ICSP2008 Proceedings

HVS-based Structural Similarity for Image Quality Assessment

Bo Wang, Zhibing Wang, Yupeng Liao, Xinggang Lin
Department of Electronic Engineering, Tsinghua University
Beijing, 100084, China
david.b.wang@gmail.com, Wangzb@gmail.com, liaoyupeng@gmail.com, xglin@tsinghua.edu.cn

ABSTRACT

Objective quality assessment is important and widely used in image processing. Recently, the metric named Structural Similarity is proposed, which is based on the assumption that human visual perception is highly adapted for extracting structural information. This metric has a better performance than PSNR in many cases but fails in case evaluating the badly blurred images. This limitation is inconsistent with the characteristics of human visual system (HVS) and leads to our inspiration of applying HVS characters to images structural similarity. Our method, HVS-based Structural Similarity(HSSIM), employs the HVS characters both in frequency domain and spatial domain. It can be concluded in our experiment that HSSIM performs better than PSNR and SSIM, especially for badly blurred images.

Key words: HSSIM, HVS, SSIM, image metric

1. Introduction

The existing image quality metric can be classified into two categories: subjective evaluation and objective evaluation. Human eyes are the ultimate judgers of images and give the most correct result. However, subjective evaluation is inconvenient, time-consuming and expensive. As a result, objective image quality metric plays an important role in image processing. Nowadays, PSNR and MSE is the most commonly adopted method due to low complexity. However, they are also widely criticized for not considering about human visual system [1,2].

In recent years, some researchers begin to focus their attention to the structural similarity of images. Wang et al [3] proposed a novel way, Structural Similarity (SSIM), for image assessment based on the HVS characteristics that human will pay more attention to the structural information while viewing an image. In the experiments performed by Wang et al [5], SSIM shows a better consistency with HVS than does PSNR. However, it is also found that SSIM often fails while evaluating badly blurred images. In this paper, we investigate the mechanism of SSIM, and we apply different attention weights to different frequency components in an image and different regions of an image. As a matter of fact, HVS is more sensitive to those high frequency components such as edges in an image. Meanwhile, human eyes do not pay an equivalent attention to different regions in a picture. In this paper, we propose an improved image quality assessment called HVS-based Structural Similarity (HSSIM) which is based on the frequency and spatial characteristic of human eyes.

The rest of this paper is organized as follows. Section 2 describes the basic idea of SSIM and analysis of its principle. In Section 3 we introduce the HVS-based Structural Similarity (HSSIM) metric. Experiment results are analyzed in Section 4. Finally, we come to a conclusion in Section 5.

2. Structural Similarity(SSIM)

The philosophy of Structural Similarity is based on the assumption that HVS is more adapted to extract structural information. SSIM proposed by Zhou Wang et al [6] is defined as:

$$SSIM(x,y) = f(l(x,y), c(x,y), s(x,y)) \quad (2.1)$$

where $l(x,y)$ is the Luminance Comparison, $c(x,y)$ is the Contrast Comparison and $s(x,y)$ is the Structure Comparison.

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad C_1 = (K_1 L)^2 \quad (2.2)$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad C_2 = (K_2 L)^2 \quad (2.3)$$

$$s(x,y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \quad C_3 = C_2 / 2 \quad (2.4)$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y) \quad (2.5)$$

978-1-4244-2179-4/08/$25.00 ©2008 IEEE

(a) (b) (c)

Figure.1 Comparison of "Church and Capitol" with different types of distortions. (a)Original image. (b) Gaussian blurred image, SSIM=0.6633, HSSIM=0.5209. (c)Noise contaminated image, SSIM=0.6085, HSSIM=0.5845.

where L is the dynamic range of the pixel values (255 for 8-bit grayscale images), μ_x, μ_y is the mean sensitivity of luminance, σ_x, σ_y is the standard deviation of image luminance and $K_1, K_2 \ll 1$ is a small constant used to avoid instability when $\mu_x^2 + \mu_y^2$ is very close to zero.

According to Wang's experimental result, SSIM has a better performance than that of PSNR [6]. However, when applying SSIM to badly blurred images, the gain sometimes isn't ideal. Three images are presented in Figure 1. (a) is the original image, (b) is a badly blurred image obtained from (a) and (c) is a noise contaminated image derived from the same image. While applying SSIM to these two distorted images, we will find that the mark of Gaussian blurred image is higher than that of the noise contaminated image. As a matter of fact, we can easily determine that the quality of the noise contaminated image is much higher than that of the Gaussian blurred image.

