A LOCAL INTENSITY ADAPTIVE STRUCTURAL SIMILARITY INDEX

Zhengguo Li, Chuohao Yeo, Yihhan Tan and Susanto Rahardja

Signal Processing Department, Institute for Infocomm Research, 1 Fusionopolis Way, Singapore

ABSTRACT

Existing structural similarity (SSIM) index comprises of one term on luminance comparison and the other term on contrast and structure comparison. In this paper, the SSIM index is first improved by introducing three weighting factors to the second term such that it is adaptive to local intensities of two images to be compared. The improved SSIM (iSSIM) index is further extended for two images with possibly different exposures. Experimental results show that the proposed indices are more robust to large intensity changes of two images from the same scene and more sensitive to two images from different scenes than the existing SSIM index.

1. INTRODUCTION

Quality metrics have been well studied for both image and video processing. Many intensity-based indices were proposed to assess the similarity of a pair of images, usually by comparing the corresponding pixel intensities [1]. One of the most important intensity-based indices is a structural similarity (SSIM) index in [2]. The fundamental principle of the SSIM index is that the human visual system is highly adapted to extract structure information from the visual scene, and thus a measurement of structure similarity should provide a good approximation of perceptual image quality. There are two terms in the SSIM index. One is on luminance comparison and the other is on contrast and structure comparison. Due to its simplicity, the SSIM index in [2] has many applications. For example, it was recently applied to video coding [3, 4]. On the other hand, the SSIM index needs to be improved which is based on the following two observations:

1) Consider a scenario that two local patches of an image have the same variance value but different mean values. Such an image can be synthesized by combining two images in Figs. 1(a) and 1(c). The same Gaussian noise is added to these two patches as shown in Figs. 1(b) and 1(d). The SSIM values of these two patches are 0.8355 and 0.8337, respectively. They are almost the same even though the noise in Fig. 1(d) is more perceivable. Thus, there is a possible mismatch between the SSIM index and the perceptual quality of an image.

(a) a bright image (b) a noisy image of (a) (c) a dark image (d) a noisy image of (c)

Fig. 1. Two pairs of images with the same level of noise.

2) The SSIM index is designed to focus on cases where the intensities of the two images are almost the same. Many real scenes possess significantly higher dynamic ranges than the dynamic range that can be captured by a single shot of digital camera. In those scenes, a single shot low dynamic range (LDR) image usually turns out to be underexposed and/or overexposed in certain regions of the image. One way to overcome this is to capture a set of differently exposed LDR images and combine all these image to have a more detailed and natural image [5, 6]. Due to different exposures, there are large intensity changes between two images to be compared. As such, the SSIM index in [2] would not be optimum. Differently exposed LDR images can be captured by automatically bracketing the exposure times for HDR scenes [5]. Even though several digital cameras provide an HDR mode, many digital cameras do not have an HDR mode but have bracketing modes. Many sets of differently exposed images will be fused offline. To simplify photographers' work, it is necessary to provide an automatic batch processing method. Thus, a quality metric is desired by the batch processing method to identify whether two differently exposed images are from the same scene.

In this paper, the SSIM index in [2] is first improved by introducing three weighting factors to the second term such that it is adaptive to local intensities of two images to be compared. As a result, the improved SSIM (iSSIM) index can remove the possible mismatch between the SSIM index and perceptual quality of an image that is mentioned above. The iSSIM index is further extended for two differently exposed LDR images by using intensity mapping functions (IMFs) between two images to be compared [7]. The IMFs are used to unify the intensities of two images from the same scene. Due to the function of IMFs, the proposed index is robust to the intensity changes between two images to be compared. Experimental results show that the proposed indices are more robust to intensity change of two images from the same scene and more sensitive to two images from different scenes than the SSIM index in [2].

The rest of this paper is organized as follow. The SSIM index in [2] is improved in Section 2. The improved SSIM index is further extended in Section 3. Experimental results are provided in Section 4 to illustrate the efficiency of the proposed indices. Finally, concluding remarks are listed in Section 5.

