Title
Quality Measurement and Use of Pre-processing in Image Compression

Authors
Algazi, Ralph
Avadhanam, Niranjan
Estes, Robert R.

Publication Date
1998

Peer reviewed

Quality Measurement and Use of Pre-processing in Image Compression

V. Ralph Algazi, Niranjan Avadhanam, Robert R. Estes, Jr. [a]

[a] *CIPIC, Center for Image Processing and Integrated Computing*
University of California, 1 Sheilds Ave., Davis, CA 95616-8553
{algazi,avadhana,estes}@ece.ucdavis.edu

Abstract

Traditional quality measures for image coding, such as the peak signal to noise ratio, assume that the preservation of the original image is the desired goal. However, pre-processing images prior to encoding, designed to remove noise or unimportant detail, can improve the overall performance of an image coder. Objective image quality metrics obtained from the difference between the original and coded images cannot properly assess this improved performance. This paper proposes a new methodology for quality metrics that differentially weighs the changes in the image due to pre-processing and encoding. These new quality measures establish the value of pre-processing for image coding and quantitatively determine the performance improvement that can be thus achieved by JPEG and wavelet coders.

1 Introduction

In image coding, the artifacts introduced by standard coders are uncontrolled or unpredictable in detail because the image representations are designed to reduce statistical redundancy, rather than to provide progressive image degradation. This characteristic may allow only a very small compression in order to maintain acceptable image quality. One approach to a better control of image quality, is to pre-process the image in an adaptive fashion so as to introduce imperceptible or controlled degradations to the image [12,14]. In previous work using this approach, an effective perceptually transparent coder was designed [6]. The key to making such an alternative approach to image coding possible is the use of methods that provide the necessary local control of image degradation. One recently developed technique, based on inhomogeneous diffusion [12,14], does so effectively.

Such pre-processing, when combined with a standard lossy or lossless coder, can also result in an improvement in overall coding performance [2,1,4]. At

lower levels of quality, the adaptive filter allows controlled image simplification so that the coder can compress it more effectively. However, if pre-processing is applied, the original image is no longer a suitable numerical reference to which the decoded image should be compared for an objective measure of image quality. Measuring and optimizing this subjective performance improvement at both high and lower levels of quality is the subject of this paper. It should be noted that similar pre-processing methods have been used to improve the performance of the MPEG coder [30].

Traditionally, objective measures of image quality are 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 an increase in mean square error that may be inferred to be a loss of subjective image quality, although no perceptible changes may have been introduced by pre-processing. As an alternative, the distortion introduced by the coder could be measured with respect to the pre-processed image. In this case, no weight is given to perceptible degradations introduced by pre-processing. Thus, both alternatives in are unsatisfactory for objective measures of the overall subjective effect of pre-processing and coding.

To assess the degradation caused by pre-processing, we use a human vision system (HVS) model for the prediction of the threshold of visibility of image degradations. In recent years, several HVS model based metrics have been proposed for evaluating objective image quality [33,10,26,29,20]. In particular, some HVS metrics are successful in predicting the perception of a wide range of simple stimuli as well as artifacts introduced by image processing and coding [20,10,9]. In this work, we use the Visible Differences Predictor (VDP) [10] developed by Scott Daly.

Using the VDP, we have developed two metrics which are applicable to the problem. The objective quality metrics we use are variations of the Peak Signal to Noise Ratio (PSNR) and the Picture Quality Scale (PQS) [23]. Using these two global metrics, we study the effect of pre-processing on the performance of JPEG and wavelet coders.

This paper is organized as follows: after this introductory section, we explain preprocessing for image coding in Section 2. The diffusion based preprocessing scheme is briefly described in Section 3. The VDP and PQS algorithms, on which the objective quality metrics are based, are then briefly outlined in Sections 4 and 5. In Section 6, we explain the methodology used to quantify coder performance. Section 7 describes our simulation results using the two metrics for the JPEG and wavelet coders. In Section 8, we compare coder performance on pre-processed images. Finally, in Section 9, we discuss the

results and conclude.

2 Pre-processing and Image Coding

In the context of image coding, desirable effects of pre-processing are to remove noise and simplify the image data in such a way that it is easier to encode. The coder will remove noise, especially at low quality, by coarsely quantizing the high frequency components. It will also simplify the image to decrease the bit rate, and in the process introduce undesirable artifacts. A suitably designed adaptive filter may provide better control, than increased compression, over the visual degradation introduced.

We use an adaptive, anisotropic filter, denoted the Corner Preserving Filter (CPF) [14,12,13] as the pre-processor, with the number of iterations controlling the degree of pre-processing. The choice of the pre-processing filter is very important. The CPF is based on a mean curvature diffusion algorithm where the local image properties control the diffusion. It has been shown that this filter preserves important image structure while effectively removing noise [12,4,14,13]. For compression, a standard JPEG coder [15] and the best performing wavelet coder obtained from a previous study [19,3] are used. The wavelet code computes a four level dyadic decomposition of the image with the biorthogonal "9-7" wavelet of Barlaud [7], quantizes the coefficients with a HVS based quantizer, and then losslessly encodes the quantized coefficients [19,3]. The coder exploits the spatial dependencies between non-zero coefficients by encoding a binary *activity mask*, using a context code similar to JBIG [17], and then encodes the coefficient magnitudes by mapping them onto a binary tree and encoding them with the same binary arithmetic encoder (the QM-code) [17].

