QUANTIFYING CODING PERFORMANCE FOR PRE-PROCESSED IMAGES

V. Ralph Algazi*, Niranjan Avadhanam and Robert R. Estes, Jr.
CIPIC, Center for Image Processing and Integrated Computing,
University of California, Davis, CA 95616

ABSTRACT

Typical objective methods for quantifying image quality, as part of evaluating coder performance, are obtained by computing a single or several numbers as a function of the difference image between the original and coded images. Pre-processing images prior to encoding can remove noise, or unimportant detail, and thus improve the overall performance of the coder. However, the error image obtained with the pre-processed image as a reference is substantially different than the one obtained if the original image is used. In particular, adaptive noise removal, that generally improves the image quality, could be interpreted as introducing noise with respect to the original. This paper addresses the issue of combining the changes in the image due to pre-processing and the degradation due to encoding. The objective is to obtain global quality measures that quantify the value of pre-processing for image coding.

Keywords : image quality, perceived distortions, objective quality, vision models, coding performance, pre-processing.

1. Introduction

In image coding, the artifacts introduced by standard coders are uncontrolled or unpredictable in detail. If high image quality is to be maintained, this characteristic may allow a very small compression, or often only error free compression, in order to maintain acceptable image quality. One option, that we have examined in previous work, is to pre-process the image in an adaptive fashion so as to introduce imperceptible or controlled degradations to the image.[1,2] Such pre-processing, when combined with a standard coder, either lossy or error free, can result in an improvement in overall coding performance.[1] Measuring this performance improvement is the subject of this paper.

Traditionally, objective measures of image quality are numerical valued functions of the difference between the original and encoded image. Preserving the numerical integrity of the original image is thus the implicit goal of the coder. If the original image is processed prior to encoding, the changes introduced by such processing will lead, for instance, to a mean square error that is considered objectively a loss of image quality. As an alternative, the performance of the coder could be measured with respect to the pre-processed image. In such a case, no weight is given to perceptible degradations introduced by pre-processing. Thus both alternatives are unsatisfactory. To correctly address this issue, we need to assess perceptually, image changes due to processing or coding.

We have developed two methods which are applicable to the problem. Both these methods are based on the Visible Differences Predictor (VDP)[3] developed by Scott Daly. The objective quality metrics we use are variations of the Peak Signal to Noise Ratio (PSNR) and a Picture Quality Scale (PQS).

This paper is organized as follows: after an introduction to the paper in section 1, we describe the problem statement in section 2. The VDP and the PQS algorithms on which the objective quality metrics are based, are briefly outlined in sections 3 and 4. In section 5, we explain the methodology used to quantify coder performance. Section 6 describes the simulation results using the two metrics. Finally, in section 7, we discuss some of the results and present conclusions.

2. Problem Statement

The main objective is to evaluate the merits of pre-processing prior to using a standard coder. We expect that the pre-processing not only does noise removal but also simplifies the data in such a way that it is easier for the coder the encode the data. At low quality levels, the coder by itself does noise removal by quantizing coarsely the high frequency components. The benefit from pre-processing is expected to be more apparent at higher levels of quality.

The Corner Preserving Filter (CPF)[2] is used as the pre-processor with the number of iterations controlling the degree of pre-processing. The CPF is based on a mean curvature diffusion algorithm that adaptively filters the image.

It has been shown previously that this filter preserves important structure in images while removing noise effectively.[1,2] A standard JPEG coder using the Independent JPEG Group software[4] is used for the coding purposes.

We break up the analysis into two parts:

1. Comparison of the coder performance on the original image and on a perceptually transparent pre-processed image. The pre-processing can be verified to be perceptually transparent on a calibrated monitor. The reduction in noise allows the coder to encode the pre-processed image more effectively. Visual inspection of the coded image at the same bitrate with and without pre-processing shows the merit of pre-processing. This gain in visual quality should be reflected by the quality metric.

2. Comparison of the coder performance for different levels of pre-processing. As the number of pre-processing iterations are increased, perceptual transparency is no longer preserved. The adaptive filtering still preserves the integrity of visually significant areas of the image. The purpose of the pre-processing here, is to control the distribution of errors, so that they are less perceptible after coding, or to simplify the image structure so that fewer bits are needed to encode it. Eventually, though, more pre-processing introduces clearly visible artifacts which degrade the overall coder performance. Again, the quality metric should reflect these observations.

