Published in final edited form as:
IEEE Trans Med Imaging. 2011 February ; 30(2): . doi:10.1109/TMI.2010.2077308.

Using a Visual Discrimination Model for the Detection of Compression Artifacts in Virtual Pathology Images

Jeffrey P. Johnson,
Siemens Corporate Research, Princeton, NJ 08540 USA

Elizabeth A. Krupinski,
Department of Radiology, University of Arizona, Tucson, AZ 85724 USA

Michelle Yan,
Siemens Corporate Research, Princeton, NJ 08540 USA

Hans Roehrig,
Department of Radiology, University of Arizona, Tucson, AZ 85724 USA

Anna R. Graham, and
Department of Pathology, University of Arizona, Tucson, AZ 85724 USA

Ronald S. Weinstein
Department of Pathology, University of Arizona, Tucson, AZ 85724 USA

Jeffrey P. Johnson: johnson.jeffrey@siemens.com; Elizabeth A. Krupinski: krupinski@radiology.arizona.edu; Michelle Yan: michelle.yan@siemens.com; Hans Roehrig: hans@radiology.arizona.edu; Anna R. Graham: agraham1@email.arizona.edu; Ronald S. Weinstein: ronaldw@email.arizona.edu

Abstract

A major issue in telepathology is the extremely large and growing size of digitized "virtual" slides, which can require several gigabytes of storage and cause significant delays in data transmission for remote image interpretation and interactive visualization by pathologists. Compression can reduce this massive amount of virtual slide data, but reversible (lossless) methods limit data reduction to less than 50%, while lossy compression can degrade image quality and diagnostic accuracy. "Visually lossless" compression offers the potential for using higher compression levels without noticeable artifacts, but requires a rate-control strategy that adapts to image content and loss visibility. We investigated the utility of a visual discrimination model (VDM) and other distortion metrics for predicting JPEG 2000 bit rates corresponding to visually lossless compression of virtual slides for breast biopsy specimens. Threshold bit rates were determined experimentally with human observers for a variety of tissue regions cropped from virtual slides. For test images compressed to their visually lossless thresholds, just-noticeable difference (JND) metrics computed by the VDM were nearly constant at the 95th percentile level or higher, and were significantly less variable than peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) metrics. Our results suggest that VDM metrics could be used to guide the compression of virtual slides to achieve visually lossless compression while providing 5–12 times the data reduction of reversible methods.

Correspondence to: Elizabeth A. Krupinski, krupinski@radiology.arizona.edu.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Index Terms

Compression; just noticeable differences; virtual pathology slides; visual discrimination model

I. Introduction

The purpose of this work was to evaluate the utility of a visual discrimination model (VDM) for predicting levels of lossy compression corresponding to visually lossless quality over a wide range of image structures and compressibility. The practical benefit of this technology would be to provide metrics of loss visibility that could guide the compression of virtual slides and expedite data transmission while ensuring consistently acceptable visual quality for telepathology.

In a previous study [1], bit rates corresponding to visually lossless JPEG2000 compression were measured with human observers for image regions selected from pathology virtual slides. Observer performance in 2AFC (alternative forced choice) trials showed that compression ratios of about 7:1 or four times the reversible compression ratio could be achieved before losses were detectable. Significant differences in visually lossless bit rates and peak signal-to-noise ratio (PSNR) were observed across test images due to normal variations in tissue structures, which affect image compressibility. VDM metrics computed for bit rates at the visually lossless thresholds were nearly constant, however, corresponding to equal just noticeable difference (JND) visibility for compression losses.

This uniformity suggested that a JND target level corresponding to visually lossless compression could be applied to adaptively compress diverse images to different bit rates and different PSNR values and still achieve uniform image quality defined in terms of compression loss visibility. The next steps reported here are the collection of threshold likelihood functions for visually lossless compression of a larger set of test images using medical students/residents and pathologists, and the use of a likelihood ratio chi-square method to test the hypothesis of equal thresholds across images for various distortion metrics.

The results of this work can likely apply to numerous situations in medical imaging where file sizes are very large and unwieldy. In this set of studies, however, we are looking specifically at virtual slides used in telepathology. Briefly, telepathology is the practice of pathology at a distance. Although various forms exist, we are concerned with the type in which traditional glass slides are digitized, stored and transmitted electronically for remote interpretation [2]. Once received by a consulting pathologist, the digital or virtual slides are viewed on a high-performance workstation. Once interpreted, the results are either phoned or faxed back to the referring clinician. Numerous studies have been carried out to demonstrate the equivalence in diagnostic accuracy between light microscopy with slides and various forms of digital slides [3]–[5]. However, there are still important technological issues that remain to be resolved [6].

One of these issues is the size of the virtual slides. Once digitized, the image files are extremely large and can cause unacceptable delays in image display and interactive visualization by pathologists retrieving the data from a server or storage device. The virtual slides used in our past and this present study are from the DMetrix (DMetrix, Inc., Tucson, AZ) scanner. This system is used in the Arizona Telemedicine Program and samples images at 0.47 μm/pixel (or 54 045 dots/in). A typical slide contains 1–4 cm^2 of tissue and a single 40X objective scan with a system scanning at 0.47 μm/pixel yields 1–5 GB of un-compressed RGB image data [2]. These requirements can increase depending on the clinical task. Some situations require only a low-resolution (20X objective) scan, but others require

significantly higher resolutions (80X or 100X objective). The DMetrix Ultrarapid Virtual Slide Processor has a high-resolution (NA = 0.65) 80-objective array that yields approximately 12 GB of scanned data. The result is significant concern among users and in the DICOM (Digital Imaging and Communications in Medicine) Pathology Working Group (WG-26) that DICOM cannot handle images larger than 64 000 pixels and 2 GB total size [7].

One way to deal with these huge volumes of virtual slide data, of course, is with compression. The problem is that it is difficult, if not impossible, to define a single "minimum" level of compression (hence image quality) for use across all clinical questions [8]. To date, most of the studies investigating compression in telepathology have been done using JPEG. However, JPEG 2000 is generally the method of choice in recent years. Both methods have advantages and disadvantages. JPEG artifacts tend to be "blocky" and JPEG 2000 artifacts tend to be "blurry." The trade-offs between the type of artifact and visibility have yet to be studied rigorously with virtual telepathology slides. Part of the reason for this is that observer studies are generally difficult to conduct. They typically require a fairly large set of carefully selected images, multiple viewing sessions and significant time and effort on the part of the pathologist readers. As clinical duties are ever-increasing, the time pathologists have to spend in such studies is limited. Therefore, alternate methods to at least narrow down the range of compression ratios to be studied are starting to be used. In particular, the use of mathematical observer models to predict human performance can be very useful in this respect [9]–[14].

Thus, we are investigating a method to improve the presentation of telepathology images for accurate diagnoses by using image compression schemes and displays that are based on information about the properties and capabilities of the human visual system [15]. With this approach we believe that we can improve image compression schemes for specific telemedicine applications, thereby enabling: 1) faster delivery of images without compromising visual quality, utility or diagnostic accuracy, 2) consistently uniform visual quality with variable compression rates optimized across images, patients, and display technologies, and 3) automatic, objective image-quality control based on simulations of visual perception.

II. Methods and Materials

Test images for visually lossless threshold experiments were selected from regions in a set of 93 uncompressed virtual slides produced by a DMetrix scanner for stained breast biopsy specimens; an example slide is provided in Fig. 1. File sizes for the slide data with *reversible* compression ranged from 0.25 to 6.05 GB, corresponding to *uncompressed* sizes of 0.37 GB (9944 × 13 304 pixels) to 7.49 GB (44 104 × 60 744 pixels). The mean file size with reversible compression was 2 GB. Diagnostically relevant regions in each slide were selected by an expert pathologist and labeled benign or malignant. From those regions, 10 benign and 10 malignant test images, each 256 × 256 pixels, were cropped. Examples of the images are shown in Fig. 2.

