EVALUATION OF EFFECTS OF JPEG2000 COMPRESSION ON A COMPUTER-AIDED DETECTION SYSTEM FOR PROSTATE CANCER ON DIGITIZED HISTOPATHOLOGY

Scott Doyle, James Monaco, Anant Madabhushi *

Rutgers University
Dept. of Biomedical Engineering
Piscataway, NJ, USA

Stefan Lindholm, Patric Ljung, Lance Ladic

Siemens Corporate Research
Princeton, NJ, USA

John Tomaszewski, Michael Feldman

University of Pennsylvania
Dept. of Surgical Pathology
Philadelphia, PA, USA

ABSTRACT

A single digital pathology image can occupy over 10 gigabytes of hard disk space, rendering it difficult to store, analyze, and transmit. Though image compression provides a means of reducing the storage requirement, its effects on CAD (and pathologist) performance are not yet clear. In this work we assess the impact of compression on the ability of a CAD system to detect carcinoma of the prostate (CaP) in histological sections. The CAD algorithm proceeds as follows: Glands in the tissue are segmented using a region-growing algorithm. The size of each gland is then extracted and modeled using a mixture of Gamma distributions. A Markov prior (specifically, a probabilistic pairwise Markov model) is employed to encourage nearby glands to share the same class (i.e. cancerous or non-cancerous). Finally, cancerous glands are aggregated into continuous regions using a distance-hull algorithm. We evaluate CAD performance over 12 images compressed at 14 different compression ratios using JPEG2000. Algorithm performance (measured using the under the receiver operating characteristic curves) remains relatively constant for compression ratios up to 1:256. After this point performance degrades precipitously. We also have an expert pathologist view the compressed images and assign a confidence measure as to their diagnostic fidelity.

1. INTRODUCTION

Digitized images of large tissue samples, such as whole-mount histological sections of the prostate, can constitute more than 10 GB of data. This places a large burden on the computational resources required for storage, transmission, and analysis. On a daily basis a large pathology lab may process hundreds of such studies. This volume of data presents several challenges to digital pathology: 1) Storage becomes prohibitively expensive. 2) Telepathology, the transmission of digital images over computer networks, is untenable. 3) It becomes impossible to employ sophisticated computer-aided diagnosis (CAD) systems. Mitigating these issues requires a method for reducing image file size while retaining diagnostic fidelity [1].

Compression algorithms are a common method for decreasing the storage size of images. The ratio of an uncompressed file size to its compressed size is known as the *compression ratio*. There are two main methods of compression: **lossless**, which are fully reversible but are limited by a low compression ratio, and **lossy**, which achieve high compression ratios at the cost of reduced image quality. In digital pathology, loss of image quality can adversely affect the ability of both a CAD system [2] and a pathologist [3] to perform analysis.

The majority of previous research into the impact of compression of histological images relied on visual quality as measured by a pathologist. For example, Foran, et al. [3] determined which compression ratios were suitable for diagnosis using telepathology. To our knowledge very few papers attempt to quantitatively measure the effects of compression on an automated system: López, et al. [2] found that the differences in nuclei counts as performed by their automated system were not significantly affected by compression ratios of up to 1:46.

In this work we evaluate the impact of JPEG2000 compression on the ability of a computer-aided diagnosis (CAD) system to detect carcinoma of the prostate (CaP) in whole-mount histological sections. We previously developed such a CAD system [4], which proceeds as follows: Step 1) glands are segmented, Step 2) the segmented glands are classified as malignant or benign, and Step 3) the malignant glands are consolidated into continuous regions. The system was shown to detect CaP regions with a sensitivity of 88% and an accompanying false positive rates of the 10% [4]. In the current study, we measure the system performance using 40 whole-mount histology compressed via JPEG2000 at 14 different ratios to determine the robustness of the CAD algorithm to lossy compression. For completeness we also perform a reader study, wherein a pathologist is asked to provide a confidence measure as to the diagnostic fidelity of the compressed images.

2. METHODOLOGY

2.1. Image Compression Algorithm

The compression scheme used is the JPEG2000 compression standard and coding system, based on the wavelet transform. In JPEG2000 images are first transformed to YUV color space and then convolved with the Cohen-Daubechies-Feauveau discrete wavelet transform. This generates sets of coefficients referred to as sub-bands. Each sub-band represents an approximation of the original image, at a corresponding resolution level, and for each additional sub-band increasing image detail is generated allowing the image to be reconstructed at higher resolution with additional sub-bands. It should be noted that the represented height and width of the image do not change between sub-bands, but the amount of information used to represent each band is increased significantly. Finally, all sub-bands are divided into code blocks of 64-by-64 pixels, which are

*Funding for this work made possible by the Wallace H. Coulter Foundation, New Jersey Commission on Cancer Research, National Cancer Institute (R01CA136535-01, R21CA127186-01, R03CA128081-01), the Cancer Institute of New Jersey, Bioimagene Inc., and the U.S. Department of Defense (W81XWH-08-1-0145).