2. AN IMPROVED STRUCTURAL SIMILARITY INDEX

Assume that two images being compared are denoted as $\mathbf{Z}_1$ and $\mathbf{Z}_2$, respectively. Their dynamic range is ϖ. For simplicity, let the number of local patches be denoted as P and the contents of $\mathbf{Z}_1$ and $\mathbf{Z}_2$ at the kth patch as $\mathbf{Z}_{1,k}$ and $\mathbf{Z}_{2,k}$, respectively. The SSIM index is defined as [2]

$$S(\mathbf{Z}_1, \mathbf{Z}_2) = \frac{1}{P}\sum_{k=1}^{P} S(\mathbf{Z}_{1,k}, \mathbf{Z}_{2,k}),$$

where $S(\mathbf{Z}_{1,k}, \mathbf{Z}_{2,k})$ is computed as

$$S(\mathbf{Z}_{1,k}, \mathbf{Z}_{2,k}) = \frac{2\mu_{\mathbf{Z}_{1,k}}\mu_{\mathbf{Z}_{2,k}} + c_1}{\mu^2_{\mathbf{Z}_{1,k}} + \mu^2_{\mathbf{Z}_{2,k}} + c_1} \frac{2\sigma_{\mathbf{Z}_{1,k}\mathbf{Z}_{2,k}} + c_2}{\sigma^2_{\mathbf{Z}_{1,k}} + \sigma^2_{\mathbf{Z}_{2,k}} + c_2}, \quad (1)$$

c_1 and c_2 are two constants, the values of c_1 and c_2 are determined as $(0.01 * \varpi)^2$ and $(0.03 * \varpi)^2$, respectively. $\mu_{\mathbf{Z}_{1,k}}, \mu_{\mathbf{Z}_{2,k}}, \sigma^2_{\mathbf{Z}_{1,k}}$ and $\sigma^2_{\mathbf{Z}_{2,k}}$ are the mean and variance values of $\mathbf{Z}_{1,k}$ and $\mathbf{Z}_{2,k}$, respectively, and $\sigma_{\mathbf{Z}_{1,k}\mathbf{Z}_{2,k}}$ is the covariance of $\mathbf{Z}_{1,k}$ and $\mathbf{Z}_{2,k}$. Their values are computed by incorporating an 11×11 circular-symmetric Gaussion weighting function with standard deviation of 1.5 samples, normalized to unit sum [2].

There are two terms in Equation (1). The first term is on the comparison of luminance of two collocated local patches and the second term is on the comparison of contrast and structure of two collocated local patches.

Suppose that two local patches $\mathbf{Z}_{1,i}$ and $\mathbf{Z}_{1,j}$ have the same variance value but different mean values. The same Gaussian noise is added to both of them. The values of $S(\mathbf{Z}_{1,i}, \mathbf{Z}_{2,i})$ and $S(\mathbf{Z}_{1,j}, \mathbf{Z}_{2,j})$ are the same. On the other hand, the noise in the patch with a smaller mean value has more chance to be perceived. There is a possible mismatch between the SSIM index and the perceptual quality of an image. Based on this observation, the SSIM index is improved as

$$\hat{S}(\mathbf{Z}_1, \mathbf{Z}_2) = \frac{1}{P}\sum_{k=1}^{P} \hat{S}(\mathbf{Z}_{1,k}, \mathbf{Z}_{2,k}), \quad (2)$$

where $\hat{S}(\mathbf{Z}_{1,k}, \mathbf{Z}_{2,k})$ is computed as

$$\hat{S}(\mathbf{Z}_{1,k}, \mathbf{Z}_{2,k}) = \frac{2\mu_{\mathbf{Z}_{1,k}}\mu_{\mathbf{Z}_{2,k}} + c_1}{\mu^2_{\mathbf{Z}_{1,k}} + \mu^2_{\mathbf{Z}_{2,k}} + c_1} \frac{2\zeta_{3,k}\sigma_{\mathbf{Z}_{1,k}\mathbf{Z}_{2,k}} + c_2}{\zeta_{1,k}\sigma^2_{\mathbf{Z}_{1,k}} + \zeta_{2,k}\sigma^2_{\mathbf{Z}_{2,k}} + c_2}, \quad (3)$$

$\zeta_{1,k}$, $\zeta_{2,k}$ and $\zeta_{3,k}$ are three weighting factors, and they are computed as

$$\zeta_{m,k} = \frac{\mu^{2\gamma}_{\mathbf{Z}_m} + \epsilon}{\mu^{2\gamma}_{\mathbf{Z}_{m,k}} + \epsilon}\ ;\ m = 1, 2, \quad (4)$$

$$\zeta_{3,k} = \frac{\mu^{\gamma}_{\mathbf{Z}_1}\mu^{\gamma}_{\mathbf{Z}_2} + \epsilon}{\mu^{\gamma}_{\mathbf{Z}_{1,k}}\mu^{\gamma}_{\mathbf{Z}_{2,k}} + \epsilon}, \quad (5)$$

$\mu_{\mathbf{Z}_1}$ and $\mu_{\mathbf{Z}_2}$ are the mean values of $\mathbf{Z}_1$ and $\mathbf{Z}_2$, respectively. $\gamma(\geq 0)$ and $\epsilon(\geq 0)$ are two parameters. When their values are 0's, the improved SSIM (iSSIM) index is the SSIM index in [2]. Clearly, smaller weighting factors are assigned to those local patches with larger mean values. The iSSIM index is adaptive to local intensities of two images to be compared.