The problem described in the previous paragraph is caused by the ignorance of an important perspective of human vision system. Human eyes pay different attentions to different frequency components and different regions in an image. In accordance to this characteristic, we propose the HVS-based Structural Similarity (HSSIM) in the next section.

3. HVS-based Structural SIMILARITY (HSSIM)

Human visual system has many characteristics and is being popularly studied at present. We incorporate two important characteristics, the frequency sensitivity and the spatial sensitivity, into our metric.

3.1 Frequency Sensitivity

Frequency components can effectively reflect the texture feature of an image. According to the study of vision characteristic in human visual system the attention human eyes pay to different frequency components varies. We can assign different weights to different frequency components in accordance with the HVS. Thus the objective judgment result can be more consistent with that of subjective judgment.

In this metric, we employ DCT transformation as our method to extract frequency components from an image. DCT transformation is defined as below.

$$F(u,v) = \frac{2}{N} C(u)C(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \cos\left[\frac{(2x+1)u\pi}{2N}\right] \cos\left[\frac{(2y+1)v\pi}{2N}\right]$$

$$\begin{cases} 0 \le u \le N-1 \\ 0 \le v \le N-1 \end{cases}, \quad C(x) = \begin{cases} \frac{1}{\sqrt{2}}, & x = 0 \\ 1, & x = 1,2,\cdots,N-1 \end{cases} \tag{3.1}$$

where $f(x,y)$ is the block of the original image and $F(u,v)$ is the coefficient after the transformation. Here we assume the size of a macro block is $N \times N$ .After the transformation, we can get the DC/AC coefficient as below.

$$F_{DC} = F_{DC}(0,0) = \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \tag{3.2}$$

$$F_{AC}(u,v) = \frac{2}{N} C(u)C(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \cos\left[\frac{(2x+1)u\pi}{2N}\right] \cos\left[\frac{(2y+1)v\pi}{2N}\right]$$

$$\begin{cases} 0 \le u \le N-1 \\ 0 \le v \le N-1 \\ u \cdot v \ne 0 \end{cases} \tag{3.3}$$

Thus, we define the frequency structure comparison function as follows:

$$freq(x,y)=\frac{2\sigma_{Fxy}+C_3}{\sigma_{Fx}^2+\sigma_{Fy}^2+C_3} \quad (3.4)$$

where σ_{Fx} , σ_{Fy} is the weighted standard error of x and y in the frequency domain, σ_{Fxy} is the weighted covariance of x and y in the frequency domain. These parameters can be calculated as follows.

$$\mu_F=\sum_{u,v} w_{uv}F_{AC}(u,v) \quad (3.5)$$

$$\sigma_F=\left(\sum_{u,v} w_{uv}(F_{AC}(u,v)-\mu_F)\right)^2 \quad (3.6)$$

$$\sigma_{Fxy}=\sum_{u,v} w_{uv}(F_{ACx}(u,v)-\mu_{Fx})(F_{ACy}(u,v)-\mu_{Fy}) \quad (3.7)$$

where w_{uv} is the visual sensitivity weight which describes the frequency focus sensitivity. The weights $\{w_{uv}\}$ used in the metric are acquired from JPEG static image compression. [7]

3.2 Spatial Sensitivity

After the analysis of the frequency characteristics, we turn to that of the spatial characteristics. It is discovered that more photosensitive cells are located in the macula lutea area on the retina. Hence the resolution of the center macula lutea area is higher while that of the remote area on the retina is lower. As a result, human eyes can only be sharply sensitive to a limited area while ignoring many details in other areas. When an image is presented, its center part catches the first attention of human eyes. After that, the focusing area will gradually expand from center to the remote areas. It means that the importance of different parts in an image usually decreases from the center to the brim. As a matter of fact, this characteristic is the same as our habit that we usually place the main object in the center of the viewfinder.

Taking the spatial focus sensitivity into consideration, a spatial affect weight is assigned to each macro block. The spatial affect weights can be calculated as follows.