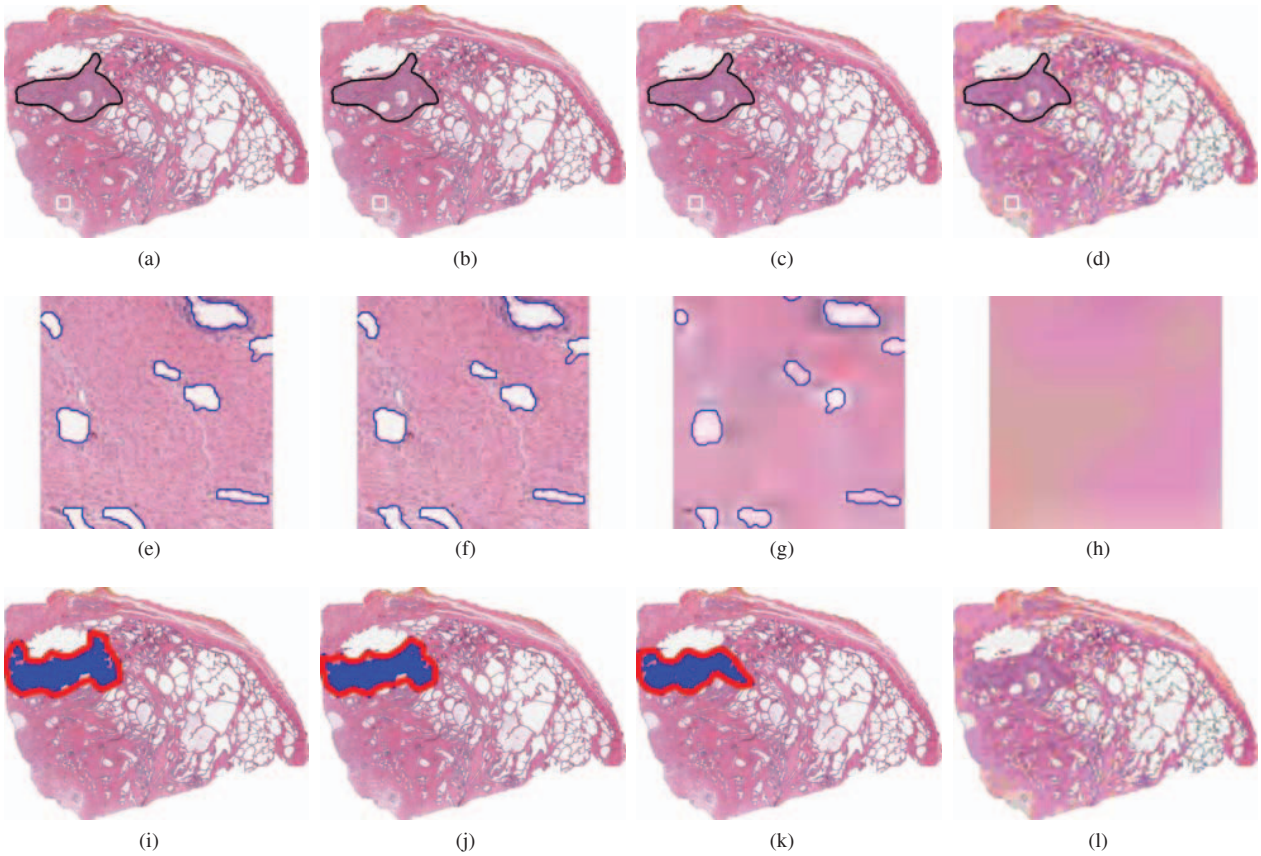


Fig. 1. JPEG2000 compression on (a) an original histopathology image at (b) 1:16 , (c) 1:256, and (d) 1:4096 compression ratios. Black contours identify the cancer region. The region of interest in a white box is magnified in (e)-(h) to illustrate differences in gland detection and segmentation at different ratios. Shown in (i)-(l) are the results of CAD on each of the compressed images. Results are fairly robust until very high compression ratios. Note that the breakdown of the CAD algorithm occurs at the gland level (h), where detection of glands is impossible.

individually encoded by a three-step Embedded Block Coding with Optimal Truncation (EBCOT) scheme. This scalar-quantization and the code block encoding together determine the amount of compression, known as the compression ratio.

