A Multicenter Observer Performance Study of 3D JPEG2000 Compression of Thin-Slice CT

Bradley J. Erickson,[1] Elizabeth Krupinski,[2] and Katherine P. Andriole[3]

The goal of this study was to determine the compression level at which 3D JPEG2000 compression of thin-slice CTs of the chest and abdomen–pelvis becomes visually perceptible. A secondary goal was to determine if residents in training and non-physicians are substantially different from experienced radiologists in their perception of compression-related changes. This study used multidetector computed tomography 3D datasets with 0.625–1-mm thickness slices of standard chest, abdomen, or pelvis, clipped to 12 bits. The Kakadu v5.2 JPEG2000 compression algorithm was used to compress and decompress the 80 examinations creating four sets of images: lossless, 1.5 bpp (8:1), 1 bpp (12:1), and 0.75 bpp (16:1). Two randomly selected slices from each examination were shown to observers using a flicker mode paradigm in which observers rapidly toggled between two images, the original and a compressed version, with the task of deciding whether differences between them could be detected. Six staff radiologists, four residents, and six PhDs experienced in medical imaging (from three institutions) served as observers. Overall, 77.46% of observers detected differences at 8:1, 94.75% at 12:1, and 98.59% at 16:1 compression levels. Across all compression levels, the staff radiologists noted differences 64.70% of the time, the resident's detected differences 71.91% of the time, and the PhDs detected differences 69.95% of the time. Even mild compression is perceptible with current technology. The ability to detect differences does not equate to diagnostic differences, although perception of compression artifacts could affect diagnostic decision making and diagnostic workflow.

KEY WORDS: 3D imaging, compression, JPEG2000, observer performance

INTRODUCTION

Imaging in medicine has seen remarkable increases in the amount of data that is produced by imaging devices. This increase presents a number of problems,[1] including the challenge of transmitting, archiving, and managing the data. Several groups have previously described findings for a number of algorithms applied to differing modalities and body parts, using many different criteria. One recent large trial is helping to establish practice guidelines for the nation of Canada.[2] Several groups have previously reported that thin-slice CT is less compressible than thick slice, primarily due to noise.[3,4] For example, Siddiqui et al.[3] used thoracic CT datasets acquired with a 16-detector MSCT scanner and reconstructed at slice thicknesses ranging from 0.75 to 10.0 mm. The images were irreversibly compressed using a 2D JPEG2000 encoder and a 3D JPEG2000 algorithm to ratios ranging from 4:1 to 64:1. We note here that 2D image compression means that only the information in one 2D slice is considered when the compression is performed. 3D compression considers all the information in a 3D volume at the time of compression, which should allow for greater compression. They computed image distortion using peak signal-to-noise ratio (PSNR) and a visual discrimination model based on just noticeable differences (JNDs). They found that for 2D compression, the thinnest sections were far less compressible than the thicker ones at equivalent levels of image quality. This was especially true at higher compression ratios.

[1]From the Department of Radiology, Mayo Clinic, Rochester, MN, USA.

[2]From the Department Radiology, University of Arizona, Tucson, AZ, USA.

[3]From the Department Radiology, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA.

Correspondence to: Bradley J. Erickson, Department of Radiology, Mayo Clinic, Rochester, MN, USA; tel: +1-507-2848548; e-mail: bje@mayo.edu

3D compression produced higher image quality in general compared to 2D for the same ratios, and the effect increased for thinner slices and higher compression ratios. It was concluded that 3D JPEG2000 compression yields better overall image quality than 2D JPEG2000. This study, however, only used PSNR and modeled JNDs to measure image quality and did not include evaluation of the images by human observers.

Several other previous studies have also been conducted on 3D compression methods, including one recent report with a similar design to the one used in the present study.[5] That study also used 3D JPEG2000 on abdominal CT. The study used thicker slices (3 or 6 mm) than the present study and had a smaller number of observers. They found 4:1 and 8:1 compression ratios indistinguishable but found that higher ratios had perceptible changes. That study also fixed the minimum window width at 350 and focused on diagnosis.

