SPATIAL RESOLUTION AND QUANTIZATION NOISE TRADEOFFS FOR SCALABLE IMAGE COMPRESSION

Soo Hyun Bae, Thrasyvoulos N. Pappas[†], Biing-Hwang Juang

Center for Signal and Image Processing, Georgia Institute of Technology, Atlanta, GA 30332
[†]EECS Department, Northwestern University, Evanston, IL 60208
{soohyun, juang}@ece.gatech.edu, pappas@ece.northwestern.edu

ABSTRACT

Most full-reference quality metrics compare the original image to a distorted image at the same level of resolution assuming a fixed viewing distance. In video streaming applications, however, the transmitted or received signal may differ from the original in compression as well as spatiotemporal resolution. For example, at low bitrate coding applications the compressed image may be too distorted, and hence the observer may prefer to reduce the resolution or increase the viewing distance in order to reduce the visibility of the compression artifacts. The selection of the best tradeoff between resolution/viewing distance and visibility of compression artifacts requires a quality metric that accounts for both image distortions and image size. Such tradeoffs are not reflected in existing quality metrics, which ignore the signal visibility and only measure the visibility of compression distortions, which decrease with image size. In order to better understand such tradeoffs, with the goal of developing better quality metrics, we conducted subjective tests using a number of existing still image coders (JPEG2000, SPIHT, and JPEG). Our results indicate that the objective quality (perceptually weighted PSNR) of the images that the viewers select decreases with resolution, that is, the viewers are willing to accept more artifacts as image size decreases.

1. INTRODUCTION

Recent advances in display and video capture technologies and digital communications have led to the development of a variety of video services. The spatiotemporal resolution of the video signals that these services provide depends on the video capture device, transmission bandwidth, and display. However, in many cases, a video sequence may be transmitted to a variety of users with different bandwidths and display devices. In such cases, each transmitted video bitstream should be adjusted according to the channel bandwidth and display device of the user. Thus, a video content provider may be required to transmit a variety of bitstreams with different signal-to-noise ratios (SNR) and spatiotemporal resolutions. Alternatively, it may employ a scalable video stream. In order to obtain sensible tradeoffs between bitrates, spatiotemporal resolutions, and compression artifacts, it is necessary to have a metric that predicts the perceived quality of each alternative.

Assuming a fixed viewing distance most full-reference quality metrics compare the original image to a distorted image at the same level of resolution. That may be sufficient for a traditional digital display setup. With the diversity in displays and image capturing devices, the displayed image may differ from the original in compression and spatiotemporal resolution. For example, at low bitrate coding applications the compressed image may be too distorted, and hence the observer may prefer to reduce the resolution or increase the viewing distance in order to reduce the visibility of the compression artifacts. The selection of the most desirable resolution/viewing distance will then depend on a tradeoff between visibility of compression artifacts and picture size. Such tradeoffs are not reflected in existing quality metrics, which ignore the signal visibility and only measure the visibility of compression distortions, which decrease with image size.

To gain a better understanding of the different tradeoffs, and with an ultimate goal of developing quality metrics for scalable video coding applications, we conducted a number of subjective experiments. The subjective experiments described in this paper explore tradeoffs between compression artifacts and spatial resolution. We used a number of images coders, including JPEG2000, SPIHT [1], and JPEG. The first two are SNR-scalable, while the last is not. We generated a series of compressed images at different bitrates, and for each of the compressed images, obtained seven different resolutions by optimal sinc-function downsampling. We fixed the viewing distance, and asked the viewers to select the resolution that they prefer at each coding rate. Analysis of the subjective results using perceptual image quality metrics indicate that the viewers tend to maintain the subjective quality as image size decreases. We are currently trying to develop performance metrics that incorporate both signal and compression noise visibility in order to predict the user preferences.

2. OBJECTIVE QUALITY METRICS

Media signal processing (speech, image and video) inevitably involves signal distortion, and this creates a need for quality evaluation metrics. These can be subjective or objective, but as the ultimate user is usually a human observer, they should reflect perceptual preferences. The most commonly used objective metric is mean squared error (MSE), but a number of perceptual metrics have been proposed. For a review, see [2].

In most of the existing metrics, the sampling rate of the reference and the distorted image is the same, and the metrics provide a measure of the degree of image distortion. However, the human visual system (HVS) perceives image quality along several dimensions, including image distortion, image size, sharpness, and colorfulness. Thus, a new class of quality metrics is needed in order to incorporate these additional dimensions. In particular, we are interested in exploring tradeoffs between visibility of compression artifacts and image size. For a given compression algorithm and bit rate, the compression distortion can be reduced by filtering and downsampling the image. Thus, the smaller the image, the less

visible the compression artifacts will be. On the other hand, as the image size decreases, so does the amount of signal information. In this paper, we propose and conduct subjective experiments in order to explore this tradeoff.

