Quality Constrained Compression Using DWT Based Image Quality Metric

Zhigang Gao, *Member, IEEE* and Yuan F. Zheng, *Fellow, IEEE*

Dept. of Electrical and Computer Engineering

The Ohio State University

Columbus, OH 43210 USA

zhgao@cisco.com, zheng@ece.osu.edu.

Abstract

A quality constrained compression algorithm based on Discrete Wavelet Transform (DWT) is proposed. The spatial-frequency decomposition property of DWT provides possibility for not only the new compression algorithm, but also a frequency-domain quality assessment method that can be executed in real-time. For this propose, a new quality metric in the wavelet domain called WNMSE is suggested, which assesses the quality of an image with the weighted sum of normalized mean square errors of the wavelet coefficients. The metric is consistent with the human judgment of visual quality as well as estimates the post-compression quality of an image efficiently. Based on the relationship among the statistic features, quantization step-sizes, and WNMSE value of a compressed image, we develop a Quality Constrained Quantization algorithm which can determine the quantization step-sizes for all the wavelet subbands for compressing the image to a desired visual quality accurately.

Index Terms

Discrete Wavelet Transform (DWT), Human Visual System (HVS), Image Quality Metric (IQM), Quality Constrained Quantization

I. INTRODUCTION

To compare the performances of any two image compression methods, both the compression ratios and the qualities of the compressed images have to be considered. An ideal compression system should represent the original image with as small amount of bits as possible while maintaining a good visual quality. In reality, it is always objective in measuring the compression ratio, but highly subjective to judge the quality. Since humans are the end user of most images, the natural way to compare the quality of two images is to have them evaluated by human observers. Typically, a group of observers examine a set of images under a controlled environment and assign a numerical score to each of them. Each image's scores are recorded and averaged later as its Mean Opinion Score (MOS) [1] that is by far the most accurate and reliable objective Image Quality Metric (IQM). Unfortunately, MOS is inconvenient and expensive to use.

In [2], ten quality metrics were evaluated against subjective human evaluation. The evaluation was conducted on five different distortion types with variant degrees of impairments. It is claimed that there still exists difference between machine and human evaluations of image quality, and it is difficult to invent a quality assessment algorithm that is superior in every distortion type. This work is motivated by the need for simple IQMs that are consistent with MOS and suitable for computer implementation. By ”consistent”, we mean that a metric should perform the same regardless of the distortion types or patterns of the images and be linearly correlated to MOS. That is, it is accurate (giving the same IQM score to images that have the same MOS scores), and increases or decreases monotonically with MOS.

According to its dependence on the original image, an IQM can be classified into three categories [4]:

1) Full-Reference (FR). A Full-Reference metric requires that the original image is available and therefore be used to evaluate the quality of the distorted image. This is the most common category.
2) Reduced-Reference (RR). A Reduced-Reference metric evaluates the quality of the distorted image with only partial knowledge of the original one.
3) No-Reference (NR). A No-Reference metric evaluates the quality of a distorted image without the knowledge of the original one.

This work will focus on Full-Reference IQMs. The most common IQMs are the Mean Squared Error (MSE) family, including MSE, root MSE (RMSE), and Peak Signal to Noise Ratio (PSNR), which are simple pixel error based and their performances are far from satisfactory [3]. Some more sophisticated pixel error based IQMs are also available, such as the method of Damera-Venkata et al. in [5], whose performance, however, is not generally better than the others [6]. The limitation of simple pixel error based metrics is also experienced in applications of medical images such as in [7], where the compressed diagnostic breast images with lower PSNR values are preferred by doctors over those with higher PSNR values. That is, the images favored by PSNR do not agree with the judgment of human eyes.

Wang and Bovik proposed a Structural SIMilarity index (SSIM) that models the total distortion of an image block as the combination of three factors: loss of correlation, luminance distortion, and contrast distortion [8]. SSIMs are measured for blocks of an image using a sliding window, and the mean value of the SSIMs (MSSIM) of all the blocks is taken as the overall quality metric of the image. In [9], Shnayderman et al. explored the feasibility of Singular Value Decomposition (SVD) in developing a new IQM that can express the quality of distorted images. An image is first divided into small blocks. The distance between the singular values of the original image block and the singular values of the distorted image block is used to indicate its quality. The overall quality of the distorted image is measured by the absolute value average of differences between these singular value distances and their median. The author claimed that better performance was achieved with smaller block size, which suggested that single pixel based measurement will have the best result. This, in fact, undermined the foundation of their work since singular value decomposition makes no sense for single pixel based measurement. In spite of the differences, these metrics have the same drawback in which they are determined in the spatial domain while compression is performed in the

frequency domain, which makes it very difficult to control the visual quality during the compression.

A great deal of effort has been made to develop IQMs that fit the Human Visual System (HVS). While some metrics yield decent results, most are not always consistent with HVS and are sometimes limited to very specific applications. Furthermore, these metrics tend to be complex for implementation. Watson et al. developed a Discrete Cosine Transform (DCT) based video quality metric that incorporates quite a few aspects of human visual sensitivity in [10], and a simple IQM was proposed by Sendashonga and Labeau for both DCT and DWT in [11].

In general, compression technologies can be classified into two categories: lossless and lossy.

- Lossless compression achieves compression by reducing the entropy of the original data, while avoiding introducing distortions into it. As a result, the original data can be perfectly recovered from the compressed bits. Examples of lossless compression technologies include Run-length coding, Huffman coding and Arithmetic coding, etc. But avoiding distortion limits its compression efficiency. So when used in multimedia compression where distortion is acceptable, lossless compression is more often applied on the output coefficients of lossy compression.
- Lossy compression technologies usually first transform an image into the frequency domain, and then quantize its coefficients. Two most common options of transformation are DCT and DWT. Compared with DCT, coefficients of DWT are localized in both spatial and frequency domains, which is desirable because HVS functions as a bandpass filter with the localization property [12]. Quantization is a process that has coefficients divided by a numeric value called the quantization step and rounds them to integers to reduce their bits. The original coefficients can not be perfectly recovered from the quantized ones because of the rounding error, i.e., distortions are introduced by quantization. As a result, an image after quantization can not be perfectly reconstructed either. Distortions can also be introduced by the transformation because of the limited precision of digital computers or the rounding of integer operations, which, however, can be ignored comparing to that caused by quantization.

Quantization, including Scalar Quantization (SQ) and Vector Quantization (VQ) [13]-[19], plays a very important role in lossy image compression. It is the primary contributor to high compression ratio, and likewise the major source of distortion. In [20], Watson et al. analyzed the DWT quantization errors and developed a quantization algorithm that is aimed to achieve visually lossless compression, but does not have the flexibility to achieve arbitrary visual quality. In [21], Liu et al. developed a quality constrained compression method for JPEG2000 that is optimized for the local profile of so called just-noticeable distortion (JND), which is similar to the distortion model in [20]. In [22], Nadenau et al. came up with a wavelet based color image compression that improved the precision of the contrast sensitive function (CSF), which is complicated though.

This work is based on our previous work in 2003 [23]. We propose a new quality metric called Weighted Normalized Mean Square Error of wavelet subbands (WNMSE), which is defined in terms of the wavelet coefficients and uses the sum of the weighted normalized mean square error of the coefficients in each wavelet subband to assess the quality of a compressed image. This metric is consistent with HVS as well as measures the post-compression quality of an image in real-time because of the simplicity of WNMSE. Taking advantage of WNMSE, we have

developed a novel compression algorithm called Quality Constrained Scalar Quantization (QCSQ) that is based on the relationship among the statistic features, quantization steps, and WNMSE value of the image. QCSQ can find the quantization steps for all the subbands efficiently for compressing the image to a desired visual quality measured by WNMSE.

The work is organized as follows. In Section 3.2, we briefly describe the DWT and define the notations that are used in the work. In Section 3.3, our new quality metric WNMSE is presented, and in Section 3.4, the innovative quality constrained quantization algorithm QCSQ is introduced. Experimental results are given in Section 3.5 to demonstrate the advantages of the new metric and compression methods. The work is concluded by Section 3.6. The detailed algorithm of QCSQ is given in Appendix A.

