UNIVERSITY OF JOENSUU
DEPARTMENT OF COMPUTER SCIENCE
Report Series A

Blockwise distortion measure for statistical and structural errors in digital images

Pasi Fränti

Report A-1996-8
(revised: 10.4.1997)

ACM I.4.2
UDK 519.712
ISSN 0789-7316
ISBN 951-708-485-4

Blockwise distortion measure for statistical and structural errors in digital images

(to appear in *Signal Processing: Image Communication*, issue 6/97)

Pasi Fränti

Department of Computer Science, University of Joensuu, P.O. Box 111, FIN-80101 Joensuu, FINLAND
Email: franti@cs.joensuu.fi; phone: +358 13 251 3104; fax: + 358 13 251 3290

Abstract

A blockwise distortion measure is proposed for evaluating the visual quality of compressed images. The proposed measure calculates quantitatively how well important visual properties have been preserved in the distorted image. The method consists of three quality factors detecting contrast errors, structural errors, and quantization errors. The proposed method outperforms PQS for a set of test images, and is much simpler to implement. The method should also be applicable to color images; properties like color richness and saturation are captured by the quantization and contrast measures respectively.

Key words: Visual quality, quality assessment, distortion measures, objective measures, image compression.

1. Introduction

The key issue in lossy image compression is how to measure the *distortion* caused by compression. *Pixelwise measures* such as *mean square error* (MSE) are widely used but unfortunately they do not coincide with the visual quality of the images very well [5, 6]. Pixelwise measures are adequate distortion indicators for random errors but cannot detect reliably typical compression artifacts such as *blockiness*, *blurring* or *jaggedness of edges*.

Several attempts have been made to analyze the spatial dependencies of pixelwise differences [9, 11, 15]. By this way it is possible to detect *correlated errors* [15] and *blockiness* [11], for example. The pixelwise differences, however, are insufficient for measuring distortion like *contrast errors* or *quantization effect*. Instead, the original pixel values should be used in the distortion measurement.

Some of the methods in literature are tailored for JPEG artifacts [9], or try to capture only certain distortion types eg. blockiness in [11]. Probably the most thorough attempt for measuring the overall distortion is known as the *picture quality scale* (PQS) [15], but also it suffers from the drawback of operating with pixelwise differences. It is also designed using prior knowledge of the current compression algorithms. Thus, it might perform poorly in case of compression artifacts that were not taken into account in the design process.

Digital image archives are widely used by newspaper editorials. The images will be object of zooming, color adjustment and other manipulation techniques which may not even be known today. Therefore the distortion measure should be independent from the factors such as compression method used, basic image processing operations like color adjustment, and viewing distance. For example, if the viewing distance is assumed to be fixed in the quality measurement it is possible to overlook certain distortion types that would be revealed after the image is zoomed.

In the present paper we propose a new blockwise distortion measure. The proposed measure is general in the sense that it measures how well important visual properties have been preserved in the distorted image instead of detecting predefined distortion types. The original motivation for designing the measure are compression artifacts. The usage of the proposed measure, however, should not be limited to compressed images only but in principle, it should be applicable to all kinds of distortion, even those whose source is not known.

In the proposed measure the image is processed by a 3×3 *sliding window* so that its center pixel hits once each pixel in the image. In each block, the three visual properties are measured: *contrast*, *number of different gray-levels*, and *spatial structure*. The distortion in each quantity is quantitatively measured and their average values over the entire image are determined. The final distortion value is summarized into a single value which is a weighted sum of these three quality factors. A preliminary version of the method was reported in [7].

The proposed measure should consider three different aspects: the properties of the *human visual system* (HVS), the quality factors (visual properties) that are measured, and the weighting of the different factors. The focus is here on the design of the quality factors. We should also not ignore the fact that the measure should be relatively easy to implement and use. Thus, the model is kept as simple as possible without too much compromising the other goals.

2. Related work

A survey of pixelwise measures is given in [5] and their correlation with the visual quality is studied in [6]. It was concluded that an evaluation across different compression techniques is not possible by pixelwise measures.