Most of the existing perceptually based image quality metrics incorporate models of human perception that account for the spatial (and temporal for video) contrast sensitivity function, contrast masking, and luminance masking. As mentioned above, they assume that the reference and distorted image are given, have the same resolution, and are viewed at a given distance. They are based on a multiscale frequency decomposition such as a discrete wavelet transform (DWT), subband decomposition or discrete cosine transform (DCT). For each coefficient of the decomposition, they derive a noise visibility threshold $\tau_{i,k}$, which indicates the amount of invisible noise that can be added to the subband. Here k denotes the subband index and i the coefficient location in the (possibly downsampled) subband image. The overall image distortion is then computed as follows:

$$D_p = \frac{1}{N} \left(\sum_{i,k \in \Re} \max \left\{ \frac{\left| \hat{b}_{i,k} - b_{i,k} \right|}{\tau_{i,k}}, 1 \right\}^{Q_s} \right)^{\frac{1}{Q_s}}, \quad (1)$$

where $b_{i,k}$ is the subband coefficient of the reference image, $\hat{b}_{i,k}$ is corresponding coefficient of the distorted image, and $\tau_{i,k}$ is the visibility threshold. The value of Q_s is determined from psychophysical experiments. Here we will use $Q_s = 2$. Note that when the difference of two coefficients is below the visibility threshold, the noise is invisible and the actual value of the difference does not matter. In order to be consistent with traditional error metrics, we express the perceptual metric in terms of "visual decibels (dB)." We define the "masked peak signal-to-noise ratio (MPSNR)" as

$$\text{MPSNR} = 10 \log_{10} \frac{255^2}{D_p^2}. \quad (2)$$

Note that the maximum value MPSNR is 48.13 dB. Here, we will use the DCT-based metric by Watson [3]. A detailed description of the metrics can be found in [2].

3. SUBJECTIVE TESTS

To investigate the subjective sensitivity to coding noise in the context of spatial resolution, we conducted subjective image quality evaluation experiments. Early psychophysical experiments for analyzing the effect of spatial resolution in image quality assessment were conducted by Westerink and Roufs in [4], and these have formed the basis of modern image quality analysis. However, the specific tradeoffs we examine in this paper were not addressed in that early paper.

As we discussed above, our goal is to exploit tradeoffs between spatial resolution and image compression noise in order to obtain the optimal display conditions for an image that has been compressed by a given algorithm at a given bitrate. The experiments were designed along two aspects of quality assessment: **absolute** perceived quality and **relative** perceived quality. In the absolute assessment, subjects were shown images independently and were asked to assign a score among 1 to 5 (corresponding to "very poor," "poor," "tolerable," "good," and "very good," respectively) to each image, on the basis of quality, *independent* of image size. This experiment was conducted as a calibration of the predictions
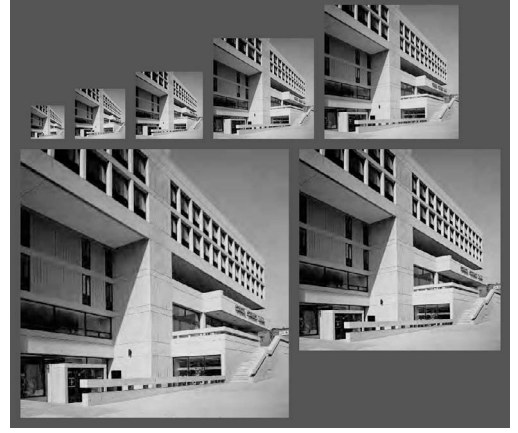
Fig. 1. Test images presented to the viewer for JPEG2000 encoded at 0.3 bits/pixel. The viewer is asked to select the most preferable image for the given viewing distance.


| (a) Bank | (b) Lena | (c) Bike |

Fig. 2. Images for subjective test.

of the objective quality metric that we used. In the relative assessment, which is the main goal of this paper, each subject was shown the same image in seven different spatial resolutions and asked to select one that they most preferred in terms of overall image quality, that is, including both distortion artifacts and image size (i.e., signal visibility). We used three test images: Lena, Bank, and cropped Bike, shown in Figure 2.

