NO-REFERENCE IMAGE BLUR ASSESSMENT USING MULTISCALE GRADIENT

Ming-Jun Chen and Alan C. Bovik

Laboratory for Image and Video Engineering (LIVE), Department of Electrical & Computer Engineering, The University of Texas at Austin, USA.

ABSTRACT

The increasing number of demanding consumer video applications, as exemplified by cell phone and other low-cost digital cameras, has boosted interest in no-reference objective image and video quality assessment (QA). In this paper, we focus on no-reference image and video blur assessment. There already exist a number of no-reference blur metrics, but most are based on evaluating the widths of intensity edges, which may not reflect real image quality in many circumstances. Instead, we consider natural scenes statistics and adopt multi-resolution decomposition methods to extract reliable features for QA. First, a probabilistic support vector machine (SVM) is applied as a rough image quality evaluator; then the detail image is used to refine and form the final blur metric. The algorithm is tested on the LIVE Image Quality Database; the results show the algorithm has high correlation with human judgment in assessing blur distortion of images.

Index Terms—no-reference blur metric, gradient histogram

1. INTRODUCTION

With the rapid and massive dissemination of digital images, people live in an era full of digitized visual information. Since many of these images are of low quality, effective systems for automatic image quality differentiation are needed. Although there are a variety of effective full-reference (FR) quality assessment (QA) methods, such as the Structural SIMilarity (SSIM) index [1] and the Visual Information Fidelity (VIF) index [2], there yet remains no definitive algorithm or approach to no-reference (NR) QA. As such, research in the area of blind or NR QA remains quite vital.

There are many artifacts that may occur in a distorted image, such as blocking, ringing, noise, and blur. Unlike FR QA, where a reference is available to test against any distortion, NR QA approaches generally seek to capture one or a few distortions. Here we are mainly concerned with NR blur assessment, which remains an important problem in many applications. Generally, humans tend to conclude that images with more detail are of higher quality. We readily acknowledge, of course, that the question is not so simple, since blur can be space-variant, may depend on depth-of-field (hence effect foreground and background objects differently), and may depend on what is being blurred in the image.

A number of NR blur metrics have already been developed. For example, Caviedes and Gurbuz [3] proposed a sharpness measurement metric based on local edge kurtosis. The method they developed first detects edges within blocks, then uses the kurtosis of the block DCT to measure the sharpness of the block. The sharpness of an image is then defined as the average sharpness over all blocks. Marziliano *et al* [4] proposed a blur measurement metric based on analyzing of the width or spread of edges in an image. Their technique an image is more likely to be blurred if the detected edges are wider. Chuang *et al.* [5] evaluate blur by fitting the image gradient magnitude to a normal distribution. The standard deviation of this distribution is then regarded as a blur measurement. It may be observed that all of these metrics attempt blur QA based on evaluation of edges only. However, the widths of edges may not reflect the true image blur, undermining the accuracy of the assessment. This paper thus aims to develop a blur assessment metric that includes more reliable features.

In our search for reliable features other than edges, we are inspired by research on natural scene statistics. It has been shown that natural images obey specific statistical laws that, in principle, can be used to distinguish natural images from artificial images [7]. From this perspective, blurred images may not belong to the category of natural images. Using this observation, a NR blur metric can be designed based on measured image statistics. Previously, Sheikh et al has shown that natural scene statistics are useful for NR QA on JPEG-2000 distorted images [8]. In this study, specific natural scene statistics drawn from the gradient histogram are used as features. A coarse blur metric using these features for evaluating video quality is developed, using a probabilistic support vector machine (SVM). Given the coarse blur assessment, a multi-resolution analysis is used to improve blur assessment. The overall algorithm is shown to effectively accord with human subjectivity.

The rest of the paper is organized as follows: Section 2 describes our use of natural scenes statistics. Section 3 presents our basic NR blur metric. Section 4 details the extended blur metric using multi-resolution analysis. We

demonstrate experiment results on our algorithm in Section 5 and conclude the paper in Section 6.