A Bayesian adaptive psychometric procedure known as QUEST, [16] as implemented in the Psychtoolbox-3 MATLAB software [17] and modified by the authors of this paper to support threshold experiments with compressed images, was used to determine the visually lossless compression threshold for each test image. Images were compressed using the Kakadu 6.0 [18] implementation of JPEG 2000 with its default rate-control algorithm based on mean squared error (MSE).

In each trial, four versions of the same image were presented as shown in Fig. 3. One of the two lower images was compressed and the other three images were uncompressed. The two-alternative, forced-choice (2AFC) task in each trial was to select which of the two bottom images appeared different from the others. Experiments were conducted with 18 observers (experienced image analysts, medical students, residents, and pathologists) in four sessions with five images per session. The Weibull psychometric function in the Psychtoolbox QUEST procedure has the form

$$\psi(x-x_t)=\{1-(1-\gamma)\exp[-10^{\beta(x-x_t+\varepsilon)}]\}(1-\delta)+\delta\gamma \quad (1)$$

where x is the log of the stimulus intensity, x_t is the estimated threshold, $\gamma = 0.5$ for 2AFC responses, δ accounts for the small probability of incorrect responses well above threshold, β influences the slope of ψ, and ε selects the value of ψ (percent correct) corresponding to threshold.

Intensity in the QUEST methodology is normally interpreted as Michelson contrast in a visual stimulus. For the detection of losses in compressed images, the stimulus variable is bit rate rather than contrast, where decreasing bit rate corresponds to increasing stimulus intensity and a higher probability of visual detection. A transformation between compression bit rate and the QUEST log threshold intensity, $x = \log_{10}$ (contrast), was developed to satisfy the inverse relationship between contrast and bit rate in the interval from 0 to 1

$$x(r)=\log_{10}\left[\frac{\alpha(1-r)}{(\alpha+r)}\right] \quad (2)$$

where r = (bit rate)/24 is the normalized bit rate and α is a curvature parameter. A reasonable estimate for β was determined by fitting the Weibull function in (1) to 2AFC data collected for several test images using the method of constant stimuli (MCS) for six compression levels spanning the visually lossless threshold. A least-squares fit to the MCS data was obtained by varying β and x_t for a given value of α. The results for one image with $\alpha = 0.01$ are shown in Fig. 4. This value of α resulted in β values near 3 that were insensitive to variations in α.

The QUEST procedure was initialized and conducted in the same manner for all of the test images. The Bayesian prior probability density function (PDF) was modeled as a Gaussian function in x with a peak at $x_t = -0.963$ (1.8 bpp) and standard deviation of 0.234 (~ 1 bpp), values that were estimated from the results of MCS and trial QUEST experiments (see Figs. 4 and 5). Psychometric parameters were set at $\beta = 3$, $\delta = 0.02$, and 75% correct at threshold.

The five images in each test session were presented in randomized order and orientation (flipped and rotated). Images were displayed on a medical-grade Barco Coronis Fusion 6 megapixel color LCD monitor with a maximum luminance of 400 cd/m^2. The Bayesian posterior PDFs after 50 trials per image were divided by the prior PDF to compute the likelihood function for the threshold variable. The visually lossless threshold for each image was determined by its maximum likelihood value.

The QUEST algorithm produced a likelihood function for each test image and observer in the log threshold intensity variable x. The likelihood functions for each image were then averaged across observers to generate a "mean observer" likelihood function with peak position and width determined by the means of the likelihood peak locations and standard deviations for all observers. The threshold variable for the mean-observer likelihood functions was then transformed to compression bit rate using the inverse of (2).

Further transformations of the threshold variable from bit rate to various distortion metrics were performed to determine if any of the threshold distortion values were sufficiently invariant across test images to provide a reliable target for achieving visually lossless compression. The threshold variable was transformed using rate-distortion curves computed for each test image and metric. The metrics included just-noticeable differences (mean, Minkowski pooled, histogram percentiles 90th, 95th, 99th, 99.9th, and maximum JND), PSNR, and structural similarity metric (SSIM) [19]. For comparison to the higher JND percentiles, we also computed the 99th percentile of squared error (PSNR is based on the *mean* squared error) and the first percentile of SSIM pixel values (the default SSIM metric is the *mean* across image pixels). Distortion metrics were computed as a function of compression bit rate from 0.7 to 4.3 bpp. Parametric functions were fit to the rate-distortion data using a least-squares criterion. A log-linear function with three parameters, $d(r) = a_1 \log(r) + a_2 r + a_3$, was fit to the JND metrics and PSNR. A five-parameter logistic function, $d(r) = a_1\{0.5 - [1 + \exp(a_2(r - a_3))]^{-1}\} + a_4 r + a_5$, was fit to the SSIM metrics [19].

JND metrics of compression loss visibility were computed using a visual discrimination model (VDM) developed at Siemens Corporate Research. The VDM simulates factors in the ocular and cortical processing of luminance stimuli by the human visual system [20], [21]. Initial stages of the model account for effects of the ocular modulation transfer function and luminance adaptation. The resulting image is processed by a 2-D Fourier transform and filtered in the frequency domain by a set of biologically inspired spatial frequency- and orientation-tuned channels represented by a set of bandpass log-Gabor filters [22]. Local band-limited contrast [23] is computed by dividing the output of each bandpass channel by the output of a low-pass, isotropic Gaussian filter applied to the image.

Channel contrasts are then normalized to 1 JND at the detection threshold using a contrast sensitivity function that depends on spatial frequency and luminance [24]. Contrast discrimination sensitivity at suprathreshold (JND > 1) contrast levels and interactions between channels are modeled by a combination of nonlinear excitatory and inhibitory (divisive suppression) factors associated with contrast or texture masking [25]. In the final stage, channel JND maps can be analyzed individually or pooled over orientation and frequency, typically by determining the maximum response at each spatial location.

Summary JND metrics can be evaluated by spatial pooling across pixel locations, typically by computing the mean, a histogram percentile, or Minkowski summation [23] with an exponent of 4. Metrics can be computed across an entire image or within regions or frequency/orientation channels containing specific features of interest. When applied to a pair of uncompressed and compressed images, the VDM generates objective measures of the visibility of compression artifacts in perceptually linear JND units. For the metric calculations in this study, only channel map JND values greater than 1 were considered.

The statistical significance of differences in threshold likelihood between test images was determined using a likelihood ratio chi-square test. We tested the hypothesis that all of the visually lossless thresholds were equal for a given threshold variable. The likelihood ratio, Λ, is defined as the maximum likelihood under the restricted (null) hypothesis, H_0, of equal thresholds to the maximum likelihood under the unrestricted hypothesis, H_1, that the thresholds are not equal

$$\Lambda = \frac{Max_{H_0} L(D|T)}{Max_{H_1} L(D|T)} \quad (3)$$

where $L(D|T)$ is the likelihood of threshold variable, T, given a set of observations, D. For each image in our threshold experiments, D consists of pairs of stimulus (bit rate) and response (correct/incorrect) values. The ratio Λ was evaluated using the mean-observer likelihood functions, $\bar{L}_i$, computed for the test images ($n = 20$)

$$\Lambda = \frac{Max\left[\prod_{i=1}^{n} \bar{L}_i(T)\right]}{\prod_{i=1}^{n} Max[\bar{L}_i(T)]}. \quad (4)$$

The test statistic defined by $\lambda = -2 \ln \Lambda$ can be approximated by a chi-square distribution with degrees-of-freedom (df) equal to the difference in df between the unrestricted and restricted hypotheses; for this test, df = 20 − 1 = 19. The probability, p, of incorrectly rejecting the restricted (null) hypothesis when it is true was computed using the chi-square cumulative distribution (MATLAB chi2cdf function), such that $p = 1-\text{chi2cdf}(\lambda, \text{df})$. A significance level of 0.05 was selected and the null hypothesis rejected if $p < 0.05$. Values of $p > 0.05$ for a given threshold variable implied that the hypothesis of equal thresholds for visually lossless compression could not be rejected.

