

IMAGE QUALITY ASSESSMENT BASED ON LOCAL ORIENTATION DISTRIBUTIONS

Yue Wang¹, Tingting Jiang², Siwei Ma², Wen Gao²

¹Graduate University of Chinese Academy of Sciences, Beijing, China

²Institute of Digital Media, Peking University, Beijing, China

wangyue@jdl.ac.cn; {ttjiang,swma,wgao}@pku.edu.cn

ABSTRACT

Image quality assessment (IQA) is very important for many image and video processing applications, e.g. compression, archiving, restoration and enhancement. An ideal image quality metric should achieve consistency between image distortion prediction and psychological perception of human visual system (HVS). Inspired by that HVS is quite sensitive to image local orientation features, in this paper, we propose a new structural information based image quality metric, which evaluates image distortion by computing the distance of Histograms of Oriented Gradients (HOG) descriptors. Experimental results on LIVE database show that the proposed IQA metric is competitive with state-of-the-art IQA metrics, while keeping relatively low computing complexity.

Index Terms—image quality assessment (IQA), human visual system (HVS), Histograms of Oriented Gradients (HOG)

1. INTRODUCTION

Image quality assessment (IQA) is a crucial technique in many digital image/video applications. To evaluate the quality of an image, the most straightforward way is using the scores directly given by human observers. However, such subjective evaluations are quite time-consuming and expensive, and could not be applied in real-time scenarios and automatic systems. Therefore, most of the IQA researches focus on designing objective quality metrics which could automatically predict image quality. The key of such metrics is that the predicted image quality should be consistent with HVS's judgment.

Most of the early IQA researches focus on pixel-based signal fidelity metrics, e.g. mean-squared-error (MSE), and related peak signal to noise ratio (PSNR). Since these metrics are simple to calculate and have clear physical meanings in term of Shannon information theory, they are widely used and have become the dominant quantitative performance metrics in the field of signal processing.

However, the pixel-based signal fidelity metrics make following assumptions which are not appropriate for natural

image signals [1]:

- 1) pixels are independent with each other
- 2) original images are independent of distortion
- 3) image fidelity is independent of the signs of the pixel errors
- 4) all pixels play roles of equal importance in the quality evaluation for the whole image

For the perception of HVS, the visual information of natural images is perceived by the relationship among the pixels instead of the intensities of individual pixels. This is because visual information contained in natural images is not reflected in the individual pixel signals, but in some basic semantic primitives of natural scenes formed by the neighboring pixels, such as edges, ridges, junctions and so on. Therefore, the image distortions perceived by HVS are not the changes happened on individual pixels, but the changes happened on some high-order statistics of the pixels.

To overcome these problems, many structural information based image quality metrics are proposed in the past years. These methods lay emphasis on extracting the image structure features which are in accord with HVS perception. The most representative work among this category is the SSIM index proposed by Wang *et al* [2]. SSIM used structural information such as mean, variance, and covariance of intensity values of pixels in local patches to evaluate image quality. Besides, Shnaydeman *et al* in [3] proposed to apply SVD on images to extract structure features since singular values could well represent structural information. An image quality metric based on the Harris response was proposed in [4], and LU factorization is used to represent the structural information in [5].

Two other famous IQA metrics besides the above category are Visual Information Fidelity (VIF) metric proposed by Sheikh *et al* [6] and Visual Signal-to-Noise Ratio (VSNR) proposed by Chandler *et al* [7]. VIF is an information-theoretic approach, in which the image fidelity is derived from a statistical model for natural scenes and a model for image distortions. VSNR is a wavelet-based IQA method, which takes into account some low-level and mid-level properties of human vision, such as contrast sensitivity, visual masking and global precedence.