Although these previous studies have been useful, they have had weaknesses as well. We believe that requiring no perceptual difference is a more appropriate metric of performance than simply using PSNR, model observers, or diagnosis only. Imaging is used to diagnose many conditions, and compression may differentially affect the ability to diagnose the various conditions so a method needs to be used that will help insure diagnostic accuracy by relying on the fact that the compression method produces no perceptually detectable differences in images compared to the original.[6–8] In recent years, there have been a number of studies conducted (6–8 for example) on medical image compression that utilize the perceptually lossless criteria for acceptable compression levels, assuming that if there are no perceptually detectable differences, then there should be no diagnostic differences. Therefore, since work is still ongoing in terms of finalizing recommendations for compression of radiographic images,[9] the present study examined the effects of 3D JPEG2000 compression in thin-slice multidetector computed tomography (MDCT) images on the ability of observers do perceptually detect differences between compressed and uncompressed images.

MATERIALS AND METHODS

This study used 80 MDCT (Siemens Medical Systems, Erlangen, Germany) 3D datasets with 0.625–1-mm thickness slices of standard chest, abdomen, or pelvis, clipped to 12 bits. It was necessary to clip the values to 12 bits because the compression algorithm available to us (Kakadu v5.2 (http://www.kakadusoftware.com) could only handle 8- and 12-bit grayscale images. This algorithm was used to compress and decompress the images creating four sets of images: lossless, 1.5 bpp (8:1), 1 bpp (12:1), and 0.75 bpp (16:1). Two randomly selected slices from each examination were used. Four observer session sequences were created with the original and four compressed images in random sequences. Each sequence had only one occurrence of each image.

A special purpose viewer application (Fig. 1) was developed that showed the images to observers using a flicker mode paradigm in which observers rapidly toggled between two images, the original and a compressed version, using the mouse scroll wheel with the task of deciding whether they could detect differences between them. Images were displayed on a Barco Coronis color 3 Mpixel LCD display (BarcoView, LLC, Kortrijk, Belgium) that was calibrated to the Digital Imaging and Communications in Medicine Grayscale Standard Display Function with an MXRT 5100 display controller. The image size was fixed at a 1:1 zoom to avoid the possible impact of interpolation on results. In addition, there were four standard presentation window width and level settings that the user could select, including lung (1500/−700), abdomen (350/40), bone (1500, 400), and liver (70/30). Ambient lighting was subdued.

Three sites participated in the study. Six staff radiologists, four radiology residents, and six PhDs experienced in medical imaging (from the three institutions) served as observers.

RESULTS

The achieved compression ratios were 16.1:1, 12.1:1, 8.04:1, and 1.87:1 (vs 12 bits). The average computation to compress and decompress was 30 min/volume on a 2-GHz desktop personal computer. Overall, as compression levels increased, there were more differences detected (χ^2=14,281.97, $p<$ 0.0001), with 77.46% of observers detecting differences at 8:1, 94.75% at 12:1, and 98.59% at 16:1 compression levels, and there were 2.44% false
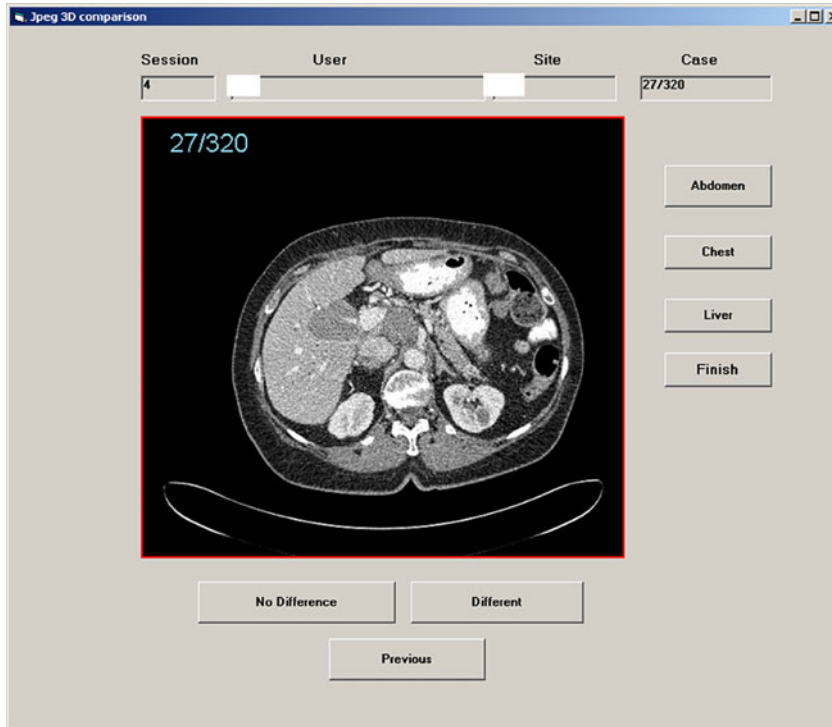
Fig 1. The user interface for the study.