II. 2-D Wavelet Transform

Subband coding, which includes wavelet coding, was first introduced by Croisier et al. for speech coding in 1976 [24]. Ten years later, 2-D subband decomposition was applied to image coding by Woods and O'Neal [26]. With the advent of the wavelet theory, wavelet coding became the dominant subband coding. Figure 1 is the diagram of a single-level 2-D wavelet decomposition system, in which four wavelet subbands are generated from the input image and labeled as LL, LH, HL and HH, respectively, where L means low pass filtering and H means high pass. From Figure 1, one can see that subband LL is the result of two low pass filtering operations in both the horizontal and vertical directions, while subband LH is the result of a low pass filtering operation in the horizontal direction and a high pass filtering operation in the vertical, respectively; so forth and so on. A balanced multilevel subband decomposition system can be constructed by applying single-level decomposition systems to all the subbands of the previous level. The wavelet transform is the extreme form of an unbalanced subband decomposition because only the subband LL of the previous level is further decomposed.

For convenience, we label subband LL as subband a (average), HL as h (horizontally high pass and vertically low pass), LH as v (vertically high pass and horizontally low pass) and HH as d (both horizontally and vertically high pass). Figure 2 is a decomposed image after three levels of 2-D Haar wavelet transform. There are totally ten subbands which can be put into 3 groups according to the levels of transformation: level-1, level-2, and level-3, respectively. After the first transformation, we get four subbands of level-1: a_1, h_1, v_1 and d_1; after applying the second wavelet transformation to a_1, we get four subbands of level-2: a_2, h_2, v_2 and d_2; finally, we get four subbands of level-3: a_3, h_3, v_3 and d_3 by applying the last wavelet transformation to a_2. The same operation can continue by applying the 2-D wavelet transform to a_n, $n = 1, 2, 3, ...$, until a_n becomes a single coefficient. However, too many levels of transformation will not contribute to the efficiency of image compression, but only increase the cost of computation.

Besides its level of transformation l, another property of subband b_l is its frequency index f_{b_l}, where b is one of $\{a, h, v, d\}$. A wavelet subband is formed by letting the coefficients passing through a series of filters which includes high pass H_H and low pass H_L, each selectively picking appropriate frequency components. If we let

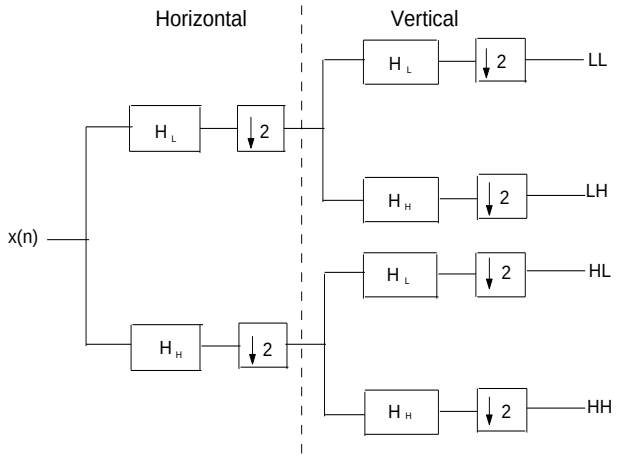


Fig. 1. The structure of a single-level 2-D subband decomposition system, where H_L represents a low pass filter and H_H represents a high pass filter.

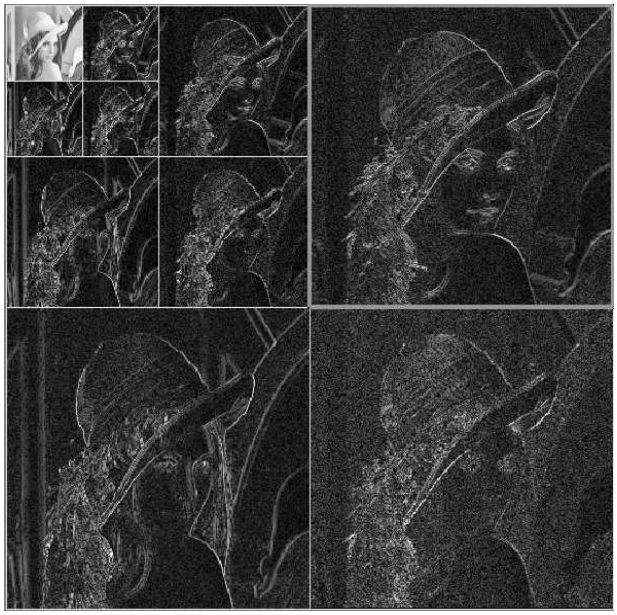


Fig. 2. The decomposed image of Lenna after three 2-D Haar wavelet transformations

the number of high pass filters that subband b_l passed through be NH_{b_l} and low pass filters NL_{b_l}, we define its frequency index as $f_{b_l} = NL_{b_l} - NH_{b_l}$. In the case above with $n = 3$, the frequency indexes of the ten subbands $\{d_1, v_1, h_1, d_2, v_2, h_2, d_3, v_3, h_3, a_3\}$ are $\{-2, 0, 0, 0, 2, 2, 2, 4, 4, 6\}$.

The main advantages of using DWT for image coding are:

1) Compared with DCT, the coefficients of DWT are well localized in not only the frequency, but also the spatial domains. This frequency-spatial localization property is highly desired for image compression [27].

2) DWT decomposes an image into spatially correlated subbands that hold different frequency components of the image. Each subband can be thought as a subset of the image with a different spatial resolution such that the visual quality and the compression ratio of the compressed image can be controlled by adjusting the distortions of different subbands [28].
3) Images coded by DWT do not have the problem of block artifacts which the DCT approach may suffer [29].
4) Compared with DCT, DWT has lower computation complexity, O(N) instead of O(NlogN) [30].

Xiong et al. claimed that, for still image compression, wavelet transform based coding systems outperform DCT by an order of 1 dB in PSNR [31]. One example of DWT's success is JPEG2000 where 2-D DWT is used instead of DCT.

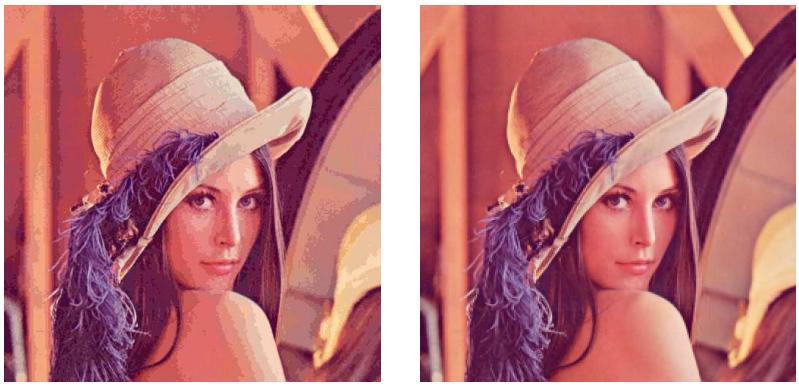
III. The New Quality Assessment Method

Human visual system takes in both frequency and spatial information following a filtering process, and different frequency portions of an image have different contributions to the visual quality. Distortions at different frequencies, even with the same magnitude, do not have the same impacts to the quality of the compressed image. We define the distortion in the spatial domain as the distance between the pixels of the original image and those of the distorted image, and the distortion in the frequency domain as the distance between the coefficients of the original image after transformation and those of the distorted image after transformation. For distortions in the spatial domain with the *same* magnitude, their corresponding distortions in the frequency domain are *combinations* of distortions of all the subbands. Although the distortion in the frequency domain is related to that in the spatial domain, given the spatial distortion, it is impossible to differentiate the contribution of each subband. An identical distortion index in the spatial domain may attribute to two compressed images which have radically different qualities. Figure 3 shows two reconstructed images with the same PSNR, among which the distortion of 3(a) is only from the a_1 subband while that of 3(b) is from the h_1, v_1 and d_1 subbands. We can see that, the quality of 3(a) is worse than that of 3(b) even though they have the same amount of distortion in the spatial domain. Since the spatial distortion is not a good indicator of the true quality for human eyes, an image quality metric which is consistent has to be developed in the frequency domain.

