WHY IS IMAGE QUALITY ASSESSMENT SO DIFFICULT?

Zhou Wang and Alan C. Bovik

Lab for Image and Video Engi., Dept. of ECE
Univ. of Texas at Austin, Austin, TX 78703-1084
zhouwang@ieee.org, bovik@ece.utexas.edu

Ligang Lu

IBM T. J. Watson Research Center
Yorktown Heights, NY 10598
lul@us.ibm.com

ABSTRACT

Image quality assessment plays an important role in various image processing applications. A great deal of effort has been made in recent years to develop objective image quality metrics that correlate with perceived quality measurement. Unfortunately, only limited success has been achieved. In this paper, we provide some insights on why image quality assessment is so difficult by pointing out the weaknesses of the error sensitivity based framework, which has been used by most image quality assessment approaches in the literature.

Furthermore, we propose a new philosophy in designing image quality metrics: *The main function of the human eyes is to extract structural information from the viewing field, and the human visual system is highly adapted for this purpose. Therefore, a measurement of structural distortion should be a good approximation of perceived image distortion.* Based on the new philosophy, we implemented a simple but effective image quality indexing algorithm, which is very promising as shown by our current results.

1. INTRODUCTION

Image quality measurement is crucial for most image processing applications. Generally speaking, an image quality metric has three kinds of applications:

First, it can be used to *monitor* image quality for quality control systems. For example, an image and video acquisition system can use the quality metric to monitor and automatically adjust itself to obtain the best quality image and video data. A network video server can use it to examine the quality of the digital video transmitted on the network and control video streaming.

Second, it can be employed to *benchmark* image processing systems and algorithms. Suppose we need to select one from multiple image processing systems for a specific task, then a quality metric can help us evaluate which of them provides the best quality images.

Third, it can be embedded into an image processing system to *optimize* the algorithms and the parameter settings. For instance, in a visual communication system, a quality metric can help optimal design of the prefiltering and bit assignment algorithms at the encoder and the postprocessing algorithms at the decoder.

The best way to assess the quality of an image is perhaps to *look* at it because human eyes are the ultimate receivers in most image processing environments. The subjective quality measurement Mean Opinion Score (MOS) has been used for many years. However, the MOS method is too inconvenient, slow and expensive for practical usage. The goal of objective image and video quality assessment research is to supply quality metrics that can predict perceived image and video quality automatically. Peak Signal-to-Nose Ratio (PSNR) and Mean Squared Error (MSE) are the most widely used objective image quality/distortion metrics, but they are widely criticized as well, for not correlating well with perceived quality measurement. In the past three to four decades, a great deal of effort has been made to develop new objective image and video quality measurement approaches which incorporate perceptual quality measures by considering human visual system (HVS) characteristics [1, 2, 3, 4, 5, 6, 7, 8, 9].

Surprisingly, only limited success has been achieved. It has been reported that none of the complicated objective image quality metrics in the literature has shown any clear advantage over simple mathematical measures such as PSNR under strict testing conditions and different image distortion environments [2, 9, 10]. For example, in a recent test conducted by the Video Quality Experts Group (VQEG) in validating objective video quality assessment methods, there are eight to nine proponent models whose performance is statistically indistinguishable [2]. Unfortunately, this group of models includes PSNR.

It is worth noting that most proposed objective image quality measurement approaches share a common error sensitivity based philosophy, which is motivated from psychological vision science research, where evidences show that human visual error sensitivities and masking effects vary in different spatial and temporal frequency and directional channels. In this paper, we try to point out the drawbacks of this framework. In addition, we propose a new philosophy for designing image quality metrics, which models image degradations as structural distortion instead of errors.

2. ERROR SENSITIVITY BASED IMAGE QUALITY MEASUREMENT

2.1. Framework of Error Sensitivity Based Methods

A typical error sensitivity based approach can be summarized as Figure 1. Although variances exist and the detailed implementations are different for different image quality assessment models, the underlying principles are the same. First, the original and test image signals are subject to preprocessing procedures, possibly including alignment, luminance transformation, and color transformation, etc. The output is preprocessed original and test signals. A channel decomposition method is then applied to the preprocessed signals, resulting in two sets of transformed signals for different channels. There are many choices for channel decomposition, such as identity transform (as the simplest special case), wavelet trans-

This research was supported in part by IBM Corporation, Inc., Texas Instruments, Inc., and by State of Texas Advanced Technology Program.

