Blind Image Quality Assessment

Xin Li

Digital Video Department, Sharp Labs of America

5750 NW Pacific Rim Blvd., Camas, WA98607, USA

Abstract

Blind image quality assessment refers to the problem of evaluating the visual quality of an image without any reference. It addresses a fundamental distinction between *fidelity* and *quality*, i.e. human vision system usually does not need any reference to determine the subjective quality of a target image. In this paper, we propose to appraise the image quality by three objective measures: edge sharpness level, random noise level and structural noise level. They jointly provide a heuristic approach of characterizing most important aspects of visual quality. We investigate various mathematical tools (analytical, statistical and PDE-based) for accurately and robustly estimating those three levels. Extensive experiment results are used to justify the validity of our approach.

1. Introduction

Most previous works on image quality assessment [1]-[4] assume the knowledge of a reference image. Their goal is to evaluate the visual difference between a target image and the reference. Strictly speaking, this is a process of evaluating the *fidelity* rather than the *quality* of the target image itself. Since human vision system never requires a reference to determine the visual quality of an image it perceives, the following fundamental problem arises, i.e. how to appraise the image quality without any reference. We call this problem "blind image quality assessment" (it has been also called *single ended* or *no reference* quality assessment in the literature).

Blind image quality assessment is useful to many image-related applications. In image interpolation, the original high-resolution image is often not available as the ground truth; therefore blind assessment of the quality of the interpolated image becomes necessary. The other example is intelligent memory management in digital cameras. Due to the limited storage space, a digital camera must choose a quality factor (e.g. parameter Q in JPEG) that does not sacrifice the visual quality but maximizes the compression ratio. Since consumers usually have no access to the original raw image but only the compressed one, a true visual quality measure would be more useful than a fidelity measure in this scenario.

In this paper, we take a heuristic approach of obtaining a few objective measures to aid the blind image quality assessment process. In spite of the heuristics, we attempt to exploit the appropriate mathematical tools to characterize most important aspects of image quality by those objective measures. We believe that the visual quality of a still image (we only consider grayscale image here) is closely related to the following three quantities: the *edge sharpness* level, the *random noise* level (including both impulse and additive white Gaussian noise) and the *structural noise* level (i.e. various artifacts). We realize that visual perception is a complicated process and other significant factors (e.g. visual masking effect [5]) are not incorporated into our model. The major reasons why we only consider the above three factors are that they can be mathematically formulated to produce objective measures and they are popular in image processing applications. Our work is also based on the assumption that the image has a normal spatial resolution (i.e. the image can be displayed on a typical computer monitor without any sampling ratio change) and is perceived at a normal viewing distance (15-20 inches [3]).

We start from a 2D parameterized edge model [6] in which the *scale* parameter in the edge model determines the sharpness of an arbitrarily oriented step edge. The scale estimation is only applied to the detected step edges that presumably capture the global structure of the image. An ad-hoc approach of estimating the scale parameter is to estimate the edge orientation first and then employ 1D estimator proposed in [6] to the edge profile across the edge orientation. However, estimation of edge orientation itself is a daunting task. We propose an alternative approach of estimating the scale parameter without the necessity of estimating the orientation. Experiment results justify that image blurring or sharpening can be measured as the evolution in the scale space.

We consider impulse noise and additive white Gaussian noise respectively because they come from different noise sources. In the case of impulse noise, a pixel is declared to be noisy if it violates some local smoothness constraint (around the edges, it becomes the geometric constraint [8]). The *percentage* of noisy pixels offers an indication of the impulse noise level. In the case of AWGN, we apply a mean curvature-based denoising algorithm [11] to the image and measure the *energy* of noise filtered out when the iterative filtering process converges. Experiment results have shown that our methods offer fairly accurate estimation of the noise level in both cases.

The evaluation of structural noise in an image often has to rely on some priori information about the characteristics of structural noise. In this paper, we investigate two types of structural noises that are well known to image coding community: block artifacts in DCT-coded (e.g. JPEG) image and ringing artifacts in wavelet-coded (e.g. JPEG2000) images. The amount of block artifacts is measured by the *likelihood* of detecting artificial horizontal or vertical edges around block borders. The amount of ringing artifacts is measured by a *ratio* indicating the deviation of the spectrum of noise filtered out by an anisotropic diffusion filter (e.g. Perona-Malik filter [7]) from the white noise spectrum). Our work aims at providing a more accurate measure than traditional PSNR for evaluating the performance of so-called post-processing techniques designed to suppress those artifacts. We leave the joint estimation of more than one objective measure (eventually we want a complete visual quality evaluation system) to the future study.

