Image compression, fidelity metrics and models of human observer

Stanley A. Klein and D. Amnon Silverstein, UC Berkeley, Berkeley, CA 94720

1. Introduction

What a tragedy. The vision research community is largely missing the fun of the fast-paced image compression research venture. It is a community that has a lot to contribute and they are missed. The purpose of our presentation is to show how vision researchers can contribute to image compression.

Why are so few of our vision research colleagues involved with fidelity metrics and picture coding (exceptions include Mannos & Sakrison, 1974; Limb, 1979; Daly, 1992; Watson, 1993)? There are a number of reasons. 1) The most important reason is vision researchers' lack of awareness of how they could contribute to image compression. They do not realize that the vision models that they have been developing for the past 100 years are very close to the fidelity metrics so needed by the compression community. 2) Vision researchers are not aware of the great impact that an improved fidelity metric will have. 3) The explosion of activity in picture coding has taken place too fast for outsiders to develop the skills to participate. 4) In the past the limited spatial and gray scale resolution associated with television and 512 x 512 images meant that images had limited quality so that an image quality (rather than fidelity) metric was desired. Quality (how pleasing is the picture?) is much more difficult to assess than fidelity (can you distinguish between pictures A and B?). Vision researchers tend to avoid research on quality. As interest in super high definition images increases, so will the interest of vision researchers since quality metrics will be replaced by fidelity metrics. In our presentation today, we attempt to clarify all of the above problems, paving the way for more vision researchers to become involved with the exciting and important picture coding enterprise.

2. Three roles for a human vision model (HVM)

Fig 1 shows how a good image fidelity (or quality) metric can assist image compression. We use MPEG as the compression standard for our discussion. However, no matter which compression scheme (DCT-based, wavelet based, vector quantization) is used, a good human vision model (HVM) can make it better. There are three points in Fig. 1 where the fidelity metric can be useful:

1) At the very first step the HVM can choose the optimal quantization matrix, Q. MPEG allows information to be passed with each macroblock (16 x 16 pixels) that specifies the scale factor of Q. A highly masked macroblock can tolerate a larger amount of quantization than a macroblock with little masking, such as text. The HVM can save bits from masked areas and place them where they are more needed. Even with JPEG which doesn't allow local specification of Q, the HVM can optimize Q for the entire specific image (Watson, 1993).

There are a number of different quantization matrices in MPEG. Each can be optimized depending on the nature of the image and the application. The optimal Q for the luminance channel differs from the chromatic channel. Furthermore, MPEG allows different matrices for intraframes and interframes. The many different end purposes involving different output devices and viewing distances provide further need for a multiplicity of quantization matrices. The HVM together with entropy coding must work together to come up with improved matrices. In Section 3 we speculate on how vision researchers can develop a framework for measuring Q for different conditions.

2) The second stage at which an HVM can be used is in what can be called a "tweaking box". The first stage of quantization (rounding off to the nearest integer after dividing by Q) is designed to minimize the mean square error (MSE) between the original and the codec. If an improved error metric were used (the HVM) then the coefficients can be "tweaked" such that visible distortion (e.g. blocking artifacts) could be reduced even while the MSE increased slightly. Girod, et al. (1993) have discussed spatial shaping methods for concentrating the MSE in the center of a block rather than at the edges. Additionally, there might be a number of coefficients that have a value of unity that could be reduced to zero with minimal loss of fidelity, but with a substantial bit savings due to entropy coding (Jayant, et al. 1993). The availability of an improved HVM will certainly improve previous schemes for tweaking the coefficients. The way in which the HVM would be used is straightforward. The multidimensional HVM would guide the tweaking of the coefficients. For example, if an artifact were detected on the top edge at a particular horizontal spatial frequency.

![Figure 1](image-url)

Three points where a human vision model can be used to improve an image compression system.
then a small number of coefficients would be singled out for tweaking. Also coefficients whose values are unity would be tested for reduction. For each choice of coefficients a rate-distortion value would be calculated, where the rate is based on entropy coding and distortion is based on the HVM. From the rate-distortion curve one chooses the preferred set of coefficients. Since similar procedures have already been considered by others (Girod, et al, 1993; Jayant, et al, 1993), nothing will be said about these here except to emphasize that using an improved HVM would improve all previous algorithms.