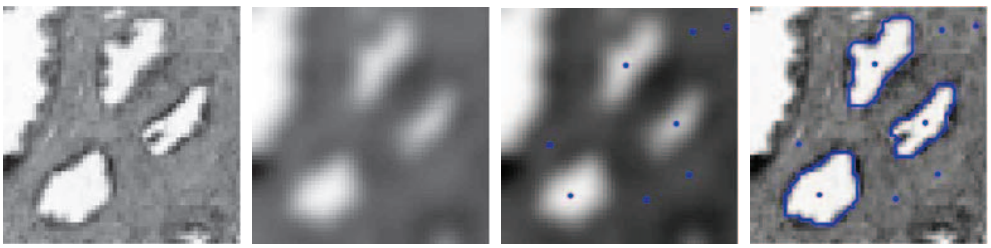
The OpenJPEG implementation of the JPEG2000 standard (http://www.openjpeg.org/) was used to produce nine different compression ratios, each containing a single image consisting of six sub-bands or levels per compressed image. The compression ratios ranged from $1:1$ (no compression) to $1:1024$ (high compression). Examples of compression can be seen in Figure 1. The original images are shown in Figure 1(a), along with subsequently higher compression ratios (Figures 1(b)-1(d), respectively).

2.2. Cancer Detection and Classification

Figure 1(a) illustrates a prostate histological (tissue) section. The pinkish hue results from the H&E staining procedure. The superimposed black line delimits the spatial extent of CaP as determined by a pathologist. The numerous white regions are the gland lumens, i.e. cavities in the prostate through which fluid flows. Our automated system identifies regions of CaP by leveraging two biological properties: 1) cancerous glands (and hence their lumens) tend to be smaller in cancerous than benign regions and 2) malignant/benign glands tend to be proximate to other malignant/benign glands.

2.2.1. Gland Segmentation

Figure 2 illustrates the gland segmentation procedure. We extract the luminance channel of the digitized section (*CIE Lab* color space), where gland regions appear as contiguous, high intensity pixels circumscribed by sharp boundaries (Figure 2(a)). We convolve the image with a Gaussian kernel at multiple scales to generate multiple smoothed images (Figure 2(b) illustrates one such image.). The local maxima (i.e. single pixel peaks) are considered to be lumen centers (Figure 2(c)), which serve as seeds for a region-growing algorithm (Figure 2(d)). We briefly outline this algorithm. First define the following: 1) *current region* (CR) is the set of pixels representing the segmented region in the current step of the algorithm, 2) *current boundary* (CB) is the set of pixels that neighbor CR in an 8-connected sense, but are not in CR, and 3) *internal boundary* (IB) is the subset of pixels in CR that neighbor CB. The growing procedure begins by initializing CR to a seed pixel assumed to lie within the gland. At each iteration CR expands by aggregating the pixel in CB with the greatest intensity. CR and CB are updated, and the process continues. The algorithm terminates when the L_∞ norm from the seed to the next aggregated pixel exceeds a predetermined threshold. That is, the L_∞ norm establishes a square bounding box about the seed; the growing procedure terminates when the algorithm attempts to add a pixel outside this box. During each iteration the algorithm



(a) (b) (c) (d)

Fig. 2. Overview of the gland detection and segmentation procedure. The luminance channel (a) is convolved with a Gaussian kernel to generate a smoothed image (b). Peaks in this image are used to detect gland centers (c). A region-growing algorithm is used in the unsmoothed image to extract gland size (d). Segmentations with poor average edge strengths are discarded.

measures the boundary strength which is defined as the average intensity of the pixels in IB minus the average intensity of the pixels in CB. After the growing procedure terminates, the region with the greatest boundary strength is selected. If the boundary strength is below an empirically-determined signal-to-noise ratio it is discarded.

2.2.2. Gland Classification

Let the set $S = \{1, 2, \ldots, N\}$ reference the N segmented glands in a histological. Each gland has an associated state $X_s \in \Lambda \equiv \{\omega_1, \omega_2\}$, where ω_1 and ω_2 indicate malignancy and benignity, respectively. The random variable $Y_s \in \mathbb{R}$ indicates the area of gland s. Let $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)$ refer to all random variables X_s and Y_s in aggregate. The state spaces of $\mathbf{X}$ and $\mathbf{Y}$ are the Cartesian products $\Omega = \Lambda^N$ and $\mathbb{R}^{D \times N}$.

