 compression of abdominal CT for primary interpretation: assessment of visually lossless threshold.," *European radiology*, vol. 17, no. 6, pp. 1529–34, 2007.

[5] P Schelkens, A Skodras, and T Ebrahimi, *The JPEG 2000 Suite*, John Wiley and Sons, 2009.

[6] D S Taubman and M W Marcellin, *JPEG2000: image compression fundamentals, standards, and practice, Volume 1*, Springer, 2002.

[7] CAR, "CAR Standards for Irreversible Compression in Digital Diagnostic Imaging within Radiology," 2008.

[8] A A Bankier, M Weber, M Prokop, C J Herold, and C Schaefer-prokop, "JPEG2000 Compression of Thin-Section CT Images of the Lung: Effect of Compression Ratio on Image Quality," vol. 240, no. 3, pp. 869–877, 2006.

[9] A Fidler and B Likar, "What Is Wrong with Compression Ratio in Lossy Image Compression?," *Radiology*, vol. 245, no. 1, pp. 299, 2007.

[10] Z Wang and A C Bovik, "Mean Squared Error : Love It or Leave It," *IEEE Signal Processing Magazine*, , no. January, pp. 98–117, 2009.

[11] D M Chandler and S S Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images.," *IEEE transactions on image processing*, vol. 16, no. 9, pp. 2284–98, Sept. 2007.

[12] A Fidler, U Skaleric, and B Likar, "The impact of image information on compressibility and degradation in medical image compression," *Medical Physics*, vol. 33, no. 8, pp. 2832, 2006.

[13] A Bradley, "A wavelet visible difference predictor," *IEEE Transactions on Image Processing*, vol. 8, no. 5, pp. 717–730, 1999.

[14] R Mantiuk, H Seidel, and K Myszkowski, "Visible difference predicator for high dynamic range images," *Proceedings of IEEE International*, vol. pp, pp. 2763–2769, 2004.

[15] Z Wang, A C Bovik, H R Sheikh, and E P Simoncelli, "Image quality assessment: from error visibility to structural similarity.," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–12, Apr. 2004.

[16] C Yang, W Gao, and L Po, "Discrete wavelet transform-based structural similarity for image quality," *15th IEEE International Conference on Image*, pp. 377–380, 2008.

[17] M P Sampat, Z Wang, S Gupta, A C Bovik, and M K Markey, "Complex wavelet structural similarity: a new image similarity index.," *IEEE transactions on image processing*, vol. 18, no. 11, pp. 2385–401, Nov. 2009.

[18] H Sheikh and A Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[19] H Sheikh, M Sabir, A Bovik, and Others, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440, 2006.

[20] L C H.R. Sheikh, Z.Wang and A Bovik, "LIVE Image Quality Assessment Database Release 2," .

[21] B Kim, K H Lee, K J Kim, R Mantiuk, V Bajpai, T J Kim, Y H Kim, C J Yoon, and S Hahn, "Prediction of perceptible artifacts in JPEG2000 compressed abdomen CT images using a perceptual image quality metric.," *Academic radiology*, vol. 15, no. 3, pp. 314–25, 2008.

[22] K J Kim, K H Lee, H S Kang, S Y Kim, Y H Kim, B Kim, J Seo, and R Mantiuk, "Objective index of image fidelity for JPEG2000 compressed body CT images," *Medical Physics*, vol. 36, no. 7, pp. 3218, 2009.

[23] B Kim, K H Lee, K J Kim, R Mantiuk, S Hahn, T J Kim, and Y H Kim, "Prediction of perceptible artifacts in JPEG 2000-compressed chest CT images using mathematical and perceptual quality metrics.," *AJR. American journal of roentgenology*, vol. 190, no. 2, pp. 328–34, Feb. 2008.

[24] K J Kim, R Mantiuk, K H Lee, and W Heidrich, "Calibration of the visual difference predictor for estimating visibility of JPEG2000 compression artifacts in CT images," *Journal Of Optics Nouvelle Revue D'Optique*, pp. 75270G–75270G–9, 2010.

[25] B Erickson, A Manduca, P Palisson, K Persons, F Earnest, V Savcenko, and N Hangiandreou, "Wavelet compression of medical images," *Radiology*, vol. 206, no. 3, pp. 599–607, 1998.

[26] E Siegel, "Compression of multislice CT: 2D vs. 3D JPEG2000 and effects of slice thickness," *Proceedings of SPIE*, vol. 5748, pp. 162–170, 2005.

[27] K Siddiqui, E Siegel, B Reiner, O Crave, J Johnson, Z Wu, J Dagher, a Bilgin, M Marcellin, and M Nadar, "Improved compressibility of multislice CT datasets using 3D JPEG2000 compression," *International Congress Series*, vol. 1268, pp. 57–62, 2004.

[28] M a Gavrielides, L M Kinnard, K J Myers, J Peregoy, W F Pritchard, R Zeng, J Esparza, J Karanian, and N Petrick, "A resource for the assessment of lung nodule size estimation methods: database of thoracic CT scans of an anthropomorphic phantom.," *Optics express*, vol. 18, no. 14, pp. 15244–55, July 2010.

[29] NBIA, "National Biomedical Imaging Archive (2010, Dec. 12). [Online]," https://imaging.nci.nih.gov/.

[30] OpenJpeg, "Open-source C-Library for JPEG 2000 (2010, Dec. 12). [Online]," http://code.google.com/p/openjpeg/.

[31] GDCM, "Grassroots DICOM (2010, Dec. 12). [Online]," http://sourceforge.net/projects/gdcm/.

[32] ITK, "Insight Segmentation and Registration Toolkit. (2010, Dec. 12). [Online]," https://www.itk.org/.