The analysis of the effect of pre-processing on coding performance is divided into two parts:

(1) *Perceptually Transparent Processing*: evaluation of pre-processing for no perceptible image degradation. The pre-processing was verified to be perceptually equivalent on a calibrated monitor. The reduction in noise by pre-processing improves the effectiveness of the coder.
(2) *Perceptually Lossy Processing*: comparison of the coder performance for different levels of pre-processing. As the number of pre-processing iterations is increased, perceptual transparency is no longer preserved, but adaptive filtering preferentially preserves the integrity of visually significant areas of the image. Here, the purpose of pre-processing is to control the distribution of errors, so that they are less perceptible after coding, and simplify the image, so that fewer bits are needed to encode it. Even-

tually, though, more pre-processing introduces clearly visible artifacts which degrade the overall coder performance. Again, the quality metric should reflect these observations, and indicate the best trade-off between distortion due to the pre-processing and distortion due to compression.

For the objective determination of quality, we use two metrics:

(1) PSNR as modified by the VDP, denoted VPSNR.
(2) A measure based on the PQS methodology and distortion factor images, also modified by the VDP, denoted VPQS.

Next, we describe the adaptive filtering technique used in this work.

3 Mean Curvature Diffusion and the Corner Preserving Filter

Selective noise removal, preservation of features and controlled degradation of the perceived image quality is not possible using space invariant linear filtering. Anisotropic diffusion based adaptive filtering achieve these objectives effectively, by making use of local image properties.

3.1 Anisotropic diffusion

In adaptive noise reduction [25], an interactive data dependent filtering algorithm is used. It can be shown that filtering with a family of Gaussian kernels with variance parameter t is equivalent to solutions of the partial differential diffusion (heat) equation

$$I_t = c\nabla^2 I = c(I_{xx} + I_{yy}) \tag{1}$$

where the subscripts denote partial derivatives and ∇^2 is the Laplacian. In anisotropic diffusion, we allow the conduction coefficient, $c(x, y, t)$, to vary with respect to space and time, so that

$$I_t = \nabla \cdot [c(x, y, t)\nabla I] = c(x, y, t)\nabla^2 I + \nabla c \cdot \nabla I \tag{2}$$

where ∇ represents the gradient operation and $\nabla\cdot$, the divergence. Typically, $c = g(\nabla I)$, where g is a nonlinear function to be specified. The mean curvature diffusion (MCD) of [12] is defined by choosing

$$c = g(\nabla I) = \frac{1}{\sqrt{1 + A\left|\nabla I\right|^2}}, \tag{3}$$

where A is a scaling parameter that controls the convergence properties of the algorithm. As such, it can be shown [12] that the local rate of diffusion is equal to twice the mean curvature, H, of the image surface about each pixel. This leads to a very effective, adaptive, iterative noise reduction technique. MCD preserves image structure, characterized by regions of consistently high gradients, and substantially reduces independent, random noise. It, however, also tends to round corners and other features characterized by higher order structure, such as edge intersections.

In a more recent work, El-Fallah [4] proposed a modification to MCD by choosing

$$c = \frac{1}{|\nabla g|\sqrt{1 + [2H(|\nabla g| - 1)]^2}} \tag{4}$$

to better preserve corners. Using this filter, denoted the corner preserving filter (CPF), more iterations are allowed (yielding more noise reduction) while maintaining high image quality and/or perceptual transparency [14].

In the continuous case, anisotropic diffusion corresponds to a solution of the heat equation and takes continuous surfaces into minimal surfaces. As such, it enhances edges and preserves their location and sharpness. This is in contrast to isotropic filtering operations which often blur and smooth important image features. In the discrete case, anisotropic diffusion is implemented as a spatially and temporally adaptive filter. Figure 1 illustrates and compares isotropic and anisotropic diffusion techniques for lossy processing. For processing close to perceptual transparency, in flat portions of the image, ten CPF iterations result in more than 10 dB of noise reduction and a significant decrease in entropy, while introducing no perceptible changes. Note that, for noisy images, adaptive noise removal may actually improve the appearance or subjective image quality, by reducing perceptible noise in flat regions and increasing the apparent edge contrast.

Next, we describe the VDP algorithm which we use to predict perceptible changes to the image introduced by pre-processing.

4 Visible Differences Predictor

The VDP algorithm proposed by Scott Daly [8] 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 error detection probabilities.

(a) 25 iterations of isotropic diffusion (Gaussian blurring).

(b) 100 iterations of CPF filtering.

Fig. 1. Comparison of isotropic and anisotropic diffusion. The original image was corrupted by Gaussian noise with variance 20 and both techniques remove the same amount of noise.

The model describes threshold perception only; all suprathreshold errors are mapped to a probability of 1.

The overall model (Figure 2) is implemented as a cascade of sub-models to incorporate the known properties of the human visual system. The main components of the model are:

(1) perceptual nonlinearity,
(2) contrast sensitivity function (CSF),
(3) orientation and frequency selective cortex bands,
(4) masking properties,
(5) psychometric function, and
(6) probability summation.

Complete details are given in [8,10]. The algorithm has been shown to be in agreement with several psychophysical experiments [9]. In addition, the model has been successfully used to measure many imaging artifacts, e.g. compression distortions, banding, blur and tone-scale changes, etc. [10].

We now summarize the main components of the algorithm. A shift invariant nonlinearity models the light adaptive property of the retina. A display calibration model is used to map the gray levels into luminance values on the monitor. These two transformations have been combined into a transformation denoted perceptual nonlinearity.