III. Results

Visually lossless compression ratios, averaged across observers for the 20 test images, ranged from 8.5:1 to 21.1:1 with a mean of 12.9:1 (see Fig. 6). Reversible compression ratios were significantly lower and varied from 1.45:1 to 1.82:1 with a mean of 1.65:1. Lossy compression to the visually lossless thresholds reduced the amount of image data by a factor of 5–12 compared to reversible compression. The range of visually loss-less ratios was correlated with image characteristics affecting compressibility. Images with relatively low ratios (<10:1) tended to have greater concentrations of high-frequency structures, e.g., stained nuclear and membrane structures in sharp focus, while those with higher ratios (>15:1) contained fewer such structures and/or were slightly out of focus. Test images representing both ends of the compressibility range are shown in Fig. 7.

Examples of threshold likelihood functions generated by the QUEST procedure for one test image are shown in Fig. 8 for the 18 observers. The mean-observer likelihood function for this image is shown as a thicker (red) curve. All of the likelihood functions were scaled for unit area. The mean-observer likelihood functions for all 20 test images are shown in Fig. 9 with the threshold variable transformed from QUEST log intensity to compression bit rate.

Further transformations of the threshold variable to distortion metrics were performed using the rate-distortion curves, illustrated in Fig. 10(a), Fig. 11(a), and Fig. 12(a) for two test images representing low and high bpp thresholds. The mean-observer likelihood functions with threshold variables converted to distortion metrics are shown in Fig. 10(b), Fig. 11(b), and Fig. 12(b) for PSNR, JND (99th percentile), and SSIM, respectively. The horizontal scales in those figures were selected to maintain the same average standard deviation as the functions plotted in Fig. 9. With PSNR as the threshold variable, the maximum likelihoods ranged from 37.4 to 41.7 dB, far exceeding the 0.5 dB difference often considered significant for PSNR, and the spread of the likelihood functions was comparable to Fig. 9 for bit rate. When the threshold variable was transformed to the 99th percentile JND, however, the likelihood functions were much more tightly grouped near a JND of 4 with maximum likelihoods varying by only 0.45 JND, well below the standard visibility threshold of 1 JND.

The spread of likelihood functions with SSIM as the threshold variable was somewhat greater than for PSNR and bit rate, indicating that SSIM would be a less reliable metric for

selecting bit rates for visually lossless compression. Results of the likelihood ratio chi-square test of the null hypothesis, which posits equal VL thresholds for all test images, are presented in Fig. 13.

For threshold metrics with $p < 0.05$, the variations across images are significant and the null hypothesis should be rejected. A value of $p > 0.05$ suggests that the threshold metric could be sufficiently constant across images to provide a reliable target for achieving visually lossless compression. The only distortion metrics with p values significantly above 0.05 were the high-percentile and maximum JND values. The greatest p value (0.84) was obtained for the 99th percentile JND. All of the threshold metrics based on spatial averaging —mean JND, PSNR, and SSIM—and JNDs below the 95th percentile, varied significantly across test images. The 99th percentile squared-error metric was only slightly better than PSNR, while the first percentile SSIM (lower SSIM corresponds to higher image distortion) was near the significance level but still performed well below the high-percentile JND threshold metrics.

These results can be understood as a consequence of the nonuniform distributions of error visibility that occurred as compression rates were increased up to and beyond the visually lossless thresholds. The default encoder rate-control algorithm used in these experiments tries to achieve a target *mean-squared error* throughout an image, a condition which does not generally correspond to uniform error *visibility*. As compression bit rate decreases and losses first become noticeable, typically as blurring or ringing artifacts for JPEG 2000 compression, they tend to most visible in regions with less masking by high-contrast features or texture.

The images compressed in these experiments had complex structures comprising cell nucleus, membrane, vessel, and other tissue features with contrasts enhanced by staining. The effects of contrast/texture masking, which tend to reduce compression error visibility, were generally strong in all but a relatively small fraction of the total areas of these images. Differences in the relatively small and scattered locations in which compression artifacts were most noticeable at threshold had to be detected in each trial by visual scanning between images (four per trial) and visual searching through complex cell/tissue patterns, many of which could have acted as distractors and effectively reduced the probability of detecting differences that would have been more noticeable if they had occurred on uniform backgrounds (much less masking) and with known locations (no search required). Consequently, the threshold metrics that were most consistent for images compressed to their visually lossless thresholds were those corresponding to the higher JND levels occurring in a relatively small fraction (~ 1%) of the image.

IV. Conclusion

This paper presents new results generated in an ongoing research project to correlate VDM metrics with human observer performance using compressed virtual slides for telepathology. Threshold likelihood functions for visually lossless JPEG 2000 compression were determined experimentally using the Bayesian adaptive QUEST psychometric procedure. Mean thresholds ranged from 8.5:1 to 21.1:1 for 20 test images with a mean of 12.9:1. The statistical significance of variations in threshold likelihood functions across test images was evaluated for various threshold metrics using a likelihood ratio chi-square test. The threshold metric showing the greatest uniformity across images (likelihood ratio chi-square $p = 0.84$ and $\Delta JND < 0.5$) was the 99th percentile JND computed by the VDM. This result is likely due to a combination of image, encoder, and observer task characteristics, primarily the spatially nonuniform emergence at threshold of noticeable artifacts embedded in highly

structured images with significant contrast/texture masking effects, and a discrimination task requiring visual search among distractors.

Further VDM refinements will be developed to account for the effects of visual search and distractor masking on observer performance. The low variability of some VDM metrics for images compressed to different visually lossless bit rates suggests that those metrics could guide the adaptive compression of virtual slides to consistently provide visually lossless image quality while achieving 5–12 times the data reduction of reversible compression methods.

From the clinical perspective, developing compression techniques tuned to human perception could have two significant benefits. First, it could lead to preferentially storing the most important image information with significant reduction in storage requirements. Second, the intelligent compression and rapid transmission of image data could enable ultrafast viewing with significant improvements in the productivity of pathologists using digital images instead of glass slides. Both of these could help remove the major impediments to increased adoption of digital-imaging technology in pathology, namely the long-term image storage requirements and the current productivity-diminishing state of the technology [6].

As noted above, it appears that benign and malignant images may be compressible to different rates because of differences in the basic cellular content. In this study we did not consider higher levels of diagnosis in which specific categories of benign and malignant conditions are differentiated or classified/graded. For example, DCIS (ductal carcinoma *in situ*) is a preinvasive carcinoma that has three grades [26]. In Grade 1, there are monotonous nuclei of about 1.5–2.0 red blood cell (RBC) diameters with finely dispersed chromatin and occasional nucleoli. Grade 3 however is characterized by markedly pleomorphic nuclei generally larger than 2.5 RBC diameters, with coarse chromatin and prominent or multiple nucleoli. Grade 2 falls in between 1 and 3. There have yet to be any studies examining the compression characteristics of breast biopsy specimens at this level, but clearly the differences in physical structures is likely to impact the degree to which they can be compressed before artifacts are induced that impact diagnostic performance.