positive differences detected with the lossless images. There were no statistically significant differences overall ($\chi^2=0.150$, $p>0.05$) as a function of body part imaged. For chest images, 67.82% of trials had detected differences versus abdomen–pelvis images that had 68.79% of trials with detected differences.

There were statistically significant differences ($\chi^2=82.12$, $p<0.0001$) overall as a function of observer experience. The residents (71.91%) and PhDs (69.95%) overall detected more differences than the staff radiologists did (64.70%) at all compression levels (Fig. 2).

Interestingly, there was a significant difference ($\chi^2=20.64$, $p<0.0001$) as a function of site (Fig. 3).

DISCUSSION

This study presents a number of significant findings. First, even mild compression is perceptible with current technology. Many studies in the past have noted that the ability to detect differences does not equate to diagnostic differences, although perception of compression artifacts could affect diagnostic decision making and diagnostic workflow. We

should note that our methodology—presentation of images using the "flicker" technique is probably the most sensitive method for human observers to detect differences. We recognize that the diagnostic task is different, and in most cases, the diagnosis can be made with mild to even substantial visual differences. However, we feel that to be conservative, a visual difference should not be considered acceptable for any medical diagnosis that might be encountered.
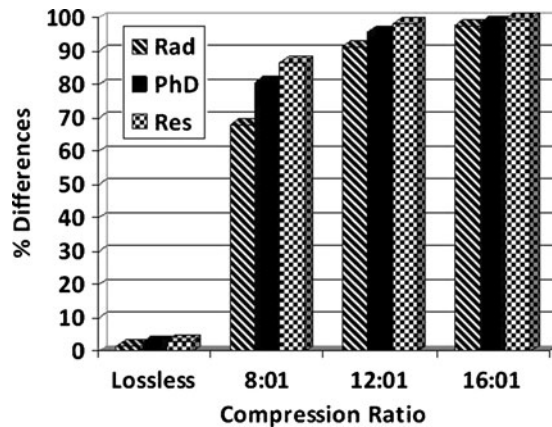

Fig 2. Percent differences detected as a function of compression ratio for the radiologists (*Rad*), *PhD*s, and residents (*Res*).

Other studies where specific diagnoses are considered cannot address the question of whether *all* diagnoses are unaffected.

Observers also quickly found that liver settings were most sensitive to compression-related changes and, thus, reviewed most images with that setting. During the design of the experiment, we recognized that if we allowed complete control of window width, the observers could set the width to 1 and make detection of changes quite easy. Fixing the width to a typical abdomen setting like 400 will make the changes much less obvious. We decided to permit a narrow "liver" width for this study because it is occasionally used in the clinical realm. A very narrow width of 1 is sometimes used to detect hemorrhage in clinical images, but that task depends much less on the appearance of the images.

Several groups have previously reported that thin-slice CT is less compressible than thick slice, primarily due to noise.[3,4] When this study was designed, we expected that true 3D image compression should be most effective in datasets with very thin slices (≤1 mm) because of the higher degree of correlation in content between slices. While this is probably a valid expectation, it appears that other factors (e.g., image noise) had a more significant impact on compressibility.

Another interesting aspect of this study is that we included staff radiologists, radiology residents, and non-physician PhDs as observers. The purpose of this arm of the study was to determine if staff radiologists were more sensitive to compression artifacts. We found that the staff radiologists were consistently the *least* sensitive to compression artifacts. This may reflect the fact that staff was more comfortable that the images contained the information they needed even if slight changes were present. This has important implications for future studies, as it suggests that it is not necessary to have MDs, let alone staff radiologists as observers, if one wishes to remain at the "safe" or conservative end of the compression spectrum. It would have been interesting to also collect the time to arrive at a decision. It is possible that this would have shown a higher correlation and that staff radiologists quickly decide on differences, without focusing on non-diagnostic components of an image.

