Influence of the Smooth Region on the Structural Similarity Index

Songnan Li and King Ngi Ngan

Department of Electronic Engineering,
The Chinese University of Hong Kong, Hong Kong SAR
{snli,knngan}@ee.cuhk.edu.hk

Abstract. The Structural Similarity (SSIM) Index is a very popular image quality assessment algorithm due to its good performance and simple calculation. This paper analyzes the influence of the smooth region on the performance of SSIM. It was found out that SSIM tends to depend on the quality of the smooth region to assess the quality of the whole image. From analysis by means of the SSIM quality map, we found out that SSIM overestimated the importance of the smooth region to visual quality. So in this paper we proposed a new weighting function as the spatial pooling method. The resultant weighted-SSIM outperforms the classic SSIM by a considerable amount.

Keywords: Image quality metric, the Structural Similarity (SSIM) index, spatial pooling, weighting function.

1 Introduction

Image quality assessment plays a fundamental role for many image processing applications, such as compression, watermarking, error protection, edge enhancement, etc. Since the Human Visual System (HVS) is the terminal of these applications, subjective evaluation is considered to be the most advisable method for gauging perceptual image quality. However subjective evaluation is not plausible in practice, because it is quite time-consuming in both setup and execution. Therefore, many objective image quality metrics have been developed and widely used for decades to automatically predict the perceptual image quality.

There are many ways to classify the image quality metrics. For example, based on the application range, image quality metrics can be classified into general purpose and application-specific ones; based on the amount of reference information used, we have full-reference, reduced-reference, and non-reference metrics; based on the ways to measure distortions, they can be classified into data metrics and picture metrics [1][1]. Data metrics measure the fidelity of the distorted image without considering its content. The representatives are PSNR and MSE, which are widely used in practice due to their simplicity. Picture metrics generally are

[1] [1] classified the video quality metrics into these two groups. The classification can also be applied to image quality metrics.

P. Muneesawang et al. (Eds.): PCM 2009, LNCS 5879, pp. 836–846, 2009.

more complicated, and depending on the design approach they can be further divided into two groups, namely HVS-based metrics and engineering-based metrics. HVS-based metrics use physiology knowledge and data from psychophysical experiments to generate a mathematic model to simulate the HVS. Engineering-based metrics, on the other hand, is based on image analysis. For example, they may extract edges from two images for comparison, or measure the strength of a particular distortion type, such as blocky, blur etc.

In this paper we investigate a popular image quality metric, namely the Structural Similarity (SSIM) index [2]. According to the above classification, SSIM belongs to general-purpose, full-reference, engineering-based metrics. Based on a subjective image database [3], we analyze how the smooth region influences the performance of SSIM, and find out that SSIM depends too much on the quality of the smooth region to gauge the quality of the entire image. So we propose a weighting function as the spatial pooling method for SSIM, which reduces the importance of the smooth region in evaluating the image quality.

The paper is organized as follows. Section 2 briefly introduces SSIM. Section 3 analyzes how the smooth region influence the performance of SSIM. Based on this analysis, a weighted-SSIM is proposed in section 4, and its performance is evaluated in section 5. Section 6 gives a conclusion.

2 The Structural Similarity (SSIM) Index

SSIM is based on the assumption that the HVS is highly adapted to extract structural information from the viewing field. Three types of similarity together constitute the SSIM index, which are Luminance Similarity $l(\mathbf{x}_i, \mathbf{y}_i)$, Contrast Similarity $c(\mathbf{x}_i, \mathbf{y}_i)$, and Structure Similarity $s(\mathbf{x}_i, \mathbf{y}_i)$:

$$SSIM(\mathbf{x}_i, \mathbf{y}_i) = [l(\mathbf{x}_i, \mathbf{y}_i)]^{\alpha} \cdot [c(\mathbf{x}_i, \mathbf{y}_i)]^{\beta} \cdot [s(\mathbf{x}_i, \mathbf{y}_i)]^{\gamma} \tag{1}$$