Various aspects of HVS can be utilized in the distortion measurement by calculating the visibility of the artifacts due to the local surrounding of the pixels [3, 11, 15, 19]. Two aspects of HVS are usually considered: *brightness* and *spatial frequency sensitivity*. The former refers to the non-linear sensitivity of human eye to the background luminance level. The spatial frequency sensitivity means that human eye is less sensitive to artifacts in areas of high activity. The activity can be measured by the total energy of spatial frequencies, which can be calculated by the *windowed Fourier transform* for example [11]. The use of HVS properties does not remove the fundamental deficiency of the pixelwise measures, but any realistic measure should analyze the spatial dependencies of the pixels also.

Several methods operate on the pixelwise differences (*error map*) by analyzing their spatial dependencies. For example, blocking artifacts are detected in [11] by analyzing the vertical and horizontal structures separately in the error map. The vertical artifacts, for example, are isolated by a highpass filtering along the rows followed by a lowpass filtering along the columns to reduce the effect of random noises. A measurement of the typical JPEG artifacts such as blockiness and *mosquito noise* is performed in [9]. The mosquito noise is measured around the contours and blockiness in the flat areas; both by calculating the spatial differences in the error map and by using masking due to HVS.

In [20] the pixelwise error values are classified according to the edges structures; *errors on own edges* (edges in the original image), *errors on false edges* (edges in the distorted image), and other errors. The hypothesis is that different types of artifacts appear only in certain classes, eg. blurring on own edges, blocking and contouring on false edges. In [13] the distortion in the spatial structure is calculated by measuring the pixel values relative to their neighboring pixels.

Much of work has been put on the picture quality scale (PQS), which has recently been summarized in [15]. PQS operates error images that are masked according to the properties of HVS. Five separate quality factors (F_1 to F_5) are measured and the overall quality of the image is determined by their weighted sum. The factors F_1 and F_2 measure random errors, F_3 blocking effect across the block boundaries (assuming 8×8 block size), F_4 correlated errors, and F_5 all errors near the high contrast transitions (edges). All these factors are calculated on the basis of the pixelwise differences.

A blockwise approach has been proposed in [10]. The images (original and distorted) are divided by quadtree segmentation and from each block the distortion in mean value and in standard deviation are measured. The main drawbacks of the method are the non-overlapping blocks of the quadtree segmentation, and that only statistical errors are detected. The method also classifies the error values according to the block size. On the basis of the classification the distortion values are graphically illustrated but the interpretation of the so-called *Hosaka plots* is left to the observer.

3. Blockwise distortion measure

Here we assume that the important visual properties of digital images are *spatial resolution*, *gray-level resolution*, *contrast* and *spatial structures*. The spatial resolution is rarely changed in compression but it is more like a parameter of the digitization process and thus not issued here. The remaining three properties, however, are all subject to change in the compression and should therefore be measured. We do not make any claims that these properties are sufficient but in our experiments no other important visual property was revealed.

The visual properties should be measured locally, but the questions of the size and shape of the block remain. In principle, the image should be decomposed into regions so that these regions would capture the same properties of the image as human eye pays attention to. The blocksize depends on the viewing distance (cycles per degree) but here we assume that the viewing distance is unknown. In fact, we make even a stronger hypothesis by assuming that the image will be viewed at all possible distances, especially at the smallest possible distance that the resolution allows. Thus, a sliding window of 3×3 pixels is chosen.

3.1 Construction of the measure

From each block, three visual properties are measured as follows. Contrast is measured by the standard deviation of the block as originally proposed in [10]. The spatial structure (shape) inside the block is described by the response of an edge detection operation adopted from [11]. Our hypothesis is that only edge structures are visually important and pixel level background texture can be ignored. Gray-level resolution is measured by calculating the number of distinct gray levels in the block; varying from 1 to 9 in the case of 3×3 blocks. The overall structure of the proposed measure is summarized in Fig. 1.

The average brightness and the activity of the background both affects the visibility of the distortion artifacts. Thus, the distortion values should be weighted differently depending on the background. This is referred here as masking. To keep the model simple we approximate the activity masking by *a contrast mask* where the distortion values are normalized to according to the contrast of the block. The contrast masking is applied to the contrast and edge detection measures but not to the quantization measure. In our experiments the quantization effect was visible in the high contrast textured areas as well. The brightness masking, on the other hand, is not applied here because the model should be independent from basic image processing operations like brightness scaling of the histogram.