These images were compressed at different bitrates using three coding algorithms (JPEG2000, SPIHT, and JPEG). For each coder and bitrate, the reconstructed images were then downsampled to obtain seven different resolutions: 512, 384, 256, 192, 128, 96, and 64, using sinc-function upsampling and downsampling in integer ratios. Since these coders have different coding efficiencies at low bitrates, the bitrates were carefully selected at 1.0, 0.75, 0.5, 0.3, 0.2 bits per pixel (bpp) for JPEG2000 and SPIHT and 1.0, 0.70, 0.5, 0.4, 0.35 bpp for JPEG. Seven subjects took part in the subjective experiments. Viewing was binocular, and all subjects had normal or corrected normal vision.

In addition to the subject ratings/orderings, numeric expressions of image quality were computed between the originals and the decoded images at the *same* resolution by different perceptual metrics. Watson's DCT-based metric is the one we chose to present here. A comparison of the relative performance of the different metrics is beyond the scope of this paper. The viewing distance is fixed to six image heights of the 512×512 image, so that the number pixels per viewing angle is the same at all resolutions (53.73 pixels per degree). As shown in Figure 1, for each compression algorithm and rate, all seven resolutions were presented to a subject at the same time, and the subject was asked to select one among seven images. The images were ordered according to size in order to facilitate pair comparisons between adjacent scales. In the actual test, the images were gray-scale and were shown in blue background. The subjects were allowed enough time to make their de-
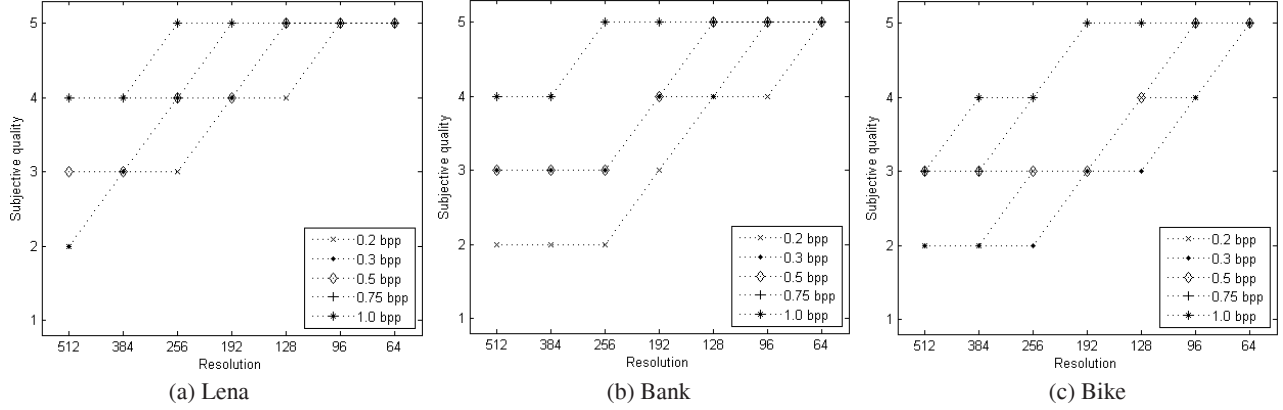
Fig. 3. Tradeoffs between spatial resolution and subjective quality. Images are coded by SPIHT.

cisions. Before the beginning of the test, they were presented with the original images and typical examples at selected compression rates, in order to become familiar with the test environment. The ordering of image coders and bitrates were randomized to avoid any biases. Again, the subjects were asked to select the image that they most preferred (as a representation of the original full resolution image) on the basis of both spatial resolution and subjective quality. The subjects were also allowed to periodically view the original test images.

4. DISCUSSION OF THE RESULTS

We begin discussion of the first test result. For the absolute perceived quality of the set of images, the median of all votes for each test condition was obtained. The results of images coded by SPIHT are given in Figure 3. It is shown that no matter what the image quality at the highest resolution, the subjective quality goes up to the highest score at the smallest resolution. This observation can be explained by the contrast sensitivity function, which is function of spatial frequency in cycles per degree of viewing angle. As spatial resolution decreases, the spatial frequency increases, hence the contrast sensitivity of the HVS decreases. When spatial frequency is infinite (infinite distance or zero resolution) the noise becomes invisible, but so does the image. Thus, the conventional description of noise visibility cannot fully incorporate the perceived image quality, making it necessary to include signal visibility in a metric that measures image quality across different viewing distances or image resolutions. Due to the nonlinearity of resolutions, we cannot obtain a formula for this relationship as a function of resolution, quantization level, and objective quality. But, there exists evidence of a strong relationship between these parameters. We note that, strictly speaking, increasing of viewing distance is not equivalent to decreasing the spatial resolution since there is an effect of the point spread function of the display device. However, we assume that this effect can be neglected in a well-calibrated display device. Selected results of computed MPSNR over bitrates and resolutions are shown in Table 1. Note that, in agreement with the results of the absolute perceived quality test, the objective quality decreases as the bitrate decreases or spatial resolution increases; conversely, it increases as the bitrate increases or spatial resolution decreases. In addition, as the resolution decreases beyond a certain level, the distortion visibility (both objective and subjective) saturates at the perceptually transparent level. Similarly, at a given (low) resolution, as the bitrate in-