where $\mathbf{x}_i$ and $\mathbf{y}_i$ are two image patches to be compared with index i indicating their local positions; $\alpha > 0$, $\beta > 0$, and $\gamma > 0$ are parameters used to adjust the relative importance of the three components. The three components of equation (1) are relatively independent to each other. In other words, the value change for one component does not necessarily mean that the value of the other components must change accordingly. This is one of the good properties of SSIM, which makes the using of α, β and γ to adjust the importance of the three components reasonable. The formula for $l(\mathbf{x}_i, \mathbf{y}_i)$, $c(\mathbf{x}_i, \mathbf{y}_i)$, and $s(\mathbf{x}_i, \mathbf{y}_i)$ are as follows:

$$l(\mathbf{x}_i, \mathbf{y}_i) = \frac{2\mu_{x_i}\mu_{y_i} + C_1}{\mu_{x_i}^2 + \mu_{y_i}^2 + C_1} \tag{2}$$

$$c(\mathbf{x}_i, \mathbf{y}_i) = \frac{2\sigma_{x_i}\sigma_{y_i} + C_2}{\sigma_{x_i}^2 + \sigma_{y_i}^2 + C_2} \tag{3}$$

$$s(\mathbf{x}_i, \mathbf{y}_i) = \frac{\sigma_{x_i y_i} + C_3}{\sigma_{x_i}\sigma_{y_i} + C_3} \tag{4}$$

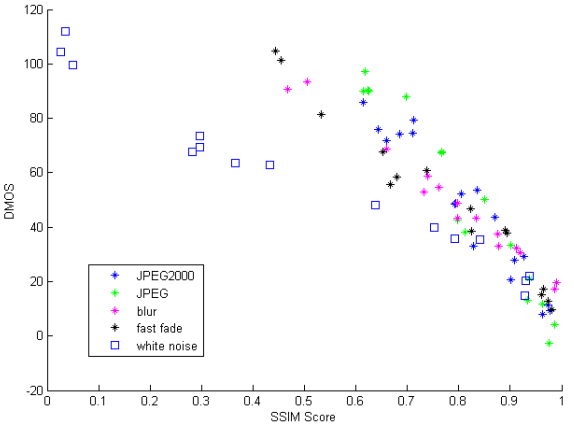


Fig. 1. Scatter plot of DMOS versus SSIM score. Results for 79 images which are generated from 3 reference images (*lighthouse*, *churchandcaptial*, and *carnivaldolls*) are shown here.

where μ_{x_i} and μ_{y_i} are the means of $\mathbf{x}_i$ and $\mathbf{y}_i$ describing their luminances; σ_{x_i} and σ_{y_i} are the variances of $\mathbf{x}_i$ and $\mathbf{y}_i$; $\sigma_{x_i y_i}$ is the covariance of $\mathbf{x}_i$ and $\mathbf{y}_i$ describing their structure similarity; C_1, C_2, and C_3 are small constants to avoid instability when the denominator is very close to zero.

The SSIM index compares two images locally. For each pair of $\mathbf{x}_i$ and $\mathbf{y}_i$ which are local patches of the whole image (sized e.g. 8x8 pixels), the $SSIM(\mathbf{x}_i, \mathbf{y}_i)$ is determined to predict the quality for that particular location which is indexed by i. In this way, a spatially varying quality map can be generated which indicates quality variations across the image. This quality map appears to be brighter at locations where the visual quality is better. In the following section, the quality map will be used to analyze SSIM's performance. The final step of SSIM is to combine the quality map into one single quality score (spatial pooling). The classic SSIM [2], or namely the mean SSIM, takes the average of the quality map as this single score to predict the image visual quality. Since in natural images smooth region is pervasive, SSIM intends to overestimate the importance of the smooth region for quality assessment, as will be explained in the next section.

3 Influence of the Smooth Region

3.1 Performance of SSIM for Different Distortion Types

The performance of SSIM is analyzed based on the Live Image Quality Assess Database [3] - a subjective image database which includes a total of 779 distorted images generated from 29 reference images with five different distortion types: JPEG, JPEG2000, white noise, Gaussian blur, and fast fade. Figure 1 shows the scatter plot of SSIM on a small portion of this database. Since SSIM is a full-reference metric, both the distorted image and its reference image are required as