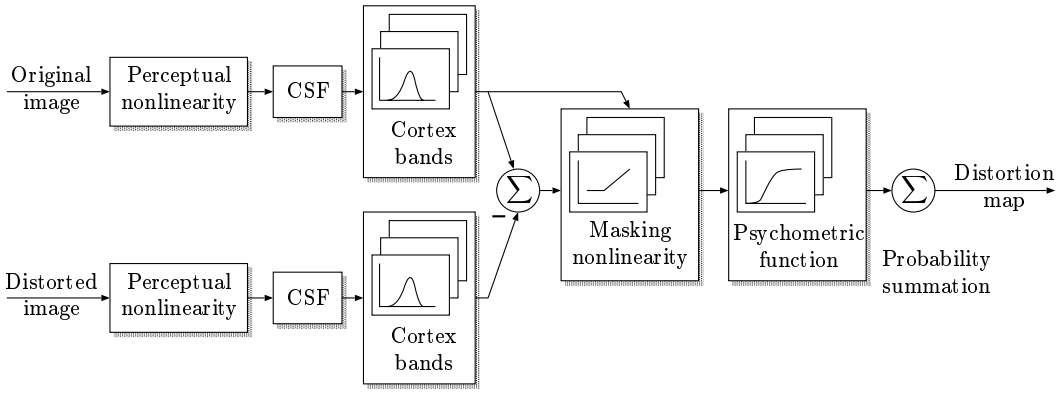


Fig. 2. VDP construction.

The CSF quantifies the visual response as a function of the spatial frequency. The cortex bands block implements the spatial frequency and orientation selectivity 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 [10]) and orientation selective fan filters based on the cortex transform [27]. 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, including the baseband. Making use of results from sinusoidal masking and noise experiments, Daly proposed a masking function of the form

$$T_e = (1 + (k_1(k_2 m_n)^s)^b)^{\frac{1}{b}}, \tag{5}$$

where the masking effect due to the image activity is evaluated for each cortex band. Threshold elevation due to masking is 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, $k_1 = 6^{-7/3}$, $k_2 = 6^{10/3}$ and $b = 4$ [10].

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

$$P(c) = 1 - e^{-(\frac{c}{\alpha})^{\beta}} \tag{6}$$

where c is the contrast of the error, α the contrast threshold, and β the slope of the psychometric function. This yields a detection probability 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 31 bands are combined to give a single map of the

error detection probability, as a function of location in the image, using

$$P_t(m,n) = 1 - \prod_{k=0..5, l=1..6} (1 - P_{k,l}(m,n)) \tag{7}$$

where $k = 0$ corresponds to the baseband.

For the present work, we ignore the sign of the error and compute a binary map where all errors that have a detection probability greater than 0.9 are mapped to a value of 1, and categorized as supra-threshold errors. It is this map that we use to drive our new quality metrics.

In addition to the threshold map provided by the VDP, our quality metrics rely on the PSNR and PQS supra-threshold metrics, which are outlined next.

5 Picture Quality Scale

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 [23,22]. Coding distortions can be typically identified as blurring, ringing, blocking, etc. The severity of these distortions to a human observer differs according to specific spatial structure of the artifacts. Hence, different distortions should be combined reflecting their degree of subjective visibility [23,31]. The PQS identifies five important coding distortions and combines them to give a global numerical quality measure [23,22]. The PQS metric has been successfully used in several applications, such as the design of an electro-optical imaging system [24], the optimization of coders [5], and the comparison of coder performance [19]. In related work, Xu *et al.* [32] proposed a segmentation based error metric using a similar method based on correlation with an impairment scale, and Davies *et al.* [11] developed a similar metric using a nonlinear neural network. All of these methods consider the distortions from a 'high level' perspective by identifying the structure in the errors. Similar methods have also been applied to video [28].

A simple overview of the algorithm is now presented. Complete details are given in [23]. Figure 3 summarizes the steps used in the construction of PQS. First, the image signal is transformed into one which is proportional to the visual perception of luminance using a power law and then the frequency weighted errors $e_w(m,n)$ are obtained by filtering with a CSF-like function $)(S_a)$. 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 lead to numerical measures in terms of distortion factors F_i. At the first stage, factor images $f_i(m,n)$

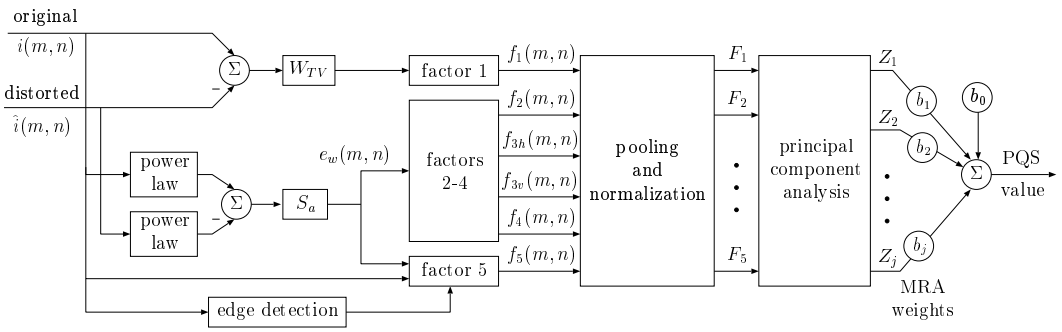


Fig. 3. PQS construction.

are obtained by performing local computations on $e_w(m,n)$. Next, distortion factors F_i are obtained by pooling the corresponding factor images. [1]

The PQS makes use of five factors, the first two factors account for random errors, the third for blocking artifacts, and the last two factors, dominant at high quality, correspond to structured errors and masked errors near edges. The global value for PQS is given by the linear combination of the principal components, $\{Z_j\}$, computed from the distortion factors, $\{F_j\}$, so that

$$PQS = b_0 + \sum_{j=1}^{J} b_j \cdot Z_j \tag{8}$$

where $\{b_j\}$ are parameters to be determined.

PQS is designed to be in good correspondence with the subjective evaluation of image quality. The subjective assessment of quality is performed under controlled viewing conditions, using the impairment scale of ITU-R recommendation 500, resulting in a Mean Opinion Score (MOS) that ranges from 1 to 5 [16]. The coefficients, $\{b_j\}$, are computed using multiple regression analysis (MRA) [18] between the distortion factors and the experimentally determined, subjective MOS ratings. 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 [23]. The correlation coefficient between PQS and the MOS scores, $R = 0.92$, which represents a great improvement as compared to a correlation of $R = 0.54$ when only the frequency weighted MSE (F_2) is used. A detailed description of PQS and its performance for the full range of quality can be found in [23].