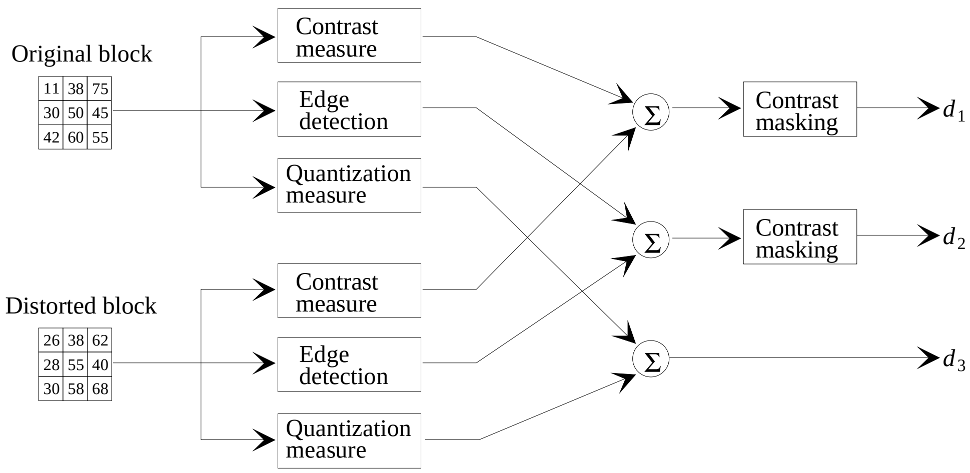

Figure 1: Flow diagram of the blockwise distortion measurement for a single block.

3.2 Calculating the quality factors

Denote the quality factors by d_1, d_2 and d_3 for the contrast errors, structural errors and quantization errors respectively. The first factor is calculated by the square difference between the standard deviation values of the original (A) and the distorted (B) blocks:

$$d_1 = \frac{(\sigma_A - \sigma_B)^2}{\max(1, \sigma_A)} \qquad \textbf{(1)}$$

The distortion value is normalized by the contrast value of the original block. The second factor is:

$$d_2 = \frac{|Gx_A - Gx_B| + |Gy_A - Gy_B|}{2 \cdot \max(1, \sigma_A)} \tag{2}$$

where *Gx* and *Gy* are the responses of the horizontal and vertical edge detectors adopt from [11]. In case of 3×3 block *Gx* and *Gy* will simplify to the filtering operations of Fig. 2. The third factor is:

$$d_3 = (Q_A - Q_B)^2 \tag{3}$$

where Q_A and Q_B are the number of distinct gray levels in the blocks. For more details of the design of d_1 and d_3, see [7, 12].

$$Gx = \frac{1}{4} \times \begin{array}{|c|c|c|} \hline -1 & -2 & -1 \\ \hline 2 & 4 & 2 \\ \hline -1 & -2 & -1 \\ \hline \end{array} \qquad Gy = \frac{1}{4} \times \begin{array}{|c|c|c|} \hline -1 & 2 & -1 \\ \hline -2 & 4 & -2 \\ \hline -1 & 2 & -1 \\ \hline \end{array}$$

Figure 2: Masks for the edge detection.

3.3 Weighting of the factors

Denote the contrast, structural and quantization errors of the entire image by D_1, D_2 and D_3 where the D_i-values are averaged d_i-values over the entire image. In order to obtain only one distortion value instead of three, the overall distortion *D* is defined as a linear function of the three quality factors:

$$D = w_1 \cdot f(D_1) + w_2 \cdot f(D_2) + w_3 \cdot f(D_3) \tag{4}$$

The scaling function *f* is defined as follows:

$$f(D_i) = 1 - \min\left(1, \ \frac{D_i}{k_i}\right) \tag{5}$$

where k_i's are set to (3, 32, 32). This will scale the distortion values D_1, D_2, and D_3 approximately to the range [0, 1]. Negative values are rounded to zero. The scaling is not necessary but it distributes the values relatively evenly into the full range of [0, 1] so that they are comparable to the subjective quality grades. The relative weights w_1, w_2, and w_3 are set to (0.45, 0.30, 0.25) so that they maximize the correlation between *D* and the subjective quality for test image *Lena* (Fig. 3). The optimization was performed using linear programming as implemented by the *Solver* in *Microsoft Excel 5.0*. The same set of weights is then used for all test images.