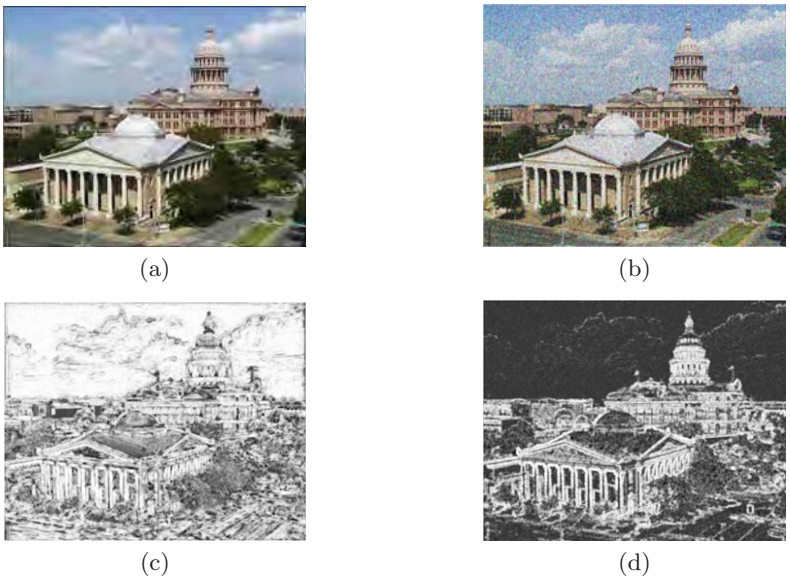


Fig. 2. Two images distorted by (a) JPEG2000 and (b) white noise. DMOS for (a) and (b) are 74.2 and 73.5, respectively; The SSIM score for (a) and (b) are 0.686 and 0.296, respectively. (c) and (d) are quality maps for (a) and (b).

its inputs. Given such an image pair, the SSIM will compute an objective score whose range is between 1 and -1 theoretically, with 1 indicating perfect visual quality of the distorted image. In Figure 1, each spot (star/square) represents such an image pair, with its horizontal index indicating the SSIM score and its vertical index indicating the subjective score — the Differenial Mean Opinion Score (DMOS). DMOS is derived from subjective experiment and provided by the database to indicate the true visual quality of the distorted image.

If the spots in Figure 1 scatter closely around a line, then it means that the objective sore and the subjective score have strong correspondences, so the performance of the image quality metric is near perfect. However in Figure 1, the distribution of the spots seems to form two lines: One for distortion types JPEG, JPEG2000, Gaussian blur, and fast fade; the other for distortion type white noise, which lies under the first one when the SSIM score is smaller than about 0.8. To see this in another way, given two distorted images with the same subjective visual quality (same DMOS) but different distortion types, SSIM tends to give a lower score to the image distorted by white noise, which means that SSIM predict it to be of less visual quality. Why does SSIM have such an inclination?

We find out the reason by observing the SSIM quality map. Figure 2 shows two images with distortion types: (a) JPEG2000 and (b) white noise. According to their DMOS, the two images possess very similar visual quality. Figure 2 (c) and (d) are their corresponding quality maps. As explained in section 2, brightness of the quality map indicates good visual quality at that location.

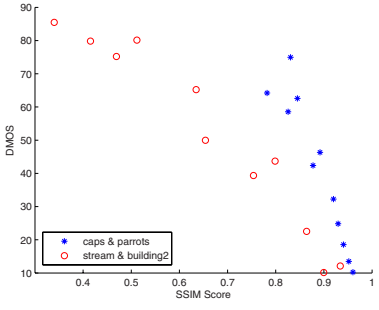
The major disparity between (c) and (d) happens to the background (the sky area), most parts of which are smooth. The quality map (c) is brighter than (d) at the background, due to the fact that JPEG2000 (also JPEG, Gaussian blur, and fast fade) does not distort the smooth region as much as the white noise does. Since the smooth region occupies a large portion of the whole image, the quality map (c) appears to be much brighter than (d), which also implies that the SSIM score of Figure 2 (a) will be much higher than that of (b). This explains the performance of SSIM in Figure 1. Firstly, since smooth region is prevalent in most natural images, SSIM tends to depend on the quality of the smooth region to assess the quality of the whole image; secondly, white noise distorts smooth region more seriously than the other distortion types. Due to the two reasons, SSIM is prone to predict images distorted by white noise with lower visual quality.