creases, the distortion visibility saturates at the perceptually transparent level. It should be obvious that, as the HVS is able to accept more distortion for low resolution or longer viewing distance, the coding bitrate should be a function of spatial resolution and viewing distance.

The *shaded* cells in the table indicate the resolution at which the evaluation score 5 is first assigned to the decoded image as the resolution decreases at a given bitrate. Recall that, in the absolute perceived quality test, 5 corresponds to "very good," which is presumably noise-transparent. The results of the relative perceived quality test are also marked on the table: The cells in which numbers are written in *bold* represent the preferred resolution for a given bitrate. Observe that the resolution of both the shaded and bold blocks decreases as the bitrate decreases, an indication that there is a need to maintain perceptual quality as the bitrate decreases. For example, in Table 1(a), the quality hovers around 47.3 dB over the various bitrates. However, there is also a notable difference between the results of the two tests. In most cases, there is a (vertical) gap between the transparent resolution and the most preferred resolution at a given bitrate. This gap indicates that the observers are willing to accept some distortion for a bigger image, and can be thought of as the perceptual tolerance over the noise-transparent condition. Our results also indicate that this gap depends on bitrate, image compression scheme, and image content.

5. TOWARD IMAGE QUALITY METRIC FRAMEWORK

This subjective test result indicates the need for a new image quality metric that accounts for viewing and display parameters as well as input signal. We believe that the first step towards deriving such a metric is is a model for image distortion as a function of resolution or viewing distance. The pseudo curves in Figure 4 show that signal distortion D_s increases as the viewing distance increases (or equivalently, image resolution decreases) while the perceptual distortion D_q decreases until it reaches the perceptual threshold (Just-Noticeable-Distortion or JND) at d_s.

The perceptual distortion D_q is in general a function of the difference between the reference image $x \in \Re^n$ and the decoded image $\hat{x} \in \Re^m$

$$D_q = H_q\{x - \hat{x}\}, \qquad (3)$$

where $H_q(\cdot)$ is a transfer function that incorporates models of the HVS and display system. Here we assume that spatial resolution and target bitrate are fixed. Several such models can be found in the literature [2] that can be used to obtain estimates of the distance

Resolution/bpp	1.00	0.75	0.50	0.30	0.20
512 × 512	47.24	46.74	45.66	44.17	43.21
384 × 384	**47.63**	**47.32**	46.44	45.17	44.23
256 × 256	47.99	47.91	**47.46**	46.65	45.88
192 × 192	48.06	48.03	47.75	**47.19**	46.49
128 × 128	48.11	48.11	48.00	47.79	**47.27**
96 × 96	48.12	48.12	48.08	47.98	47.69
64 × 64	48.13	48.13	48.12	48.12	48.05

(a)

Resolution/bpp	1.00	0.75	0.50	0.30	0.20
512 × 512	**47.10**	46.72	45.55	44.41	43.24
384 × 384	47.47	**47.23**	46.29	45.37	44.23
256 × 256	47.81	47.76	**47.27**	46.74	45.87
192 × 192	47.88	47.86	47.57	**47.21**	**46.44**
128 × 128	47.95	47.94	47.79	47.66	47.20
96 × 96	47.98	47.97	47.92	47.86	47.51
64 × 64	48.00	47.99	47.96	47.96	47.85

(b)

Resolution/bpp	1.00	0.70	0.50	0.40	0.35
512 × 512	47.69	47.17	45.95	45.03	44.40
384 × 384	**47.78**	47.15	46.02	44.98	44.38
256 × 256	48.04	**47.84**	47.26	46.74	46.06
192 × 192	48.09	48.01	**47.59**	**47.19**	46.68
128 × 128	48.12	48.08	47.93	47.73	**47.39**
96 × 96	48.12	48.10	48.02	47.94	47.71
64 × 64	48.13	48.13	48.09	48.09	48.00

(c)