Similar to the SSIM index in [2], it can be proved that 1) $\hat{S}(\mathbf{Z}_1, \mathbf{Z}_2) = \hat{S}(\mathbf{Z}_2, \mathbf{Z}_1)$; 2) the value of $\hat{S}(\mathbf{Z}_1, \mathbf{Z}_2)$ is not greater than 1; and 3) the value of $\hat{S}(\mathbf{Z}_1, \mathbf{Z}_2)$ is 1 if and only if two images $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are the same.

Two simple examples are given to compare the proposed iSSIM index and the SSIM index in [2].

Example 1: Assume that there are two pairs of $(\mathbf{Z}_{1,k}, \mathbf{Z}_{2,k})(k = i, j)$ satisfying

$$\mu_{\mathbf{Z}_{1,k}} = \mu_{\mathbf{Z}_{2,k}}\ ;\ \sigma^2_{\mathbf{Z}_{2,k}} = \sigma^2_{\mathbf{Z}_{1,k}} + \varsigma\ ;\ \sigma_{\mathbf{Z}_{1,k}\mathbf{Z}_{2,k}} = \sigma^2_{\mathbf{Z}_{1,k}}. \quad (6)$$

The values of $\hat{S}(\mathbf{Z}_{1,k}, \mathbf{Z}_{2,k})(k = i, j)$ are then computed as

$$\hat{S}(\mathbf{Z}_{1,k}, \mathbf{Z}_{2,k}) = \frac{2\sigma^2_{\mathbf{Z}_{1,k}} + c_2\frac{\mu^{2\gamma}_{\mathbf{Z}_{1,k}}+\epsilon}{\mu^{2\gamma}_{\mathbf{Z}_1}+\epsilon}}{2\sigma^2_{\mathbf{Z}_{1,k}} + \varsigma + c_2\frac{\mu^{2\gamma}_{\mathbf{Z}_{1,k}}+\epsilon}{\mu^{2\gamma}_{\mathbf{Z}_1}+\epsilon}}. \quad (7)$$

Example 2: Assume that there are two pairs of $(\mathbf{Z}_{1,k}, \mathbf{Z}_{2,k})(k = i, j)$ satisfying

$$\mu_{\mathbf{Z}_{1,k}} = \mu_{\mathbf{Z}_{2,k}}\ ;\ \sigma^2_{\mathbf{Z}_{2,k}} = (1 + 2\alpha)\sigma^2_{\mathbf{Z}_{1,k}} + (1 + 2\beta)\varsigma, \quad (8)$$

$$\sigma_{\mathbf{Z}_{1,k}\mathbf{Z}_{2,k}} = (1 + \alpha)\sigma^2_{\mathbf{Z}_{1,k}} + \beta\varsigma, \quad (9)$$

where $\alpha(> -1)$ and β are two constants. The values of $\hat{S}(\mathbf{Z}_{1,k}, \mathbf{Z}_{2,k})(k = i, j)$ are then given as

$$\hat{S}(\mathbf{Z}_{1,k}, \mathbf{Z}_{2,k}) = \frac{2(1+\alpha)\sigma^2_{\mathbf{Z}_{1,k}} + 2\beta\varsigma + c_2\frac{\mu^{2\gamma}_{\mathbf{Z}_{1,k}}+\epsilon}{\mu^{2\gamma}_{\mathbf{Z}_1}+\epsilon}}{2(1+\alpha)\sigma^2_{\mathbf{Z}_{1,k}} + (1+2\beta)\varsigma + c_2\frac{\mu^{2\gamma}_{\mathbf{Z}_{1,k}}+\epsilon}{\mu^{2\gamma}_{\mathbf{Z}_1}+\epsilon}}. \quad (10)$$

It follows from Equations (7) and (10) that if $\mu_{\mathbf{Z}_{1,i}} > \mu_{\mathbf{Z}_{1,j}}$ and $\sigma^2_{\mathbf{Z}_{1,i}} = \sigma^2_{\mathbf{Z}_{1,j}}$, $\hat{S}(\mathbf{Z}_{1,i}, \mathbf{Z}_{2,i})$ is then greater than $\hat{S}(\mathbf{Z}_{1,j}, \mathbf{Z}_{2,j})$. Therefore, the proposed iSSIM index is consistent with the following fact on the perceptual quality of an image: *When two local patches have the same complexity, the brighter local patch can tolerate more distortion.* On the other hand, it can be verified that the SSIM index in [2] is not consistent with the above fact on the perceptual quality of an image.