3.2 Performance of SSIM for Specific Distortion Type

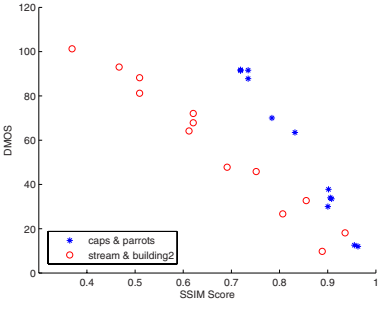
To analyze the performance of SSIM for specific distortion type, distorted images which originate from 4 reference images as shown in Figure 3 (a) are used here. Two of the reference images (*caps* and *parrots*) contain much smooth region, while the other two (*stream* and *building*) consist of structural and textural region mostly. Figure 3 (b), (c), (d), (e), (f) are scatter plots of SSIM on distortion types JPEG2000, JPEG, Gaussian blur, white noise, and fast fade, respectively. Again we observe that for each distortion type the distribution of the spots (stars and circles) seems to form two separated lines. Stars are for the smoother images, while circles are for the more textured images. Figure 3 tells us that: given two images distorted by the same distortion type with the same visual quality but containing different contents, SSIM tends to give a different score to them depending on the smoothness of their contents. Again, why does SSIM have such an inclination?

Figure 4 shows two images (a) *parrots* and (b) *stream* both distorted by Gaussian blur. According to their DMOS, they have a very similar visual quality. But their SSIM quality maps given in Figure 4 (c) and (d) clearly have different brightness, with (c) much brighter than (d), which means SSIM will give (c) a much higher score instead of a similar one. To explain this, again we need to investigate the influence of the smooth region. Since Gaussian bur distorts smooth region less than non-smooth region, and image *parrots* contains far more smooth areas than *stream* does, so the quality map of *parrot* appears to be brighter, which means that SSIM predicts it to be of higher visual quality. For distortion types JPEG2000, JPEG, and fast fade, we can conduct a similar analysis and get the same conclusion: since these distortion types do not distort smooth areas much, then the larger smooth region contained, the higher SSIM score will be given. On the other hand, white noise distorts smooth areas so much that if it contains more smooth areas, the image will be given a lower SSIM score, as shown in Figure 3(e).

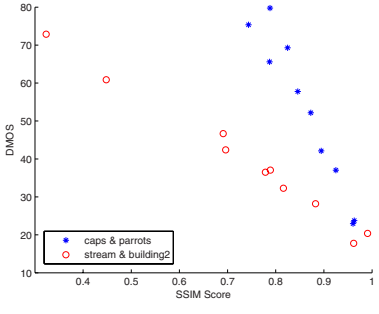
(a) reference images



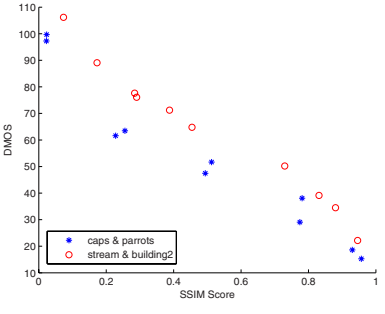
(b) JPEG2000



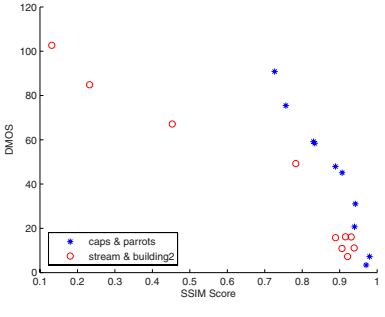
(c) JPEG



(d) Gaussian blur



(e) white noise



(f) fast fade

Fig. 3. a) 4 reference images which from the left to the right are *caps*, *parrots*, *stream*, and *building*. Distorted images originated from these images are used to generate (b), (c), (d), (e), (f), which shows scatter plots for distortion types JPEG2000, JPEG, Gaussian blur, white noise, and fast fade, respectively.