3) Our favorite role for the HVM is in a final stage of image enhancement shown near the bottom of Fig. 1. One or more enhancement algorithms could be applied to the image and the HVM would be used block by block to measure which of the enhanced or non-enhanced blocks is perceptually closest to the original. One can think of many schemes by which a codec could be improved. Here we list two and Section 4 presents details on a third method: a) Suppose it is known that within a region only two gray levels are present (as in simple text). Then one could develop an enhancement algorithm that forces the two levels plus an interpolating border separating the regions of the two levels. b) Suppose it is known that there is no discontinuity at the border between two blocks. Then the DCT coefficients could be delicately tweaked (using floating point values rather than the integer compressed values) to minimize the border discontinuity. The information about which image to use for a given block must be hidden within the bit stream in a manner that doesn’t increase the bit-rate. Section 4 presents the details of this approach using a crude low-pass filtering method (no HVM) for codec enhancement.

3. Determining the optimal quantization matrix, Q.

Very little has been written on a theoretical basis for calculating the elements of the quantization matrix (see, however, Peterson, Ahumada & Watson, 1993). As pointed out by Nill (1985) and as discussed below it is useful to know the Fourier transform of the DCT basis functions (closely related to the basis functions’ visibility), Klein, Silverstein & Carney (1992) have made some progress in this direction. They start with the mth one-dimensional DCT coefficient:

$$DCT(m, i) = \cos((i + 0.5)m/N)$$

where N is the number of pixels per block (8 for JPEG and MPEG) and i (an integer from 0 to 7) is the spatial position within the block. We will assume that DCT(m, i) vanishes for i ≠ 7 and i ≠ 0 so that we are dealing with a single block surrounded by a uniform field. For simplicity, we ignore the normalization factors (2/N)² for m ≠ 0 and (1/N)² for m = 0.

The Fourier transform of a single basis function, with pixels sampled at the very center of each pixel, was derived by Klein et al. (1992):

$$F_{DCT}(m, f) = \sum_{i=0}^{3} \cos((i + 0.5)m/N) \cos((i + 0.5)m/N) \text{ for even } m$$

and

$$F_{DCT}(m, f) = \sum_{i=0}^{3} \sin((i + 0.5)m/N) \sin((i + 0.5)m/N) \text{ for odd } m$$

The spatial frequency, f, is a continuous variable since the basis functions are restricted to a single block. If the basis functions were not restricted to a single block then the Fourier transform would be trivial in that it would only be nonzero only when f = m (see the circles in Fig. 2). Eqs. 2 were shown (Klein et al., 1992) to equal:

$$F_{DCT}(m, f) = \frac{2}{N} \left[ \sin((f-m)\pi/2) \sin((f+m)\pi/2) \right] \text{ for all } m$$

If instead of pixels being sampled at their centers, they are sampled uniformly (rectangular rather than delta function pixels) then Eq. 1 must be convolved with a rectangular pulse going from i = -3 to +3. The Fourier transform (Eqs. 2 and 3) would then be multiplied by the sinc function given by:

$$\text{sinc}(f) = \sin((f\pi/2N)/(f\pi/2N))$$

The combination of Eqs. 3 and 4 are shown in Fig. 2.

Now for two tricky questions. 1) Given the Fourier transform of a basis function what is its visibility? 2) Given the visibility of each basis function what should be the value of the corresponding quantization matrix element? Klein, et al. (1992) address these questions but do not provide final answers. For question 1, first consider the one-dimensional basis functions as in Eq. 1. To first order their visibility is given by the product of $F_{DCT}(m, f)$ with the contrast sensitivity function CSF(f), integrated over f. An improved calculation would replace the CSF with medium bandwidth mechanisms and
then do probability summation across the mechanisms. For the two-dimensional basis functions an orientation tuning factor should be included (Ahumada & Peterson, 1992; Klein, et al., 1992) since the two-dimensional functions are equivalent to the sum of two off-axis one-dimensional functions. If the pair have similar orientations then they will sum for the detection task.