2. NATURAL IMAGE STATISTICS

Recent research on natural image statistics have shown that natural scenes belong to a small set in the space of all possible image signals [6]. An example of a natural scene property is the greater prevalence of strong image gradients along the cardinal (horizontal and vertical) orientations, in both indoor and outdoor images. A number of researchers have developed statistical models that describe generic natural images [7] (including images of man-made scenes).

According to the research conducted by Field [9], although images of real-world scenes vary greatly in their absolute color distributions, image gradients generally have heavy tailed distributions. Natural image gradient magnitudes are mostly small or zero, yet take large values significantly more often than a Gaussian distribution. This corresponds to the intuition that images often contain large sections of smoothly-varying intensities, interrupted by occasional abrupt changes at edges or occlusive boundaries. However, blurred images usually do not have sharp edges, so the gradient magnitude distribution should more of its mass at small or zero values. By example, Fig. 1 shows one sharp and one blurred image. Figure 2 shows the distribution of their respective gradients.

Liu *et al.* [10] and Levin [11] have demonstrated that the heavy tailed distributions of gradients can be used for blur detection. Liu *et al.* used the gradient histogram span as a feature in their classification model. Levin describes the observed histogram using a mixture model.

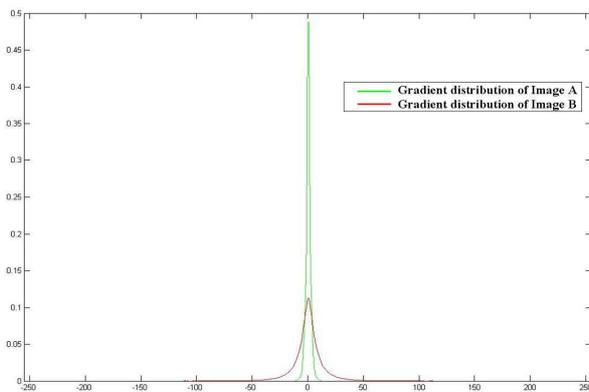**Fig. 1 Left: blurred image A.   Right: Sharp image B.**

Fig. 2. Gradient distributions of images A and B in Fig. 1.

3. PROBABILISTIC SVM FOR BLUR ASSESSMENT

Based on our observation on natural scene statistics, we seek to evaluate the distance between the gradient statistics of an image and a corresponding statistical model of natural scenes. This distance can then be used for image QA.

A classification method is used to measure the distance. We classify images into two groups. One is tagged as "sharp" and the other as "blurred." Using the probabilistic SVM classification model, confidence values represent the distance between the test image and the training set. A higher confidence value implies a higher certainty of the classification result. In this case, this means that the test sample is closer to the assigned class center, i.e., the statistic of the test image is closer to that of "sharp" or "blurred" images.

We chose to use a support vector machine (SVM) as our classification model. The main reason for using SVM is that it works well for classifying a few classes with few training samples. This is highly suitable in our case of having only two classes. Moreover, SVM allows substitution of kernels to achieve better classification results. Although here we only use the default kernel, the possibility of changing the kernel leaves the room for performance improvement.

In choosing features, instead of using the features mentioned in Liu *et al.* [10] or Levin [11], we use the entire gradient histogram as the feature, which contains more information than just the mean or slope of the histogram.

After applying probabilistic SVM classification [12] on an image, a label that indicates its class and a confidence score that indicates the degree of confidence in the decision are obtained. Then the single-scale quality score of the image is defined as:

$$\text{QS-SVM}(x) = \begin{cases} 50 + 50 \cdot \text{confidence if } x \text{ is classified as sharp} \\ 50 \cdot (1\text{-confidence}) \text{ if } x \text{ is classified as blurred} \end{cases} \quad (1)$$

4. MULTI-RESOLUTION NR QA OF BLUR

As with most other areas of image processing and analysis, multi-resolution methods have been shown to afford improved performance relative to single-resolution methods for FR QA [13]. Next we explore the possibility of modifying QS-SVM using information from a multi-resolution decomposition.