The 2-D wavelet transform decomposes an image into subbands that represent different frequency components of the image. Let $x_{b_l,i,j}$ denote a wavelet coefficient before compression and $y_{b_l,i,j}$ the coefficient after compression at position (i, j) in subband b_l. The distortion on this coefficient is $D = |x_{b_l,i,j} - y_{b_l,i,j}|$. In the remaining part of this section, we will analyze how the distortions from different subbands affect the quality of the reconstructed image. For convenience, our analysis is based on the example using Haar wavelet, but the conclusion is applicable to all types of wavelets.

Before we introduce the new quality metric, the following observations are in order.

1) The subbands with higher transformation levels hold more structural or global information, such as shape and luminance, than those with lower transformation levels. So the distortions from the subbands of higher transformation levels degrade the quality of an image more significantly. For example, each coefficient in a



(a) With distortion in only a_1 subband

(b) With distortion in h_1, v_1 and d_1 subbands

Fig. 3. Two reconstructed images with the same amount of spatial distortions: (a) only has distortion in a_1 subband while (b) has distortions in h_1, v_1 and d_1 subbands. The visual quality of (a) is much worse.

level-1 subband comes from four image pixels. If one coefficient has a distortion, it is very likely that those four pixels will all have distortion after reconstructing. Similarly, each coefficient in a level-2 subband comes from sixteen pixels and its distortion will affect those sixteen pixels in the reconstructed image. In a word, any distortion on a coefficient in a level-l subband will generate distortion on each of the 4^l pixels in the reconstructed image, and smaller distortions in a higher level subband may have more negative impact on the quality of an image than the larger ones in a lower level subband.

2) The subbands of lower frequency (larger frequency indexes) hold more structural or global information than those of higher frequency (smaller frequency index). Since the structural information plays a more important role in maintaining the fidelity of an image, a subband with larger frequency index has more visual impact than that with smaller frequency index. Figure 4 shows that the same amount, Normalized Mean Square Error (NMSE) = 10%, of distortion produced by subbands with nonidentical frequency indexes has different impact on image quality. Figure 4(a) only has distortion in subband a_3 ($f_{a_3} = 6$) while 4(b), 4(c) and 4(d) in h_3 ($f_{h_3} = 4$), v_3 ($f_{v_3} = 4$) and d_3 ($f_{d_3} = 2$), respectively. The quality of Figure 4(a) is the worst, 4(b) and 4(c) next, and 4(d) the best.

In light of the above discussion, we believe that a good IQM should be defined in the frequency domain in order to utilize this subband dependent feature. Our new quality metric chooses to use the weighted sum of normalized mean square errors of the coefficients in all the wavelet subbands as the quality metric of an image, which is called the Weighted Normalized Mean Square Error of wavelet subbands (WNMSE):

$$WNMSE_1 = \sqrt{4^{(L-1)} \times 2^{f_{a_L}/2}} \times NMSE_{a_L} + \sum_{b\in\{h,v,d\}} \sum_{l=1}^{L} \sqrt{4^{l-1} \times 2^{f_{b_l}/2}} \times NMSE_{b_l} \quad (1)$$

where $\sqrt{4^{l-1} \times 2^{f_{b_l}/2}}$ is the weight factor for subband b_l whose transformation level is l and frequency index f_{b_l},

(a) Distortion only in a_3 subband, $f_{a_3} = 6$

(b) Distortion only in h_3 subband, $f_{h_3} = 4$

(c) Distortion only in v_3 subband, $f_{v_3} = 4$

(d) Distortion only in d_3 subband, $f_{d_3} = 2$

Fig. 4. Four reconstructed images with the same amount of frequency distortions (NMSE equal to 10%): Figure (a) only has distortion in subband a_3, (b) in h_3, (c) in v_3, and (d) in d_3, respectively. The quality of (a) is the worst, (b) and (c) next, and (d) the best.

L is the highest transformation level, $NMSE_{b_l}$ is the NMSE of subband b_l, and

$$NMSE_{b_l} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{b_l,i,j} - y_{b_l,i,j})^2}{\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{b_l,i,j})^2} \tag{2}$$

where m is the number of pixels in the horizontal direction and n vertical. In the case that $\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{b_l,i,j})^2 = 0$, let $NMSE_{b_l} = 0$ if $\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{b_l,i,j} - y_{b_l,i,j})^2 = 0$, and $NMSE_{b_l} = 1$ otherwise. For convenience, we define $WNMSE$ as:

$$WNMSE = 20 \times \log 10 \frac{100}{WNMSE_1}. \tag{3}$$

In this way, a better quality image will have a higher value of WNMSE which is similar to PSNR and MSSIM.

In this equation, each $NMSE_{b_l}$ is calculated and weighted individually and separately, which reflects the

contribution of each subband to the total distortion. *NMSE*, instead of *MSE*, is used because the absolute amount of the distortion is not a good indicator of the contribution of a subband towards the overall quality loss. As discussed above, with the transformation level going up, the number of supporting pixels of a coefficient and its impact to the global structure both increase. By putting 4^{l-1} in the weight factor for subband b_l, its weight goes along with its level. Similarly, by putting $2^{f_{b_l}/2}$ in the weight factor, the impact of frequency is considered accordingly. A subband with higher transformation level and lower frequency will have larger weight. These weights loyally represent the contribution of each wavelet subband to the overall visual quality.

Unlike the conventional quality metrics, WNMSE evaluates the quality of an image in the wavelet domain, which possesses the following two advantages:

1) WNMSE is HVS optimized. Using the weighted contributions of different subbands in the wavelet domain, WNMSE does not simply evaluate the quality of an image by its total distortion, but treats subbands discriminatingly because different subbands have non-uniform impacts to visual quality. By using different weights, the contribution of each wavelet subband to the overall quality is considered accordingly. In this way, the impacts of distortions to both global structure and local details are more likely to be balanced, which leads to a more objective quality assessment.
2) WNMSE is real-time suitable. By defining WNMSE in the wavelet domain, the quality can be easily assessed during the process of compression. In contrast to those quality metrics in the spatial domain, WNMSE can measure the quality of an image right after quantization without a new computation in the spatial domain. Computation is thus more efficient, especially when iteration is necessary to adjust the quality of the image.

Our research shows that WNMSE is much more consistent with the results of MOS, compared with PSNR and MSSIM. WNMSE is thus a better quality indicator of an image by HVS. In addition, it enables us to link the two operations, quality assessment and quantization during compression, because both of them operate in the frequency domain. With the linkage established, accurate quality constrained compression becomes possible. The experimental results which compare the performance of WNMSE with that of PSNR and MSSIM are provided in Section 3.5.1.

IV. Quality Constrained Compression

Image compression is usually treated as a bit-rate constrained problem, i.e., compression ratio is on the top of consideration while quality is secondary. Since the features of images may vary significantly, image qualities can be different for the same bit-rate. Consequently, bit-rate constant compression is not always desired.

We call a compression method which prioritizes the quality *quality constrained compression*. Unfortunately, quality constrained compression has been difficult because of the following two reasons:

1) Quality assessment, such as PSNR and MSSIM, and image compression, such as DCT or DWT based, are pursued in the spatial and frequency domains, respectively, and there is no direct and simple link between them.
2) The reliability of current IQMs still have to be improved to satisfy the need of the quality assessment.