For the analysis, we use two objective quality metrics:

1. PSNR based on the VDP

2. PQS based on factor images obtained using the VDP

3. Visible Differences Predictor

The VDP algorithm proposed by Scott Daly is a multichannel human vision model which takes an image processing approach to quality prediction. The inputs to the algorithm are the original and distorted images and the viewing conditions. The output is a map showing the probabilities of detection of the errors. The original model describes threshold perception only; all suprathreshold errors are mapped to a probability of 1.

The overall model is implemented as a cascade of submodels to incorporate the known properties of the visual system. The main components of the model are:

1. retinal nonlinearity

2. contrast sensitivity function

3. orientation and frequency selective cortex bands

4. masking properties

5. psychometric function and probability summation.

Figure 1 gives an overall view of the algorithm. Complete details are given in references.[3,5] A number of psychophysical tests have been performed to test the validity of this algorithm.[6] A shift invariant nonlinearity models the light adaptive property of the retina. A display calibration model is needed to map the gray levels into luminance values on the monitor.

The CSF quantifies the visual response as a function of the spatial frequency. The spatial frequency and orientation selectivity which reflects the properties of simple cells in the cortex.The decomposition into multiple spatial frequency and orientation tuned channels is achieved by a cascade of frequency selective filters (denoted as difference of mesa (dom) filters[3]) and orientation selective fan filters based on the cortex transform.[7] This selectivity yields specific frequency and orientation tuned bands called cortex bands.

The dom filters have octave bandwidths and are symmetric on a log frequency axis. The fan filters have a tuning bandwidth of 30 degrees. The present implementation has 5 doms and 6 fans yielding a total of 31 cortex bands

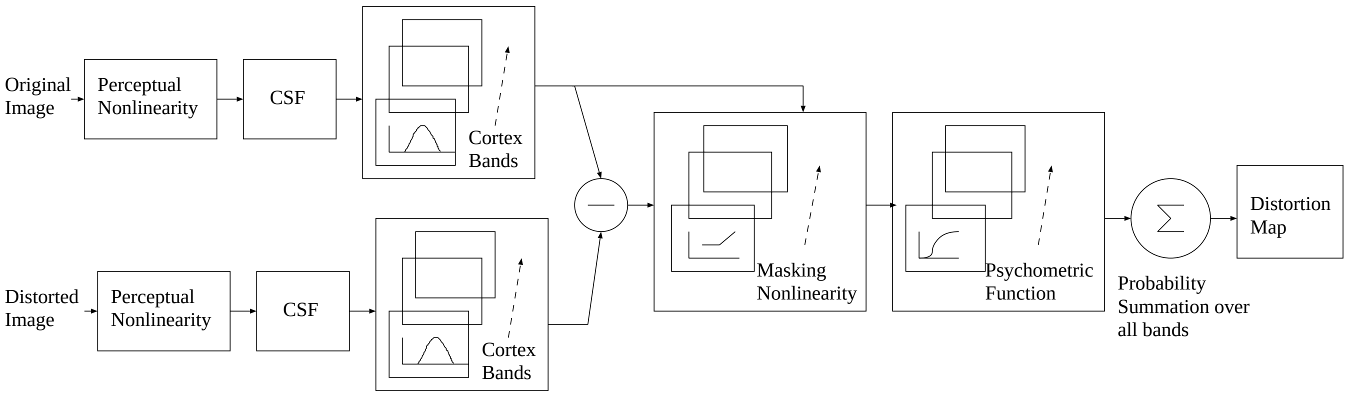

Figure 1. VDP construction

including the baseband. Making use of results from sinusoidal masking and noise experiments, Daly proposed a masking function of the form

$$T_e = (1 + (k_1(k_2 m_n)^s)^b)^{\frac{1}{b}} \tag{1}$$

where the masking effect due to the image activity is evaluated for each cortex band and results in a threshold elevation due to masking calculated as a nonlinear function of the normalized mask contrast m_n. Here s is the learning effect slope and takes values between 0.7 and 1, $k_1=6^{\frac{-7}{3}}$, $k_2=6^{\frac{10}{3}}$ and b=4. The choice of these parameter values is explained in the reference.[3]

The contrast difference of the errors for each location in a band is mapped through a psychometric function of the form

$$P(c) = 1 - e^{-(\frac{c}{\alpha})^{\beta}} \tag{2}$$

where c is the contrast of the error, α the contrast threshold, and β the slope of the psychometric function. This yields a probability of detection map for each band. Since the channels are assumed to be independent, an error above threshold in any of the cortex bands would be perceivable. Hence the probability maps for all the 31 bands are combined as below to give a single map of the probability of detection of the errors as a function of location in the image.

$$P_t(m,n) = 1 - \prod_{k=0..5, l=1..6} (1 - P_{k,l}(m,n)) \tag{3}$$

where $k = 0$ corresponds to the baseband.