[1] Note that factor F_1 does not make use of the same source transformations as the other factors. It is included in PQS because it is a commonly used CSF weighted distortion.

Fig. 4. Error visibility map for Lena: (left) CPF50 preprocessed image, (right) overlayed VDP mask.

6 New Quality Metrics

We now develop two new objective quality metrics that combine threshold visibility maps, as indicated by the VDP output, with the supra-threshold objective quality measures, PSNR and PQS.

The VDP predicts areas of an image where distortions will be perceptible. This suggests that the binary 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 with respect to the original. Figure 4 shows the Lena image after 50 iterations of the CPF (lossy pre-processing) and the resulting VDP mask. Areas where the pre-processing introduces suprathreshold errors (mask values of 1) are mapped to white and overlaid on the original image for comparison. Note that the striped appearance of the VDP mask is mainly due to ignoring the sign of the supra-threshold errors.

6.1 VPSNR: extension of PSNR using VDP

Let $M(m,n)$ be the binary 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, $I_p(m,n)$ the pre-processed image, and $I_e(m,n)$ the encoded

image. We can now compute two error images

$$I_{do}(m,n) = M(m,n)[I_e(m,n) - I_o(m,n)] \tag{9}$$

and

$$I_{dp}(m,n) = [1 - M(m,n)][I_e(m,n) - I_p(m,n)] \tag{10}$$

where $[1-M(m,n)]$ is the complement of $M(m,n)$ and the subscripts indicate the reference image used.

$I_{do}(m,n)$ 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, $I_{dp}(m,n)$, evaluated in areas where pre-processing does not introduce perceptible errors, uses the pre-processed image as a reference.

We now evaluate the performance of the coders by adding the contributions due to $I_{do}(m,n)$ and $I_{dp}(m,n)$. Thus, a modified mean-square error is given by

$$\text{MSE} = \sum_{m,n} I_{do}^2(m,n) + I_{dp}^2(m,n), \tag{11}$$

and the (V)PSNR is computed as usual.

6.2 VPQS: extension of PQS using VDP

Using a similar approach, we compute the PQS value based on the two domains $M(m,n)$ and $[1 - M(m,n)]$. As described in Section 4, the factor images are are obtained from local computations on the error image. Here, two such error images are computed, $[I_e(m,n) - I_o(m,n)]$ and $[I_e(m,n) - I_p(m,n)]$, from which two sets of factor images are obtained, respectively, $\{f_{io}(m,n)\}$ and $\{f_{ip}(m,n)\}$. Each pair of factor images is combined to obtain

$$f_i'(m,n) = M(m,n)f_{io}(m,n) + [1 - M(m,n)]f_{ip}(m,n), \tag{12}$$

which are normalized and pooled to obtain the global PQS distortion factors and the VPQS value, using exactly the same methodology as used in PQS (and shown in Figure 3).

7 Experiments and Results: Distinct Coders

First, for clarity, we characterize the performance of the JPEG and wavelet coders separately. In the next section, we compare the performance of the two coders in the pre-processing framework.

In our study, we considered the four images shown in Figure 5, i.e. Bldg, Lena, Wheel and Smile. The subjective evaluation of images used a Super Match monitor which had a gamma of 2 and a peak luminance of 72 cd/m^2. Ambient light levels were reduced to a minimum. The viewing distance was 4 times the picture height and all other viewing conditions met the ITU-R recommendations [16]. In this experiment, the original and pre-processed images were compared side by side. The presentation time was 12 seconds, and the three subjects were asked if the two images were perceptually equivalent. The number of iterations for perceptual transparency was consistent across subjects, but was image dependent. However, 5 iterations resulted in perceptual transparency for all images and observers. The perceptual equivalence of the CPF5 images with the originals was also verified by computing the corresponding VDP maps which, in all cases, indicated that the errors were at, or below, the visual threshold.

Hence we consider CPF5 to be a noise reduced image which is perceptually equivalent to the original image (CPF0). We also performed 10, 20, 30 and 50 iterations of the CPF in our study of the effect of lossy pre-processing on coder performance. In all the experiments, the JPEG coder was used at quality settings between 5 and 90 in steps of 5. We roughly characterize the range for quality settings 5 to 40 as low quality and 45 to 90 as high quality. For the wavelet coder, we use a scaling parameter ranging from 1 to 70. This scales the HVS based quantization matrix [19,3] and results in a range of bit-rates. Visually, we roughly characterize the range from 5 to 25 as high quality and 30 to 70 as low quality.