2. Edge Sharpness Level

Edges are presumably the most important features in the image source. An ideal two-dimensional step edge can be characterized by the following model [6]

$$f(x,y) = b + \frac{c}{2}[1 + erf\ (\frac{x\cos\theta + y\sin\theta}{\sqrt{2}w})] \quad (1)$$

where c, θ ,w are the contrast, orientation and scale parameters of the edge. We only consider the class of step edges because they usually characterize the global structure of an image (local details correspond to tightly packed edges). Any existing edge detection operator can be employed to identify the edge location first. Then we use the first-order derivative characteristics to separate step edges from tightly packed edges. The first-order derivatives of tightly packed edges tend to fluctuate around zero; while those of step edge are uniformly bounded away from zero. A morphological filter can be used to clean up the obtained step edge map. Fig. 1 shows the step edge detection results for the Lena image.

Fig. 1. Left: Lena image; Right: detected step edge map.

The scale parameter w is only estimated for the detected step edges. Here, we propose a direct scale estimation method that does not need to estimate edge orientation. We firstly estimate the contrast parameter as

$$\hat{c} = \max - \min \quad (2)$$

where *max* and *min* are the local maximum and minimum. Then we sample the image intensity field (I) along x and y axis at the unit distance respectively, and get

$$d_x = f(1,0) - (b + c/2) = \frac{c}{2} erf\ (\frac{x\cos\theta}{\sqrt{2}w}), \quad (3)$$

$$d_y = f(0,1) - (b + c/2) = \frac{c}{2} erf\ (\frac{x\sin\theta}{\sqrt{2}w}) \quad (4)$$

From (3) and (4), we can eliminate θ and obtain the following estimate of w

$$\hat{w} = \frac{1}{\sqrt{2[erf^{-1}(\frac{2d_x}{c})^2 + erf^{-1}(\frac{2d_y}{c})^2]}} \quad (5)$$

It should be noted that the above estimation is only accurate for those pixels along the contour of the middle value ($m=b+c/2$). Therefore we choose to estimate the scale parameter only for the detected edge pixel whose deviation from the estimated m

$$\hat{m} = (\max + \min)/2 \quad (6)$$

is below a pre-selected threshold. From the histogram of scale parameter, we can calculate the average scale parameter (s_{avg}) as the objective measure for the edge sharpness. To justify the validity of the above approach, we use the proposed method to estimate the scale parameters for the original and smoothed Lena images after Gaussian blurring. Fig. 2 shows the evolution of the scale-parameter histograms as image gets blurred. It clearly shows that image blurring can be accurately characterized by the increase of the scale parameter. On the other hand, we synthesize a disk image and apply the edge sharpening technique [16] to enhance the image (see Fig. 3). The average scale parameter is found to be 1.17 for the original image and 0.46 for the sharpened one.

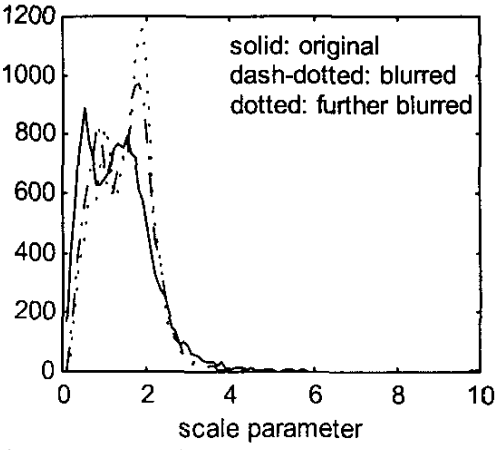

Fig. 2. The effect of Gaussian blurring on the histogram of scale-parameter (s_{avg}=1.60, 1.74, 1.82).

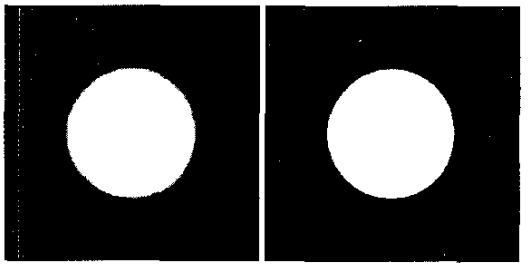

Fig. 3. Left: original disk image (s_{avg}=1.17); right: sharpened disk image (s_{avg}=0.46).