For the present work, we ignore the sign of the error. We obtain a binary map where all suprathreshold errors are mapped to a value of 1.

4. Picture Quality Scale

The PQS for the evaluation of coded images is obtained by considering perceptual properties and the errors that can be observed in the coding of images.[8,9] First, the image signal is transformed into one which is proportional to the visual perception of luminance according to a power law and then the frequency weighted errors $e_w(m,n)$ are obtained by filtering with a CSF like function. Perceived image disturbances are identified and the corresponding objective quality factors which quantify each image degradation are computed as functions of $e_w(m,n)$. The perceived disturbances are D_i and the defined numerical measures of them are the corresponding distortion factors F_i. At the first stage, factor images $f_i(m,n)$ are obtained by performing local computations on $e_w(m,n)$. Next, distortion factors F_i are obtained as summations on the corresponding factor images. The PQS makes use of five factors, the first two factors accounting for random errors and the third for blocking artifacts. The last two factors, dominant at high quality, correspond to structured errors and errors near edges. It is assumed that the mean opinion of scores (MOS) is approximated by a linear combination of the distortion factors. Reference[8] gives a detailed description of

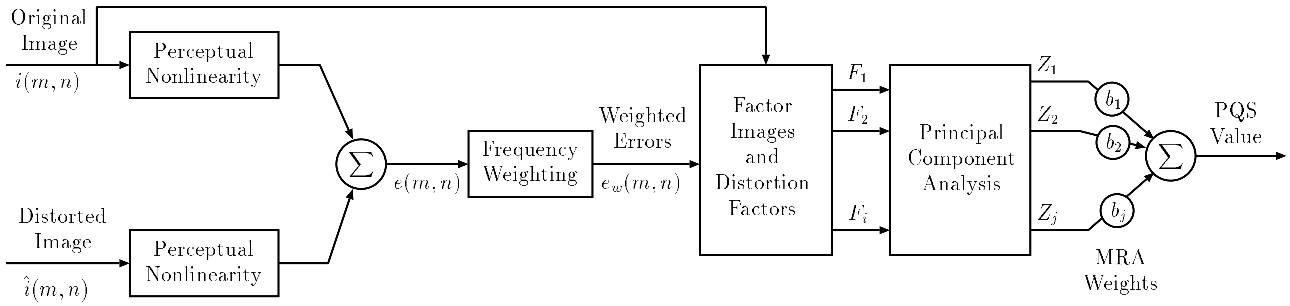

Figure 2. PQS construction

Figure 3. "Lena". (a) left CPF50 (d) right VDP mask

PQS and its performance for the full range of quality. Figure 2 summarizes the steps used in the construction of PQS.

The global value for PQS is given by the linear combination of the factors F_j as

$$PQS = b_o + \sum_{j=1}^{J} bj \cdot Fj \tag{4}$$

where b_O is a fixed parameter and b_j are the partial regression coefficients which are computed by Multiple Regression Analysis (MRA)[10] between the factors and the MOS values of observers obtained by experimental quality assessment tests. The correlation coefficient R between PQS and MOS is 0.92, which is a great improvement when compared to the correlation of R=0.47 of the conventional WMSE scale which is calculated by using F_1 only.

5. Methodology

The VDP allows the prediction of areas of an image where distortions will be perceived. This suggests that the 0-1 mask produced by the VDP be integrated into the quality evaluation methodology of encoders for pre-processed images. We apply the VDP to the difference between the original and the pre-processed images. The VDP mask will thus indicate portions of the image where the processing has introduced perceptible changes from the original. Figure 3 shows the "Lena" image after 50 iterations (lossy pre-processing) of the CPF have been performed and the VDP mask obtained for 50 iterations. The areas where the pre-processing introduces suprathreshold errors (mask values of 1) are mapped to a gray level of 255 in this figure. These areas have been overlayed on the original image for comparison.