7.1 Perceptually Transparent Pre-Processing

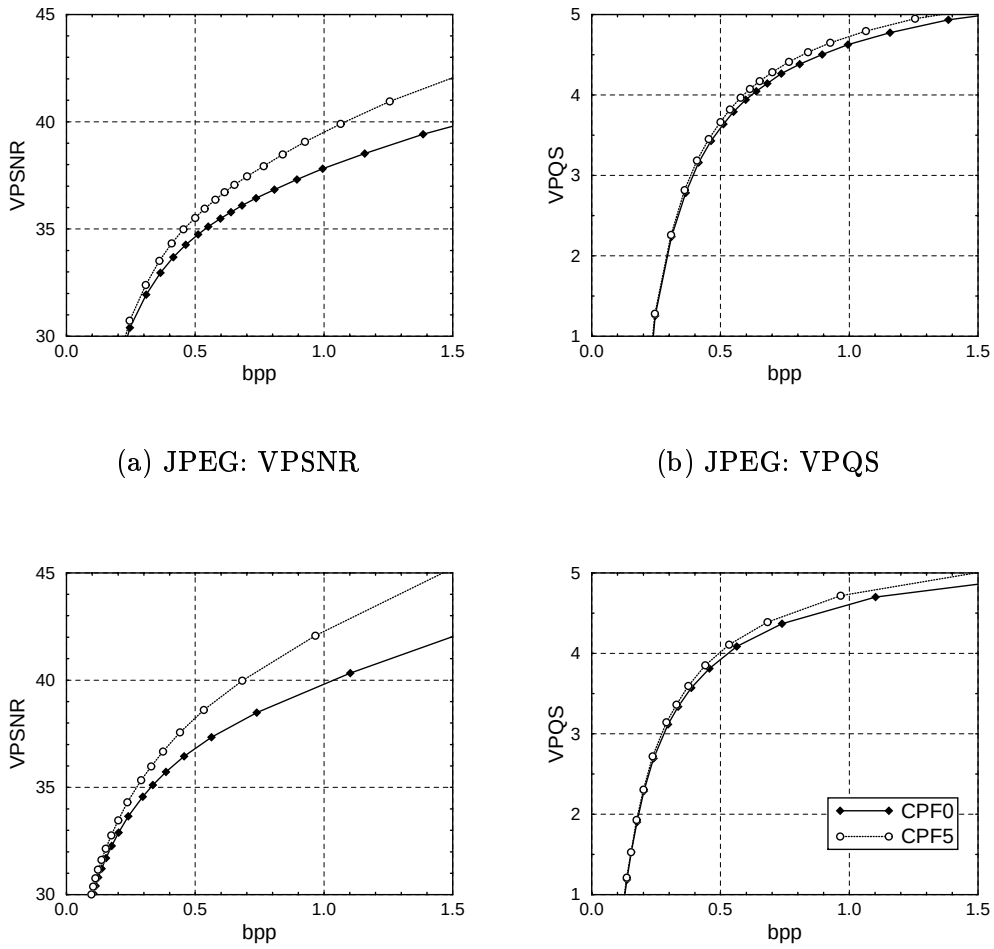
We first verified that the conventional PSNR and PQS using the original image as a numerical reference show an increase in distortion, and thus do not illustrate the value of pre-processing. However, as discussed, we also visually verified that CPF5 pre-processing is perceptually transparent. Thus, the new quality metrics are necessary to quantify the gain achieved by pre-processing for such imperceptible changes to the image. Figure 6 illustrates the benefits of pre-processing for both coders, using CPF0 and CPF5, as measured using VPSNR and VPQS, for the Lena image. It is apparent from the graphs that

Fig. 5. Test Images. From left to right, then top to bottom: Bldg, Lena, Wheel, and Smile.

pre-processing is beneficial at higher levels of quality for both coders. For example, at 40 *dB* VPSNR, the bit rate required by the wavelet coder is reduced from 1.04 bpp to 0.69 bpp. Using the more perceptually correct VPQS measure, at a VPQS value of 4.6, indicates a decrease from 1 bpp to 0.83 bpp.

7.2 Comparison for Increased Pre-Processing

We now investigate whether pre-processing beyond perceptual transparency provides an improvement in coder performance. As mentioned, we use the adaptive filter for controlled image simplification so as to obtain better coding performance. However, we expect that, as we increase the number of iterations, pre-processing will eventually degrade the perceived image quality as it introduces visually significant artifacts without a commensurate gain in