In the evaluation of the impact of compression on diagnostic performance in pathology as well as any other medical imaging task, it is also important to note that the degree of compression achievable and acceptable will depend greatly on the original quality of the images. It has already been demonstrated that the quality of the tissue section thickness and consistency affect whole slide imaging capture and quality [27] and that steps such as appropriate use of optical filters can improve the appearance and resolution of virtual slides [28], although the impact of these measures on compression has yet to be evaluated. There are even efforts underway to standardize the technological aspects of acquiring, transmitting and displaying virtual slides [29], [30], as well as associated methods to measure image quality of the scanned specimens [31]. Interestingly, the quality of the scan may depend on the type of staining performed. In the present study, the specimens were H&E stained (the most common type) but there are also fluorescent and immunohistochemical stains that are used depending on the tissue and task, and these stains influence the quality of the scans and thus have the potential to influence the quality of the compression [31].

There is therefore a significant amount of work yet to be done before we fully understand the intricate relationships between original specimen quality, scanned image quality, the amount and type of compression applied and the impact on diagnostic interpretation and other related diagnostic tasks (e.g., grading).

Finally, it is important to note that although the present study used JPEG2000, there are other compression schemes available including JPEG2000's predecessor JPEG. JPEG2000 solves some of the artifact problems that JPEG suffers from (e.g., blocking and color distortion, which is particularly relevant for pathology), but does not solve all of them. JPEG2000 is generally considered an advantage over JPEG at lower bit rates, but at higher bit rates the differences in image quality are smaller and JPEG2000 may not work as well as JPEG in images with noisy textures. Most virtual slide scanners (or whole slide imagers) already incorporate some type of compression (usually proprietary) in the scanning process. In some cases it can be turned off but in most it cannot so it is important to consider how the additional compression of an image by JPEG2000, JPEG or any other scheme is affected by the initial scanning compression. Again, there has been little to no study on this topic to date.

Acknowledgments

This work was supported by National Institutes of Health/National Institute of Biomedical Imaging and Bioengineering (NIH/NIBIB) under Grant R01 EB008055.

References

1. Johnson JP, Krupinski EA, Nafziger JS, Yan M, Roehrig H. Visually lossless compression of breast biopsy virtual slides for telepathology. Proc SPIE Med Imag. 2009; 7263:72630N-1–72630N-8.
2. Weinstein RS, Descour MR, Liang C, Barker G, Scott KM, Richter L, Krupinski EA, Bhattacharyya AK, Davis JR, Graham AR, Rennels M, Russum WC, Goodall JF, Zhou P, Olszak AG, Williams BH, Wyant JC, Bartels PH. An array microscope for ultrarapid virtual slide processing and telepathology. Design, fabrication, and validation study. Human Pathol. 2004; 35:1303–1314. [PubMed: 15668886]
3. Weinstein RS, Descour MR, Liang C, Bhattacharyya AK, Graham AR, Davis JR, Scott KM, Richter L, Krupinski EA, Szymus J, Kayser K, Dunn BE. Telepathology overview: From concept to implementation. Human Pathol. 2001; 32:1283–1299. [PubMed: 11774159]
4. Weinstein RS, Lopez AM, Barker GP, Krupinski EA, Descour MR, Scott KM, Richter LC, Beinar S, Holcomb MJ, Bartels PH, McNeely RA, Bhattacharyya AK. The innovative bundling of teleradiology, telepathology, and teleoncology services. IBM Syst. 2007; 46:69–84.
5. Kayser, K.; Szymas, J.; Weinstein, RS. Telepathology and Telemedicine. Berlin, Germany: VSV Interdisciplinary Medical; 2005.
6. Krupinski EA. Virtual slide telepathology workstation-of-the-future: Lessons learned from teleradiology. Sem Diag Path. 2009; 26:194–205.
7. DICOM WG-26 Pathology: Minutes. Jan 5. 2007 [Online]. Available: http://www.conganat.org/digital/index.htm
8. Yagi Y, Gilberson JR. Digital imaging in pathology: The case for standardization. J Telemed Telecare. 2005; 11:109–116. [PubMed: 15901437]
9. Johnson JP, Lubin J, Krupinski EA, Peterson HA, Roehrig H, Baysinger A. Visual discrimination model for digital mammography. Proc SPIE Med Imag. 1999; 3663:253–263.
10. Johnson JP, Krupinski E, Nafziger J, Lubin J, Wus J, Roehrig H. Perceptually optimized compression of mammograms. Proc SPIE Med Imag. 2002; 4686:256–262.
11. Johnson JP, Nafziger JS, Krupinski EA. Effects of grayscale window/level parameters on lesion detectability. Proc SPIE Med Imag. 2003; 5034:462–473.
12. Krupinski EA, Johnson JP, Roehrig H, Lubin J. Using a human visual system model to optimize soft-copy mammography display: Influence of display phosphor. Acad Radiol. 2003; 10:161–166. [PubMed: 12583567]
13. Krupinski EA, Johnson JP, Roehrig H, Lubin J, Engstrom M. MTF correction for optimizing softcopy display of digital mammograms: Use of a vision model for predicting observer performance. Proc SPIE Med Imag. 2003; 5034:323–327.

14. Krupinski EA, Johnson J, Roehrig H, Nafziger J, Lubin J. On-axis and off-axis viewing of images on CRT displays and LCDs: Observer performance and vision model predictions. Acad Radiol. 2005; 12:957–964. [PubMed: 16023384]
15. Johnson JP, Krupinski EA, Yan M, Roehrig H. Use of a visual discrimination model to detect compression artifacts in virtual pathology images. Proc SPIE Med Imag. 2010; 7627:762704–762704.
16. Watson AB, Pelli DG. QUEST: A Bayesian adaptive psychometric method. Perception Psychophys. 1983; 33:113–120.
17. Brainard DH. The psychophysics toolbox. Spatial Vis. 1997; 10:433–436.
18. Kakadu JPEG. 2000. [Online]. Available: http://www.kakadusoftware.com
19. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. IEEE Trans Image Process. Apr; 2004 13(4):600–612. [PubMed: 15376593]
20. Lubin, J. A visual discrimination model for imaging system design and evaluation. In: Peli, E., editor. Visual Models for Target Detection and Recognition. Singapore: World Scientific; 1995. p. 245-283.
21. Daly, S. The visual differences predictor: An algorithm for the assessment of image fidelity. In: Watson, A., editor. Digital Images and Human Vision. Cambridge, MA: MIT Press; 1993. p. 179-206.
22. Field DJ. Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am. 1987; 4:2379–2394.
23. Peli E. Contrast in complex images. J Opt Soc Am. 1990; 7:2032–2040.
24. Barten, PGJ. Contrast Sensitivity of the Human Eye and Its Effects on Image Quality. Bellingham, WA: SPIE Press; 1999. p. 27-60.
25. Foley JM. Human luminance pattern-vision mechanisms: Masking experiments require a new model. J Opt Soc Am. 1994; 11:1710–1719.
26. Rosen, PP. Rosen's Breast Pathology. Philadelphia, PA: Lippincott Williams & Wilkins; 2009.
27. Yagi Y, Gilbertson JR. A relationship between slide quality and image quality in whole slide imaging (WSI). Diag Path. 2008; 3:S12–S12.
28. Yagi Y, Gilbertson JR. The importance of optical optimization in whole slide imaging (WSI) and diogital pathology imaging. Diag Path. 2008; 3:S1–S1.
29. Kayser K, Gortler J, Goldmann T, Vollmer E, Hufnagl P, Kayser G. Image standards in tissue-based diagnosis (diagnostic surgical pathology). Diag Path. 2008; 3:3–17.
30. Lundin M, Szymas J, Linder E, Beck H, deWilde P, van Krieken H, Rojo MG, Moreno I, Ariza A, Tuzlali S, Dervisoglu S, Helin H, Lehto VP, Lundin J. A European network for virtual microscopy —Design, implementation and evaluation of performance. Virchows Arch. 2009; 454:421–429. [PubMed: 19280223]
31. Kayser K, Gortler J, Metze K, Goldmann T, Vollmer E, Mireskandari M, Kosjerina Z, Kayser G. How to measure image quality in tissue-based diagnosis (diagnostic surgical pathology). Diag Path. 2008; 3:S11–S11.