Figure 4. Images used. (a) top left "Building" (b) top right "Lena" (c) bottom left "Wheel" (d) bottom right "Smile"

5.1. Use of VDP with PSNR (VDP+PSNR)

Let $M(m,n)$ be the 0-1, mask produced by the VDP, where 1 is assigned to areas of the image where the errors are perceptible. Let $I_o(m,n)$ be the original image, $I_p(m,n)$ be the pre-processed image, and $I_{en}(m,n)$ be the encoded image. We can now compute two error images

$$I_{d1}(m,n) = M(m,n)[I_{en}(m,n) - I_o(m,n)] \tag{5}$$

and

$$I_{d2}(m,n) = [1 - M(m,n)][I_{en}(m,n) - I_p(m,n)] \tag{6}$$

where [1- $M(m,n)$] is the complementary set to the $M(m,n)$ mask.

$I_{d1}(m,n)$ only includes errors in areas where the VDP indicates pre-processing has introduced perceptible changes. In such areas, the encoding distortions should be evaluated with respect to the original image as indicated. On the other hand, $I_{d2}(m,n)$, evaluated in areas where pre-processing does not introduce perceptible errors, uses the pre-processed image as a reference.

We can now evaluate the performance of the coders by adding the contributions due to $I_{d1}(m,n)$ and $I_{d2}(m,n)$. Thus, a modified mean-square error can be evaluated as :

$$MSE = \sum_m \sum_n (I_{d1}^2(m,n) + I_{d2}^2(m,n)) \tag{7}$$

and the corresponding computation of the PSNR follows.

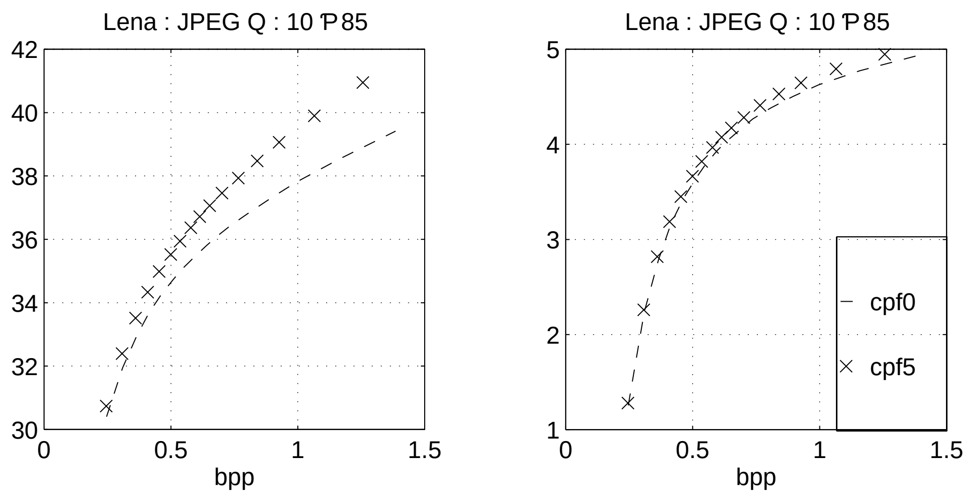

Figure 5. Coder Performance using the two metrics for "Lena". (a) left VDP+PSNR for CPF5 and PSNR for CPF0 (b) right VDP+PQS for CPF5 and PQS for CPF0

5.2. Use of VDP with PQS (VDP+PQS)

Using an approach similar to the previous sub-section, we compute the PQS value based on the two domains $M(m,n)$ and [1-$M(m,n)$]. As described in section 4, five factor images are to be obtained by performing local computations on the error image. We first compute the two error images [$I_{en}(m,n)$ - $I_o(m,n)$] and [$I_{en}(m,n)$ - $I_p(m,n)$]. Next, we obtain two factor images $f_{d1i}(m,n)$ and $f_{d2i}(m,n)$ for each distinct factor F_i by performing identical computations on the two error images. The two factor images are then combined as :

$$f_{d1_i}(m,n) = M(m,n) f_{d1i}(m,n) + [1 - M(m,n)] f_{d2i}(m,n) \tag{8}$$

These factor images are then integrated into global values for the factors which are in turn, combined to give the global PQS value.

6. Experiments and Results

In this paper, we consider the four images, "Building", "Lena", "Wheel" and "Smile" shown in figure 4. After careful verification on a calibrated monitor, it was determined that 5 iterations of CPF (CPF5) preserves the detail in the images while cleaning up the noise. Hence we consider CPF5 to be a noise free image which is perceptually equivalent to the original noisy image (CPF0). We also consider 10, 20, 30 and 50 iterations of the CPF for lossy pre-processing of the four images. That will allow a study of lossy pre-processing on overall coder performance. In all the experiments, the JPEG coder is used at quality settings of 5 to 90 in steps of 5. We can characterize roughly the range for quality settings 5 to 40 as low quality and 45 to 90 as high quality.