Now for question 2. Based on the above assumptions, Klein, et al. (1992) calculated quantization matrices appropriate for single DCT basis functions for pixel sizes of 2.0, 1.5 and 1.0 min/pixel. There are several problems with their calculation. Entropy coding must also be taken into account. The rate-distortion curve would imply that high spatial frequencies should be truncated more severely than is indicated by the CSF calculations outlined above, because there is substantially greater advantage in reducing a coefficient from 1 to 0 (typical of high spatial frequencies) than from 8 to 4 (typical of low spatial frequencies). However, the presence of the "Tweaker box" in Fig. 1 would allow the quantization matrix to be based solely on basis function visibility with the effects of entropy coding taken into account at the Tweaker box stage. A further complication relevant to question 2 is the issue of summation across DCT basis functions. Klein, et al. (1992) discussed situations whereby many basis functions sum. This summation can make the quantization errors quite visible even though the contribution from each individual basis function was invisible. The "killer dots" show by Klein, et al. (1992) are an example of a quite visible high-pass dot that is clearly visible in the original image but becomes a uniform field after quantization. With an improved HVM and the MPEG scale factor, the killer dot would have been seen in the original and bits could be added to ensure that the dot is visible in the transmitted codec. The availability of a robust HVM would allow researchers to determine optimal quantization matrices for different images and viewing conditions.

4. Image enhancement using hidden bits.

Several researchers have been developing methods for enhancing codec images. A new method for improving the codec enhancement (and hence increasing the degree of compression) can incorporate an HVM. Before an image is stored or transmitted, we have access to the original image and the distorted codec image. We can enhance the codec with one or more different enhancement schemes. The efficacy of each enhancement scheme can be measured by a fidelity comparison of the enhanced images to the original. The process is shown schematically in the last four stages of Fig. 1, and an example of the process is shown in Fig. 3 (see Silverstein & Klein, 1994, for details).

![Figure 3.](image)

An example of map guided image enhancement.
The HVM can construct a fidelity map for each enhanced image. The map shows how many JNDs of perceptual distance the enhanced image is from the original, region by region. That is, an enhancement scheme may improve the fidelity of the codec image in one spatial location, and degrade the fidelity in other locations. The map can be included with the compressed image file. The receiver can then apply the appropriate enhancement schemes to only the image regions where that scheme is effective. In this example we used an simple algorithm to predict where the enhancement scheme would fail instead of a HVM. The results would be superior if a very successful HVM had constructed the map.

Fig. 3 shows how such a marking system can be used to improve JPEG while maintaining compatibility with the standard. Image A is the original in Fig. 1. Image B is a JPEG compatible file that has been compressed severely and invisibly marked with a map for post-processing (corresponding to the output of the tweaker in Fig. 1). Images C and D show two enhancement schemes (in this example just two low pass filters). Image E is the output of the HVM which is used by the composer to make the final image F, shown at the bottom of Fig. 1. Image F is constructed of C wherever the map is white, D wherever the map is gray and B wherever the map is black.

The map was hidden in image B by using a parity scheme that distorts the image by only a tiny amount. A single JPEG coefficient is adjusted so that the sum of the coefficients contains the flag for post-processing as the parity of the block. Half of the blocks already have the correct parity. In the other blocks, a coefficient that is close to being half way between two values (before quantization) will be chosen and rounded in the other direction. The end result is a compressed image that can be decompressed on any standard JPEG decompressor, but that can be enhanced by a sophisticated decompressor.

In summary, Section 2 outlined three places in which a human vision model can be used to improve the quality of a compressed image while maintaining the same bit-rate. One could also use the HVM to compare different compression methods. Although in this paper we focused on DCT based methods, all methods will benefit from a high quality HVM to measure distortion. The image enhancement example of the present section showed that decent quality images could be transmitted at very low bit-rates (2 bits/pixel). In our presentation at PCS-94 we intend to also apply our methods to the super high definition sample test images that were selected for this conference.

5. Acknowledgments This work was supported by AFOSR F49620-92-1-0359.

6. References.