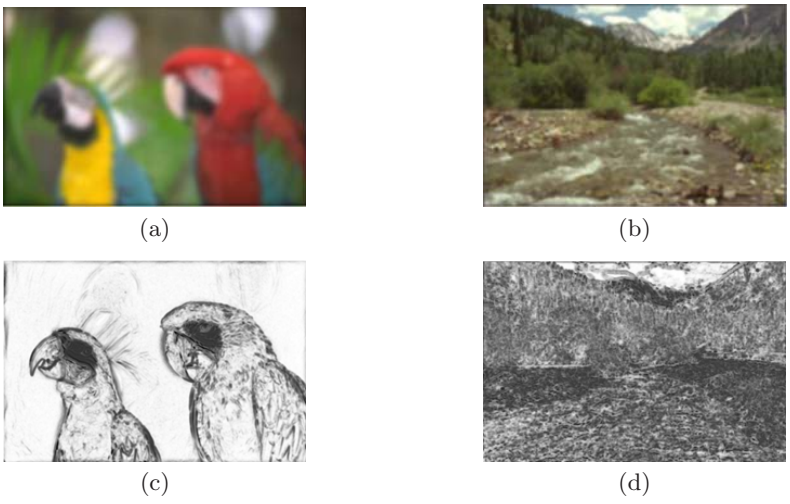


Fig. 4. Two images (a) *parrot* and (b) *stream* both distorted by Gaussian blur. DMOS for (a) and (b) are 79.8 and 72.8, respectively; The SSIM score for (a) and (b) are 0.788 and 0.323, respectively. (c) and (d) are quality maps for (a) and (b).

To summarize above, SSIM tends to depend on the quality of the smooth region to assess the quality of the whole image: firstly, different distortion types distort the smooth region to different extents, and SSIM is prone to give images whose smooth region is less distorted a higher score; secondly, different images distorted by a specific distortion type may have different smoothness levels, and SSIM is pone to give images with larger smooth region a higher score. The above behaviors of SSIM make the SSIM prediction does not accord with the subjective evaluation results very well.

4 The Proposed Weighted-SSIM

From the above analysis, we found that the problem of SSIM may lie in the fact that it excessively depends on the quality of the smooth region to gauge the quality of the whole image. If we reduce the contribution from the smooth region to the SSIM score, the performance of SSIM can be improved. So we modify SSIM by using the following spatial pooling function, which was proposed in [4] and further investigated in [5]:

$$SSIM(\mathbf{X},\mathbf{Y}) = \frac{\sum_{i=1}^{N_s} W(\mathbf{x}_i,\mathbf{y}_i)\cdot SSIM(\mathbf{x}_i,\mathbf{y}_i)}{\sum_{i=1}^{N_s} W(\mathbf{x}_i,\mathbf{y}_i)} \tag{5}$$

where $\mathbf{x}_i$ and $\mathbf{y}_i$ are the i-th image patches from the reference image $\mathbf{X}$ and the distorted image $\mathbf{Y}$, respectively; N_s is the total number of image patches; $W(\mathbf{x}_i,\mathbf{y}_i)$ is the weight given to $\mathbf{x}_i$ and $\mathbf{y}_i$. If $W(\mathbf{x}_i,\mathbf{y}_i)\equiv 1$, then equation (5) is

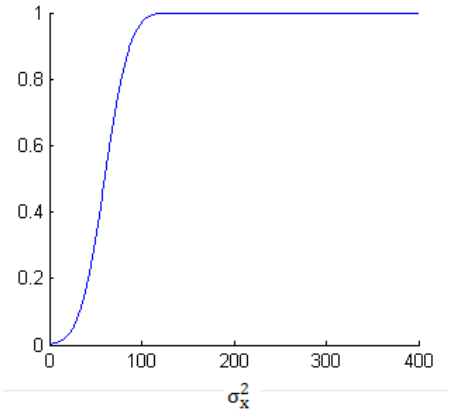


Fig. 5. The proposed weighting function

the classic SSIM introduced in [2]. Due to the above analysis, we propose to use the following weighting function to reduce the influence of the smooth region on the SSIM score:

$$W(\mathbf{x}_i, \mathbf{y}_i) = \frac{erf(\sigma_{x_i}^2 - C_a)}{2 \cdot C_b} + 0.5 \tag{6}$$

where erf is the error function in MATLAB; $\sigma_{x_i}^2$ as in equation (3) is calculated by SSIM which indicates the smoothness of $\mathbf{x}_i$; C_a and C_b are constants which are set to 60 and 30, respectively, based on a training using about 10% images from the database. Figure 5 shows the proposed weighting function (6) using σ_x^2 as input. Smaller σ_x^2 corresponds to smoother region. So by using equation (6) the weight given to the smooth region is less than 1.