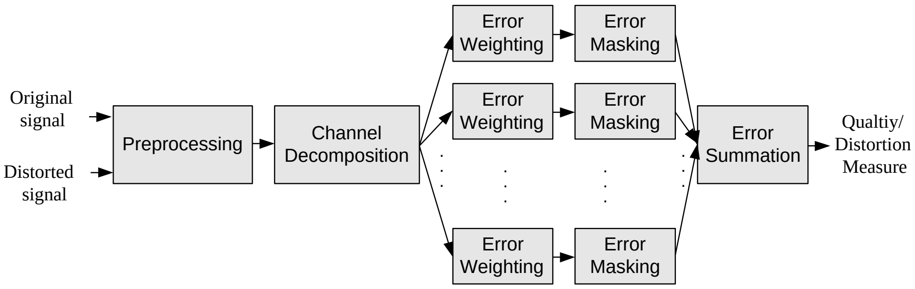


Fig. 1. Error sensitivity based image quality measurement.

forms, discrete cosine transform (DCT), and Gabor decompositions. The decomposed signal is treated differently in different channels according to human visual sensitivities measured in the specific channel. The errors between the two signals in each channel are calculated and weighted, usually by a Contrast Sensitivity Function (CSF). The weighted error signals are adjusted by a visual masking effect model, which reflects the reduced visibility of errors presented on the background signal. Finally, an error pooling method is employed to supply a single quality value of the whole image being tested. The summation usually takes the form:

$$E = \left(\sum_{l} \sum_{k} |e_{l,k}|^{\beta} \right)^{1/\beta} , \qquad (1)$$

where $e_{l,k}$ is the weighted and masked error of the *k-th* coefficient in the *l-th* channel, and β is a constant typically with a value between 1 and 4. This formula is commonly called Minkowski error pooling.

2.2. Weaknesses of Error Sensitivity Based Methods

The above error sensitivity based framework can be viewed as a simplified representation of the HVS. Such simplification implies the following assumptions:

1. The reference signal is of perfect quality.

2. There exist visual channels in the HVS and the channel responses can be simulated by an appropriate set of channel transformations.

3. CSF variance and intra-channel masking effects are the dominant factors that affect the HVS's perception on each transformed coefficient in each channel.

4. For a single coefficient in each channel, after CSF weighting and masking, the relationship between the magnitude of the error, $|e_{l,k}|$, and the distortion perceived by the HVS, $d_{l,k}$, can be modelled as a non-linear function: $d_{l,k} = |e_{l,k}|^{\beta}$.

5. In each channel, after CSF weighting and masking, the interaction between different coefficients is small enough to be ignored.

6. The interaction between channels is small enough to be ignored.

7. The overall perceived distortion is monotonically increasing with the summation of the perceived errors of all coefficients in all channels.

8. The perceived image quality is determined in the early vision system. Higher level processes, such as feature extraction, pattern matching and cognitive understanding happening in the human brain, are less effective.

9. Active visual processes, such as the change of fixation points and the adaptive adjustment of spatial resolution because of attention, are less effective.

The first assumption is reasonable for image/video coding and communication applications. The second and third assumptions are also practically reasonable, provided the channel decomposition methods are designed carefully to fit the psychovisual experimental data. However, all the other assumptions are questionable. We give some examples below.

Notice that most subjective measurement of visual error sensitivity is conducted near the visibility threshold, typically using a 2 Alternative Forced Choice (2AFC) method. These measurement results are not necessarily good for measuring distortions much larger than just visible, which is the case for most image processing applications. Therefore, Assumption 4 is weak, unless more convincing evidence can be provided.

It has been shown that many models work appropriately for simple patterns, such as pure sine waves. However, their performance degrades significantly for natural images, where a large number of simple patterns coincide at the same image locations. This implies that the inter-channel interaction is strong, which is a contradiction of Assumption 6.

