Summary and Review of
Suprathreshold Visual Psychophysics and Applications to Image Compression
Professor Sheila S. Hemami
April 30, 2003

Lav Varshney
lrv2@cornell.edu
(607)-269-0449
ECE 311 – Honors Seminar Summary Paper
May 2, 2003
Choice of Seminar for Summary Paper

Professor Hemami’s lecture on *Suprathreshold Visual Psychophysics and Applications to Image Compression* was the first lecture in the Honors Seminar series to deal directly with communications and signal processing, my main areas of interest. I found the talk to be very interesting, and hence decided to write a summary paper about it. Having taken Fundamental Information Theory this semester, I found the material that she covered to relate very nicely with my study of Rate Distortion Theory and the fundamental limits on lossy compression that it establishes. Professor Hemami’s lecture gave me a good idea of the types of psychophysically meaningful distortion measures that can be used for image compression. After all, “the ultimate purpose of a distortion measure is to reflect the distortion between $x$ and $\hat{x}$ as perceived by the user” [1]. It was good to see that, contrary to Cover and Thomas [2], there are in fact real alternatives to the mean squared error distortion measure for image coding.

Additionally, I had participated in an experiment performed by Professor Hemami’s research group about the visual perception of different video coding schemes, and the lecture helped explain some of the theory behind the actual implementation that I had seen. In particular, the lecture helped me understand the relationship between the video sequences that I had seen and the philosophy of developing robust compression algorithms, which are effective for a wide range of bit rates and video content.

Summary of Seminar

When images must be compressed into low bit rates for transmission or storage, loss of information and degradation of image quality above the visibility threshold is inevitable. The goal of an image compression technique is to represent an image with minimal distortion for a given bit rate. Professor Hemami and her research group have developed a visual psychophysics foundation for the development of compression techniques that maximize the visual quality of compressed images. Building on this foundation, they have developed robust compression algorithms, independent of image content or required compression ratio, that produce visually optimal compressed images for a given information rate.

Psychophysics is the branch of psychology that deals with the relationship between physical stimuli and human perception. Visual psychophysics, in particular, relates light stimulation and the visual response that it induces in humans. In order to understand the visual
response in humans, it is first necessary to understand the stimuli that induce the response. For computer images, the relationship between the pixel value that is stored in the computer and the light stimulus, called luminance, that is produced on a cathode ray tube monitor is given by

$$L = (b + kp)^\gamma,$$

where $L$ is the luminance, $b$ is the black level offset of the monitor, $k$ is the voltage to pixel scaling factor, $p$ is the pixel value, and $\gamma$ is a parameter that relates to the monitor’s luminance to voltage response curve.

Humans do not perceive absolute luminance, but perceive changes in luminance, known as contrast. Contrast is given by

$$C = \frac{\Delta L}{\bar{L}},$$

where $C$ is the contrast, $\Delta L$ is the change in luminance, and $\bar{L}$ is the mean background luminance. Sensitivity is defined as the inverse of contrast, and is used as a measure of the sensitivity of the human visual system. In engineering applications, the physical stimulus is used to measure the sensitivity of the human visual system, rather than the actual perception, due to the ease of measurement. The Michelson contrast measure is commonly used for sine-wave and Gabor-patch targets presented against well-defined backgrounds, and is defined in a peak-to-peak manner, namely

$$C_{\text{mich}} = \frac{L_{\text{max}} - L_{\text{min}}}{L_{\text{max}} + L_{\text{min}}}.\quad (3)$$

The Michelson contrast measure, however, is unsuitable for measuring contrast in complex images. Consequently, the root mean square contrast is used instead and is defined as

$$C_{\text{rms}} = \frac{1}{\bar{L}} \sqrt{\frac{1}{N} \sum_{i=0}^{N} (L_i - \bar{L})^2},$$

where $L_i$ is the luminance of the $i^{th}$ pixel and $N$ is the total number of pixels.

The visibility threshold is defined as the point at which a stimulus just becomes visible. The contrast at the visibility threshold is the contrast threshold, and has a corresponding sensitivity. Contrast thresholds have been measured experimentally, and a human contrast sensitivity function (CSF) has been developed. The CSF measures the sensitivity threshold as a function of spatial frequency. Unfortunately, the CSF has major limitations, restricting its usefulness for the development of optimal compression schemes. Specifically, the CSF is not
orientation specific, whereas the human visual system has channels tuned to different orientations as well as frequencies. In addition, the CSF is based on experiments with simple gratings with single frequencies, rather than natural images that contain a large number of frequencies. Finally, the CSF represents subthreshold perception, rather than suprathreshold perception, as is needed for low bit rate image compression. Despite these limitations, the CSF has been used to develop compression schemes.

The human visual system is often modeled as a bank of dimensionally-selective channels, channels tuned to spatial frequency and orientation, which decompose an image into its dimensional components. To obtain visually superior representations, many compression techniques attempt to follow this multi-channel model, decomposing an image into a series of subbands prior to compression. The most commonly used such decomposition is the wavelet decomposition. Although psychologists convincingly state that wavelet decomposition is significantly different than the human visual system multi-channel decomposition, engineers feel that it is close enough to produce a psycho-visually meaningful result. The process of one-dimensional wavelet decomposition is shown in Figure 1. The $x_l[n]$ are called the scaling coefficients, and the $x_h[n]$ are called the wavelet coefficients. Effectively, the wavelet decomposition halves the frequency range of individual coefficients. The two-dimensional wavelet decomposition acts similarly in both the horizontal and vertical directions, producing 4 sets of outputs. This process is applied recursively to images, resulting in many frequency-selected bands. In image compression schemes, after an image has been decomposed into frequency bands, these frequency bands are quantized, introducing distortion.

Psychophysical foundations are used to decide how the frequency bands should be quantized in a compression scheme, and how much distortion will be perceived after this quantization. Perceived distortion is measured experimentally, using human subjects. Two kinds of tests are performed. The first type is unmasked detection, which involves testing with only the distortion presented. The subject must determine the point at which he/she can see the
distortion. The second type is masked detection, in which the original image and a distorted image are presented next to each other, and the task is to detect a difference between the two. As expected, unmasked distortion is detected at lower distortion levels than masked distortion. The response of the human visual system to a combination of multiple distortions in both orientation and frequency is measured using the Minkowski metric. The exponent $\beta$ in the Minkowski metric parameterizes the amount of summation. When $\beta$ is less than 1, there is superadditivity; when $\beta$ is greater than 1, there is subadditivity; and when $\beta$ is 1, the additivity is linear [3]. Experimental results for the additivity of distortions in suprathreshold quantization show that $\beta \approx 1$, implying that distortions have linear additivity in terms of human visual system sensitivity. This result corresponds to psychological studies that have shown that distortion for object recognition also added linearly.

Since humans are less sensitive to high frequency than low frequency, high frequencies can be discarded without much loss in perceived image quality. In fact the notion of global precedence in human visual processing, in which image structure and content is processed by combining information from coarse to fine scales, has been shown experimentally. Consequently, in image compression, information should always be discarded in a fine-to-coarse-scale order, thereby minimizing the perceived distortion.

Using these principles, coding schemes were developed that minimize the visual distortion of a compressed image. It has been found that the compressed images using this contrast-based quantization procedure produces visually superior images at low rates, as compared to coding schemes that minimize the mean squared error of the compressed image. In particular, there are significantly fewer high frequency clipping artifacts and blurring is reduced in textured regions. The contrast-based quantization scheme can be used at all bit rates, and for any image content, while still showing the improved visual quality.

References