These two problems can be solved by using the new index WNMSE. From Equation (1), the WNMSE of a compressed image can be controlled if the distortion of each wavelet subband can be manipulated. This could be done through a brutal-force searching method, but an applicable solution has to be more efficient. Ideally, we want to be able to predict the distortion caused by a given quantization step. This appears to be a challenging task because it requires a highly accurate statistical description of the subband. Many efforts have been made to develop statistic models of wavelet coefficients and employ them in image compression. Unfortunately, they are often inaccurate in the modeling, and not easy to use [32]-[35]. From the discussion of the previous section, one can see that choosing of the step for a particular subband must be related to its contribution to the quality of the image. Large contributors should have less distortions, i.e., smaller steps. The question is who are the large contributors? We propose to predict the contribution of a subband by a set of features, and use these features to select the initial step and subsequently tune it to reach the desired quality. These features are *transformation level*, *frequency index*, *energy level*, *standard deviation*, and *complexity*, respectively. While the definitions of the transformation level and frequency characteristic have been described earlier and that of the standard deviation is trivial, the other two features are defined below.

1) Since the energy of a subband is calculated as the sum of the squares of each coefficient, it depends only on the absolute magnitude of each coefficient. So we use the absolute mean value m_{b_l} to represent the *energy level* of subband b_l.
$$m_{b_l} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}|x_{b_l,i,j}|}{n \times m} \tag{4}$$
where $x_{b_l,i,j}$ is a coefficient of subband b_l at position (i, j), and m and n is the dimensions of subband b_l.
2) At the first glance, the standard deviation σ_{b_l} of the wavelet coefficients in subband b_l is the only parameter needed to represent the complexity of the subband.
$$\sigma_{b_l} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{b_l,i,j} - \overline{x_{b_l}})^2}{n \times m - 1}} \tag{5}$$
where
$$\overline{x_{b_l}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} x_{b_l,i,j}}{n \times m}. \tag{6}$$
It is not enough because the energy levels of subbands could be different. For example, two subbands with identical standard deviations of 10, may have absolute means of 50 and 5, respectively. In this situation, the two subbands do not have the same complexity level. So we use the "relative" standard deviation $vm_{b_l} = \sigma_{b_l}/m_{b_l}$ (std/mean) to represent the *complexity* of subband b_l.

It is well known that the subbands of the wavelet transformation are projections of the original image to various resolutions, and their energy levels and complexities are related to each other. We can simply use the energy level and complexity of subband a_1, i.e., m_{a_1} and vm_{a_1}, to uniformly represent those of all the subbands.

The impact to the image quality by a particular step is affected by the five features just mentioned. It is not possible to deduct a quantitative relationship between the step and the features for a desired image quality, but it is

not difficult to understand the qualitative relationship between the two. Based on these observations, we introduce the following equation for defining the quantization step of subband b_l:

$$s_{b_l} = C_l \cdot V_{b_l} \tag{7}$$

where C_l is a variable that is only dependent on the transformation level l, and V_{b_l} is a variable whose value is derived from a function of σ_{b_l} while the function itself is determined by the other four features of subband b_l. Accordingly, to get a high compression ratio while satisfying a quality constrain, C_l and V_{b_l} can be determined using the following rules:

1) V_{b_l} should increase as m_{a_1} increases.
2) V_{b_l} should increase as vm_{a_1} increases.
3) V_{b_l} should decrease as f_{b_l} increases.
4) C_l should decrease as the transformation level increases.
5) V_{b_l} should be proportional to σ_{b_l}.
6) The quality and compression ratio of a compressed image can be tuned by adjusting its quantization steps to achieve an optimal result.

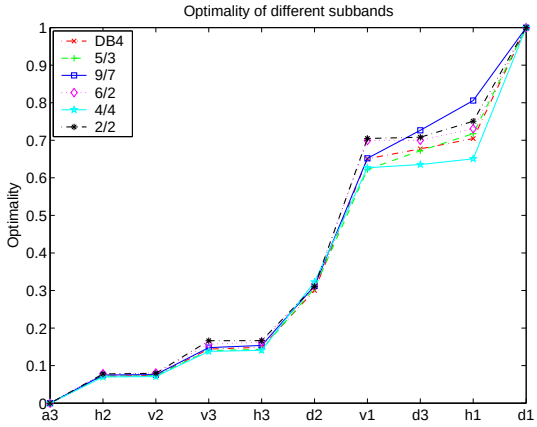
Using the rules just defined, a process has been found to search for quantization steps for compressing an image. This process is called Quality Constrained Scalar Quantization (QCSQ) which takes two steps: first, find the initial set of steps which is nearly optimal in the compression ratio with a uniform quality metric WNMSE ≈ 28, where 28 was empirically selected as the lower bound of an acceptable visual quality based on our observations. Secondly, tune the initial steps to increase the quality of an image to a desired value.

A. Find the Initial Set of steps

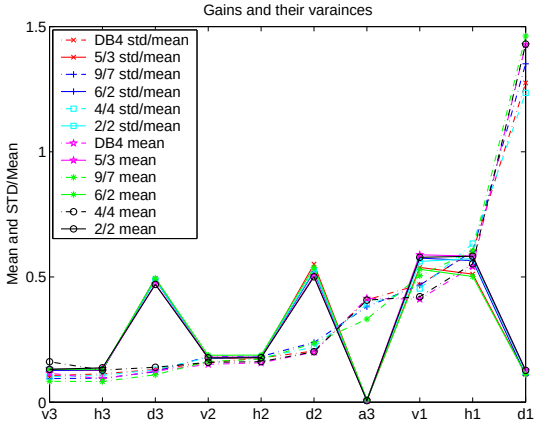
When calculating the WNMSE indexes, we multiply the NMSE of a subband b_l by $\sqrt{4^{l-1} \times 2^{f_{b_l}/2}}$, where 4^{l-1} is dependent on its transformation level and $2^{f_{b_l}/2}$ is dependent on its frequency. Here the dependence of step on the transformation level is reflected by defining the variable C_l = $4^{(L-l)}$. The impact of the frequency index is reflected by multiplying a factor $2^{-f_{b_l}/2}$ when calculating V_{b_l}. The detailed implementation of this algorithm is in Appendix VII-A.

B. Tune the Initial Set of steps

An image quantized by its initial set of steps only achieves the lower bound of visual quality. By further tuning its steps, one can improve the quality of the image to a desired level. Figure 5 shows how the variations of the steps of different subbands alternate the quality of images. We use two empirical parameters to evaluate the efficiency of the step tuning of a subband: quality gain and optimality. The quality gain of subband b_l is reduced from the quality improvements of images with different features by reducing the step of subband b_l by half. By optimality, we mean the ratio between the quality increment (ΔQ) and the compression ratio decrement (ΔR): (ΔQ) / (ΔR).



(a) The optimality level of different subbands when reducing their quantization steps by half to improve the quality of the image. Sorted as the ascending order of their optimality levels, the higher the better.



(b) The quality gains of different subbands when reducing their quantization steps by half to improve the quality of the image. Sorted as the ascending order of their normalized standard deviations, the lower the better.

Fig. 5. When reducing their quantization steps, each subband has different quality gains and different optimality levels. Some of them have higher optimality levels and lower invariance of quality gains, which can be used to adjust the quality of the compressed image.

Tuning Order	1	2	3	4	5	6	7	8
Subband	h_1	d_3	v_1	d_2	h_3	v_3	v_2	h_2
Quality Gain (Haar)	0.58	0.47	0.58	0.50	0.14	0.13	0.18	0.18
Quality Gain (5/3)	0.57	0.49	0.58	0.52	0.13	0.13	0.18	0.18
Quality Gain (9/7)	0.57	0.49	0.56	0.53	0.13	0.13	0.18	0.18
Quality Gain (DB4)	0.51	0.49	0.54	0.55	0.13	0.13	0.17	0.18
Quality Gain (4/4)	0.50	0.49	0.53	0.54	0.13	0.13	0.19	0.18
Quality Gain (6/2)	0.58	0.47	0.59	0.50	0.13	0.13	0.18	0.19

TABLE I

ORDER OF THE SUBBANDS FOR FINE-TUNING (STARTING FROM THE LEFT SIDE FIRST) AND THEIR PREDICTED QUALITY GAINS BY REDUCING THEIR QUANTIZATION STEPS BY HALF.

Since we want to maintain as high a compression ratio as possible when increasing the quality, ΔR should be as small as possible; therefore, the higher the ratio (ΔQ) / (ΔR) is, the higher the optimality level is.