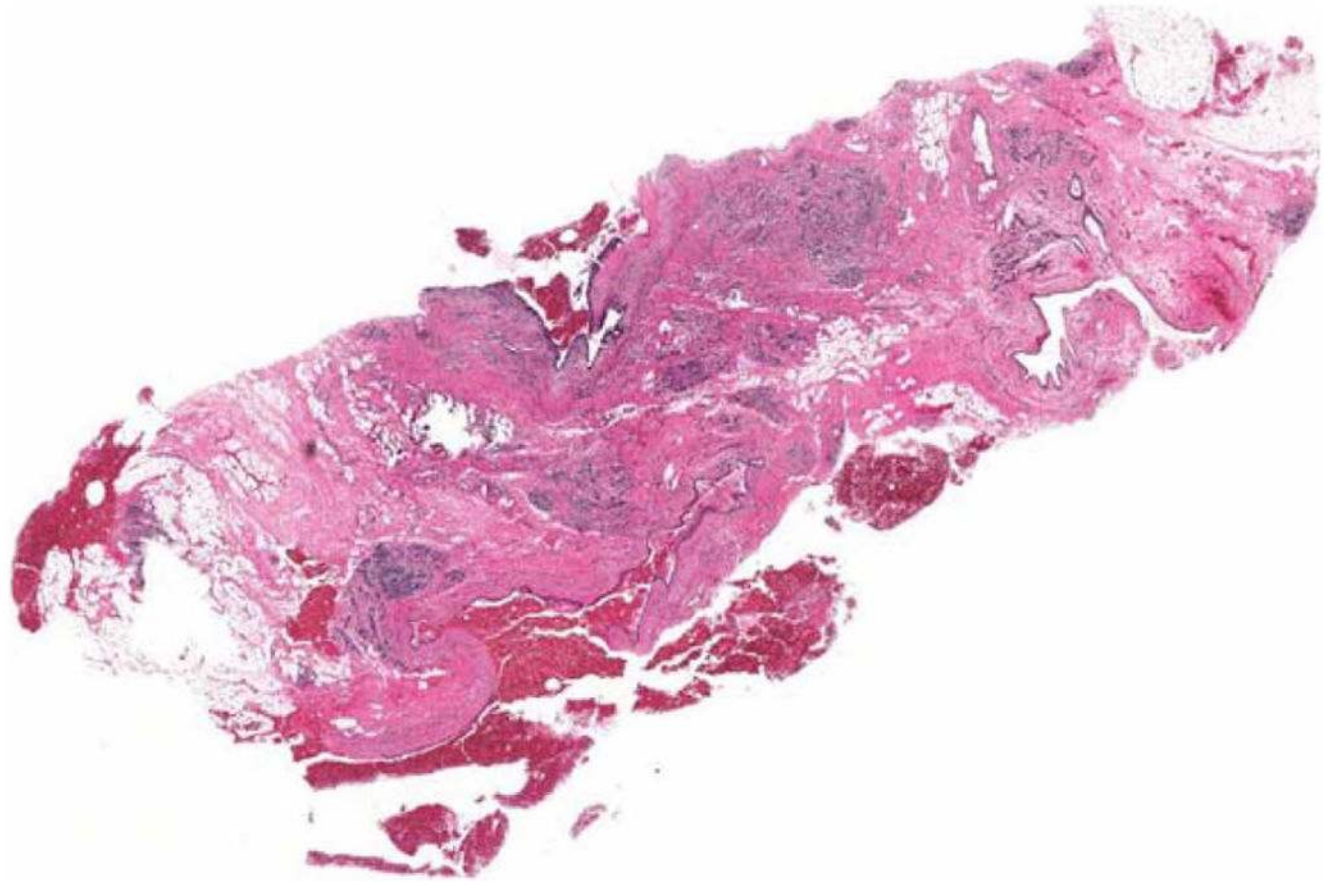

Fig. 1.
Virtual slide of a stained breast biopsy specimen.

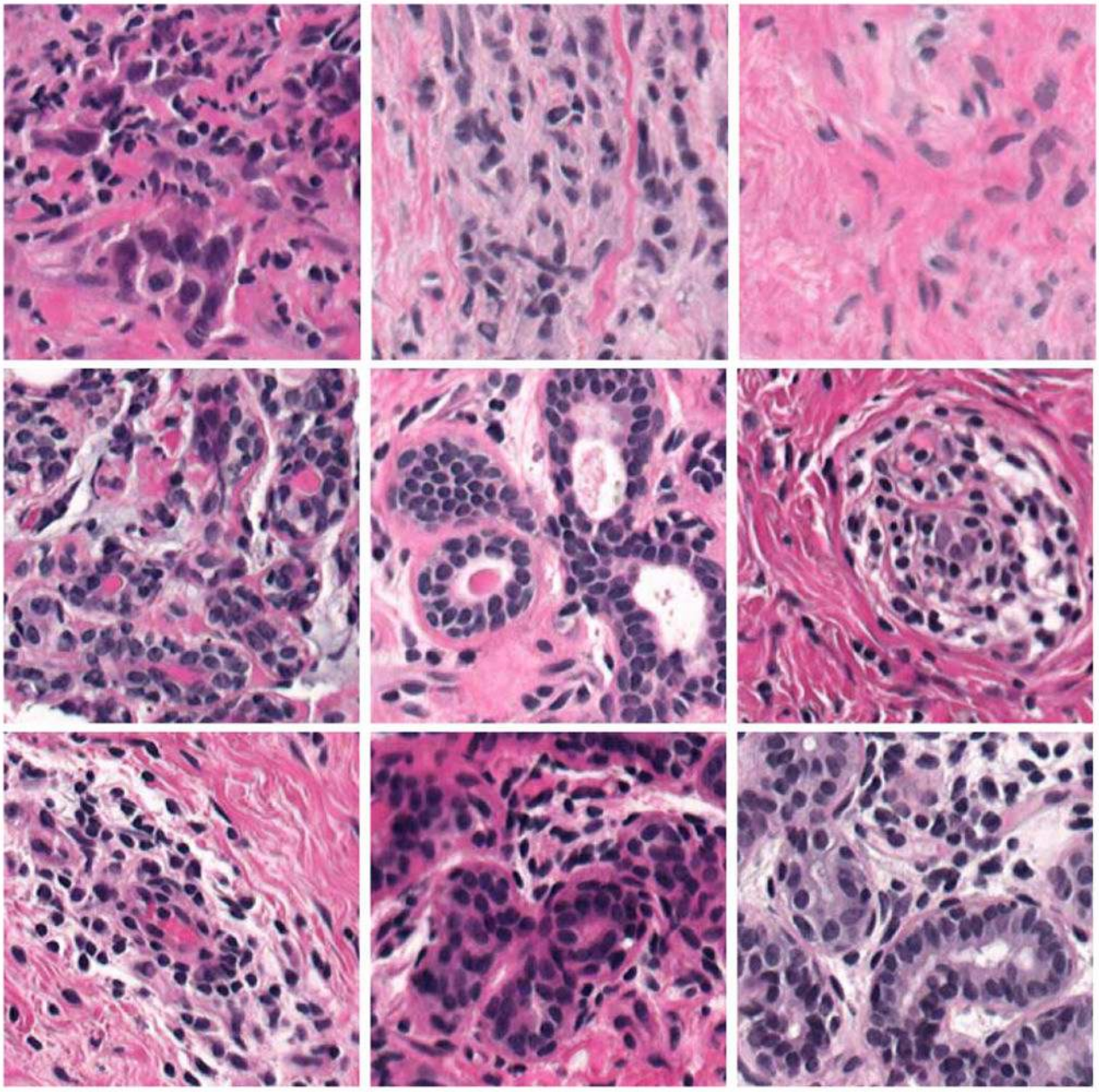

Fig. 2.
Examples of test images (256 × 256 pixels) cropped from breast biopsy virtual slides.