6.1. Perceptually Transparent Pre-Processing

We have verified that the conventional PSNR and PQS based on the original image as a reference show an increase in distortion, and thus do not show the benefits of pre-processing. However, it has been verified visually that CPF5 cleans up the image while preserving perceptual transparency. Although we cannot assess reliably the improvement directly from the graphs, figure 5 shows, for the "Lena" image, the comparison for the metrics VDP+PSNR for CPF5 and PSNR for CPF0 as the bit-rate varies. The comparison for the metrics VDP+PQS for CPF5 and PQS for CPF0 with respect to the bit rate for the same image in figure 5 indicates also the benefit of pre-processing. Figure

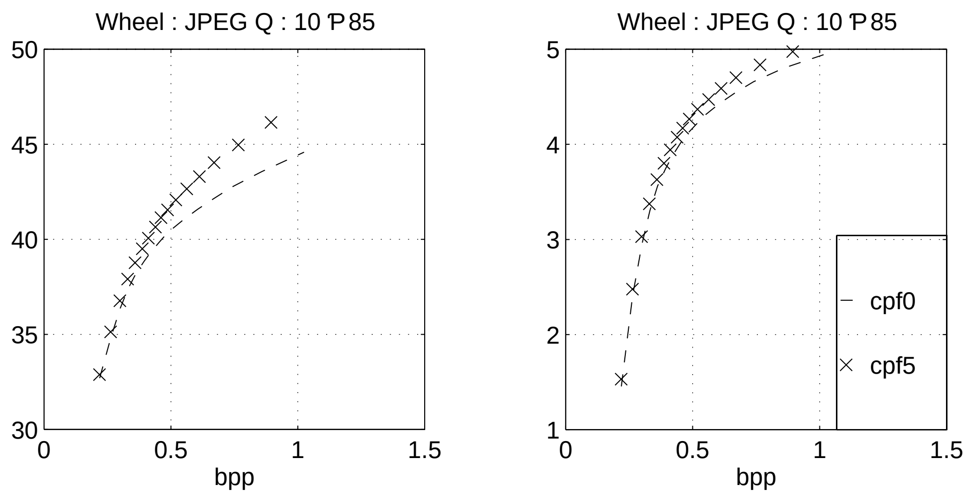

Figure 6. Coder Performance using the two metrics for "Wheel". (a) left VDP+PSNR for CPF5 and PSNR for CPF0 (b) right VDP+PQS for CPF5 and PQS for CPF0

6 shows similar plots for the "Wheel" image. It is apparent from the graphs that pre-processing is beneficial at a quality corresponding to perceptual transparency, which is achieved at bit rates of approximately 1 bpp. If a PSNR of 44 dB is used as quality indicator, then pre-processing would decrease needed bit rate from 1 bpp to .65 bpp. PQS is more pertinent and at a PQS of 5 indicates a decrease from 1 bpp to .85 bpp.

6.2. Comparison for different degrees of Pre-Processing

Here, we investigate whether pre-processing beyond perceptual transparency provides an improvement in coder performance. We expect that, as we increase the number of iterations beyond a certain value, pre-processing will degrade the image quality as it introduces visually significant artifacts without a commensurate gain in bitrate.

6.2.1. VDP+PSNR

We observe that the effect of pre-processing depends on both the number of iterations of inhomogeneous diffusion and on the target bit rate. Note in particular that in Figure 7, for the "Lena" image, that the graphs for different number of iterations intersect. As a general rule, for the four images used, CPF 20 gives the best overall performance. This can be considered to be the best compromise between the pre-processing simplifying the data prior to encoding while introducing only visually insignificant errors. Table 1 gives a summary of the results obtained using the best CPF filter. The performance improvement is over CPF5. The first four rows correspond roughly to lower quality levels and the bottom four rows to higher quality levels. Table 2 shows the highest percentage improvement in the bitrate (fourth column) for a fixed value of VDP+PSNR using the best CPF filter. The results show that the gain is dependent on the image, with significant gain seen for the "Lena", "Wheel" and "Smile" images at higher quality levels.. This metric correctly shows the benefit of pre-processing for the "Lena" and the "Wheel" images, and the slight improvement of performance for the "Building" image.