Applying a wavelet decomposition on an image is natural way to reveal local spatio-spectral properties that may reveal whether the image has been modified. For example, Fig. 3 shows a sharp image decomposed using a two-level of wavelet decomposition, while Fig. 4 a decomposed blurred image. The sharp image is a high-quality image from the LIVE Database that appears sharp on close visual inspection. The blurred image was modified by a Gaussian low-pass filter. Here, the 2D analysis filter bank designed by Farras [14] is used to analyze the image. From Figs. 3 and 4, it is

apparent that the sharp image contains significant horizontal and vertical energy in the high bands, while the blurred image does not. As a simple measure of sharpness, we sum the horizontal and vertical responses in the high band, which we term a *detail map*. The detail map is used to improve the QA process. Figure 5 shows the detail map of the sharp image in Fig. 3.

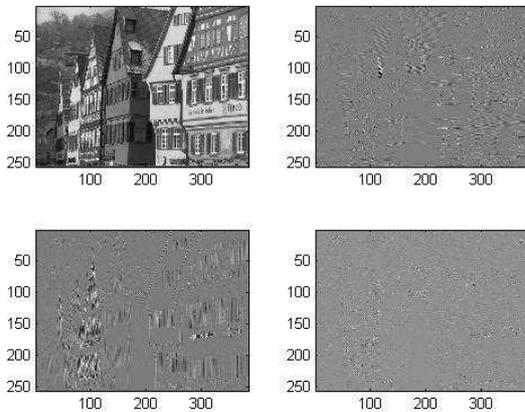A multi-resolution quality (or sharpness) score can then be calculated as follows:

$$Blur\,Quality\,Score = \left(\text{QS-SVM}\right)^{r_0} \prod_{i=1}^{N}\left(\text{DS}_i\right)^{r_i} \qquad (2)$$

where N is the number of layers in the wavelet decomposition, and QS-SVM is the score obtained by analyzing the original image with the probabilistic SVM model, as described in the preceding section. Further, DS_i is a *detail score* obtained from the detail map of layer i. The detail score (omitting indices indicating the wavelet level, since it is applied identically) is defined as:
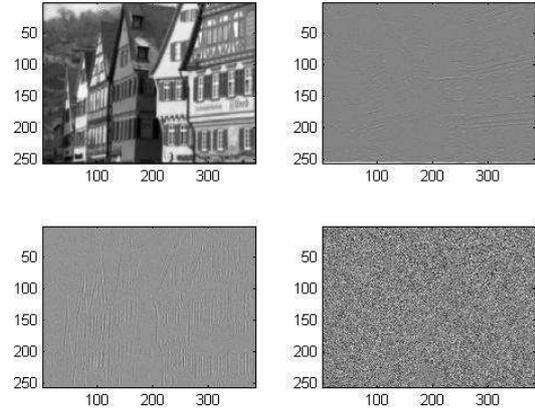
$$\text{DS} = \frac{\sum_{m=1}^{width}\sum_{n=1}^{height}\text{gradient}(m,n)}{width \cdot height} \qquad (3)$$

where *width* and *height* are the dimensions of the image that DS is applied to, and gradient(m, n) is the gradient magnitude value of the image at coordinate (m, n).

Blur Quality Score is the final blur evaluation result, which is the product of the full-resolution score QS-SVM and the values of DS from each layer. The parameters r_i are normalized exponents, where $\sum_{i=0}^{N} r_i = 1$.

Fig. 3. Wavelet decomposition result of sharp image.
Top Left: low low band responses;
Top Right: horizontal high band responses;
Bottom Left: vertical high band responses;
Bottom Right: high high band responses.

Fig. 4. Wavelet decomposition result of blurred image.
Top Left: low low band responses;
Top Right: horizontal high band responses;
Bottom Left: vertical high band responses;
Bottom Right: high high band responses.