Also, we find that Minkowski error pooling (1) is not a good choice for image quality measurement. An example is given in Figure 2, where two test signals, test signals 1 (up-left) and 2 (up-right), are generated from the original signal (up-center). Test signal 1 is obtained by adding a constant number to each sample point, while the signs of the constant number added to test signal 2 are randomly chosen to be 1 or -1. The structural information of the original signal is completely destroyed in test signal 2, but preserved pretty well in test signal 1. In order to calculate the Minkowski error metric, we first subtract the original signal from the test signals, leading to the error signals 1 and 2, which have very different structures. However, applying the absolute operator on the error signals results in exactly the same absolute error signals. The final Minkowski error measures of the two test signals are equal, no matter how the β value in (1) is selected. This example not only demonstrates that structure-preservation ability is an important factor in image quality assessment, but also shows that Minkowski error pooling (1) is very inefficient in capturing the structures of errors. By the observation that the frequency distributions of the test signals 1 and 2 are very different, one might argue that the problem can be solved by transforming the error signals into different frequency channels and measure the errors differently in different channels. This argument is seemingly reasonable, but if the above example signals are extracted from certain frequency bands instead of the spatial domain, then repeated

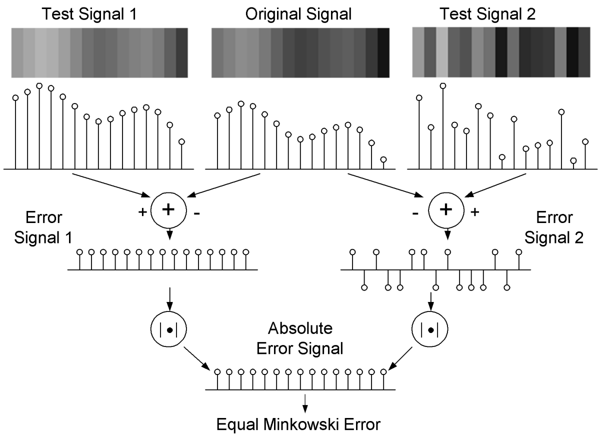


Fig. 2. Minkowski error pooling.

channel transformation is needed to further decompose the transformed signal (possibly iterative transformations will be involved), and finally the multiple time-transformed error signal will still be measured by the Minkowski error summation. In this sense, the weaknesses of Minkowski error pooling still cannot be avoided.

There are some other weaknesses of the framework. For example, channel decompositions usually lead to very high computational complexity, especially for well-designed visual channel transformations such as the Gabor tranforms.

3. STRUCTURAL DISTORTION BASED IMAGE QUALITY MEASUREMENT

3.1. New Philosophy

We believe that one of the most important reasons that the error sensitivity based methods cannot work effectively is that they treat any kind of image degradation as certain type of *errors*. Our new philosophy in designing image quality metrics is:

> *The main function of the human eyes is to extract structural information from the viewing field, and the human visual system is highly adapted for this purpose. Therefore, a measurement of structural distortion should be a good approximation of perceived image distortion.*

As exemplified by Figure 2, large errors do not always result in large structural distortions. The key point of the new philosophy is the switch from *error* measurement to *structural distortion* measurement.

3.2. A New Image Quality Index

Given the new philosophy above, the next problem is how to define and quantify structural distortions. This is a challenging but interesting research topic that needs thorough investigations.

As a first attempt, we developed a simple but effective quality indexing algorithm [11]. Let $\mathbf{x} = \{x_i \,|\, i = 1, 2, \cdots, N\}$ and $\mathbf{y} = \{y_i \,|\, i = 1, 2, \cdots, N\}$ be the original and the test image signals, respectively. The proposed quality index is defined as

$$Q = \frac{4\,\sigma_{xy}\,\bar{x}\,\bar{y}}{(\sigma_x^2 + \sigma_y^2)\,[(\bar{x})^2 + (\bar{y})^2]}\ , \qquad (2)$$

where

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i\,, \qquad \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i\,,$$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2\,,\ \sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2\,,$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})\,.$$