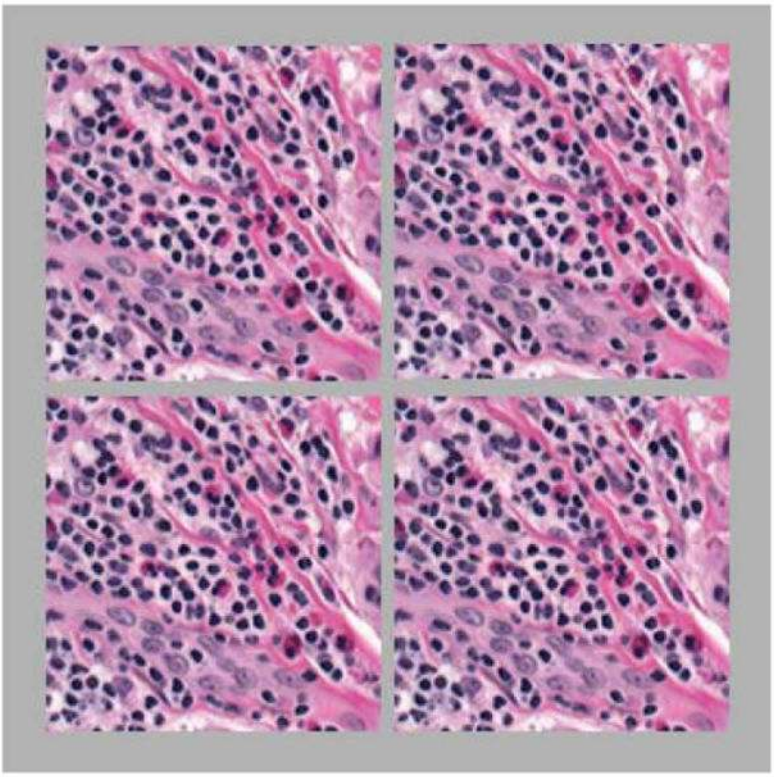

Fig. 3.
Array of four images displayed in each QUEST trial to determine the visually lossless threshold.

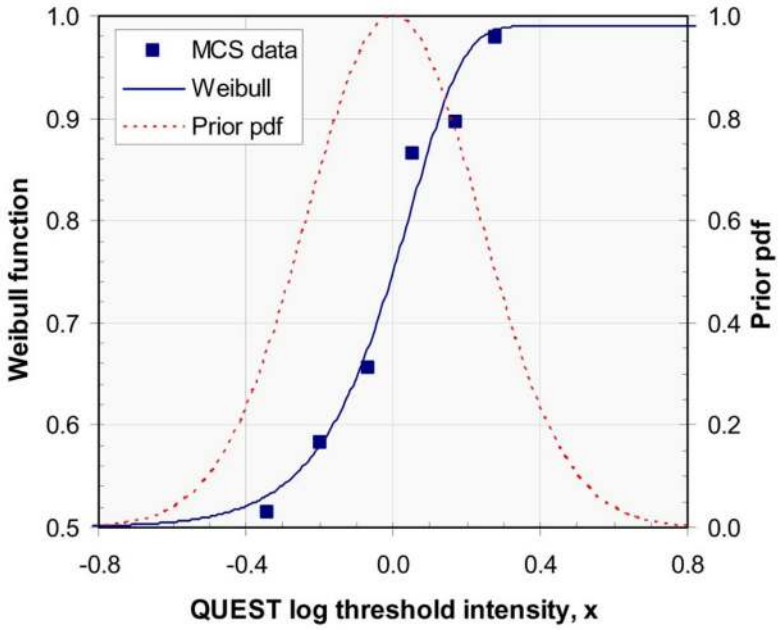

Fig. 4.
Weibull psychometric function fit to MCS observer data for one test image. Bayesian prior PDF used in the QUEST experiments.

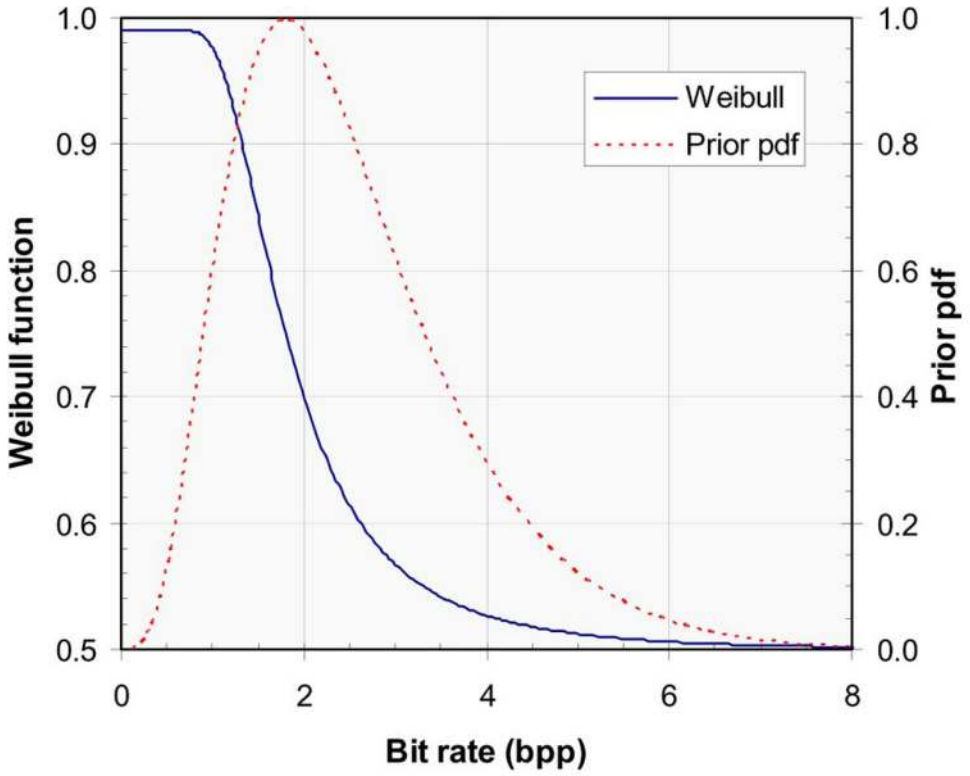

Fig. 5.
Weibull psychometric function and prior PDF with threshold variable transformed to bit rate and shifted to an initial threshold estimate of 1.8 bpp.