For a macro block whose center is (x,y),

$$w_{spatial}(x,y)=1-C\times\frac{\sqrt{(x-x_c)^2+(y-y_c)^2}}{d_{\max}} \quad (3.8)$$

where x_c , y_c represent the center of the image center; $d_{\max}$ is the max distance from each pixel to the center; C, ranging from $0\sim1$, is a parameter determines the affect caused by the spatial focus sensitivity. The spatial affect weight is assigned to each block as follows.

$$h(x,y)=w_{spatial}(x,y)\cdot freq(x,y) \quad (3.9)$$

Here, $h(x,y)$ is the HVS-based structure comparison function. The HVS-based SSIM for a single macro block is described as follows.

$$HSSIM=[l(x,y)]^{\alpha}\cdot[c(x,y)]^{\beta}\cdot[h(x,y)]^{\gamma} \quad (3.10)$$

Here we define $\alpha=\beta=\gamma=1$ .Finally, we use a mean value (MHSSIM) to represent the overall image quality.

$$MHSSIM(X,Y)=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}HSSIM(x_i,y_j) \quad (3.11)$$

4. Experimental Result

The evaluation of the performance of HSSIM is based on the Live Image Quality Assess Database Release2 [8] of the Laboratory for Image & Video Engineering in the University of Texas at Austin. A total of 633 images are tested in the experiment, including 168 JPEG2000 compressed images, 175 JPEG compressed images, 145 Gaussian blurred images, and 145 fast-fading distorted images.

We compare the performance of the proposed HSSIM against PSNR and SSIM. For HSSIM and SSIM, each image is partitioned into non-overlapping 8x8 blocks.

Table 1 shows the quantitative measures of the performance of HSSIM, SSIM and PSNR, and four metrics are used to measure these three objective models. The correlation coefficient (CC) means the correlation degree between each model and DMOS, they provide the prediction accuracy evaluation, and the large CC value means the better accuracy. The mean absolute error (MAE), root mean squared error (RMS) and Spearman's Rank-Order Correlation. Coefficient (SROCC) are measures of prediction consistency. HSSIM is better than MSSIM and PSNR in all the criteria on Gaussian blurred images.

Table 1. Performance comparison of image quality assessment models (PSNR, MSSIM, and MHSSIM) on Gaussian blurred images.

	CC	MAE	RMS	SROCC
PSNR	0.6501	9.4652	11.9465	0.6550
SSIM	0.9060	5.0918	6.6544	0.9206
HSSIM	0.9231	4.6095	6.0451	0.9319

In addition, we tested all the images. The results are shown in Figure 3 and Table 2. HSSIM is also better than MSSIM and PSNR in all the criteria.

Table 2. Performance comparison of image quality assessment models (PSNR, MSSIM, and the MHSSIM) on all the images

	CC	MAE	RMS	SROCC
PSNR	0.8089	7.3181	9.4847	0.8102
SSIM	0.9171	5.0173	6.4296	0.9186
HSSIM	0.9312	4.5217	5.8812	0.9330

5. Conclusion

In this paper, we proposed an HVS-based structural similarity (HSSIM) for image quality assessment. This metric is based on the frequency and spatial characteristic of human visual system. The experiment we conducted proves that our assessment can perform better than PSNR and SSIM, especially for badly blurred images.

6. References

[1] B. Girod, "What's wrong with mean-squared error," in Digital Images and Human Vision, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 207–220.

[2] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult," in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 4, Orlando, FL, May 2002, pp. 3313–3316.

[3] W. Xu and G. Hauske, "Picture quality evaluation based on error seg-mentation," in Proc. SPIE, vol. 2308, 1994, pp. 1454–1465.