The dynamic range of Q is $[-1, 1]$. The best value 1 is achieved if and only if $y_i = x_i$ for all $i = 1, 2, \cdots, N$. This quality index models any distortion as a combination of three different factors: loss of correlation, mean distortion, and variance distortion. In order to understand this, we rewrite the definition of Q as a product of three components:

$$Q = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\,\bar{x}\,\bar{y}}{(\bar{x})^2 + (\bar{y})^2} \cdot \frac{2\,\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\,. \qquad (3)$$

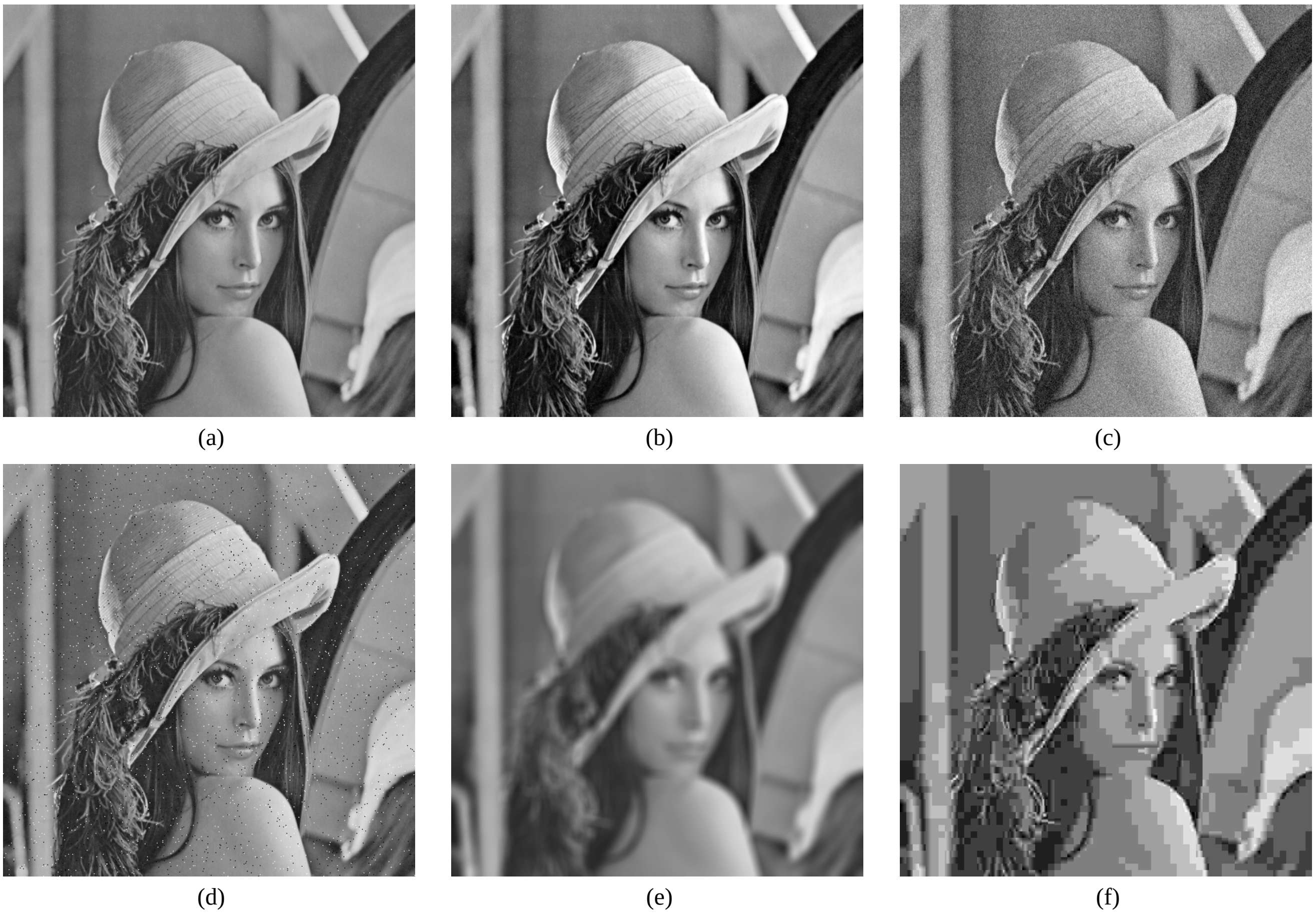
The first component is the linear correlation coefficient between **x** and **y**, whose dynamic range is $[-1, 1]$. The second component, with a value range of [0, 1], measures how close the mean values are between **x** and **y**. It equals 1 if and only if $\bar{x} = \bar{y}$. The third component measures how similar the variances of the signals are. Its range of values is also $[0, 1]$, where the best value 1 is achieved if and only if $\sigma_x = \sigma_y$.