In this paper, we propose a new structural information based

image quality metric. Considering that the distributions of local orientations in one image reflect the high-order statistics of image primitives well, we use Histograms of Oriented Gradients (HOG) [8] as an image structure feature to evaluate image distortions. This metric has clear physical meaning in accord with visual perceptions of HVS. The proposed metric can be calculated easily and the experimental results demonstrate that its efficiency is competitive with other state-of-the-art IQA metrics.

The rest of the paper is organized as follows. In Section 2, we will focus on presenting the details of the proposed image quality metric. Simulation results are presented in Section 3. Finally, Section 4 concludes this paper.

2. THE PROPOSED M-HOG METRIC

From the HVS perspective, natural images are not random collections of pixels, but have strong structural and statistical dependencies. Understanding the properties of psychological and physiological perception of HVS allows the development of better algorithms for image quality assessment. In this work, we consider that the most important characteristic of natural scenes is the presence of strongly oriented features. This is because the basic primitives of natural scenes such as edges, ridges, junctions as well as textures can be described by the distribution of intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions. To characterize the local orientation feature, Histograms of Oriented Gradients (HOG) descriptor was proposed in [8], which is used for the purpose of human detection. In our work, with a little modification, this feature is utilized for evaluating the structural distortion of natural images.

2.1. The HOG descriptor

The first step of calculating HOG descriptor is to derive image gradient and orientation information. In our work, for each pixel $p(i, j)$, Sobel Operator is used to compute its horizontal gradient $f_x(i, j)$ and vertical gradient $f_y(i, j)$.

Then, the orientation map of the image is computed as

$$Dir(i, j) = \tan^{-1}(f_y(i, j) / f_x(i, j)) + \pi / 2 \quad (1)$$

After the gradient map and orientation map are derived, the images are divided into non-overlapped blocks. Then HOG descriptors are calculated on each block. Each pixel within the block casts a weighted vote based on its orientation for an orientation-based histogram. The weight of a vote is designed as a function of the gradient magnitude at the pixel because an image region with larger contrast is more salient to visual perception. Furthermore, the weight function is designed as a constant when the gradient magnitude is above a threshold in order to represent soft presence/absence of an edge at a pixel.

Based on the above consideration, a clipped version of square root of the gradient magnitude $G(i, j)$ is used in this

work:

$$Weight(i, j) = \min(threshold, \sqrt{G(i, j)}) \quad (2)$$

where

$$G(i, j) = \sqrt{f_x(i, j)^2 + f_y(i, j)^2} \quad (3)$$

Now, for each block of the original image and distorted image, a HOG descriptor is obtained. The distance between two HOG descriptors is used to represent the structural distortion between the corresponding blocks.

2.2. Graphical measure

To compute the distance between two histograms, the Euclidean distance is used in our work. However, we found other possible distance measures between histograms, such as χ^2 -distance and city block distance give the similar performance. Specifically, the designed distance function is as follows:

$$D_i(H_A, H_B) = \sqrt{\sum_{j=0}^{b-1} (H_A(j) - H_B(j))^2} \quad (4)$$

where H_A and H_B denote the two compared histograms, $H(j)$ represents the j -th element of a histogram, and b is the number of histogram bins.

The computed block distances generate a distortion map between the original image and the distorted image. An ideal image quality metric should not only be able to give a numerical score of the distorted image, but also be able to give a graphical measure indicating the distribution of error. Fig. 1 illustrates examples of reference and distorted images and their distortion maps. We could observe that the indicated distortion distributions are consistent with HVS.

The most annoying artifacts introduced by JPEG compression at high compression ratios are blocky artifacts. Since blocky artifacts introduce some visible non-existent edges (see block A and A'), local orientation distributions would change dramatically around these regions (see HOG of A and A' respectively). Therefore, the distortion map has high values around blocky regions (around the face and shoulder). As for Gaussian blurring, this type of distortion would severely affect edges and textures. The blurring artifacts would smooth out the peaky areas of local orientation distributions which correspond to edges or textures in the original image (see block B and B') and cause largely diverged orientation distributions for the corresponding blocks in the distorted image (see HOG of B and B' respectively). Accordingly, these regions would have higher distortion values. However, the pixel-based distortion metrics usually fail to capture these distortions.