4. Subjective quality assessment of test images

The subjective quality assessments were performed for the test images of Fig. 3. Three different image sets (one from each original) were generated each having 14 distorted versions of the same image. Each image was compressed by the same set of algorithms.

Lena (512×512) *Airplane* (512×512) *Bridge* (256×256)

Figure 3: The set of test images.

4.1 Compression methods

The following compression algorithms were applied to generate compression errors: *block truncation coding*, *hierarchical block truncation coding*, *quadtree compression*, *recursive plane decomposition*, *JPEG* image data compression standard, and fractal compression by *weighted finite automata*.

Block truncation coding (*BTC*) divides the image into 4×4 blocks and performs two-level quantization to the blocks [4]. The quantization levels are chosen so that the average value and variance of the block would be preserved. Another variant of BTC, called *absolute moment BTC* (*AMBTC*), selects the quantization levels as the mean values of the pixels within the two partitions [14]. Hierarchical block truncation coding (*HBTC*) refers here to an improved BTC variant [8], where the basic BTC algorithm has been augmented with *quadtree segmentation*, *interpolation*, and *entropy coding*. The idea of Quadtree compression (*QT*) is to divide the image into sufficiently homogeneous areas like in the HBTC algorithm. In QT, however, the blocks are quantized to one level only so that each block will be represented by its average value.

Recursive plane decomposition (RPD) is a quadtree based compression technique [17] where the blocks are represented by a two-dimensional plane described by a linear model $f(x, y) = a + bx + cy$. We implemented the basic version of the RPD algorithm and avoided the other features (such as quantizer optimization and entropy coding) which are not vital here. In the baseline JPEG (*Joint photographic experts group*) [16], the image is segmented into 8×8 blocks, which are then transformed to frequency domain by the *fast discrete cosine transform* (FDCT). The transformed coefficients are quantized and then entropy coded. Weighted finite automata (WFA) [2] is a quadtree based fractal compression algorithm, where the blocks are described as linear combinations of the other blocks in the image. The WFA is applied here in spatial domain.

4.2 Test arrangements

Subjective quality assessment in normalized conditions has been defined by CCIR [1]. The CCIR recommendations, however, were originally designed for television pictures and do not take into account the bivariate nature of the distortion measurement. Our intention is to measure the distortion between two images instead of the quality of a single image. Therefore we only partially follow the CCIR recommendations. Furthermore, digital images are rarely viewed at normalized viewing conditions but they are edited, enlarged and they will most likely be viewed at the smallest possible viewing distance in a sense. The test conditions implemented here are summarized in Table 1.

Photographic samples of the test images were prepared with the dimensions of 13×13 centimeters (≈5.12×5.12 inches). Offset quality with 1500 lines per inch were used with halftone frequency of 150 dots per inch. In our experiments, all compression errors that were visible on the screen were also visible in the photographs. The advantages of using photographic samples are: (1) original image is accessible as a reference image, (2) the images can be easily compared against each other during the evaluation, (3) the viewing distance is not fixed but the observer may look at the images from various viewing distances, and/or focus only on parts of the image.

Test persons (both experts and non-expert) were instructed to rank the images in a descending order from worst to best, and then assign the images with a grade from 0 to 10. The grades are essential; the ranking itself is not interesting except helping the evaluation process. Besides the grades, the test persons were encouraged to give verbal comments to describe the distortion in the images. The interpretation of the scale was similar to the MOS scale (*mean opinion score*) but an extended version of the original five-grade scale was used.

Before analyzing the results, the grades given by each person in each image set were normalized to have zero mean and unit standard deviation. The normalized grades in each set were then rescaled to the range [0, 1] according to the minimum and maximum of the normalized values. For example, the average grades (for *Lena*) given by different persons varied from 3.00 to 7.43 before normalization. The standard deviation varied from 1.80 to 3.55. After normalization the corresponding average values and standard deviation were 0.54 and 0.23 for all test persons.

Table 1: Conditions of the subjective quality assessment.

Image samples:	Offset quality photographs
Viewing conditions:	Normal office environment
Viewing distance:	Free of choice for the observer
Time of evaluation:	Unlimited
Number of observers:	15-39
Grading:	Extended MOS scale (0-10).