Figure 5(a) shows the normalized optimality of each subband, which is sorted in the ascending order, and Figure 5(b) shows the magnitude and variance of the quality gain of each subband. To achieve accuracy, efficiency, and high compression ratio, only those subbands that have low quality gain variances, high quality gains, and high optimality values are used for quality tuning. Since the initial steps give the lower bound of the visual quality of an image, only the tuning for quality increase is considered. Combining the results of Figure 5(a) and Figure 5(b), the following rules of fine-tuning are obtained:

1) If there is more than one choice satisfying the quality requirement, choose the one which has the maximum compression ratio.
2) Tune the steps of the subbands with higher optimality first.
3) Tune only subbands whose quality gains are more than 0.1.
4) Tune only subbands whose variance of quality gains is less than 0.66.
5) Reduce the step by half when tuning it (because of the binary property of digital data).

The resulting order of tuning and the expected quality gain for each fine-tuning are listed in Table I, where the values shown are the average of 31 different images. We can see that the tuning orders are identical for all the wavelets and the quality gains show little difference. For a specific image, the quality gain may be slightly different, but the order of tuning is universally true.

Since WNMSE is defined in the wavelet domain, we can easily measure it after quantizing an image with the initial steps. Let the initial WNMSE be Q_0 and the objective WNMSE be Q, the difference is ΔQ = Q - Q_0. To increase the quality metric by ΔQ, we should tune the steps following the rules above. The detailed implementation of this algorithm is in Appendix VII-B.

V. Experimental Results

In this section, we first compare the quality assessment performance of WNMSE with that of PSNR and MSSIM, and then use an example to show how to achieve quality constrained compression using QCSQ. The Haar, DB4, 5/3 and 9/7 wavelets are used in our experiments to show the generalization of the algorithm.

A. Compare the Performance of WNMSE with PSNR and MSSIM

We used two experiments in this paper to compare the performance of WNMSE with that of PSNR and MSSIM. In each experiment, the objective quality indexes of the impaired images were measured in WNMSE, PSNR and MSSIM first. Then these objective IQMs were evaluated with regard to the subjective quality scores of those images. The subjective scores of the first experiment were obtained from the subjective image quality test we designed and executed, which will be described in details; the subjective scores of the second experiment were from the *LIVE Image Quality Assessment Database Release 2* [39].

The first subjective image quality test was done under controlled lab environment that complied with ITU-R BT.500-11 [36].

- Laboratory environment. The test was done in our lab that has normal indoor luminance. The lab only has windows on the east side of the walls. The monitors displaying images was put against the west wall, where the luminance is stable and low compared to the monitor.
- Monitors. We used two 19" CRT computer monitor to present the images. Their resolutions were set to 1024×768, with fairly high brightness and contrast ratio compared to the white wall behind it.

Score	Description
5.0	Perfect. The distortion is imperceptible
4.0	Good. The distortion is perceptible, but not annoying
3.0	Fair. The distortion is slightly annoying
2.0	Bad. The distortion is annoying
1.0	Very bad. The distortion is very annoying
0.0	Unidentifiable. The image is totally ruined

TABLE II
THE REFERENCE TABLE OF RANKING SCORES IN THE MOS TEST.

- Test materials. we used two sets of images, which includes twenty four and twenty five degraded images each. The impairments of the degraded images are either from compression with JPEG or JPEG2000, or various amounts of additive noise, including Gaussian, Speckle and Salt-pepper.
- Observers. These images are independently evaluated by 18 persons who come from different backgrounds. Two of them are considered as experts since they work in the image processing field, and the others are non-experts. All of the observers have normal visual acuity and normal color vision.
- Test procedure. The observer sits right in front of the monitor, with a view angle less than 30 degree. The distances from the eyes of observers to the screens were between two to three feet. Images are stored in a html file. The first page of the file is instructions that introduces the method of assessment, the types of impairment and the grading scale (Table II). From the second page, each page has four images, including the original image and three degraded images with different types of impairment. So the observer can compare the images under exactly same conditions. The order of the degraded images are random, i.e., not ordered or grouped by their impairment.
- Scoring. To evaluate the quality of a degraded image, a person compared it with the original one and gave it a score using Table II as a reference. Each score is from 0.0 to 5.0 including a decimal fraction of one digit. The average score of an image is taken as the Mean Opinion Score (MOS) of it.

We used four popular criteria to evaluate the accuracy of the quality metrics. Among them, the first three are the standard criteria used by the Video Quality Expert Group (VQEG) [37], and the fourth is straight "Sum of Squared Errors". In the following definitions, "X" can be "PSNR", "MSSIM" or "WNMSE", X_i is the normalized "X" and MOS_i the normalized MOS of the ith image, and n is the number of images.

1) Pearson Linear Correlation Coefficient ($PLCC$) is used to evaluate the *accuracy* of an IQM. The $PLCC$ of "X" with regard to MOS is

$$PLCC_X = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(MOS_i - \overline{MOS})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(MOS_i - \overline{MOS})^2}}. \quad (8)$$

The larger the $PLCC_X$ is, the more accurate X will be with regard to MOS.

	Evaluation Metrics	WNMSE (Haar)	WNMSE (DB4)	WNMSE (5/3)	WNMSE (9/7)	PSNR	MSSIM
Image	SSE	0.606	0.750	0.873	0.791	1.347	1.453
series	PLCC	0.7604	0.7548	0.7095	7380	0.6044	0.5558
A	SROCC	0.7457	0.7439	0.6896	0.7300	0.5809	0.5326
n=24	OR	0.0420	0.1250	0.1250	0.1250	0.1250	0.1250
Image	SSE	0.549	0.589	0.693	0.615	0.842	1.089
series	PLCC	0.8008	0.7888	0.7584	0.7835	0.7040	0.6078
B	SROCC	0.8485	0.8354	0.7632	0.8285	0.7573	0.6562
n = 25	OR	0.0800	0.0800	0.0800	0.0800	0.0800	0.0800

TABLE III

COMPARING THE OVERALL PERFORMANCES OF WNMSE (IMPLEMENTED WITH FOUR DIFFERENT WAVELET TRANSFORMS), PSNR AND MSSIM MEASURED BY FOUR EVALUATION METRICS WITH REGARD TO MOS ON IMAGE "LENNA" AND "PEPPERS". FOR PLCC AND SROCC METRICS, THE LARGER VALUES MEAN BETTER PERFORMANCE. FOR SSE AND OR METRICS, THE SMALLER VALUES MEAN BETTER PERFORMANCE.

2) Spearman Rank Order Correlation Coefficient ($SROCC$) is used to evaluate the *monotonicity* of an IQM. The $SROCC$ of "X" with regard to MOS is

$$SROCC_X = 1 - \frac{6\sum_{i=1}^{n}(d_i)^2}{n^3 - n} \tag{9}$$

where d_i is the difference between each rank of corresponding values of X and MOS. The larger the $SROCC_X$ is, the better monotonicity X will have with regard to MOS.

3) Outlier Ratio (OR) is used to evaluate the *consistence* of an IQM. The OR of "X" OR_X with regard to MOS is defined as the number of outliers divided by n, where twice of the standard error of MOS was used as the threshold for defining outliers. From the definition we can see that the smaller the OR_X is, the more consistent X will be with regard to MOS.

4) Sum of Squared Errors (SSE). We also use SSE of an index with regard to MOS to measure the overall performance of it. The values of every index are normalized so that they all have the same range between [0, 1].

$$SSE_X = \sum_{i=1}^{n}(X_i - MOS_i)^2. \tag{10}$$

The smaller the SSE_X is, the better X will perform with regard to MOS.

The evaluation results of the quality metrics are listed in Table III. WNMSE implemented with four DWTs are compared with PSNR and MSSIM. Looking at the table, we can see that all the WNMSEs outperform PSNR and MSSIM in every aspect while the Haar WNMSE is the best.