Although there was a perceptible difference, one cannot safely conclude that important information was lost because of compression. Indeed, studies have shown that at these lower compression ratios, the primary effect is to reduce noise, and improved diagnostic performance could result.[10] Others have pointed out that 3D lossy compression can be useful for specific applications such as the creation of 3D renderings.[11] Additionally, as we did not ask the observers to report on exactly which organs or body parts they noticed the compression artifacts in, we cannot at this time address the question of whether different organs or body parts compress differentially. This would be an important topic for future investigation. Nevertheless, it appears that one cannot apply 3D JPEG2000 compression to thin-slice CT studies at 8:1 or higher without considering possible effects on diagnosis.
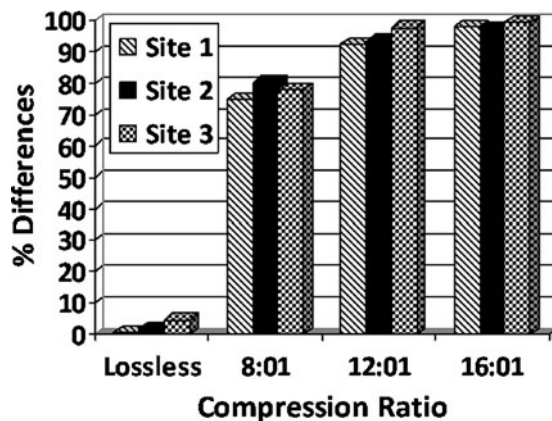
SUMMARY

There are perceptible changes in an important fraction of thin CT images compressed using 3D JPEG2000, even at low ratios (8:1). It is unclear whether these changes might degrade the diagnostic value. Experienced radiologists appear to be less sensitive to compression-related changes than other observers.

Fig 3. Percent differences detected as a function of compression ratio for the three sites.

ACKNOWLEDGMENTS

REFERENCES

1. Andriole K, Morin R, Arenson R, et al: Addressing the coming radiology crisis-the Society for Computer Applications in Radiology transforming the radiological interpretation process (TRIP) initiative. J Dig Imag 17:235–243, 2004

2. Koff D, Bak P, Brownrigg P, et al: Pan-Canadian evaluation of irreversible compression ratios ("Lossy" Compression) for development of national guidelines. J Dig Imag 2008

3. Siddiqui K, Siegel E, Reiner B, Johnson J, Crave O, Nadar M: Improved image compression at various slice thickness for multi-slice CT using 3D JPEG 2000 (part 2) in comparison with conventional 2D compression. In:Society for Computer Applications in Radiology Vancouver, BC: SCAR, 87–88, 2004.

4. Woo H, Kim K, Kim T, et al: JPEG 2000 compression of abdominal CT: difference in tolerance between thin- and thick-section images. AJR 189:535–541, 2007

5. Ringl H, Schernthaner RE, Kulinna-Cosentini C, et al: Lossy three-dimensional JPEG2000 compression of abdominal CT images: assessment of the visually lossless threshold and effect of compression ratio on image quality. Radiol 245:467–474, 2007

6. Wu D, Tan DM, Baird M, Decampo J, White C, Hong RW: Perceptually lossless medical image coding. Trans IEEE Med Imag 25:335–344, 2006

7. Fukatsu H, Naganawa S, Yum S: Development and evaluation of a novel lossless image compression method (AIC: artificial intelligence compression method) using neural networks as artificial intelligence. Radiat Med 26:120–128, 2008

8. Kim KL, Kim B, Lee KH, Kim TJ, Mantiuk R, Kang HS, Kim YH: Regional difference in compression artifacts in low-dose chest CT images: effects of mathematical and perceptual factors. AJR 291:W30–W37, 2008

9. National Electrical Manufacturers' Association: Directory for the standardization on medical informatics. Accessed at: http://www.med.kyushu-u.ac.jp/info/std/archives/cat_dicom.html. Last accessed February 11, 2009

10. Savcenko V, Erickson BJ, Palisson PM, et al: Detection of subtle abnormalities on chest radiographs after irreversible compression. Radiol 206:609–616, 1998

11. Krishnan K, Marcellin MW, Bilgin A, Nadar MS: Efficient transmission of compressed data for remote volume visualization. IEEE Trans Med Imag 25:1189–1199, 2006