We use maximum *a posteriori* (MAP) estimation to find the optimal $\mathbf{X}$ given the feature vector $\mathbf{Y}$, i.e. we maximize the *a posteriori* probability $P(\mathbf{X}|\mathbf{Y})$. This probability is proportional to the product of the conditional probability $P(\mathbf{Y}|\mathbf{X})$ and the prior distribution $P(\mathbf{X})$. The conditional probability models the area of the glands, as cancerous glands tend to be smaller in size than benign glands [5]. The prior distribution incorporates the biological tendency for cancerous/benign glands to appear near other cancerous/benign glands. More specifically, $P(\mathbf{X})$ is modeled using probabilistic pairwise Markov model (PPMM) [4], a novel Markov prior which is both more flexible and intuitive than typical Markov priors (such as the Potts model). Both the conditional and prior distributions can be learned via training. This approach allows us to generate Receiver Operating Characteristic (ROC) curves for quantitative evaluation as opposed to setting a hard threshold.

2.2.3. Gland Consolidation

Glands determined to be cancerous are consolidated into continuous regions. To perform this consolidation we use a modified form of the convex hull called distance hull or Dhull [4]. Unlike the convex hull, Dhull places a restriction on the maximum distance between consecutive points on the hull, thus allowing the formation of non-convex boundaries which can better conform to the true CaP regions.

3. EXPERIMENTAL SETUP AND EVALUATION

The dataset consists of 40 prostate histology sections stained with hematoxylin and eosin (H&E), obtained from radical prostatectomies at the University of Pennsylvania and Queens University in Canada. Each sample contains regions of CaP ranging in malignancy from Gleason scores six to eight, and is digitized at 1.25x optical magnification (8 μm per pixel) using an Aperio slide scanner. The CaP regions on each digitized sample are manually delineated by a pathologist using a black contour in an image editor.

3.1. Experiment 1: Automated Cancer Detection via CAD

Twenty eight of the histological sections (uncompressed) were used to train the CAD system described in Section 2.2. The remaining 12 images were each compressed at 1:1, 1:2, ..., 1:8192, yielding a test set of 168 images. To assess system performance we define the following measure: true positives (TP) indicate the area of the HSs denoted as cancerous by both the pathologist and CAD, and similarly we define true negatives (TN), false positives (FP), and false negatives (FN). From these we obtain two additional measures: the true positive rate TP/(TP+FN) and the false positive rate FP/(TN+FP).

The performance of the CaP detection system with respect to all preceding measures is influenced by the probability that a gland is malignant (or one minus the probability it is benign). This probability can be varied by the user from zero to one, which yields a receiver operator characteristic (ROC) curve. To arrive at a measure that is independent of the prior probability we can calculate the total area under the ROC curve (AUC). Therefore, to evaluate the impact of compression ratio on the performance of the CaP detection system we choose to measure the AUC for each group of 12 images sharing the same compression ratio. This produces 14 total AUCs (one for each compression ratio).

3.2. Experiment 2: Pathologist Reader Visual Inspection

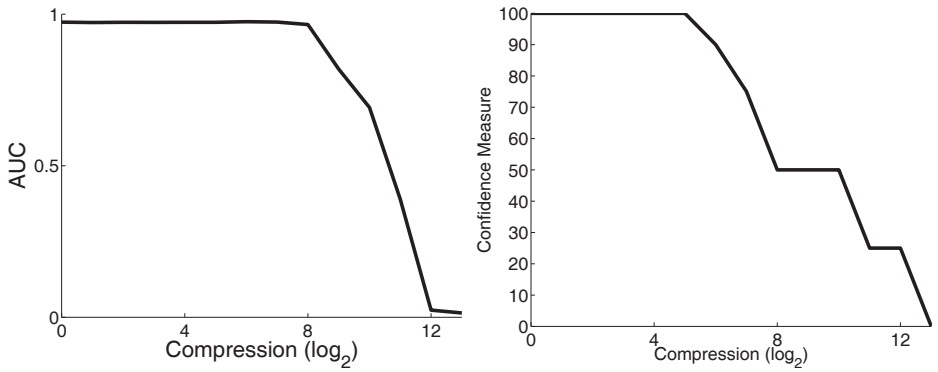
An expert pathologist was instructed to state the confidence in their ability to identify the regions of CaP. The confidence measure ranges from 0 (absence of diagnostic information) to 100 (absolute certainty). To prevent previously-viewed images from influencing subsequent confidence measures, the images were considered serially from the most- to the least-compressed.