Fig. 5. Detail map of image in Fig. 3.

5. EXPERIMENTS AND RESULTS

The LIVE database [16] of images was used in the following experiment. The database includes DMOS subjective scores for each image and several types of distortions. The distortions include JPEG2000 compression distortion, JPEG compression distortion, white noise, gaussian blur, and fast fading channel noise. The experiment was performed only on the gaussian blur images (174 images).

5.1. Performance of SVM Classification

To train the SVM classification model, we used 240 training samples which were marked as "sharp" or "blurred." The training samples were randomly chosen and some of them are out-of-focus photos. Due to the unbalanced quality of the natural training samples (there were more sharp images than naturally blurred images), we applied a gaussian blur to some of some sharp samples to generate additional blurred samples. The final training set included 125 sharp samples and 115 blurred samples.

While tagging samples, if an original image's quality was mediocre, the image was duplicated; one copy marked as "blurred" and the other marked as "sharp," with both images used for training. This procedure prevents misclassifications arising from marking mediocre image as "sharp" or "blurred." This duplication was applied to lower confidence when classifying mediocre samples.

The algorithm was evaluated against the LIVE DMOS scores using the Spearman rank order correlation coefficient (SROCC). The results are shown in Table 1

Prediction Model	SROCC
QS-SVM	0.6136
PSNR (FR)	0.7729
MS-SSIM (FR)	0.9425

Table 1. Comparison of the performance of VQA algorithms.

In Table 1, QS-SVM means probabilistic SVM, PSNR means peak signal to noise ratio, and MS-SSIM means multi-scale structure similarity index. In order to obtain an objective evaluation result, we compared our method to the FR methods [15] tested on the same database.

QS-SVM delivered lower SROCC scores than the FR indices, although the results are promising.

5.2. Performance with Multi-Resolution Decomposition

We began by estimating which layers of the wavelet decomposition achieve the best QA result. We found the correlations between the DS scores and human subjectivity for each layer. The performance numbers are shown in Table 2.

Prediction Model	SROCC
QS-SVM	0.6136
DS_0	0.6583
DS_1	0.8884
DS_2	0.7733
DS_3	0.5587

Table 2. QA performance using different layers.

In Table 2, DS_0 is the detail score computed from the original image. The experiment shows the SROCC score of DS_1 to be significantly higher than for the other layers. The detail map at this middle scale appears to deliver a high correlation with human impression of image quality.

Next we studied combining the QA performance of different layers, omitting level 3 because of its poor performance. Table 3 shows the results of several combinations of algorithms. The parameters r_i of each combination were determined by regression on the training samples.

Prediction Model	SROCC
$DS_0 \cdot DS_1$	0.8884
$DS_1 \cdot DS_2$	0.8884
QS-SVM $\cdot DS_1$	0.9105
QS-SVM $\cdot DS_1 \cdot DS_2$	0.9105
QS-SVM $\cdot DS_2$	0.8428

Table 3. QA performance using different combinations of layers.

Table 3 shows that, except for combination with QA-SVM, all other combinations with DS_1 didn't achieve higher performance than using only DS_1. This result is consistent with our other work in FR QA, where we have found that mid-band QA scores tend to score higher than low-band or high-band scores. Adding more layers did not improve performance here. The highest performance occurs by combining DS_1 with P-SVM (r_0=0.999, r_1=0.001), yielding an impressive SROCC score of 0.9105. Combination QS-SVM with DS_2 (r_0=0.996, r_2=0.004) also improved the performance relative to DS_2, suggesting that QS-SVM and the DS scores offer complementary measurements. Overall, these results confirm our hypothesis that adding more reliable features to the problem of blur assessment can improve performance.

Table 4 illustrates the final performance of our algorithm compared with full-reference algorithms.

Prediction Model	SROCC
QS-SVM	0.6136
PSNR	0.7729
QS-SVM $\cdot DS_1$	0.9105
MS-SSIM (FR)	0.9425

Table 4. Summary of QA performance of different algorithms.