3. Random Noise Level

We consider two popular types of noise source: impulse noise and additive white Gaussian noise (AWGN). Impulse noise usually comes from noisy sensor or transmission errors; while AWGN is appropriate for modeling electronic noise and film grain noise.

A. Impulse Noise

Impulse noise removal is a widely studied problem. Recent works in this area (e.g. [10]) include a preliminary stage of detecting noisy pixels and only applying the filtering process to detected noisy pixels. Therefore, identification of noisy pixels plays the critical role in the overall performance. Here, we propose a nearly optimal detection strategy based on the Least-Square (LS) adaptive predictive model [8]. A pixel is declared to be noisy if it violates the local smoothness constraint (around edge area, smoothness constraint only holds along the edge orientation). As proposed in [8], LS estimation provides an appropriate mathematical tool of exploiting such geometric constraint of edges in the spatial domain.

Let us denote the eight nearest neighbors of pixel $x(i,j)$ as set N_x ($N=|N_x|$) and the local window centered at $x(i,j)$ as set M_x ($M=|M_x|=(2T+1)\times(2T+1)$). We note that both N_x and M_x include non-causal neighbors of $x(i,j)$. Suppose $\vec{y}_{M\times 1} = [y_1 \cdots y_M]^T$ denotes the M elements in M_x and $C_{M\times N} = [\vec{c}_1 \cdots \vec{c}_M]^T$ is a matrix whose k-th column vector consists of the set N_x of pixel y_k. Then the optimal filtering coefficients are given by

$$\vec{a}_{N\times 1} = (C^T C)^{-1}(C^T \vec{y}) \qquad (7)$$

and the filtered value of x is given by

$$\hat{x} = \vec{a}^T \vec{x} \qquad (8)$$

where $\vec{x}_{N\times 1} = [x_1 \cdots x_N]^T$ is the column vector containing N elements in N_x. If $| x - \hat{x} | > Th$, the local smoothness constrain is violated and x is likely to be a noisy pixel. To reduce the effect of noisy on LS estimation, a pre-filtering stage (e.g. median filter) can be used before applying LS estimation.

We add different amount of impulse noise to the Lena image and detect the noise percentage using the proposed method. Table 1 includes the ground truth and the estimated noise level. We observe that the estimated results are highly accurate. It should be noted that careful inspection with the original Lena image reveals that it also contains a small amount of impulse noise (e.g. around the hat and shoulder areas), which explains the positive noise level even if we do not add any noise.

Ground truth p	0	0.01	0.05	0.10	0.15
Estimated p	0.001	0.011	0.050	0.099	0.148

Table 1. Impulse noise estimation results at different noise level p.

B. Additive White Gaussian Noise

Fundamentally speaking, detection of AWGN is also related to the violation of smoothness constraint. However, due to different noise characteristics, LS based approach is not suitable here due to the destructive role of noisy observations. Therefore, a more appropriate mathematical tool is necessary for handling this type of noise and we have found that PDE-based models lend themselves to our task. We propose to iteratively apply mean curvature based filter [11] to the image until it converges and then calculate the noise power filtered out by the diffusion process. An alternative approach suggested in [9] is to take the median of high-band coefficients in the wavelet domain. But we have found it is far less accurate than our method. Table 2 includes the ground truth noise variance and the estimated variance by PDE method. Fairly accurate estimation has been achieved.

Ground truth σ	5	10	15	20	25
Estimated σ	5.08	9.54	14.35	19.10	23.85

Table 2. AWGN variance estimation results at various noise levels.

4. Structural Noise Level

Since it is generally impossible to exhaustively search all possible structural noise patterns, we have to rely on priori knowledge about the degradation process to estimate the structural noise level. In particular, we consider two types of structural noise that are widely known from image compression practice: block artifacts and ringing artifacts.