In [5], a different weighting function is proposed which is given by the following equation:

$$W(\mathbf{x}_i, \mathbf{y}_i) = log[(1 + \frac{\sigma_{x_i}^2}{C})(1 + \frac{\sigma_{y_i}^2}{C})] \tag{7}$$

where C is constant, and is considered as an estimate of the intrinsic noise in the visual system. In the following section, performance of the classic SSIM, weighted-SSIM (WSSIM) from [5] and the proposed WSSIM will be evaluated and discussed.

5 Experiment

5.1 Experiment Results

The proposed WSSIM is based on the latest version of the SSIM program downloaded from [6]. The Live Image Quality Assess Database [3] introduced in section 3 is used as the subjective image database to evaluate performance. Figure 6

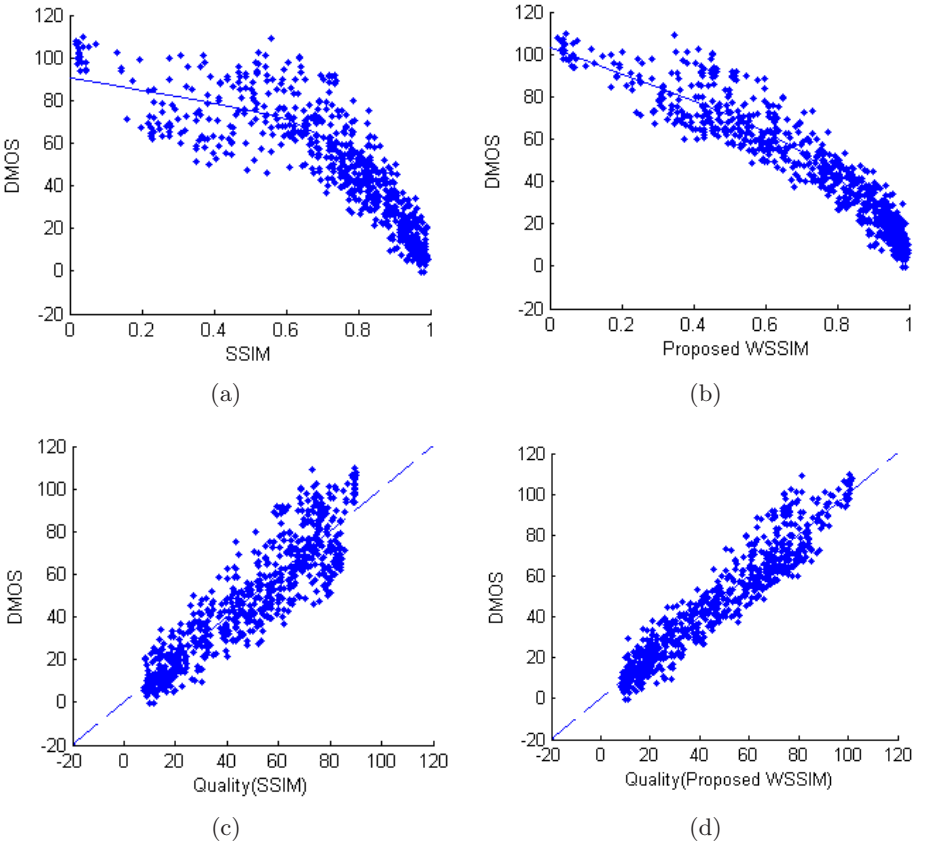


Fig. 6. (a) and (b) are scatter plots for SSIM and the proposed WSSIM, respectively. (c) and (d) are their non-linear mapped versions.

shows the scatter plots of the classic SSIM and the proposed WSSIM. A total of 700 distorted images excluding 79 images for training the proposed weighting function (7) are used here. Comparing Figure 6 (a) with (b), intuitively we can see that the spots in Figure 6 (b) scatter more closely around a line, indicating that the modification improves the performance of SSIM. Adopted from VQEG Phase-I [7] and Phase-II [8] testing, a nonlinear mapping between objective scores and subjective scores is applied to Figure 6 (a) and (b), generating (c) and (d) which shows the improvement even clearly.