It also verifies correctly that more pre-processing beyond perceptual transparency is not beneficial for the "Smile" image as clearly visible artifacts are produced beyond 10 iterations. For the "Smile" image, the performance for larger pre-processing steps deteriorates rapidly after 0.2 bpp as seen in the figure 8. CPF10 gives the best overall performance, verifying that for this image, more pre-processing is not desirable. It also indicates that for the "Building" image, CPF30 is detrimental to the performance. This metric, though, seems to overestimate the value of pre-processing.

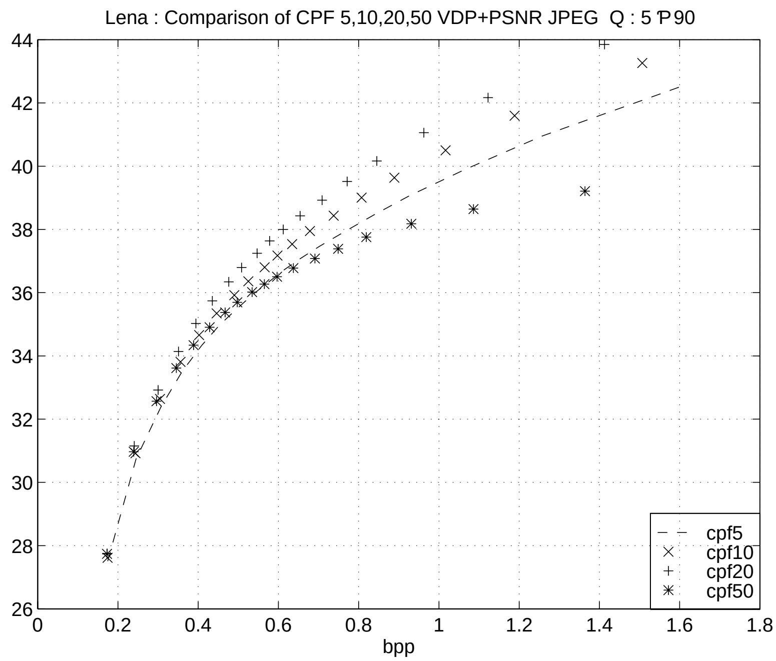

Figure 7. Coder Performance for different CPF iterations using VDP+PSNR for "Lena".

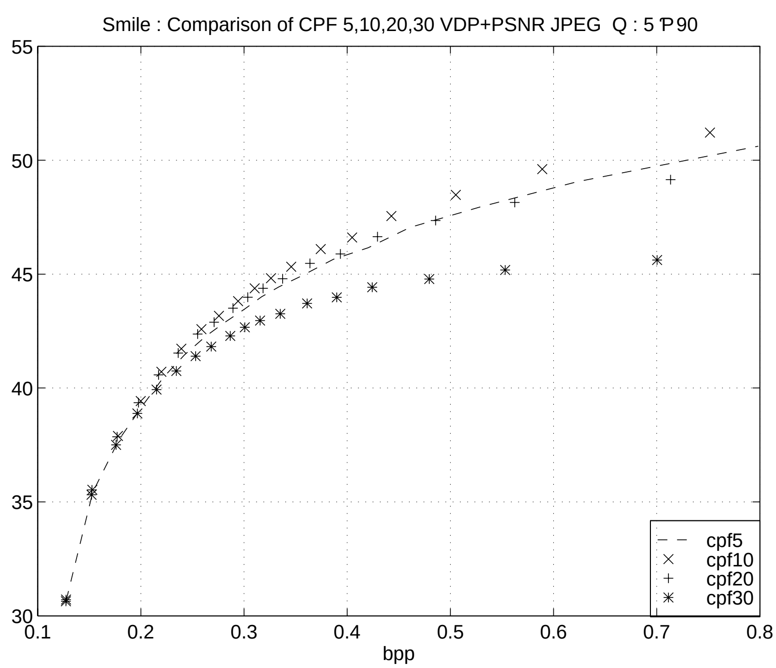

Figure 8. Coder Performance for different CPF iterations using VDP+PSNR for "Smile".