5. Test results

Error maps produced from the three quality factors are illustrated in Fig. 4. JPEG was able to retain the colors and the contrast of the image rather well; only casual distortion spots appear in the error map of d_1. Most of the distortion originates from structural errors shown in d_2. The dominating distortion type in *BTC* is the jaggedness of the edges which reflects to the error map of d_2. BTC has also the effect of increasing the contrast which is shown in d_1. Quantization errors in *BTC* are shown throughout the error map of d_3. In *QT*, the blockiness and quantization effect appear only in the low contrast areas of the image. They are captured by d_2 and d_3, respectively.

The overall performance of the proposed method is tested next by comparing its results with the subjective grades. Before making any comparisons the values are scaled into the range [0, 1]. PQS results are also included in the comparison. They have been calculated by the PQS-software available at http://info.cipic.ucdavis.edu. The original PQS values (in the range [0, 5]) have been rescaled by dividing them by 5.

Overall, the proposed method correlates well to the subjective quality throughout the image quality range, see Fig. 5. PQS, on the other hand, gives consistently too high grades for *QT*, *BTC* and *AMBTC* compressed images, see Fig. 6. Especially the blockiness in the low contrast regions in *QT* seems to be problematic. The proposed method also tends to give slightly too high grades for the same images but with lesser amount. The corresponding correlation coefficients for each image are summarized in Fig. 7.

The contribution of the three quality factors are shown in Table 2 where the correlation coefficients are shown for all set of combinations of the weighted quality factors. The results are twofold: in case of *Lena* and *Airplane* (512×512) D_1 and D_3 contributed most to the measure; whereas D_2 was the most important factor in case of *Bridge* (256×256). It is also noted that the contrast measure alone works reasonably well giving comparable results to PQS even though it is unable to detect correctly structural errors and the quantization effect in *BTC*, *AMBTC* and *QT* images.

We studied both the proposed measure and PQS further by optimizing their weights for each image separately. This is not a realistic way to design any measure but it helps us understand the potential advantages and problems of these measures. The optimized correlation coefficients were (0.995, 0.962, 0.964) for the proposed measure and (0.989, 0.916, 0.927) for PQS in case of (*Bridge*, *Lena*, *Airplane*). The correlation between the proposed measure and the subjective quality was relatively high, and better than that of PQS. From the five PQS quality factors only F_4 (correlated errors) and F_5 (errors nears the edges) seem to be important. In case of *Bridge*, the weights of all other factors were set to zero due to the optimization. The relative importance of F_4 and F_5 was approximately equal. In case of *Lena* and *Airplane*, also the factor F_5 was masked out and only F_4 contributed to PQS. The problem of the quality factor F_3 seem to be its assumption of 8×8 block size which is not the case in *BTC*, *AMBTC* and *QT*. Quality factors F_1 and F_2 did not contribute to PQS at all.