To test the generality of WNMSE's superiority over other IQMs, we conducted the second experiment. It measured the WNMSE, PSNR and MSSIM values of the images in the *LIVE Image Quality Assessment Database Release 2* and used the subjective quality scores the authors provided. The evaluation is the same as that in the first experiment

Evaluation Metrics	WNMSE	PSNR	MSSIM
SSE	4.47	7.01	15.72
PLCC	0.862	0.803	0.757
SROCC	0.887	0.812	0.879
OR	0	0	0.0155

TABLE IV

COMPARING THE OVERALL PERFORMANCES OF WNMSE, PSNR AND MSSIM MEASURED BY FOUR EVALUATION METRICS WITH REGARD TO DMOS SCORES ON IMAGES IN [39]. FOR PLCC AND SROCC METRICS, THE LARGER VALUES MEAN BETTER PERFORMANCE. FOR SSE AND OR METRICS, THE SMALLER VALUES MEAN BETTER PERFORMANCE.

Image	WNMSE Haar	WNMSE DB4	WNMSE 5/3	WNMSE 9/7	PSNR	MSSIM	MOS
b)	19.16	16.99	14.48	15.87	23.06	0.496	3.26
c)	17.28	15.07	12.54	13.93	23.10	0.498	3.07
d)	13.08	10.80	10.03	10.34	24.37	0.605	0.85

TABLE V

QUALITY INDEXES OF THE DISTORTED IMAGES OF LENNA IN FIGURE 6

except that the subjective scores here are in DMOS (Difference of Mean Opinion Score) [2] that needs to be treated differently than MOS. The evaluation results are listed in Table IV. WNMSE implemented with Haar wavelet was compared with PSNR and MSSIM. Looking at the table, we can see that WNMSE still outperformed PSNR and MSSIM in every aspect.

Figures 6 and 7 show how WNMSE outperforms both PSNR and MSSIM. Image (a) is the original image and the other three are degraded by Gaussian noise, Salt-Pepper noise, and JPEG2000 compression, respectively, which are listed in the descending order of their MOS values. The measured quality metrics are listed in Table V and Table VI. From the measured quality metrics, we can see that the WNMSE indexes are in the same order as those of MOS, while PSNR and MSSIM give the reverse results. This proves that WNMSE functions more like human

Image	WNMSE Haar	WNMSE DB4	WNMSE 5/3	WNMSE 9/7	PSNR	MSSIM	MOS
b)	18.90	16.45	13.76	15.24	23.04	0.542	3.25
c)	18.60	16.18	13.16	14.80	24.01	0.601	3.24
d)	17.64	15.16	12.60	14.20	24.35	0.691	1.02

TABLE VI

QUALITY INDEXES OF THE DISTORTED IMAGES OF PEPPERS IN FIGURE 7

(a) The original ”Lenna” image. (b) Gaussian noise added.

(c) Salt-Pepper noise added. (d) Compressed by JPEG2000.

Fig. 6. WNMSE outperforms both PSNR and MSSIM for Lenna images. Image qualities measured by MOS indexes are in the order as from (b) to (d), while (b) is the best. WNMSE gives the same order as MOS, but PSNR and MSSIM give the reverse order (Table V).

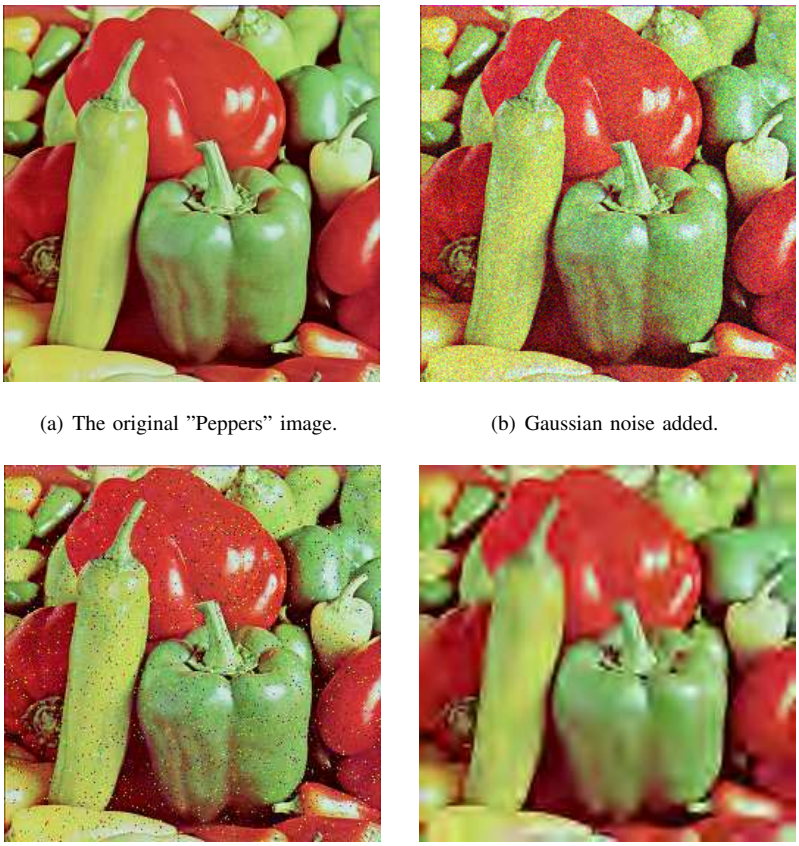
eyes.

B. QCSQ Examples

In this example, we use 9/7 wavelet which has been used in the JPEG2000 standard for lossy compression. We first apply three levels of 2-D 9/7 wavelet transform to an image, and then use QCSQ to determine the quantization steps for all its wavelet subbands. After quantization, we use Zig-zag sorting followed by Stack-run [38] and Arithmetic coding to code quantized bits.

Six images are used in the experiment, where the target quality index is chosen as Q = 30 in WNMSE with an acceptable error of 0.3. So the final quality metrics of the six images should be between 29.7 and 30.3 in WNMSE. 0.3 was chosen as the acceptable error because it is small enough (1% of the target index 30) in value, and makes little visual difference.

1) Find the initial steps and compute the initial quality metrics. According to our algorithm, an image quantized by its initial steps should have an initial quality index $Q_0 = 28$ measured in WNMSE. The results are listed



(a) The original "Peppers" image. (b) Gaussian noise added.

(c) Salt-Pepper noise added. (d) Compressed by JPEG2000.

Fig. 7. WNMSE outperforms both PSNR and MSSIM for Peppers images. Image qualities measured by MOS indexes are in the order as from (b) to (d), while (b) is the best. WNMSE gives the same order as MOS, but PSNR and MSSIM give the reverse order (Table VI).

in Table VII, from which we can see that the initial quality indexes of all the other five images are distributed closely around 28.0 except Mige171 which has a WNMSE value of 28.31.

2) Tune the steps. We know that $\Delta Q = Q - Q_0 \approx 30 - 28 = 2$ for the other five images and the sum of the quality gains of the first four most optimal subbands: h_1, d_3, v_1 and d_2 (Table I), are 0.57 + 0.49 + 0.56 + 0.53 = 2.15. So we first reduce the steps of these four subbands by half for the five images. As for Mige171 whose $\Delta Q = 1.69$, we only need to reduce the steps of the first three subbands: h_1, d_3 and v_1 which will give a quality gain of 1.62. The resulting quality metrics and compression ratio are listed in Table VIII. We can see that the WNMSEs of the five images (Lenna, Lethal, Tree, Mige171 and Building) already fall into the desired range. For the Peppers image, the WNMSE is a little too high, which will cause unnecessary loss in compression ratio. If we recover the steps of subband d_2 to the initial setting, its predicted quality metric is $30.70 - 0.53 = 30.17$ that is in the desired range. The measured quality metric after tuning is 30.14 that is only slightly different from the predicted value. Table IX lists the final results, along with the compression

Image	s_{a_3}	s_{h_3}	s_{v_3}	s_{d_3}	s_{h_2}	s_{v_2}	s_{d_2}	s_{h_1}	s_{v_1}	s_{d_1}	WNMSE	bpp
Lenna	42	8	15	16	13	22	22	80	128	128	28.05	0.56
Lethal	45	16	20	21	24	32	30	176	144	128	27.97	0.53
Peppers	46	16	21	21	27	31	30	176	192	224	28.03	0.67
Tree	48	13	10	14	22	18	28	160	112	128	28.08	0.57
Mige171	63	24	18	19	35	30	32	192	176	128	28.31	0.47
Building	46	22	14	17	39	29	32	256	192	160	27.86	0.62