Fig. 2. "Avion". The 1st image is the original image and the remaining 5 images contain different distortions caused by the compression of JPEG 2000.

3. EXTENSION OF THE PROPOSED ISSIM INDEX

Let $\Lambda_{1,2}(z)$ and $\Lambda_{2,1}(z)$ be the intensity mapping functions (IMFs) from $\mathbf{Z}_1$ to $\mathbf{Z}_2$ and vice versa, respectively [7]. $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are first

Fig. 3. "Barbara". The 1st image is the original image and the remaining 5 images contain different distortions caused by the compression of JPEG.

bidirectionally mapped by using the IMFs $\Lambda_{1,2}(z)$ and $\Lambda_{2,1}(z)$ as

$$\tilde{Z}_1(p) = \begin{cases} \Lambda_{1,2}(Z_1(p)); & \text{if } w(Z_1(p)) \geq w(Z_2(p)) \\ Z_1(p); & \text{otherwise} \end{cases},$$

$$\tilde{Z}_2(p) = \begin{cases} Z_2(p); & \text{if } w(Z_1(p)) \geq w(Z_2(p)) \\ \Lambda_{2,1}(Z_2(p)); & \text{otherwise} \end{cases},$$

where $p(= (x, y))$ represents a pixel point, the weighting function $w(z)$ is a triangular function as [5]:

$$w(z) = \begin{cases} z + 1; & \text{if } z \leq 127 \\ 256 - z; & \text{otherwise} \end{cases}.$$

An IMF based iSSIM index denoted as extended SSIM (ESSIM) is then defined for two differently exposed images $\mathbf{Z}_1$ and $\mathbf{Z}_2$ as follows:

$$\tilde{S}(\mathbf{Z}_1, \mathbf{Z}_2) = \frac{1}{P}\sum_{k=1}^{P} \hat{S}(\tilde{\mathbf{Z}}_{1,k}, \tilde{\mathbf{Z}}_{2,k}). \tag{11}$$

4. EXPERIMENTAL RESULTS

4.1. Performance of the Proposed iSSIM Index

The proposed iSSIM index and the SSIM index in [2] are compared by testing two sets of images from the VQEG [8] and they contain different noises as shown in Figs. 2 and 3. The values of γ and ϵ are selected as 1 and $c_1/2$, respectively. It is shown from Tables 1 and 2 that the values of $\hat{S}(\mathbf{Z}_1, \mathbf{Z}_2)$ are usually smaller than the values of $S(\mathbf{Z}_1, \mathbf{Z}_2)$.

Table 1. Comparison of SSIM and iSSIM for the image sequence in Fig. 2.

Pair	1	2	3	4	5
SSIM	0.9468	0.9056	0.8722	0.8524	0.8015
iSSIM	0.9561	0.9034	0.869	0.848	0.7952

Table 2. Comparison of SSIM and iSSIM for the image sequence in Fig. 3.

Pair	1	2	3	4	5
SSIM	0.8938	0.8543	0.8024	0.7633	0.7094
iSSIM	0.8755	0.8352	0.7833	0.7444	0.6906

Fig. 4. Sequence of "Fusionopolis" with different exposures.

4.2. Performance of the Proposed ESSIM Index

The SSIM index in [2] and the proposed ESSIM index are first compared by testing two differently exposed images of static HDR scenes [5], namely Fusionopolis and Memorial. The Fusionopolis is shown in Figs. 4 and the Memorial is given in [5]. The experimental results are shown in Tables 3 and 4, respectively. They are also compared by studying two HDR scenes with moving objects, as illustrated in Figs. 5 and 6. The experimental results for these two image sequences are illustrated in Tables 5 and 6, respectively. It is shown in Tables 3-6 that the gap between 1 and $\tilde{S}(\mathbf{Z}_1, \mathbf{Z}_2)$ is smaller than the gap between 1 and $S(\mathbf{Z}_1, \mathbf{Z}_2)$. Thus, the proposed ESSIM index is more robust than the SSIM index in [2] with respect to large intensity change between two images to be compared. Finally, these two indices are compared by testing five pairs of images that are are captured from different scenes but look somewhat similar as in Fig. 7. The experimental results are shown in Table 7. The values of $\tilde{S}(\mathbf{Z}_1, \mathbf{Z}_2)$ are smaller than the values of $S(\mathbf{Z}_1, \mathbf{Z}_2)$. Therefore, the proposed ESSIM index is more sensitive than the SSIM index in [2] to two images from different scenes.