Image	Bit-rate	SNR Improvement	Best CPF
bldg	0.6	0.4dB	20
lena	0.4	1dB	30
wheel	0.4	0.5dB	20
smile	0.25	0.55dB	10
bldg	1.2	0.5dB	20
lena	1	2dB	20
wheel	1	1.4dB	20
smile	0.5	1dB	10

Table 1. The performance improvement over CPF5 measured by VDP+PSNR

Image	VDP+PSNR	Bit-rate	Rate decrease %
bldg	42dB	1.52bpp	3.95
lena	44dB	1.96bpp	27.55
wheel	48dB	1.17bpp	19.47
smile	50dB	0.74bpp	16.21

Table 2. The performance improvement over CPF5 measured by VDP+PSNR

Image	VDP+PQS	Bit-rate	Rate decrease %
bldg	4.6	1.32bpp	4.82
lena	4.8	1.07bpp	11.5
wheel	4.8	0.73bpp	17.27
smile	4.6	0.6bpp	10.41

Table 3. The performance improvement over CPF5 measured by VDP+PQS

6.2.2. VDP+PQS

The results show the performance improvement due to pre-processing. In general, this metric indicates that the gain is more modest, as shown in table 3, as compared to the predictions by the VDP+PSNR metric. The table shows the highest percentage improvement (fourth column) in the bitrate for a fixed value of VDP+PQS using the best CPF filter. The results confirm a significant gain for the "Lena", "Wheel" and "Smile" images at higher quality levels. Figure 9 shows the performance for the "Lena" image. The problem is that the VDP+PQS metric underestimates in some cases the adverse effect of pre-processing (for example, for CPF30 and CPF50). This is likely to be due to the regression weights and the distortion factor normalization used in the computation of the global value of PQS.

Because of their small regression weights, the factors quantifying random errors, F1 and F2 are essentially overcome by the other factors, hence they do not play a major role in the final PQS value at all quality levels. The factor quantifying blocking, F3 is reasonably significant at very low quality (this being a JPEG coder, it introduces a lot of blocking errors at low quality) but the regression weights again underestimate this effect. Correlated errors quantified by F4 and errors at high transition regions quantified by F5 determine to a large extent, the PQS values (especially at higher quality ranges). This becomes clearer when we consider figure 10 which shows the variations of the raw values of factor 3 for the "Building" image. Although the raw values show correctly that CPF30, for instance, degrades the performance, this loss is not reflected well in the weighted values and hence in the global PQS value.

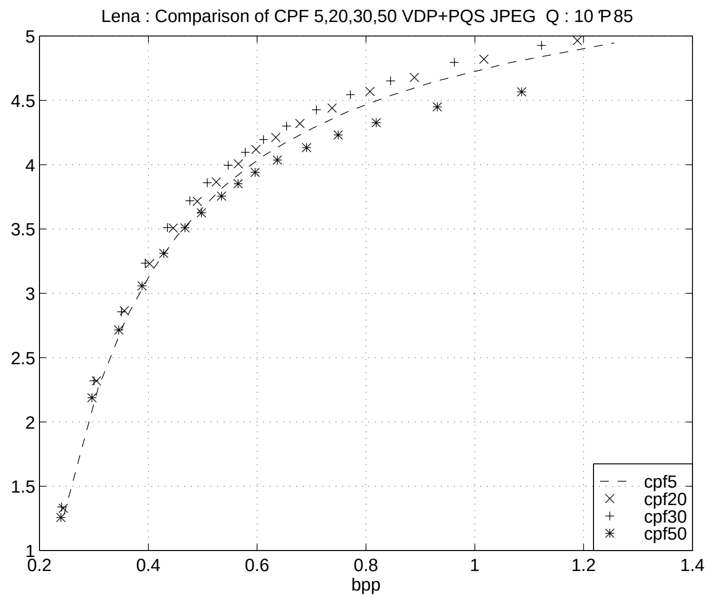

Figure 9. Coder Performance for different CPF iterations using VDP+PQS for "Lena".

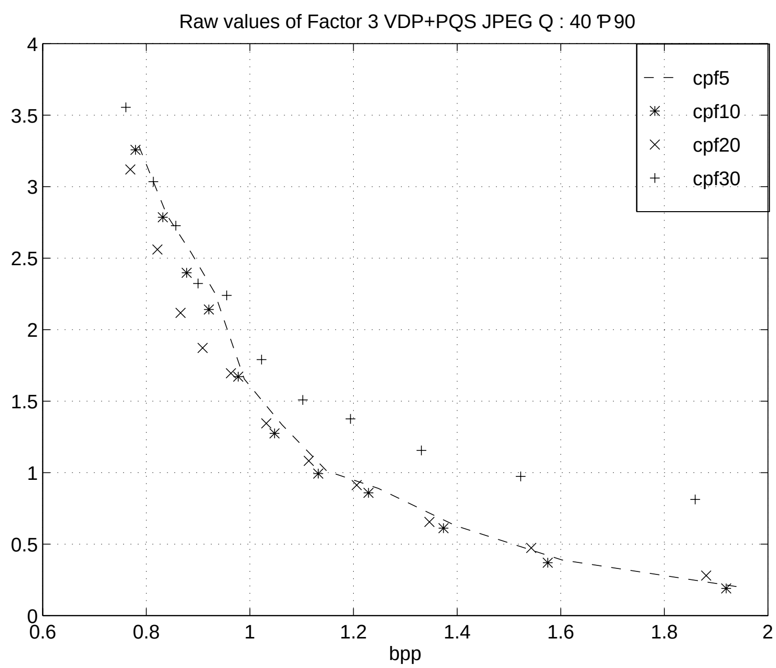

Figure 10. Raw factor 3 for different CPF iterations using VDP+PQS for "Bldg".