TABLE VII

INITIAL QUANTIZATION STEPS AND COMPRESSION RESULTS BEFORE FINE-TUNING

Image	s_{a_3}	s_{h_3}	s_{v_3}	s_{d_3}	s_{h_2}	s_{v_2}	s_{d_2}	s_{h_1}	s_{v_1}	s_{d_1}	WNMSE	bpp
Lenna	42	8	15	8	13	22	11	40	64	128	29.97	0.66
Lethal	45	16	20	11	24	32	15	88	72	128	30.17	0.62
Peppers	46	16	21	11	27	31	15	88	96	224	30.70	0.78
Tree	48	13	10	7	22	18	14	80	56	128	30.21	0.65
Mige171	63	24	18	10	35	30	32	96	88	128	30.24	0.52
Building	46	22	14	9	39	29	16	128	96	160	29.83	0.76

TABLE VIII

INTERMEDIATE QUANTIZATION STEPS AND COMPRESSION RESULTS OF FINE-TUNING

results of JPEG2000, which also compressed the images to WNMSE≈30. Figure 8 shows the compressed images.

The experimental results have shown that the desired quality metric is achieved with no or only one iteration. QCSQ was designed to demonstrate the application of WNMSE in quality constrained compression. Compared with JPEG2000, QCSQ is very simple and not compression ratio optimized. Subjective evaluation showed that the

Image	s_{a_3}	s_{h_3}	s_{v_3}	s_{d_3}	s_{h_2}	s_{v_2}	s_{d_2}	s_{h_1}	s_{v_1}	s_{d_1}	WNMSE	bpp	J2K WNMSE	J2K bpp
Lenna	42	8	15	8	13	22	11	40	64	128	29.97	0.66	29.98	0.43
Lethal	45	16	20	11	24	32	15	88	72	128	30.17	0.62	30.22	0.55
Peppers	46	16	21	11	27	31	30	88	96	224	30.14	75	29.91	0.71
Tree	48	13	10	7	22	18	14	80	56	128	30.21	0.65	30.01	0.58
Mige171	63	24	18	10	35	30	32	96	88	128	30.24	0.52	29.97	0.42
Building	46	22	14	9	39	29	16	128	96	160	29.83	0.76	30.01	0.79

TABLE IX

FINAL QUANTIZATION STEPS AND COMPRESSION RESULTS OF FINE-TUNING. THE LAST TWO COLUMNS LIST THE COMPRESSION RESULTS OF JPEG2000 (J2K), WHICH ALSO COMPRESSED THE IMAGES TO WNMSE≈30.

(a) Lenna: WNMSE = 29.97.

(b) Tree: WNMSE = 30.17.

(c) Mige171: WNMSE = 30.14.

(d) Building: WNMSE = 30.21.

(e) Peppers: WNMSE = 30.24.

(f) Lethalweapon: WNMSE = 29.83.

Fig. 8. Images compressed by QCSQ to the target quality: WNMSE $\approx$ 30.00.

images compressed by QCSQ and JPEG2000 had the same visual qualities when their WNMSE values were the same.

VI. Conclusions

In this work, we have proposed a new quality metric WNMSE and an innovative quantization algorithm QCSQ. WNMSE uses the weighted sum of the normalized mean square errors of wavelet coefficients to assess the quality of an image. According to the concepts of HVS, the weight for each subband is chosen to reflect its perceptual impact on the image, which measures the distortions in the global structure and local details of an image in a more balanced way automatically. Because WNMSE is defined in the wavelet domain, it can be calculated in the middle of compression without reconstructing the image. Furthermore, it facilitates the link between the quantization steps and the quality metric. Our experiments show that WNMSE has better performance than both the legacy PSNR and the well referenced new IQM SSIM.

The features of a subband can be represented by its transformation level, frequency, energy, standard deviation, and complexity, which alternate the effect of the quantization step to the WNMSE of a compressed image. Based on the analysis of the relationship among the subband features, steps, and WNMSE values, we have invented a quality constrained compression algorithm QCSQ which can identify the quantization step for every subband of an image. With these steps, the image can be compressed to a desired visual quality measured by WNMSE.

This work shows that, by developing the quality metric and the quantization algorithm in the same wavelet domain, we have made the quality constrained image compression possible, while pushing the compression ratio as high as possible.

VII. Appendix: QCSQ Algorithm Implementation Details

A. Find the Initial Set of Step-Sizes

$s_{b_l} = C_l \cdot V_{b_l}$

1) Find the initial step-size for subband b_l in the level-l, $x \in \{h, v, d\}$ and $l \in \{1, 2, ..., L\}$:

C_l = $4^{(L-l)}$;

if $vm_{a_1} > (0.7 + 0.1 \times l)$

V_{b_l} = ceil(σ_{b_l}) $\times$ $2^{-f_{b_l}/2}$;

else if $vm_{a_1} < 0.2$

V_{b_l} = floor(σ_{b_l}) $\times$ $2^{-f_{b_l}/2}$;

else if $vm_{a_1} > 0.6$

if ($m_{a_1} > 96$)

V_{b_l} = ceil(σ_{b_l}) $\times$ $2^{-f_{b_l}/2}$;

else

V_{b_l} = round(σ_{b_l}) $\times$ $2^{-f_{b_l}/2}$;

end

else if ($m_{a_1} > 96$)

V_{b_l} = round(σ_{b_l}) $\times$ $2^{-f_{b_l}/2}$;

else

$V_{b_l} = \text{floor}(\sigma_{b_l}) \times 2^{-f_{b_l}/2}$;

end.

$s_{b_l} = C_l \times V_{b_l}$;

if $s_{b_l} < 1$

$s_{b_l} = 1$;

else if $s_{b_l} > 256$

$s_{b_l} = 256$;

end.

2) Find the initial step-size for subband a_L:

$C_L = 4^{(L-L)} = 1$;

if $m_{a_1} < (160 - 32 \times L)$

$V_{a_L} = \text{floor}(\sigma_{a_L}) \times 2^{(NH_{a_L} - NL_{a_L})/2}$;

else if $vm_{a_1} > (0.7 + 0.1 \times l)$

$V_{a_L} = \text{ceil}(\sigma_{a_L}) \times 2^{(NH_{a_L} - NL_{a_L})/2}$;

else if $vm_{a_1} < 0.2$

$V_{a_L} = \text{floor}(\sigma_{a_L}) \times 2^{(NH_{a_L} - NL_{a_L})/2}$;

else if $vm_{a_1} > 0.6$

if ($m_{a_1} > 96$)

$V_{a_L} = \text{ceil}(\sigma_{a_L}) \times 2^{(NH_{a_L} - NL_{a_L})/2}$;

else

$V_{a_L} = \text{round}(\sigma_{a_L}) \times 2^{(NH_{a_L} - NL_{a_L})/2}$;

end

else if ($m_{a_1} > 96$)

$V_{a_L} = \text{round}(\sigma_{a_L}) \times 2^{(NH_{a_L} - NL_{a_L})/2}$;

else

$V_{a_L} = \text{floor}(\sigma_{a_L}) \times 2^{(NH_{a_L} - NL_{a_L})/2}$;

end.

$s_{a_L} = C_L \times V_{a_L}$;

if $s_{a_L} < 1$

$s_{a_L} = 1$;

else if $s_{a_L} > 256$

$s_{a_L} = 256$;

end.

The specific values used in this algorithm are first chosen according to the rules described above, and finally determined after adjustments using experiments. Our experimental results show that these values are independent of wavelets used and suitable for all kinds of natural images.