2.2. Numerical M-HOG metric

Finally, the numerical expression for the proposed image quality metric called *M-HOG* is computed as the average of the graphical distortion map:

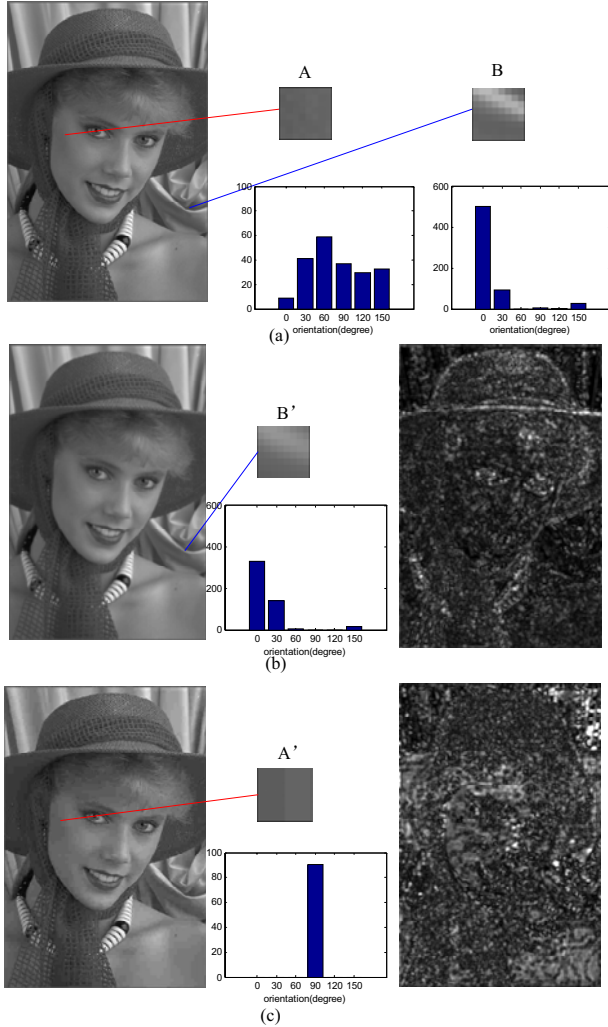


Fig. 1. (a) Original image, two blocks A and B in it, and their corresponding HOG (b) image distorted by Gaussian blurring, block B' with the same position of B, the HOG of B' and the distortion map of the image (c) image distorted by JPEG compression, block A' with the same position of A, the HOG of A' and the distortion map of the image

$$M-HOG = \frac{\sum_{i=1}^{(W/bs) \times (H/bs)} D_i^2}{(W/bs) \times (H/bs)} \quad (5)$$

where W and H denote width and height of the images, respectively, bs denotes the size of the image block, and D_i represents the HOG distance of block i .

3. SIMULATION RESULTS

To evaluate the performance of the proposed IQA metric, we used the laboratory for image and video engineering (LIVE) data set [9], which consists of five subsets with different distortion types including JPEG2000 compression, JPEG compression, White Gaussian noise, Gaussian Blur

and Fast fading (Bit error). For comparison, these same sets of images were evaluated by the following metrics: (1)PSNR; (2)M-SVD; (3)SSIM; (4)VSNR and (5)VIF.

For the whole database, M-HOG metric was computed with the following parameter configuration: $Blocksize = 8 \times 8$ pixels, $histogram\ bins = 6$, $Sobel\ window\ size = 5$ pixels, $weight\ threshold = 50$. For M-SVD, blocksize was set as 8×8 pixels, as suggested in [3]. SSIM, VIF and VSNR were computed using their default implementations provided at [10], [11] and [12], respectively.