A. Block Artifacts

In JPEG compression, an image is structured into non-overlapping blocks and each block is coded independently. Consequently artificial block boundaries often appear in the decompressed images at low bit rate, which is well known as "block artifacts". Among the handful of deblocking algorithms, explicit blocky detection strategy exists (e.g. [12]). However, for the task of noise level estimation, we have found that simple edge detection around the block borders is often sufficient. The likelihood of the event that a horizontal or vertical edge happens to *exactly* locate at the block border is small in typical natural images. Such likelihood could be significantly increased in JPEG decoded images.

The detection of horizontal and vertical artificial edges can employ classical gradient operators (we use Prewitt in this paper). To suppress the disturbance from true edge features in the image, we throw away all gradients above a specified threshold (the artificial edges are usually weak edges). The existence of an artificial edge near the block border is declared by a majority vote on the significance of local gradients. The overall edge detection ratio on block borders indicates the blocky noise level. Fig. 3 compares the local gradient maps (after strong edge suppression) for two JPEG compressed Lena images. We can observe that the blocky noise level dramatically increases as the bit rate drops.

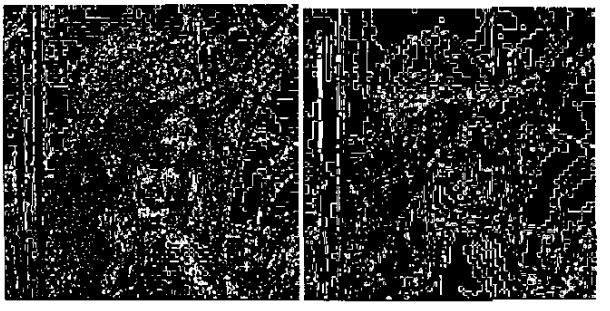

Fig. 4. The local gradient map of JPEG decoded image at 0.43bpp (left, blocky noise level=0.069) and at 0.23bpp (right, blocky noise level =0.297).

B. Ringing Artifacts

In spite of the success of wavelet-based image coders (e.g. SPIHT [13] and EBCOT [14]), they suffer from annoying ringing artifacts around sharp edges at low bit rates. Such artifacts are caused by the violation of smoothness constraint along the edge orientation. Furthermore, the violation along the edge orientation occurs in a periodic pattern and makes the artifacts more striking to human eyes. The works on ringing artifacts removal (e.g. [15]) are relatively fewer than those on block artifacts removal. Objective measurement of ringing artifacts is more challenging than that of block artifacts due to the uncertainty of edge location and orientation. In this paper, we propose to use the powerful Perona-Malik scheme [7] to filter the image and measure the noise spectrum filtered out by the anisotropic diffusion process. For completeness, we include the iterative diffusion equation of P-M filter here

$$X_{i,j}^{t+1} = X_{i,j}^{t} + \lambda[c_N \nabla_N X + c_S \nabla_S X + c_E \nabla_E X + c_W \nabla_W X]$$

where $c_k=g(|\nabla_k X|),k=N,S,E,W$ and ∇ is the local gradient operator. The function $g(.)$ used here is the exponential function suggested in

$$e_{i,j} = X_{i,j}^{t} - X_{i,j}^{0}$$

[7]: $g(x)=exp\{-(x/K)^2\}$. Then the filtered noise can be represented as

The key motivation behind our approach is that due to the effectiveness of anisotropic diffusion on deringing, the artifacts would be mostly assimilated into the spectrum of filtered noise. That is, the noise spectrum would be nearly white for an image containing no artifacts; but colored otherwise. Therefore, we propose to use R, the percentage of energy measured at high frequency (e.g. $|w_x|>\pi/8$, $|w_y|>\pi/8$), as the indicator of structural noise level (it should be close to 1 for white noise spectrum). When comparing the spectrum of the filtered noise for two test images (Fig. 5), we find that R is close to 0.8; while for the SPIHT decoded image, R drops to about 0.548.