Resolution/bpp	1.00	0.75	0.50	0.30	0.20
512 × 512	**47.73**	**47.58**	47.18	46.20	45.54
384 × 384	47.86	47.76	**47.46**	**47.67**	46.16
256 × 256	47.97	47.94	47.80	47.38	**47.08**
192 × 192	47.99	47.98	47.89	47.61	47.40
128 × 128	48.00	48.00	47.96	47.81	47.70
96 × 96	48.01	48.01	47.99	47.92	47.89
64 × 64	48.01	48.01	48.00	47.95	47.97

(d)

Table 1. MPSNR computed by DCT-based metric above threshold from images over different coding bit rates and spatial resolutions. Bike by JPEG2000 (a), Bike by SPIHT (b), Bank by JPEG (c), Lena by SPIHT (d)   (unit:decibels)
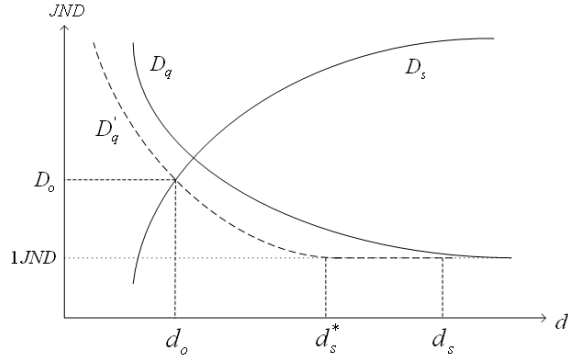
Fig. 4. The effect of scale change on processing/compression distortion and signal distortion. The x-axis is viewing distance, and the y-axis is JND. D_s is amount of signal loss with spatial resolution, D_q is perceptual distortion according to an objective metric, and $D_{q'}$ is an ideal measure of perceptual distortion.

d_s that corresponds to 1 JND.

The signal distortion D_s is the amount of signal loss with viewing distance or spatial resolution, and should primarily be a function of the reference image $x \in \Re^n$

$$D_s = H_s\{x\}, \qquad (4)$$

where $H_s(\cdot)$ should also depend on models of the HVS and the display device.

Depending on the accuracy of the underlying models, the objective perceptual distortion function D_q may not agree with the subjective experiments (absolute test). The ideal perceptual distortion D'_q is shown in dashed lines in Figure 4, and in this particular case, is below the curve obtained by the objective metric.

The total distortion is

$$D_t = \mathcal{W}\left(D_s\left(x, \vec{p}\right), D_{q'}\left(x - \hat{x}, \vec{p}\right)\right) \qquad (5)$$

where $\mathcal{W}(\cdot)$ is an arbitrary function and $\vec{p}$ is a parameter vector. Given the two distortion curves, for each spatial resolution and bitrate, there exists an optimal distance d_o (see Figure 4) that provides the optimal distortion D_o

$$d_o = \underset{d}{\operatorname{argmin}} \ \mathcal{W}\left(D_s, D_{q'}\right). \qquad (6)$$

6. CONCLUSION

The perceptual visibility of compression distortions over different resolutions and bitrates is investigated for a framework of image quality metrics. Since viewer selection of the most desirable resolution/viewing distance depends not only on the visibility of distortions but also on the visibility of the signal (i.e., image size), an analysis of the tradeoff between the two considerations plays an important role in scalable image compression. We conducted an experiment involving subjective assessment of images, which explored such tradeoffs using three image coders, JPEG2000, SPIHT and JPEG at seven resolutions and five bitrates. The experiment was designed along two aspects of quality assessment, absolute perceived quality and relative perceived quality. Our results indicate that there is a need to decrease image size or increase viewing distance as the bitrate decreases in order to maintain the subjective quality. At the same time, we found that as the bitrate decreases, viewers are willing to tradeoff some visible distortion for larger image size. Our current efforts are focused in development of image quality metrics that incorporating signal visibility as well as noise visibility.

7. REFERENCES

[1] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hiearchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, June 1996.

[2] T. N. Pappas, R. J. Safranek, and J. Chen, "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Processing*, Alan C. Bovik, Ed., pp. 939–959. Academic Press, second edition, 2005.

[3] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in *Proc. SPIE, Human Vision, Visual Proc., and Digital Display IV*, Jan P. Allebach and Bernice E. Rogowitz, Eds., San Jose, CA, Feb. 1993, vol. 1913, pp. 202–216.

[4] J. H. D. M. Westerink and J. A. J. Roufs, "Subjective image quality as a function of viewing distance, resolution, and picture size," *SMPTE Journal*, vol. 98, pp. 113–119, Feb. 1989.