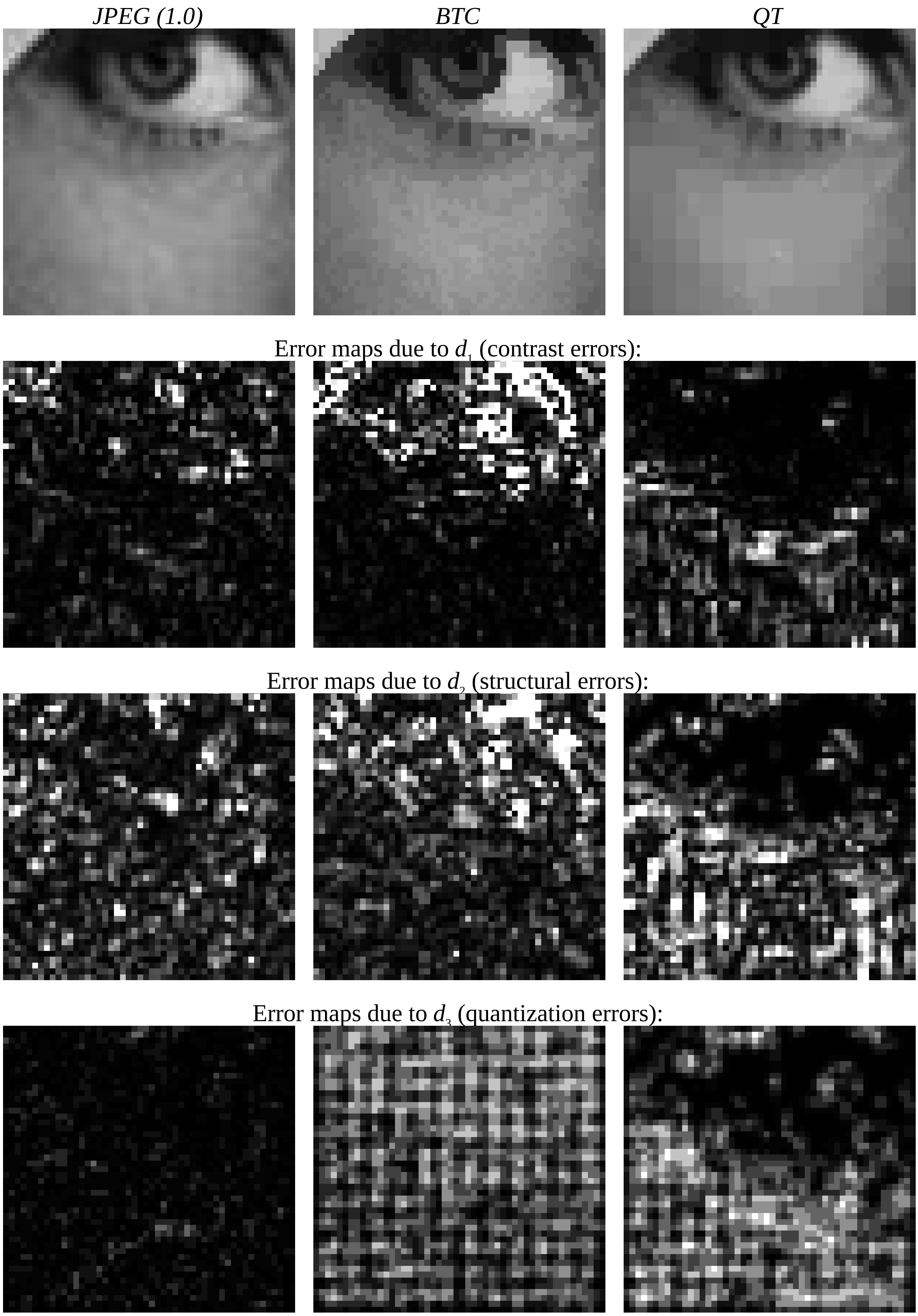

Figure 4: Magnifications of the compressed *Lena* image and the corresponding error maps.

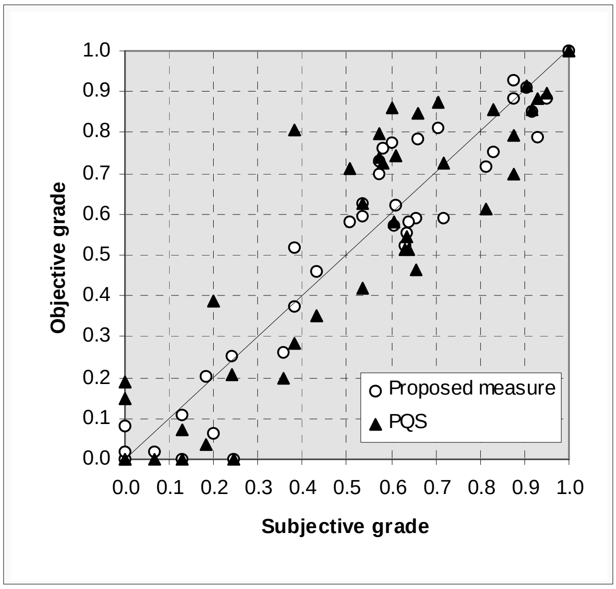

Figure 5: Scattering of the quality grades of the proposed model and PQS.