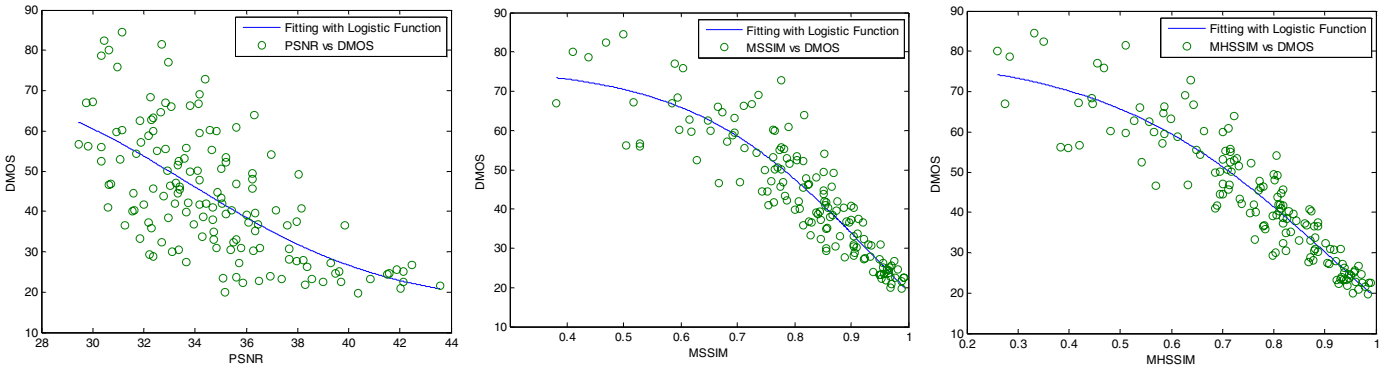
[4]C. J. van den Branden Lambrecht and O.Verscheure, "Perceptual quality measure using a spatio-temporal model of the human visual system," in Proc. SPIE, vol. 2668, 1996, pp. 450–461.

[5]Claudio M Privitera, Lawrence W.Stark, "Algorithms for Defining Visual Regions of Interest Comparison with Eye Fixation," IEEE trans on PAMI, Vol.22, No.9, pp.970-980, 2000

[6] Z. Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P.Simoncelli, "Image Quality Assessment: From Error Measurement to Structural Similarity", IEEE Transactions on Image Processing,Vol. 13, No.4, pp600-613, April 2004.

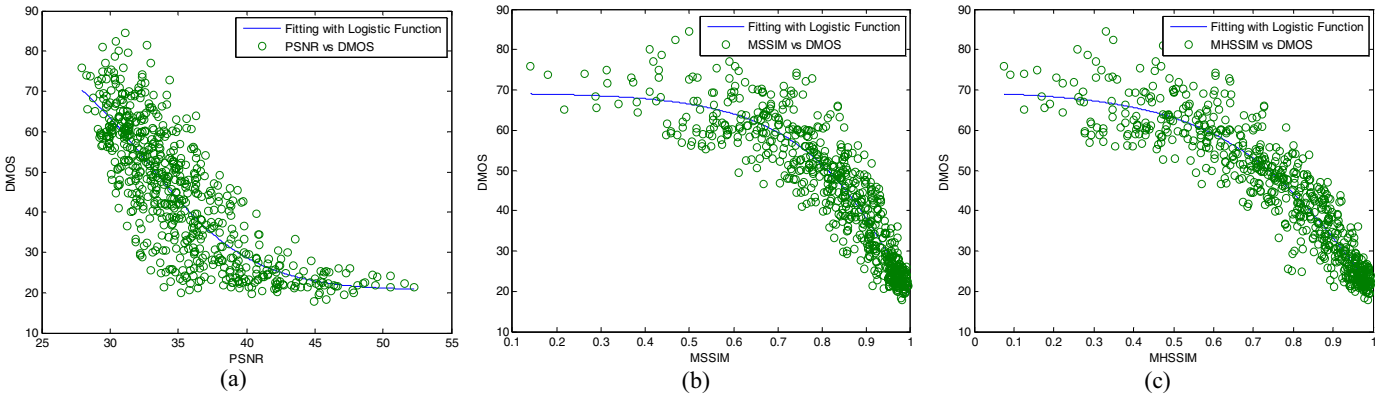
[7] JPEG (ISO/IEC JTC1/SC2/WG8). Digital compression and coding of continuous tone still pictures [S]. ISO CD10918-1, 1991.

[8] H. R. Sheikh, Z. Wang, A. C. Bovik, and L. K. Cormack. Image and Video Quality Assessment Research at LIVE. [Online]Available: http://live.ece.utexas.edu/research/quality/



(a) (b) (c)

Figure.2 Scatter plots of DMOS versus model prediction for Gaussian blur distorted images. (a) PSNR, (b) MSSIM and (c) MHSSIM



(a) (b) (c)

Figure.3 Scatter plots of DMOS versus model prediction for all the images. (a) PSNR, (b) MSSIM and (c) MHSSIM