6. CONCLUSION AND FUTURE WORK

We found that the statistics of the image gradient histogram and a detail map from the image wavelet decomposition can be combined to yield good performance for NR blur QA. The performance was demonstrated using the LIVE database of images. Using measured natural scene statistics is useful and there are a few directions worthy of further exploration. We plan to refine our research in the following ways:

1. Revise the classification. Currently the QS-SVM yields an SROCC score of 0.6136 when assigning samples into two classes. Further experiments applying other classification methods may contribute to the algorithm.
2. Expand the algorithm to evaluate video with multiple distortions. Compression distortions may occur with

block, blur, and ringing in the same frame, and these distortions also may be regarded as un-natural images. Therefore, NR QA indices for more general compression-related distortions will be an important and interesting for future study.

3. Explicit statistical modeling. The NR QA algorithm(s) developed here have utilized measured natural scene statistics. A further step in this direction will be to deploy quantitative models of natural scene statistics, and how they are modified by the various distortions, leading to improved classification and overall QA performance.

7. REFERENCES

[1]   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, no. 4, pp. 600-612, Apr. 2004.

[2]   H.R. Sheikh and A.C. Bovik, "Image information and visual quality," *IEEE Trans Image Process,* vol. 15, no. 2, pp. 430-444, Feb. 2006.

[3]   J.E. Caviedes, S. Gurbuz, "No-reference sharpness metric based on local edge kurtosis," *IEEE Int'l Conf Image Process.,* Rochester, NY, Sept 2002.

[4]   P. Marziliano, F. Dufaux, S. Winkler and T. Ebrahimi, "A no-reference perceptual blur metric," *IEEE Int'l Conf Image Process.,* Rochester, NY, Sept 2002.

[5]   Y. Chung, J. Wang, R. Bailey, S. Chen, and S. Chang. "A nonparametric blur measure based on edge analysis for image processing applications," *IEEE Conference on Cybernetics and Intelligent Systems,* 1, 2004.

[6]   D.L. Ruderman, "The statistics of natural images," *Network: Computation in Neural Systems,* vol. 5, no. 4, pp. 517–548, Nov. 1994.

[7]   E.P. Simoncelli, "Statistical models for images: compression, restoration and synthesis," *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers,* Nov. 1997.

[8]   H.R. Sheikh, A.C. Bovik and L.K. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans Image Process,* vol. 14, no. 11, pp. 1918-1927, Nov 2005.

[9]   D. Field, "What is the goal of sensory coding?" *Neural Computation,* vol. 6 pp. 559-601, 1994.

[10] L. Renting, L. Zhaorong, J. Jiaya, "Image partial blur detection and classification," *IEEE Int'l Conf. Computer Vision and Pattern Recognition,* pp. 1-8, 2008.

[11] A. Levin, "Blind motion deblurring using image statistics," *Advances in Neural Information Processing Systems*, Dec. 2006

[12] T.-F. Wu, C.-J. Lin, and R. C. Weng. "Probability estimates for multi-class classification by pairwise coupling." *Journal of Machine Learning Research,* 5:975-1005, 2004.

[13] Z. Wang, E.P. Simoncelli, and A.C. Bovik, "Multi-scale structural similarity for image quality assessment," *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers,* Nov. 2003.

[14] A.F. Abdelnour, I.W. Selesnick. "Nearly symmetric orthogonal wavelet bases", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, May 2001.

[15] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E. H. Yang, and A. C. Bovik, "Quality-aware images," *IEEE Trans. Image Process.,* vol, 15, no. 5, pp. 1680-1689, June 2006.

[16] H.R. Sheikh, Z. Wang, L.K. Cormack and A.C. Bovik, "LIVE Image Quality Assessment Database," Release 2, [online]: Available at: http://live.ece.utexas.edu/research/quality/subjective.htm