4. RESULTS AND DISCUSSION

4.1. Experiment 1: CAD Performance on Compressed Images

4.1.1. AUC vs. Compression Ratio

Quantitative classification results are shown in Figure 3(a), with the AUC for each compression ratio plotted as a function of compression ratio. The independent axis is plotted using a log (base 2) scale.



(a) (b)

Fig. 3. (a) Plot of evaluation metric (AUC) as the compression level increases. As compression increases, performance of the CAD algorithm decreases due to a loss of diagnostically useful information. (b) Plot of pathologist confidence in diagnosis as compression level increases. Note that a decrease in pathologist confidence does not indicate incorrect diagnosis, but simply a lack of diagnostically useful information.

For compression ratios up to 1:256 there is very little degradation in classifier performance. At higher compression ratios performance decreases rapidly; as seen in Figure 1(h), the gland detection algorithm can no longer identify the lumens.

4.1.2. Qualitative Evaluation of CaP Regions

Figure 1(e) shows a portion of an uncompressed image that contains several glands. Notice that the number and relative sizes and shapes of the gland segmentations are very consistent up to a compression ratio of 1:256 (Figures 1(e)-(g)). As the compression ratio reaches increasingly higher levels, details become lost and the algorithm can no longer find the lumen regions (Figure 1(h)).

4.2. Experiment 2: Reader Inspection of Compressed Images

The reader confidence in classification is plotted in Figure 3(b) as a function of compression ratio. The pathologist is quite confident in classifying the cancerous regions in the image until compression ratio reaches around 1:64, at which point confidence decreases to 0%. Note that we are making a distinction between confidence and accuracy: although the pathologist becomes much less confident at ratios exceeding 1:64, this may not necessarily signify a commensurate reduction in detection performance.

5. CONCLUDING REMARKS

Since digitized histological samples can be several gigabytes in size, image compression is necessary component of digital pathology. Unfortunately, the effects of lossy compression on the analysis of histology images is not well understood. In this paper, we evaluated the impact of image compression with respect to the ability of a CAD algorithm to identify CaP regions on whole-mount histology sections. Specifically, we applied our previously-developed CAD system to images compressed at 14 different compression ratios using JPEG2000. System performance was shown to be very robust for compression ratios up to 1:256. Beyond this level performance dropped off sharply. As can be easily seen in the images in Figure 1, this drop-off results from the inability of the CAD system to detect the individual glands. Local high frequency information is lost at high compression rates, thus we should expect a decline in CAD performance when gland size becomes small in relation to remaining high frequency information (less such information will remain for higher compression rates).

An expert pathologist evaluated the effects of compression on diagnostic image quality. Interestingly, degradation was perceived at compression ratios that did not affect CAD performance. This is not unexpected. Whereas the CAD algorithm only considers the size of the glands, a pathologist interprets additional information such as glandular morphology and the coloring from the H&E stain. Perhaps these attributes degrade more quickly with compression than does glandular area. This suggests that it might be useful to store images at one compression ratio for visual analysis and at another for automated CAD analysis. Additionally, different CAD systems (for the same task) would likely vary in their robustness to compression. For example, those using co-occurrence matrices to extract textural features would likely be very sensitive to the removal of high frequency information. In general, the impact of compression is a function of many factors such as the compression scheme, the general task, and the specific algorithmic implementation. Further research is needed to better understand these dependencies.

6. REFERENCES

[1] M. Gao, P. Bridgman, and S. Kumar, "Computer aided prostate cancer diagnosis using image enhancement and jpeg2000," *SPIE*, vol. 5203, pp. 323–334, 2003.

[2] C. Lopez, M. Lejeune, P. Escriva, R. Bosch, M.T. Salvado, L.E. Pons, J. Baucells, X. Cugat, T. Alvaro, and J. Jaen, "Effects of image compression on automatic count of immunohistochemically stained nuclei in digital images," *Jnl. of the Amer. Med. Inf. Assoc.*, vol. 15, no. 6, pp. 794–798, 2008.

[3] D. Foran, P.P. Meer, T. Papathomas, and I. Marsic, "Compression guidelines for diagnostic telepathology," *IEEE Trans. on Inf. Tech. in Biomed.*, vol. 1, no. 1, pp. 55–60, 1997.

[4] J. Monaco, J. Tomaszewski, M. Feldman, M. Mehdi, P. Mousavi, A. Boag, C. Davidson, P. Abolmaesumi, and A. Madabhushi, "Probabilistic pair-wise markov models: Application to prostate cancer detection," *SPIE*, vol. 7260, 2009.

[5] V. Kumar, A. Abbas, and N. Fausto, *Robbins and Cotran Pathologic Basis of Disease*, Saunders, 2004.