(a) JPEG: VPSNR

(b) JPEG: VPQS

(c) wavelet: VPSNR

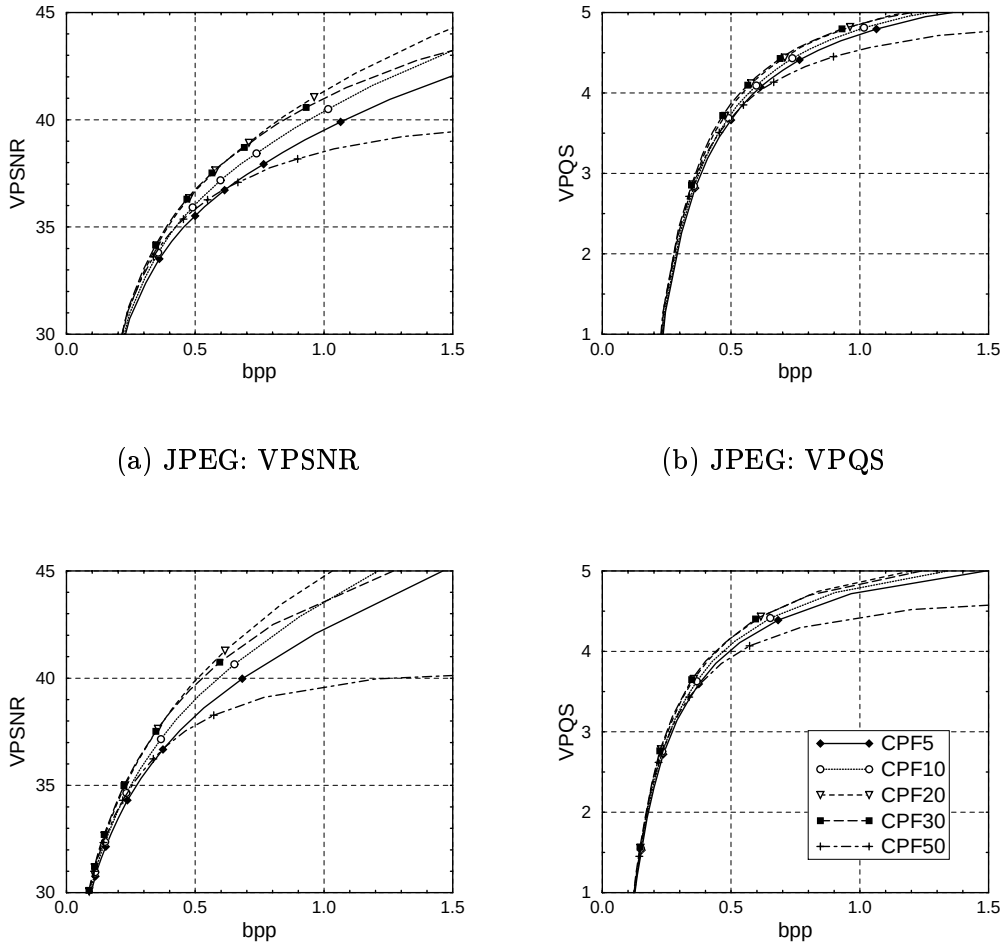
(d) wavelet: VPQS

Fig. 6. Coder performance for perceptually transparent pre-processing (Lena image).

bitrate.

7.2.1 VPSNR

We observe that for both coders, the effect of pre-processing depends on both the number of CPF iterations and on the target bit rate. Note, in particular, that the graphs for different levels of pre-processing intersect (Figures 7(a) and 7(c)). As a general rule, 20 iterations (CPF20) gives the best overall performance. This can be considered to be the best compromise between the controlled distortion and simplification due to pre-processing and the additional degradations introduced by the coder. Table 1 summarizes of the VPSNR improvement results obtained for both coders using the best CPF filter. The performance improvement is computed with respect to the CPF5 result. The



(a) JPEG: VPSNR

(b) JPEG: VPQS

(c) wavelet: VPSNR

(d) wavelet: VPQS

Fig. 7. Coder performance for perceptually lossy pre-processing (Lena image). For clarity, only every third data point is marked.

first four rows correspond roughly to lower quality images and the bottom four rows to higher quality levels.

Concentrating on the wavelet coder, Table 2 shows representative percentage decreases in the bitrate for a fixed value of VPSNR using the best CPF filter. These results indicate that the gain is image dependent, and that significant gains are possible for Lena, Wheel and Smile, at higher quality levels.

Note that for Smile, the best results are obtained at 10 CPF iterations, as indicated in Figure 8. For this image, we have verified that additional pre-processing produced clearly visible artifacts.

Image	Wavelet		JPEG		Best
	rate (bpp)	Improv. (dB)	rate (bpp)	Improv. (dB)	CPF iterations
Bldg	0.32	0.22	0.60	0.4	20
Lena	0.20	0.84	0.40	1.0	30
Wheel	0.15	0.23	0.40	0.5	20
Smile	0.07	0.60	0.25	0.55	10
Bldg	1.00	0.46	1.20	0.5	20
Lena	0.80	2.25	1.00	2.0	20
Wheel	0.55	1.25	1.00	1.4	20
Smile	0.25	0.84	0.50	1.0	10

Table 1
Performance improvement over CPF5, as measured by VPSNR.

Image	VPSNR (dB)	rate (bpp)	rate decrease (%)
Bldg	40.0	0.99	4.3
Lena	40.9	0.80	28.8
Wheel	45.2	0.55	14.2
Smile	46.7	0.25	12.5

Table 2
Performance improvement over CPF5 for the wavelet coder, as measured by VP-SNR.

7.2.2 VPQS

The results using this metric again establish the performance improvement due to pre-processing (Figures 7(b) and 7(d) and Table 3). In general, this metric indicates that the gain is more modest, as compared to the predictions by the VPSNR metric. Table 3 shows the highest percentage improvement in bitrate for a fixed value of VPQS using the best CPF filter and the wavelet coder, and confirms significant gain for Lena, Wheel and Smile, at higher quality levels.

8 Experiments and Results: Comparison of coders

In this section, we compare the performance of both coders, using the best CPF pre-processing results from the previous section. We use both metrics,

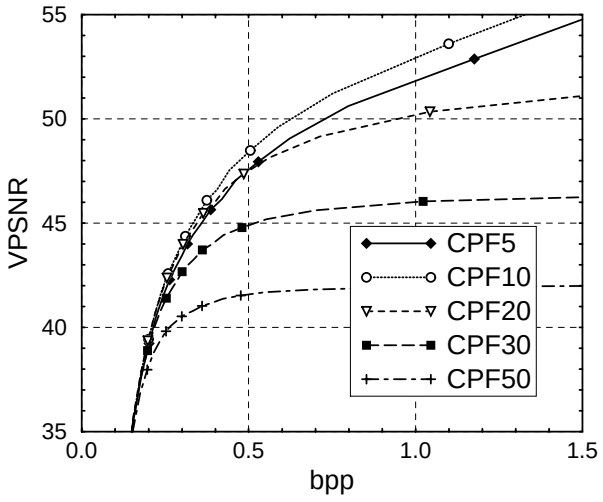


Fig. 8. JPEG coder performance for perceptually lossy processing (Smile image). For clarity, only every third data point is marked.