As for performance criteria, Pearson correlation coefficient (CC), Spearman rank order correlation coefficient (SROCC), and root mean square error (RMSE) are used as performance indicators. For each metric, a nonlinear mapping between the objective scores and subjective quality ratings was used [13]. In this work, the mapping function chosen for regression for each of the metrics was a 4-parameter logistic function:

$$f(x) = \frac{\tau_1 - \tau_2}{1 + \exp(-\frac{x - \tau_3}{\tau_4})} + \tau_2 \quad (6)$$

Table 1 shows the results on the whole LIVE database with 982 images. It can be seen that the proposed M-HOG metric outperforms PSNR, M_SVD, VSNR and SSIM according to all the three indicators, and is comparable to VIF metric.

Table 2~4 shows the performance comparison in each subset of the LIVE database. It can be seen that the proposed M-HOG metric achieves very good performance in all of the subsets, especially in JPEG and JPEG2000 compressed images.

Fig.2 shows the scatter plots of the DMOS (scaled to the full range of 1~100) versus the objective prediction (after logistic regression) by the proposed and other IQA metrics. The experimental results demonstrate the effectiveness of the proposed M-HOG metric.

Table 1. Performance comparison in the whole LIVE database

	CC	SROCC	RMSE
M-HOG	0.974	0.970	5.26
VIF	0.966	0.972	5.99
VSNR	0.952	0.939	7.07
SSIM	0.939	0.926	7.95
M-SVD	0.923	0.917	8.92
PSNR	0.930	0.909	8.49

Table 2. Performance comparison in each subset of the LIVE database (CC indicator)

	Jpeg2000	Jpeg	wn	gblur	fastfading
M-HOG	0.983	0.980	0.981	0.974	0.963
VIF	0.972	0.963	0.974	0.978	0.965
VSNR	0.974	0.974	0.985	0.967	0.948
SSIM	0.966	0.970	0.953	0.932	0.955
M-SVD	0.971	0.964	0.965	0.920	0.945
PSNR	0.959	0.952	0.991	0.912	0.946

Table 3. Performance comparison in each subset of the LIVE database (SROCC indicator)

	Jpeg2000	Jpeg	wn	gblur	fastfading
M-HOG	0.981	0.962	0.981	0.973	0.960
VIF	0.972	0.955	0.989	0.982	0.977
VSNR	0.967	0.953	0.985	0.963	0.941
SSIM	0.972	0.958	0.978	0.943	0.966
M-SVD	0.978	0.945	0.988	0.885	0.958
PSNR	0.954	0.932	0.991	0.873	0.936

Table 4. Performance comparison in each subset of the LIVE database (RMSE indicator)

	Jpeg2000	Jpeg	wn	gblur	fastfading
M-HOG	4.513	4.891	4.248	4.900	5.932
VIF	5.714	6.551	4.968	4.542	5.789
VSNR	5.529	5.508	3.803	5.566	7.008
SSIM	6.309	5.917	6.684	7.864	6.580
M-SVD	5.879	6.498	5.754	8.508	7.254
PSNR	6.889	7.457	2.922	8.939	7.140