Overall, the proposed ESSIM index is more robust to intensity changes of two images from the same scene and more sensitive to two images from different scenes than the SSIM index in [2].

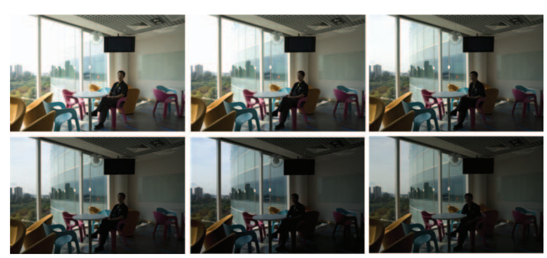

Fig. 5. Sequence of "Pantry" with different exposures and a moving human subject.

5. CONCLUSION AND DISCUSSION

In this paper, an existing structural similarity (SSIM) index is first improved in the sense that the improved SSIM (iSSIM) index is adaptive to local intensities of two images to be compared. The iSSIM index is further extended by using intensity mapping functions (IMFs) between two images to be compared. Experimental results show that the extended index is more robust to intensity changes of two images from the same scene and more sensitive to two images from different scenes.

The iSSIM index can be adopted to address rate distortion optimization (RDO) and rate control of video coding. In addition, it would be more convincing to compare the proposed indices and the SSIM index in [2] by finding the correlation coefficient between them and the mean opinion score (MOS). All these problems will be studied in our future research.

References

[1] R. C. Gonzalez and R. E. Woods, *Digital image processing*, Prentice Hall, 2002.

[2] Z. Wang, A. C. Bovik, H. Sheikh, and E. Simonceli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp.600-612, Apr. 2004.

[3] Y. H. Huang, T. S.Ou, P. Y. Su, and Homer H. Chen, "Perceptual rate-distortion optimization using structural similarity index as quality metric," *IEEE Trans. Circuits and System for Video Technology*, vol. 20, no. 11, pp. 1614-1624, Nov. 2010.

[4] T. S. Ou, Y. H. Huang, and Homer H. Chen, "SSIM-based perceptual rate control for video coding," *IEEE Trans. Circuits System for Video Technology*, vol. 21, no.5, pp. 682-691, May 2011.

[5] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photograph," In *Proceedings SIGGRAPH 1997*, pp.369-378, 1997.

[6] T. Mertens, J. Kautz, and F. V. Reeth, " Exposure fusion: a simple and practical alternative to high dynamic range photography," *Computer Graphics Forum*, Vol. 28, pp.161-171, 2009.

[7] M. Grossberg and S. Nayar, "Determining the camera response from images: what is knowable?" *IEEE Trans. on pattern analysis and machine intelligence*, pp. 1455-1467, Vol. 25, No. 11, Nov. 2003.

[8] http://www.its.bldrdoc.gov/vqeg/

Table 3. Comparison of SSIM and ESSIM for "Memorial".

$\frac{\Delta t_1}{\Delta t_2}$	2	2^2	2^3	2^4	2^5
SSIM	0.8536	0.6794	0.488	0.3481	0.2769
ESSIM	0.9375	0.9493	0.9383	0.937	0.9409

Fig. 6. Sequence of "Sky Garden" with different exposures and moving human subjects.

Table 4. Comparison of SSIM and ESSIM for "Fusionopolis".

$\frac{\Delta t_1}{\Delta t_2}$	2	2^2	2^3	2^4	2^5
SSIM	0.8661	0.56	0.3389	0.2176	0.1328
ESSIM	0.9416	0.904	0.8398	0.7874	0.7619

Table 5. Comparison of SSIM and ESSIM for "Pantry".

$\frac{\Delta t_1}{\Delta t_2}$	1.6	1.6^2	1.6^3	1.6^4	1.6^5
SSIM	0.8765	0.7662	0.5334	0.4149	0.4521
ESSIM	0.9095	0.8817	0.7455	0.7721	0.8131

Table 6. Comparison of SSIM and ESSIM for "Sky Garden".

$\frac{\Delta t_1}{\Delta t_2}$	2	2^2	2^3	2^4	2^5
SSIM	0.9013	0.813	0.6806	0.6142	0.496
ESSIM	0.937	0.9237	0.8637	0.892	0.8812

Fig. 7. Five pairs of images from different scenes.

Table 7. Comparison of SSIM and ESSIM for images in Fig. 7.

Pair	1	2	3	4	5
SSIM	0.4772	0.4771	0.4687	0.6971	0.4507
ESSIM	0.3997	0.4123	0.4489	0.5409	0.304