The quality index is applied to natural images using a sliding window approach, with a window size of 8×8. The quality indices are calculated within the sliding window, leading to a quality map of the image. The overall quality index value is the average of the quality map. Some test images are shown in Figure 3, where the original "Lena" image is distorted by a wide variety of corruptions: contrast stretching, additive Gaussian noise, impulsive salt-pepper noise, blurring, and JPEG compression. The MSE and the new quality index values are also given. In this experiment, the performance of MSE is extremely poor in the sense that images with nearly identical MSE are drastically different in perceived quality. By contrast, the new quality index exhibits very consistent correlation with subjective measures. More demonstrative images and an efficient MATLAB implementation of the proposed algorithm are available online at: *http://anchovy.ece.utexas.edu/˜zwang/research/quality_index/demo.html.*

4. CONCLUSIONS AND DISCUSSIONS

In this paper, we provide some insights on why image quality assessment is so difficult by showing the weaknesses of the traditional error sensitivity based image quality measurement approaches. A new philosophy is proposed, which models image degradation as structural distortions instead of errors. A simple implementation of the new philosophy exhibits very promising results.

As pointed out by Watson in [12]: "Much of the theoretical and experimental work in spatial vision in the last thirty years has focussed upon spatial channels; on their existence and on their detailed shape and number." We believe that the issues raised in this paper are more critical for the future development of successful image quality assessment methods.



(a) (b) (c)

(d) (e) (f)

Fig. 3. Evaluation of "Lena" images distorted by different means. (a) Original "Lena" image, 512×512, 8bits/pixel; (b) Contrast stretched image, MSE = 225, Q = 0.9372; (c) Gaussian noise contaminated image, MSE = 225, Q = 0.3891; (d) Impulsive noise contaminated image, MSE = 225, Q = 0.6494; (e) Blurred image, MSE = 225, Q = 0.3461; (f) JPEG compressed image, MSE = 215, Q = 0.2876.

5. REFERENCES

[1] T. N. Pappas and R. J. Safranek, "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Processing* (A. Bovik, ed.), Academic Press, May 2000.

[2] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," *http://www.vqeg.org/*, Mar. 2000.

[3] C. J. van den Branden Lambrecht, Ed., "Special issue on image and video quality metrics," *Signal Processing*, vol. 70, Nov. 1998.

[4] B. Girod, "What's wrong with mean-squared error," in *Digital Images and Human Vision* (A. B. Watson, ed.), pp. 207–220, the MIT press, 1993.

[5] J. Lubin, "A visual discrimination mode for image system design and evaluation," in *Visual Models for Target Detection and Recognition* (E. Peli, ed.), pp. 207–220, Singapore: World Scientific Publishers, 1995.

[6] S. Daly, "The visible difference predictor: An algorithm for the assessment of image fidelity," in *Proc. SPIE*, vol. 1616, pp. 2–15, 1992.

[7] A. B. Watson, J. Hu, and J. F. III. McGowan, "Digital video quality metric based on human vision," *Journal of Electronic Imaging*, vol. 10, no. 1, pp. 20–29, 2001.

[8] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Trans. Image Processing*, vol. 6, pp. 1164–1175, Aug. 1997.

[9] J.-B. Martens and L. Meesters, "Image dissimilarity," *Signal Processing*, vol. 70, pp. 155–176, Nov. 1998.

[10] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Communications*, vol. 43, pp. 2959–2965, Dec. 1995.

[11] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, 2002.

[12] A. B. Watson, "Visual detection of spatial contrast patterns: Evaluation of five simple models," *Optics Express*, vol. 6, pp. 12–33, Jan. 2000.