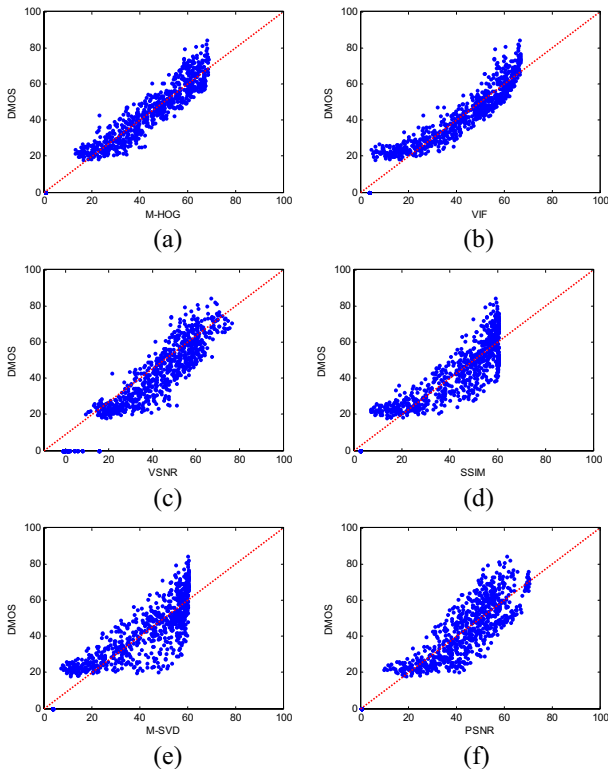


Fig. 2. Scatter plots of the DMOS values versus the objective image quality metrics (after logistic regression) (a) M-HOG, (b) VIF, (c) VSNR, (d) SSIM, (e) M-SVD, and (f) PSNR.

4. CONCLUSION

In this paper, a novel image quality metric based on image

local orientation distributions is proposed. We assess the image quality by computing the difference between the HOG descriptors of image patches. Experimental results on the *LIVE Database Release 2* show that the proposed M-HOG metric outperforms conventional quality metrics such as PSNR, SSIM, M-SVD, VSNR and performs competitively with VIF metric. Moreover, it also shows that the proposed M-HOG metric is rather robust to various types of image distortions and quite efficient in terms of computational complexity. Further work will focus on extending the proposed algorithm by incorporating motion information of videos, and proposing a spatial-temporal local orientation based video quality metric.

5. ACKNOWLEDGEMENTS

This work was supported in part by National Science Foundation (60833013, 60803068) and National Basic Research Program of China (973 Program, 2009CB320903).

6. REFERENCES

- [1] Z. Wang and A.C. Bovik, "Mean squared error: Love it or leave it? - A new look at signal fidelity measures", *IEEE Signal Processing Magazine*, vol. 26, no. 7, pp: 98-117, Jan. 2009
- [2] Z. Wang, A.C. Bovik, H.R. Sheikh and E. Simoncelli, "IQA: From error visibility to structural similarity", *IEEE Trans. Image Process*, vol. 13, no. 4, pp: 600-612, Apr. 2004
- [3] A. Shnayderman, A. Gusev, and A. M. Eskicioglu, "An SVD-based grayscale image quality measure for local and global assessment", *IEEE Trans. Image Processing*, vol. 15, no. 2, pp. 422-429, Feb. 2006.
- [4] Dong-O Kim and Rae-Hong Park, "New Image Quality Metric Using the Harris Response", *IEEE Signal Processing Letters*, vol. 16, no. 7, pp: 616-619, Jun. 2009
- [5] Ho-Sung Han and Dong-O Kim, "Structural Information-Based Image Quality Assessment Using LU Factorization", *IEEE Trans. Consumer Electronics*, vol. 55, no. 1, pp: 165-171, Jan. 2009
- [6] 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
- [7] D.M.Chandler and S.S.Hemami, "VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images", *IEEE Trans. Image Process*, vol.16, no.19, pp: 2284-2298, Sep. 2007
- [8] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", *Computer Vision and Pattern Recognition*, 2005, CVPR 2005, IEEE Computer Society Conference on, vol. 1, pp: 886-893, Jun. 2005
- [9] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, *LIVE Image Quality Assessment Database Release 2* [Online]. Available: <http://live.ece.utexas.edu/research/quality>
- [10] SSIM website. [Online]. Available: <http://www.cns.nyu.edu/zwang/files/research/ssim/index.html>
- [11] VIF website. [Online]. Available: <http://live.ece.utexas.edu/research/quality/VIF.htm>
- [12] VSNR website. [Online]. Available: http://foulard.ece.cornell.edu/gaubatz/metrix_mux.htm
- [13] A.M.Rohaly, J. Libert, P. Corriveau, and A. Webster, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment", Eds. 2000