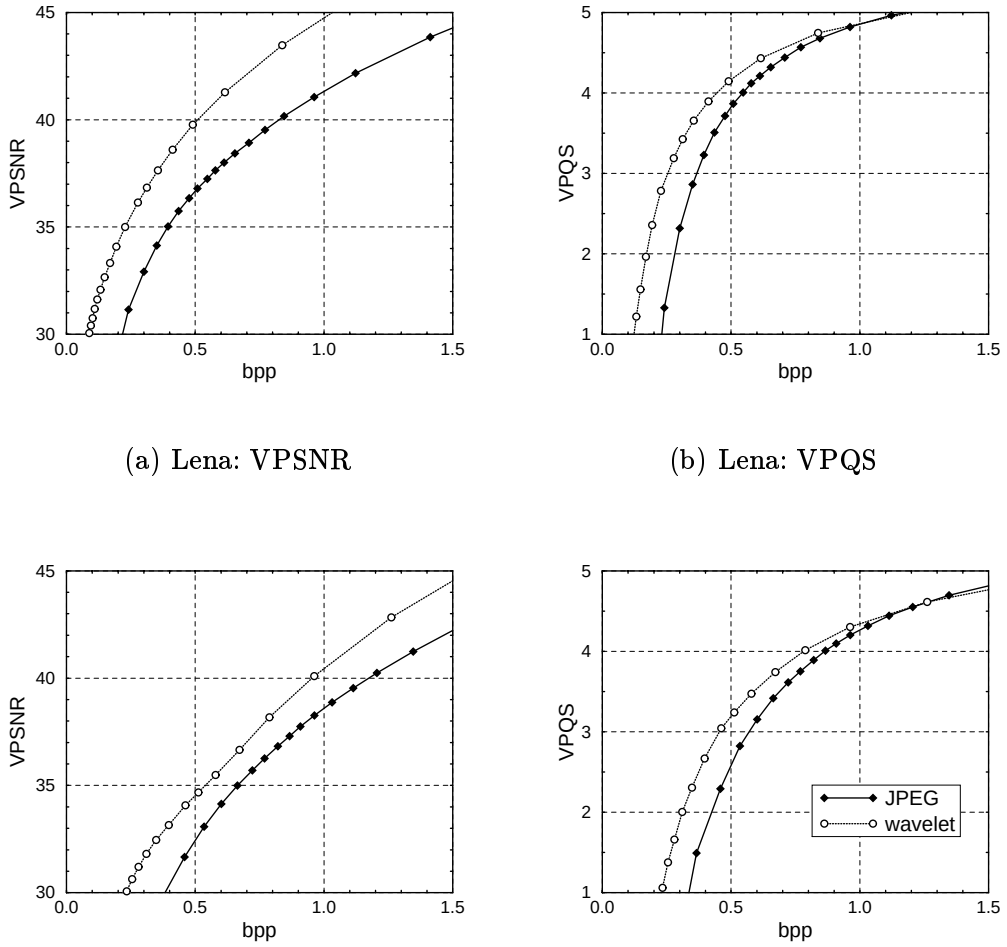
Image	VPQS	bitrate (bpp)	rate decrease (%)
Bldg	4.7	1.0	5
Lena	4.5	0.8	17.5
Wheel	4.6	0.5	15.8
Smile	4.0	0.25	10

Table 3
Performance improvement over CPF5 for the wavelet coder, as measured by VPQS.

VPSNR and VPQS, in the comparison.

In general, the wavelet coder performs better than the JPEG coder, specially at lower bit-rates. Figures 9(a) and 9(c) compare the two coders, for the Lena and Bldg images, using the VPSNR metric. Using this metric, we observe an improvement of 42% for Lena at a bitrate of 0.40 bpp. For Bldg, the gain is 22% at a bitrate of 0.60 bpp. Figures 9(b) and 9(d) show the same comparison using the VPQS metric. The gains, in this case, are more modest, an improvement of 27% at 0.40 bpp for Lena and 20% at 0.40 bpp for Bldg.

It can be seen that although both metrics are able to demonstrate the improvement in performance, the VPQS measure more accurately characterizes the wavelet coder gain over JPEG (as verified visually) at lower levels of quality.



(a) Lena: VPSNR

(b) Lena: VPQS

(c) Bldg: VPSNR

(d) Bldg: VPQS

Fig. 9. Coder comparison using VPSNR and VPQS. Results are based on optimal number of CPF iterations (for these images, 20).

9 Discussion and Conclusions

The methodology presented in this paper addresses, in part, the problem of numerical evaluation of image quality resulting from an image filtering operation, by differentiating between changes to the image that are perceptually benign, from modifications that may be detrimental. For image coding at high quality, large portions of the image do not undergo perceptually important changes in either processing or encoding, and the method we have presented identifies such regions. For lower quality levels, although we take the modifications introduced by the filtering operation into account, we observe that there is a coding gain due to the controlled image simplification prior to encoding.

The use of an indicator mask for perceivable errors provides 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 VPSNR metric is, in general, 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-based distance metric, which ignores the perceptual properties of the human visual system (HVS) [21,23]. The HVS is more sensitive to disturbances in areas of structure in the image. The VPSNR metric, however, gives importance to the magnitude of the error and ignores the information about its location in the image.

The VPQS 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 depends on the suitability of the dataset used in the regression. In the present version of VPQS, the dataset consists of five different images which were distorted using JPEG and standard wavelet and subband coders [23].

Note that we did not consider other suprathreshold objective quality metrics, based directly on HVS models, that have been published recently [33,26,29,20]. The choice of the objective quality metric used is not crucial in the sense that the methodology can be used for any objective metric which uses a reference image to compute image quality.

In conclusion, 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.

Acknowledgement

This work was supported in part by the UC MICRO program of the University of California, Lockheed Martin, Cornerstone Imaging, and Qualimage, Inc. The authors also wish to thank the anonymous reviewers, whose comments led to significant improvement of the paper.

References

[1] V. R. Algazi, N. Avadhanam, and R. R. Estes, Jr. Optimizing the performance of wavelet coders by image processing. In *Proceedings of SPIE, Applications of Digital Image Processing XX*, 1997.

[2] V. R. Algazi, N. Avadhanam, and R. R. Estes, Jr. Quantifying coding performance for pre-processed images. In *Proceedings of SPIE, Very High Resolution and Quality Imaging*, pages 123–133, 1997.

[3] V. R. Algazi and R. R. Estes, Jr. Comparative performance of wavelet and JPEG coders at high quality. In *Proceedings of SPIE, Very High Resolution and Quality Imaging*, pages 71–82, 1997.