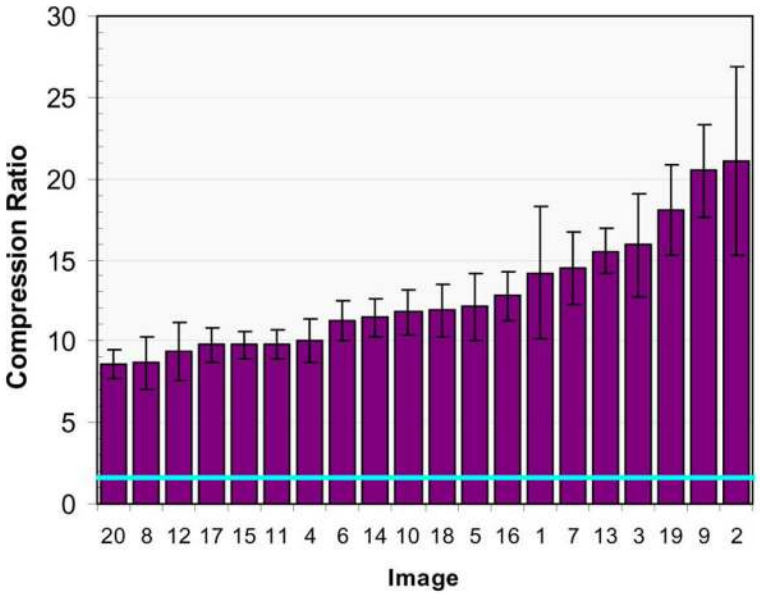

Fig. 6.

Visually lossless ratios for JPEG 2000 compression of 20 images, averaged across observers (bars show 95% confidence intervals). The horizontal line shows the mean ratio for *reversible* compression (1.65:1).

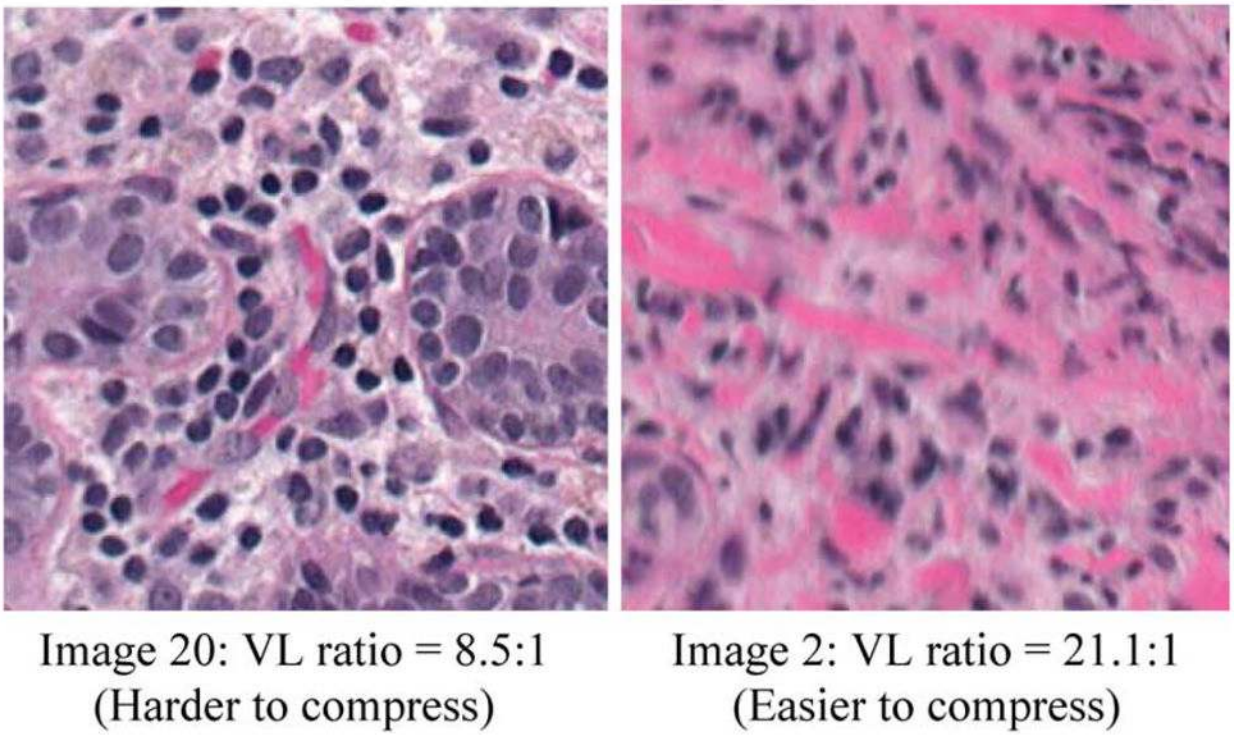

Fig. 7.
Test images with minimum and maximum visually lossless (VL) compression ratios.

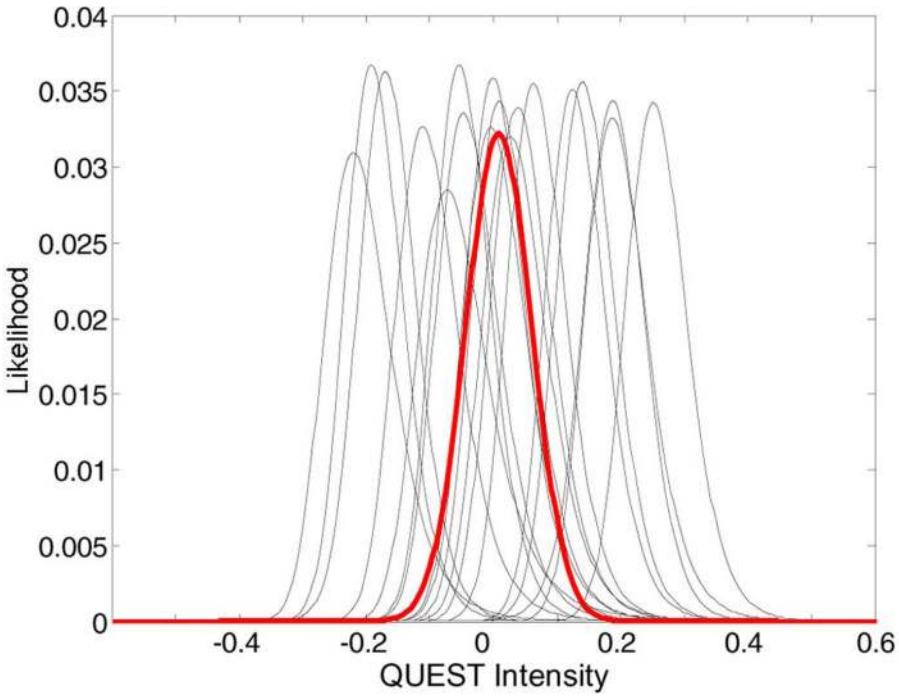

Fig. 8.

Threshold likelihood functions for 18 observers viewing the same test image. The mean-observer likelihood function is shown as a thicker (red) curve.

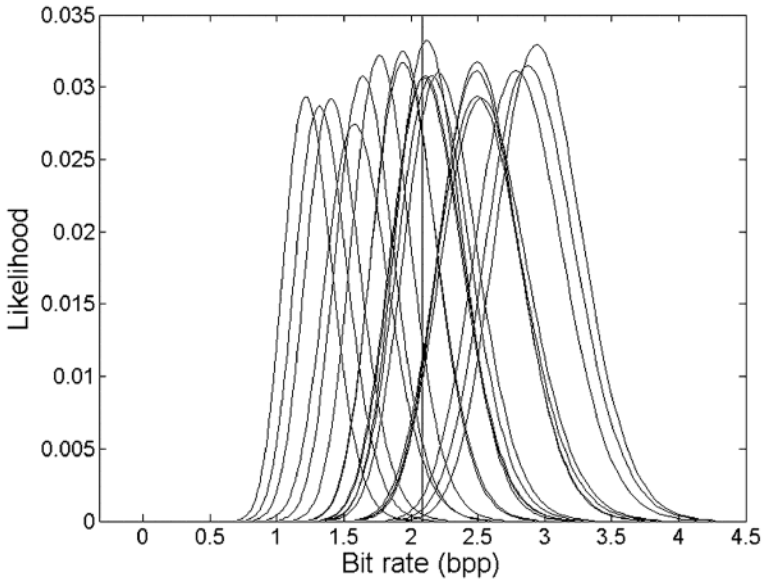

Fig. 9.
Mean-observer likelihood functions for 20 test images with the threshold variable transformed to compression bit rate.