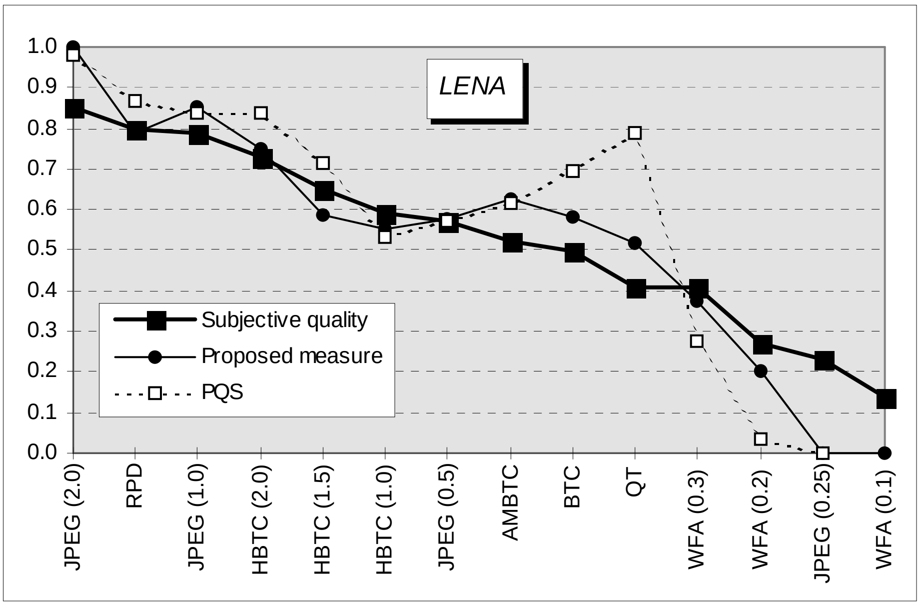

Figure 6: Quality grades of the proposed measure and PQS.

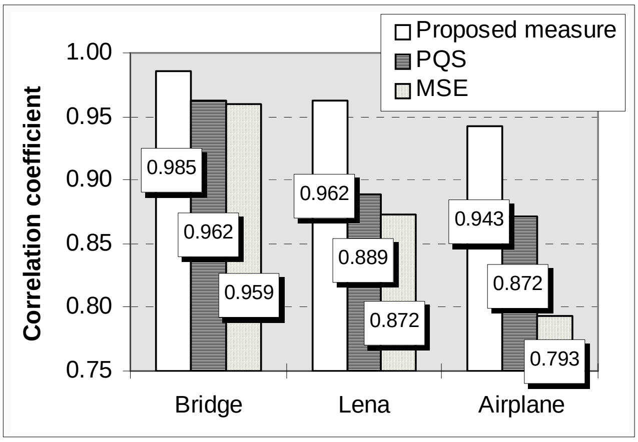

Figure 7: Correlation coefficients between the objective measures and subjective quality.

Table 2: Correlation coefficients of the different combinations of the quality factors.

Factor	Bridge	Lena	Airplane
D_1	0.966	**0.900**	**0.916**
D_2	**0.985**	0.871	0.870
D_3	0.464	0.290	0.787
D_1+D_2	**0.982**	0.908	0.910
D_1+D_3	0.964	**0.960**	**0.952**
D_2+D_3	0.892	0.627	0.947
$D_1+D_2+D_3$	**0.985**	**0.962**	**0.943**

6. Conclusions

A blockwise distortion measure was proposed for evaluating the visual quality of compressed images. The method outperformed PQS and is much simpler to implement. The proposed measure is general in the sense that it measures how well important visual properties have been preserved in the distorted image. It does not make any strict assumptions on the viewing distance or the compression methods used. The method should also be applicable to color images; properties like color richness and saturation are captured by the quantization and contrast measures respectively. In video images, however, there are aspects like temporal distortion [18] which cannot be measured by our model.

The main emphasis has been here on the design of the quality factors. Even though the proposed method performed well for the set of test images, the weighting was somewhat problematic. The optimal weighting was found to be different for each image and it is unlikely that there are any globally optimal weighting for the quality factors. Thus, the generality of the method leaves something to be desired. A more general (but also more complex) measure could be designed by defining better quality factors, better weighting, and by taking account HVS more thoroughly. Nevertheless, the proposed method still outperformed both PQS and MSE.

References

[1] CCIRR, Method for the Subjective Assessment of the Quality of Television Pictures, *Recommendations and reports of the CCIR*, Rec. 500-2, Geneva, 1982.