[4] 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 *Proceedings of SPIE, Applications of Digital Image Processing XVIII*, pages 22–31, 1995.

[5] V. R. Algazi, G. E. Ford, M. Mow, and A. Najmi. Design of subband coders for high quality based on perceptual criteria. In *Proceedings of SPIE, Application of Digital Image Processing XVI*, volume 2028, pages 40–49. SPIE, 1993.

[6] V. R. Algazi, G. E. Ford, A. Najmi, A. El-Fallah, and R. Estes. Progressive, perceptually transparent coder for very high quality images. In *Applications of Digital Image Processing XVII*, volume 2298 of *Proc. of the SPIE*, pages 13–24, San Diego, CA, Jul. 1994.

[7] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. *IEEE Transactions on Image Processing*, 1(2):205–220, April 1992.

[8] S. Daly. The visible differences predictor: An algorithm for the assessment of image fidelity. In *Human Vision, Visual Processing and Digital Display III*, volume 1666 of *Proceedings of the SPIE*, pages 2–15. SPIE, 1992.

[9] S. Daly. A visual model for optimizing the design of image processing algorithms. In *Proceedings of the IEEE International Conference on Image Processing*, volume 3, pages 16–20, 1994.

[10] Scott Daly. The visible difference predictor: An algorithm for the assessment of image quality. In *Digital Images and Human Vision*. MIT Press, Cambridge, Mass., 1993.

[11] I. R. L. Davies, D. Rose, and R. J. Smith. Automated image quality assessment. In *Human Vision, Visual Processing, and Digital Display IV*, volume 1913 of *Proceedings of the SPIE*, pages 27–36, 1993.

[12] A. I. El-Fallah and G. E. Ford. Mean curvature evolution and surface area scaling in image filtering. *IEEE Transactions on Image Processing*, 6(5):750–753, May 1997.

[13] A.I. El-Fallah and G.E. Ford. The evolution of mean curvature in image filtering. In *ICIP Proceedings, First International Conference on Image Processing*, pages 298–302. IEEE, 1994.

[14] 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 *Proceedings of SPIE, Image and Video Processing III*, pages 2–14, 1995.

[15] Independent JPEG Group. JPEG software available at `ftp://ftp.uu.net/graphics/jpeg`.

[16] ITU/R Rec. BT.500-8: Methodology for the subjective assessment of the quality of television pictures. `http://www.itu.ch/`, 1998.

[17] ITU/T Rec. T.82: Information technology - coded representation of picture and audio information - progressive bi-level image compression. `http://www.itu.ch/`, 1993. Also appears as ISO/IEC, International Standard 11544: 1993, "Progressive Bi-level Image Compression".

[18] M. G. Kendall. *Multivariate Analysis.* Charles Griffin, 1975.

[19] Jian Lu, V. Ralph Algazi, and Robert R. Estes, Jr. A comparative study of wavelet image coders. *Optical Engineering*, 35(9):2605–19, Sep. 1996.

[20] J. Lubin. A visual discrimination model for imaging system design and evaluation. In E. Peli, editor, *Vision Models for Target Detection and Recognition.* World Scientific, Singapore, 1995.

[21] H. Marmolin. Subjective MSE measures (picture processing). *IEEE Transactions on System, Man, and Cybernetics*, SMC-16(3):486–489, May 1986.

[22] M. Miyahara. Quality assessments for visual service. *IEEE Communications Magazine*, pages 51–60, October 1988.

[23] M. Miyahara, K. Kotani, and V. R. Algazi. Objective picture quality scale (PQS) for image coding. *IEEE Transactions on Communications*, 1998. In press. Also available as CIPIC report 98-01, `http://info.cipic.ucdavis.edu/scripts/reportPage?98-01`.

[24] J. L. Olives, B. Lamiscarre, and M. Gazalet. Optimization of electro-optical imaging system with an image quality measure. In *Proceedings of SPIE, Very High Resolution and Quality Imaging*, pages 158–167, 1997.

[25] P. Perona and J. Malik. Scale–space edge detection using anisotropic diffusion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(7):629–39, 1990.

[26] P. C. Teo and D. J. Heeger. A model of perceptual image fidelity. In *Proceedings of the IEEE International Conference on Image Processing*, volume 2, pages 343–5, Oct. 1995.

[27] A. B. Watson. The cortex transform: Rapid computation of simulated neural images. *Computer Vision, Graphics and Image Processing*, 39:311–327, 1987.

[28] A. A. Webster, C. T. Jones, M. H. Pinson, S.D. Voran, et al. An objective video quality assessment system based on human perception. In *Proceedings of SPIE, Human Vision, Visual Processing, and Digital Display IV*, volume 1913, pages 15–26. SPIE, 1993.

[29] S. J. P. Westen, R. L. Lagendijk, and J. Biemond. Perceptual image quality based on a multiple channel HVS model. In *Proceedings, International Conference on Acoustics, Speech, and Signal Processing*, volume 4 of *ICASSP 95*, pages 2351–2354. ICASSP, 1995.

[30] Y. Wong, E. Viscito, and E. Linzer. Preprocessing of video signals for MPEG coding by clustering filter. In *ICIP Proceedings, International Conference on Image Processing*, volume 2, pages 129–132. IEEE, 1995.

[31] W. Xu and G. Hauske. Image quality models for extended sets of distortions. In *IEE International Conference on Image Processing and its Applications*, volume 354, pages 262–265. IEE, 1992.

[32] W. Xu and G. Hauske. Picture quality evaluation based on error segmentation. In *Proceedings of SPIE, Visual Communications and Image Processing*, volume 2308, pages 3:1454–65. SPIE, 1994.

[33] C. Zetzsche and G. Hauske. Multiple channel model for the prediction of subjective image quality. In *Proc. SPIE*, volume 1077, pages 209–216. SPIE, 1989.