7. Discussion and Conclusions

The use of a indicator mask for perceivable errors provides, to the first order, the proper methodology for the evaluation of the effect of pre-processing on the performance of coders. Both the PSNR and the PQS metrics, when combined with the VDP mask, are successful in predicting the performance improvement due to pre-processing. However both metrics have specific limitations. The VDP+PSNR metric in general, is less robust, and overestimates the improvement due to noise removal, specially at high quality. This can be associated with the difficulty commonly seen when using the PSNR as a quality metric; since it is a pixel wise distance metric, which ignores the perceptual properties of the human visual system (HVS).[11,8] The HVS is more sensitive to disturbances in areas of structure in the image. The PSNR metric, however, gives importance to the magnitude of the error and ignores the information about its location in the image.

The PQS metric is based on the perceptual properties of human vision and on extensive engineering experience with the observation of image disturbances due to image coding. The PQS identifies five important coding distortions and combines them to give a global value for quality based on weights obtained from the MRA. The applicability of the MRA weights depend on the suitability of the dataset used in the regression. In the present version of PQS, the dataset consists of five different images which were distorted using JPEG and standard Wavelet and Subband coders.[8] These regression weights are not fully suitable to the present problem, where significant removal of random noise occurs.

Even though we have discussed the problems associated with specific objective quality metrics, this paper not only gives a methodology for quantifying coding performance for pre-processed images but shows that pre-processing prior to using a standard coder gives considerable improvement in coding performance. In future work, improvements will be made to the two objective quality metrics; in particular, the normalizations for the individual factors and the regression for the combining weights will be recomputed for the PQS to give better performance.

8. ACKNOWLEDGMENTS

We acknowledge the support of this research by the UC MICRO program, Lockheed Martin and Hewlett Packard.

REFERENCES

1. V. R. Algazi, G. E. Ford, A. I. El-Fallah, and R. R. Estes, Jr., "Preprocessing for improved performance in image and video coding," in *Applications of Digital Image Processing XVIII, Proc. SPIE*, pp. 22–31, 1995.
2. A. I. El-Fallah, G. E. Ford, V. R. Algazi, and R. R. Estes, Jr., "The invariance of edges and corners under mean curvature diffusions of images," in *Image and Video Processing III, R. L. Stevenson and S. A. Rajala, Eds., Proc. SPIE*, pp. 2–14, 1995.
3. S. Daly, "The Visible Difference Predictor: An Algorithm for the Assessment of Image Quality," in *Digital Images and Human Vision* (A. B. Watson, ed.), Cambridge, Mass.: MIT Press, 1993.
4. "Independent JPEG group's free JPEG software." Home archive is ftp.uu.net:/graphics/jpeg.
5. S.Daly, "The visible differences predictor : an algorithm for the assessment of image fidility," in *Proceedings of SPIE*, vol. 1666, pp. 2–15, SPIE, 1992.
6. S.Daly, "A visual model for optimizing the design of image processing algorithms," in *International Conference on Image Processing*, vol. 3, pp. 16–20, ICIP, 1994.
7. A. Watson, "The cortex transform: rapid computation of simulated neural images," *Computer Vision and Graphics and Image Processing*, vol. 39, pp. 311–27, Sept. 1987.
8. M. Miyahara, K. Kotani, and V. R. Algazi, "Objective Picture Quality Scale (PQS) for image coding." Technical Report, CIPIC, UC Davis. Submitted to IEEE Trans. on Communications, 1996.
9. M. Miyahara, "Quality assessments for visual service," *IEEE Communications Magazine*, pp. 51–60, October 1988.
10. M. G. Kendall, *Multivariate Analysis.* Charles Griffin, 1975.
11. H. Marmolin, "Subjective MSE measures (picture processing)," *IEEE Transactions on System, Man, and Cybernetics*, vol. SMC-16, pp. 486–489, May 1986.