B. Tune the Initial Quantization Steps

The initial set of quantization steps has quantized the image to a WNMSE equal to Q_0, and the objective WNMSE is Q. The difference is ΔQ = Q - Q_0. Assuming that subband α has the highest optimality opt_α and subband β has the second highest optimality opt_β, and their quantization steps, the averages of the quality gains, and the standard deviations of the quality gains are (s_α, m_α, std_α) and (s_β, m_β, std_β), respectively. The error threshold is δ, that is, we call it a successful tuning if the difference between the achieved and the target quality indexes is less then δ.

The process includes three iterative steps:

1) Adjust quantization steps to improve image quality.

   if $\Delta Q > m_\alpha$

   reduce s_α by half;

   ΔQ = ΔQ - m_α;

   else if m_β / $opt_\beta > m_\alpha$ / opt_α

   reduce s_α by half;

   ΔQ = ΔQ - m_α;

   else

   reduce s_β by half;

   ΔQ = ΔQ - m_β;

   end.

2) Check whether the target quality is achieved. If not, go back to step 1); if yes, go to step 3).

   if $\Delta Q < 0$

   done;

   else

   $\alpha \leftarrow$ the subband (α or β) whose quantization step was not modified;

   $\beta \leftarrow$ the subband whose optimality level is next to β;

   repeat 1;

   end.

3) Calculate the current predicted quality metric Q'. If it is too big compared with Q, tune it down; if it is too small, go back to step 1).

   while $Q' - Q > \delta$

   recover the quantization step of the subband x that was the last being modified;

   $Q' = Q' - m_x$, where m_x is the average quality gain of the subband x;

end.

if $Q - Q' < \delta$

$\alpha \leftarrow$ the most optimal subband among those whose quantization step was never modified;

$\beta \leftarrow$ the second most optimal subband among those whose quantization step was never modified;

repeat 1;

end.

REFERENCES

[1] T. Betchaku, N. Sato and H. Murakami, "Subjective evaluation methods of facsimile image quality," in *Proceedings of IEEE International Conference on Communications*, vol. 2, pp. 966-970, May 1993.

[2] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, pp. 3441-3452, November 2006.

[3] B. Girod, "What's wrong with mean-squared error," in *Digital Images and Human Vision,* pp. 207-220, the MIT press, 1993.

[4] R. Dosselmann and X. Yang, "Existing and Emerging Image Quality Metrics," in *Proceeding of CCECE/CCGEI*, pp. 1906-1913, May 2005.

[5] N Damera-Venkata T. D. Kite, W. S. Geisler, B.L. Evans and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Transactions on Image Processing*, vol. 9, pp. 636-650, April 2000.

[6] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communication*, vol. 43, pp. 2959-2965, December 1995.

[7] Private Communication, Nancy Obuchowski to Kim Powell, "Results of preference studies of compressed diagnostic breast images," Technical Report, *Biomedical Engineering, Cleveland Clinic Foundation*, February 3, 2000.

[8] Z. Wang, A. C. Bovik, H.R. Sheikh and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600-612, April 2004.

[9] A. Shnayderman, A. Gusev, and A. M. Eskicioglu, "An svd-based grayscale image quality measure for local and global assessment," *IEEE Transactions on Image Processing*, vol. 15, pp. 422-429, February 2006

[10] A. B. Watson, J. Hu, and J. F. McGowan III, "DVQ: A Digital Video Quality Metric Based on Human Vision," *Journal of Electric Imaging*, vol. 10, pp. 20-29, January 2001.

[11] M. Sendashonga and F. Labeau, "Low complexity image quality assessment using frequency domain transforms," in *Proceeding of IEEE International Conference on Image Processing*, pp. 385-388, October 2006.

[12] R. L. De Valois and K. K. De Valois, Spatial vision, *Oxford University Press*, 1988.

[13] K. Bao and X. Xia, "Image compression using a new discrete multiwavelet transform and a new embedded vector quantization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, pp. 833-842, September 2000.

[14] N. B. Karayiannis, P. Pai and N. Zervos, "Image compression based on fuzzy algorithms for learning vector quantization and wavelet image decomposition," *IEEE Transactions on Image Processing*, vol. 7, pp. 1223-1230, August 1998.

[15] S. Kasaei, M. Deriche and B. Boashash, "A novel fingerprint image compression technique using wavelets packets and pyramid lattice vector quantization," *IEEE Transactions on Image Processing*, vol. 11, pp. 1365-1378, December 2002.

[16] Y. Huh, J. J. Hwang and K. R. Rao, "Block wavelet transform coding of images using classified vector quantization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, pp. 63-67, February 1995.

[17] P.C. Cosman, R. M. Gray and M. Vetterli, "Vector quantization of image subbands: a survey," *IEEE Transactions on Image Processing*, vol. 5, pp. 202-225, February 1996.

[18] X. Wang, L. Chang, M. K. Mandal and S. Pancharathen, "Wavelet-based image coding using nonlinear interpolative vector quantization," *IEEE Transactions on Image Processing*, vol. 5, pp. 518-522, March 1996.

[19] E. A. B. da Silva, D. G. Sampson and M. Ghanbari, "Super high definition image coding using wavelet vector quantization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, pp. 399-406, August 1996.

[20] A. B. Watson, G. Y. Yang, J. A. Solomon and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Transactions on Image Processing*, Vol. 6, pp. 1164-1175, August 1997.

[21] Z. Liu, L. J. Karam and A. B. Watson, "JPEG2000 encoding with perceptual distortion control," *IEEE Transactions on Image Processing*, vol. 15, pp. 1763-1778, July 2006.

[22] M. J. Nadenau, J. Reichel and M. Kunt, "Wavelet-based color image compression: exploiting the contrast sensitivity function," *IEEE Transactions on Image Processing*, Vol. 12, No. 1, pp. 58-70, January 2003.

[23] Z. Gao and Y. F. Zheng, "Variable quantization in subbands for optimal compression using wavelet transform", in *Proceeding of SCI*, July 2003.

[24] A. Croisier, D. Esteban and C. Galland, "Perfect channel splitting by use of interpolation/decimation/tree decomposition tichniques," in *Proceeding of ICIS*, pp. 443-446, August 1976.

[25] J. W. Woods and S. D. O'Neil, "Subband coding of images," *IEEE Transactions on Communication*, vol. COM-31, pp. 532-540, April 1983.

[26] J. W. Woods and S. D. O'Neil, "Subband coding of images," *IEEE Transactions on Communication*, vol. COM-31, pp. 532-540, April 1983.

[27] A. S. Lewis and G. Knowles, "Image compression using the 2-D wavelet transform," *IEEE Transactions on Image Processing*, vol. 9, pp. 244-250, April 1992.

[28] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Transactions on Image Processing*, vol. 1, pp. 205-220, April 1992.

[29] H. Jozawa, H. Watanabe and S. Singhal, "Interframe video coding using overlapped motion compensation and perfect reconstruction filter banks," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 649-652, March 1992.

[30] H. Guo and C. Burrus, "Wavelet Transform based Fast Approximate Fourier Transform," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1973-1976, April 1997.

[31] Z. Xiong, K. Ramchandran, M. T. Orchard and Y. Zhang, "A comparative study of DCT-based and wavelet-based image coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 692-695, August 1999.

[32] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 8, pp. 1688-1701, December 1999.

[33] J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Transactions on Image Processing*, vol. 10, pp. 1647-1658, November 2001.

[34] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, vol. 6, pp. 300-303, December 1999.

[35] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, "Image coding using wavelet transform," *IEEE Transactions on Image Processing*, vol. 1, pp. 205-220, April 1992.

[36] "RECOMMENDATION ITU-R BT.500-11: Methodology for the subjective assessment of the quality of television pictures," *http://www.itu.int*.

[37] "Video Quality Expert Group, Final Report from the Video Quality Expert Group on The Validation of Objective Models of Video Quality Assessment," *http://www.vqeg.org*, March 2000.

[38] M. J. Tsai, J. D. Villasenor and F. Chen, "Stack-run image coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, pp. 519-521, October 1996.

[39] H. R. Sheikh, Z. Wang, L. Cormack and A. C. Bovik, "LIVE Image Quality Assessment Database Release 2," *http://live.ece.utexas.edu/research/quality*.