[2] K. Culik and J. Kari, Image Compression Using Weighted Finite Automata, *Computers and Graphics*, Vol. 17 (3), 305-313, 1993.

[3] S. Daly, The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity, *Proc. SPIE*, *Human Vision, Visual Processing, and Digital Display III*, Vol. 1666, 2-14, 1992.

[4] E.J. Delp and O.R. Mitchell, Image Coding Using Block Truncation Coding. *IEEE Transactions on Communications*, Vol. 27 (9), 1335-1342, September 1979.

[5] A.M. Eskicioglu and P.S. Fisher, A Survey of Quality Measures for Gray Scale Image Compression, *Proc. Space and Earth Science Data Compression Workshop*, 49-61, Snowbird, Utah, 1993.

[6] A.M. Eskicioglu and P.S. Fisher, Image Quality Measures and Their Performance, *IEEE Transactions on Communications*, Vol. 43 (12), 2959-2965, December 1995.

[7] P. Fränti, T. Kaukoranta and O. Nevalainen, Blockwise Distortion Measure in Image Compression, *Proc. Very High Resolution and Quality Imaging*, San Jose, CA, SPIE vol. 2663, 78-87, 1996.

[8] P. Fränti, O. Nevalainen and T. Kaukoranta, Compression of Digital Images by Block Truncation Coding: A Survey, *The Computer Journal*, Vol. 37 (4), 308-332, 1994.

[9] H. Hamada and S. Namba, A Study on Objective Picture Quality Scales for Pictures Digitally Encoded for Broadcast. *IEICE Transactions on Communications*, Vol. E77-B (12), 1480-1487, December 1994.

[10] K. Hosaka, A New Picture Quality Evaluation Method, *Proc. Picture Coding Symposium*, 17-18, Tokyo, Japan, 1986.

[11] S.A. Karunasekera and N.G. Kingsbury, A Distortion Measure for Blocking Artifacts in Image Based on Human Visual Sensitivity. *IEEE Transactions on Image Processing*, Vol. 4 (6), 957-971, June 1995.

[12] T. Kaukoranta, P. Fränti and O. Nevalainen, Empirical Study on Subjective Quality Evaluation of Compressed Images, *Proc. Very High Resolution and Quality Imaging*, San Jose, CA, SPIE vol. 2663, 88-99, 1996.

[13] P. Kuosmanen, P. Koivisto, H. Huttunen and J. Astola, Shape Preservation Criteria and Optimal Soft Morphological Filtering, *Journal of Mathematical Imaging and Vision*, Vol. 5 (4), 319-335, December 1995.

[14] M.D. Lema and O.R. Mitchell, Absolute Moment Block Truncation Coding and Its Application to Color Images, *IEEE Transactions on Communications*, Vol. 32 (10), 1148-1157, October 1984.

[15] M. Miyahara, K. Kotani and V.R. Algazi, Objective Picture Quality Scale (PQS) For Image Coding. *CIPIC Report 96-12*, Univ. of California, Davis, 1996. (to appear in *IEEE Transactions on Communications*).

[16] W.B. Pennebaker and J.L. Mitchell, *JPEG Still Image Data Compression Standard*. Van Nostrand Reinhold, New York, 1993.

[17] P. Strobach, Quadtree-Structured Recursive Plane Decomposition Coding of Images, *IEEE Transactions on Signal Processing*, Vol. 39 (6), 1380-1397, June 1991.

[18] A.A. Webster, C.T. Jones, M.H. Pinson, S.D. Voran and S. Wolf, An Objective Video Quality Assessment System Based on Human Perception, *Proc. Human Vision, Visual Processing, and Digital Display IV*, Bellingham, WA, SPIE vol. 1913, 15-26, 1993.

[19] S.J.P. Westen, R.L. Lagendijk and J. Biemond, Perceptual Image Quality Based on A Multiple Channel HVS Model, *Proc. ICASSP*, *Int. Conf. on Acoustics, Speech, and Signal Processing*, Vol. 4, 2351-2354, 1995.

[20] W. Xu and G. Hauske, Picture Quality Evaluation Based on Error Segmentation, *Proc. SPIE*, *Visual Communications and Image Processing*, Vol. 2308, 1454-1465, 1994.