Adopted from [9], three evaluation criteria are used to measure the performance of the image quality metric, namely the Correlation Coefficients (CC), the Spearman Rank Order Correlation Coefficient (SROCC), and the Root Mean Squared Error (RMSE). A larger CC or SROCC means that the correlation between the objective scores and the subjective scores is higher, that is to say, a better performance of the image quality metric; oppositely, a smaller RMSE indicates a better performance. In Table 1, the performances of the classic SSIM,

Table 1. Performance comparison between SSIM, WSSIM from [5], and the proposed WSSIM

METRIC	CC	SROCC	RMSE
SSIM	0.905	0.909	11.66
WSSIM [5]	0.934	0.937	9.78
WSSIM (proposed)	0.949	0.952	8.64

WSSIM from [5] and the proposed WSSIM are listed. Both WSSIM from [5] and the proposed WSSIM outperforms MSSIM by a considerable amount, and compared with each other the proposed WSSIM has a little better performance than WSSIM from [5] for all three evaluation criteria.

5.2 Discussion

By reducing the contribution from the quality of the smooth region to the SSIM score, we successively improved the performance of SSIM. There are several ways to explain this performance improvement. The most straightforward one is that generally visual attention will not be paid to the smooth region. Another possible explanation is proposed as follows. While evaluating visual signal quality, two aspects need to be considered: increment of redundant visual information (distortions) and loss of useful visual information (which existed in the original image once). Distortion which happens to the smooth region only introduces redundant visual information but does not cause useful information loss. On the other hand, non-smooth region, especially the structural region, usually contains important visual information which plays key role in higher-level human tasks (e.g. object recognition). So distortion on non-smooth region will result in useful information loss, which we believe will cause more quality reduction compared with mere redundant information increment caused by distortions in the smooth region.

As shown in Table 1, the WSSIM from [5] has comparable performance with the proposed WSSIM. As explained in [5], equation (7) is an information content-weighted pooling function, which allocates higher weights to locations where information (measured by variances) in both original and distorted image is high. Although explained in a different way, actually equation (7) also possesses the same character that it weights smooth region less.

6 Conclusion and Future Work

In this paper, we analyzed how the smooth region influenced the performance of SSIM based on a popular subjective image database. It was found out that SSIM tends to depend on the quality of the smooth region to gauge the quality of the whole image. By proposing a weighting function which reduces the contribution

from the smooth region to the image quality, we improved the performance of SSIM by a considerable amount.

In the current work we did not distinguish the importance of structural and texture region, both of which are non-smooth. It is well believed that structural region is more important than texture region for visual quality evaluation. We will further investigate this difference in the future works. Also, the concept about the redundant information increment and useful information loss and their influences to visual signal quaity will be fully explored.

Acknowledgement

This work was partially supported by a grant from the Chinese University of Hong Kong under the Focused Investment Scheme (Project 1903003).

References

1. Winkler, S.: Video Quality and Beyond. In: Proceedings of European Signal Processing Conference, Poznan, Poland (September 2007)
2. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004)
3. Sheikh, H.R., Wang, Z., Cormack, L., Bovik, A.C.: LIVE Image Quality Assessment Database (Release February 2005), `http://live.ece.utexas.edu/research/quality`
4. Wang, Z., Simoncelli, E.P.: Stimulus synthesis for efficient evaluation and refinement of perceptual image quality metrics," Human Vision and Electronic Imaging IX. In: Proc. SPIE, San Jose, January 2004, vol. 5292 (2004)
5. Wang, Z., Shang, X.: Spatial Pooling Strategies for Image Quality Assessment. In: IEEE Inter. Conf. Image Proc., Atlanta, GA (September 2006)
6. Wang, Z.: The Structural Similarity (SSIM) Index program, `http://www.ece.uwaterloo.ca/~z70wang/research/ssim/`
7. VQEG, Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment (2000), `http://www.vqeg.org/`
8. VQEG, Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment Phase II (2003), `http://www.vqeg.org/`
9. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. IEEE Trans. Image Proc. 15(11), 3440–3451 (2006)