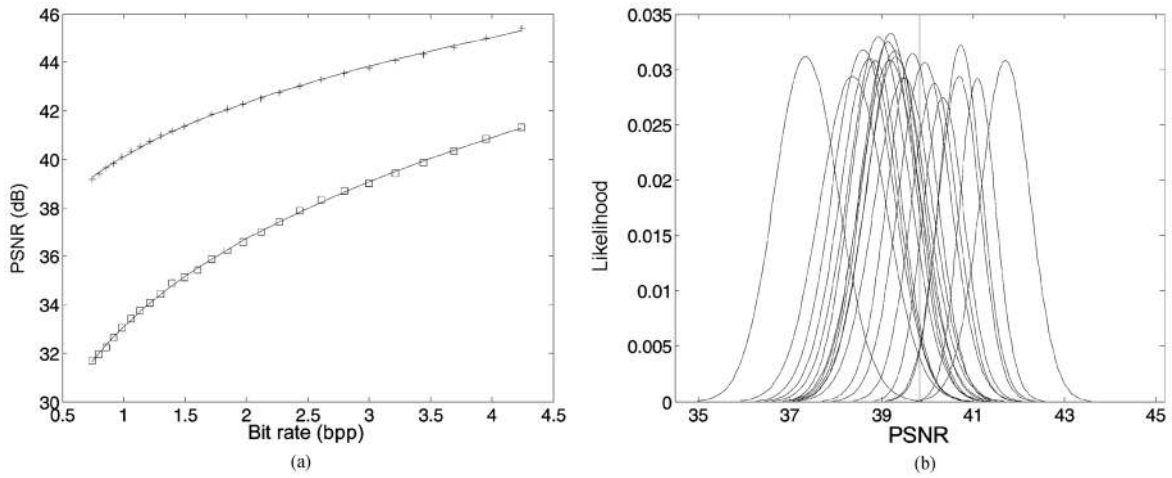

Fig. 10.

(a) PSNR rate-distortion data and best-fit curves for test images 8 (□) and 9(+) with extreme VL thresholds in bit rate. (b) Mean-observer likelihood functions for 20 test images with threshold variable transformed from bit rate to PSNR.

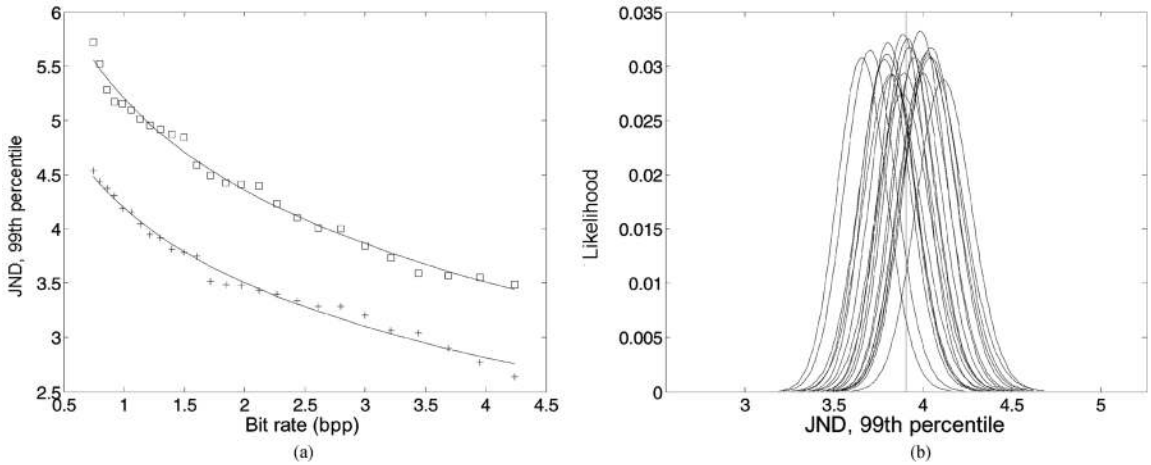

Fig. 11.

(a) JND rate-distortion data and best-fit curves for test images 8 (□) and 9(+) with extreme VL thresholds in bit rate. (b) Mean-observer likelihood functions for 20 test images with threshold variable transformed from bit rate to 99th percentile JND.

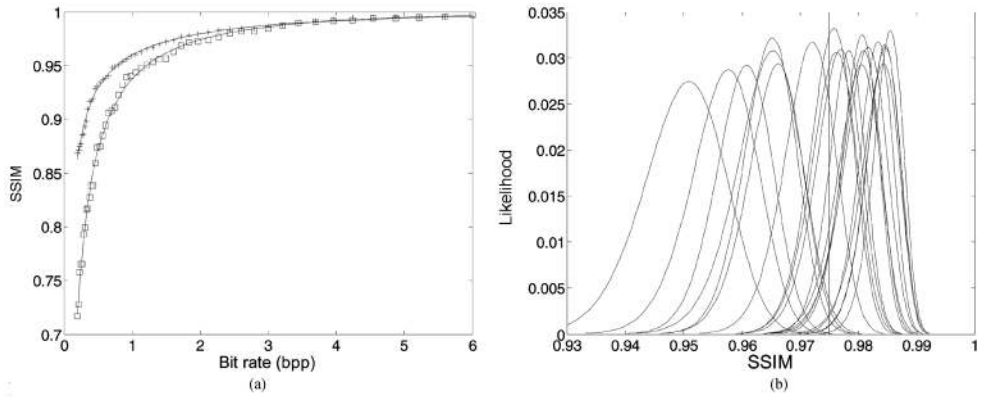

Fig. 12.

(a) SSIM rate-distortion data and best-fit curves for test images 8 (□) and 9(+) with extreme VL thresholds in bit rate. (b) Mean-observer likelihood functions for 20 test images with threshold variable transformed from bit rate to SSIM.

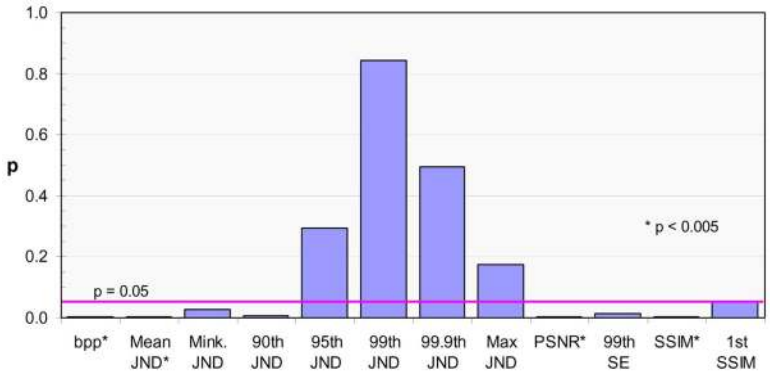

Fig. 13.

From the likelihood ratio chi-square tests for various threshold metrics, the probability *p* of incorrectly rejecting a true null hypothesis that assumes equal thresholds for all images.