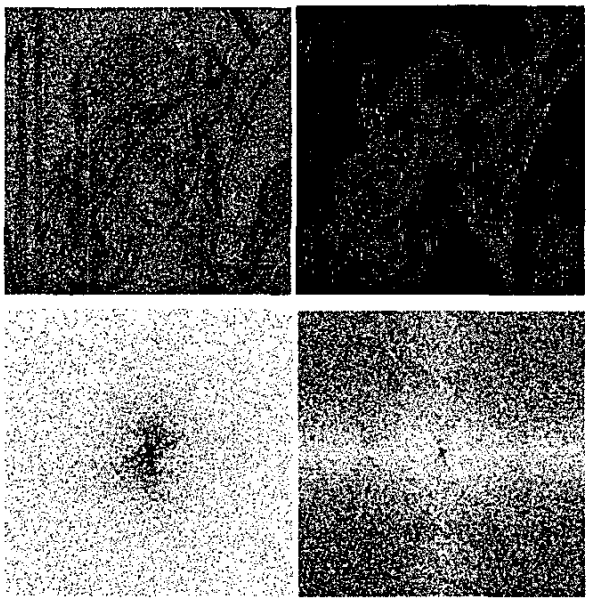

Fig. 5. Filtered noise image (top) and their spectrum (bottom). Left: original Lena image; Right: SPIHT decoded Lena image at 0.125bpp. Note the different noise characteristics after P-M filter.

5. Conclusion and Perspective

In this paper, we study how to objectively measure the edge sharpness level, the random noise level and the structural noise level of an image without any reference. Analytical, statistical and PDE-based models are employed to characterize important image features (e.g. edges) and distinguish them from various noise sources. Several mathematical tools are developed to transform the heuristics into sound methodology of probing signal or noise structure in a variety of situations. Some combination of the proposed measures would provide a more accurate quality measure (e. g. edge sharpness level + structural noise level) than PSNR for many image-related applications such as post-processing and intelligent memory management in digital cameras. Currently, we are investigating how to handle the interference between image features and noise levels when attempting to simultaneously estimate the proposed three objective measures.

6. References

[1] N. Damera-Venkata et al., "Image quality assessment based on a degradation model", *IEEE Trans. on Image Process.*, vol. 9, no. 4, pp. 636-650, Apr. 2000.

[2] S. Daly, "The visible differences predictor: an algorithm for the assessment for image fidelity", in *Proc. SPIE Conf. on Human Vision, Visual Proc., Digital Disp.*, vol. 1666, pp. 2-15, Feb. 1992.

[3] P. Teo and D. Heeger, "A model of perceptual image fidelity", *Proc. of ICIP*, vol. 2, pp. 343-345, Oct. 1995.

[4] J. Lubin, "A visual discrimination model for imaging system design and evaluation", in *Vision Models for Target Detection and Recognition*, Singapore: World Scientific, pp. 245-283, 1995.

[5] A. B. Watson, "DCT quantization matrices visually optimized for individual images, *Proc. SPIE Conf. on Human Vision, Visual Processing, Digital Display*, pp. 202-216, Feb. 1993.

[6] P. van beek, "*Edge-based image representation and coding*", Ph.D dissertation, Delft Univ. of Tech., the Netherland, 1995.

[7] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion", *IEEE Trans. on Pattern Analy. and Mach. Intell.*, vol. 12, no. 7, pp. 629-639, 1990.

[8] X. Li and M. Orchard, "Edge directed prediction for lossless compression of natural images", *IEEE Trans. on Image Processing*, vol. 10, no. 6, pp. 813-817, June 2001.

[9] D. L. Donoho and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage", *Biometrika*, vol. 81, pp. 425-455, 1994.

[10] G. Pok and J. Liu, "Decision-based median filter improved by predictions", *Proc. of ICIP*, vol. 2, pp. 410-413, 1999.

[11] A.I. El-Fallah and G.E. Ford, "The evolution of mean curvature for image filtering", Proc. of ICIP, vol. 1, pp. 298-302, 1994.

[12] J. Chou, M. Crouse and K. Ramchandran, "A simple algorithm for removing blocking artifacts in block-transform coded images", IEEE *Signal Processing Letters*, 1998.

[13] A. Said and W. Pearlman, " A new fast and efficient image codec based on set partitioning in hierarchical trees", IEEE Trans. *CSVT*, vol. 6, pp. 243-250, June 1996.

[14] D. Taubman, "High-performance scalable image compression with EBCOT", IEEE Trans. on *Image Processing*, vol. 9, no. 7, pp. 1158-1170, July 2000.

[15] S. Yang et al., "Maximum-Likelihood parameter estimation for image ringing artifacts removal", IEEE *CSVT*, vol. 11, no. 8, pp. 963-973, Aug. 2001.

[16] X. Li, "Low bit rate image coding in the scale space", Proc. of *Data Comp. Conf.*, pp. 33-42, Mar. 2002.