FAST STRUCTURAL SIMILARITY INDEX ALGORITHM

Ming-Jun Chen and Alan C. Bovik

Laboratory for Image and Video Engineering (LIVE)
The University of Texas at Austin, USA

ABSTRACT

The development of real-time image quality assessment algorithms is an important direction on which little research has focused. This paper presents a design of real-time implementable full-reference image quality algorithms based on the SSIM index [2] and multi-scale SSIM (MS-SSIM) index [3]. The proposed algorithms, which modify SSIM/MS-SSIM to achieve speed, are tested on the LIVE image quality database [13] and shown to yield performance commensurate with SSIM and MS-SSIM but with much lower computational complexity.

***Index Terms*—** Real time, MS SSIM

1. INTRODUCTION

With the increasing prevalence of digital images and videos, people live in an era full of digitized visual information. Effective systems for automatic image quality differentiation are thus urgently needed to help people manage the abundance of presented digital visual content. In the field of full-reference (FR) image quality assessment research, several algorithms have been proposed and studied for assessing image or video quality. For instance, "Yonsei" has been recommended by the VQEG group as an FR quality assessment method in the J.144 document [1], while the Structural SIMilarity (SSIM) index [2] is widely used algorithm in FR image quality assessment applications. A number of algorithms have been derived from SSIM: Multi-scale SSIM (MS-SSIM) [3], Percentile Pooling SSIM (P-SSIM) [3], Complex-Wavelet SSIM index (CW-SSIM) [5], Gradient-based Structural Similarity (G-SSIM) [6], and Three-Component Weighted SSIM [7]. All these derivative algorithms aim to improve the accuracy but inevitably increase the computational complexity.

The surge of mobile applications has created a demand for low complexity algorithms that can run on mobile devices. Multimedia services, such as Video On Demand (VOD) and IPTV are already available on mobile devices, further necessitating algorithms that have low complexity. Here we focus on reducing the computational complexity of SSIM, and propose a low complexity SSIM index, which we term Fast SSIM, which performs at a level comparable to the SSIM index. We also extend the Fast SSIM concept to the Multi-Scale SSIM index, which has better performance than single scale SSIM. The resulting Fast MS-SSIM algorithm performs commensurate with that of MS-SSIM. The rest of this paper is organized as follows. Section 2 reviews the SSIM and MS-SSIM indices. Section 3 describes the luminance term of Fast SSIM. In Section 4, the contrast term and the structural term of FAST are elaborated. Section 5 explains how optimization can be applied to Fast SSIM and Fast MS-SSIM. Experiments are presented in Section 6, and concluding remarks are offered in Section 7.

2. STRUCTURAL SIMILARITY INDEX

2.1. Single Scale Structural Similarity Index

Based on the hypothesis that the HVS is highly adapted for extracting structural information, the SSIM algorithm assesses three terms between two non-negative signals x and y: luminance $l(x, y)$, contrast $c(x, y)$, and structure $s(x, y)$:

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

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

$$s(x,y)=\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3}$$

where $C_1 = (K_1L)^2$, $C_2 = (K_2L)^2$, and $C_3 = C_2/2$ are small constants; L is the dynamic range of the pixel values, and $K_1 \ll 1$ and $K_2 \ll 1$ are scalar constants. The constants C_1, C_2 and C_3 provide spatial masking properties and ensure stability when the denominator approaches zero. Combining the three terms, the general form of SSIM is:

$$SSIM(x,y)=\left[l(x,y)\right]\ \left[c(x,y)\right]\ \left[s(x,y)\right]$$

$$=\frac{\left(2\mu_x\mu_y+C_1\right)\left(2\sigma_{xy}+C_2\right)}{\left(\mu_x^2+\mu_y^2+C_1\right)\left(\sigma_x^2+\sigma_y^2+C_2\right)}$$

The overall SSIM index value between x and y is defined as the average of all SSIM index values calculated within an 11x11 isotropic Gaussian weighting window passed over image, although other "pooling" strategies exist [4]. The Gaussian weighting window prevents artifacts arising from a discontinuous truncation of the image when computing the local values of the SSIM index map.

2.2. Multi-scale Structural Similarity Index

The distance between the image and the observer affects the observer's perceived image quality. The results of subjective image tests will vary as the viewing distance changes. In addition, images are naturally multi-scale, and both distortions and image features possess multi-scale attributes. For these reasons, the Multi-scale SSIM (MS-SSIM) index was developed.

In MS-SSIM, quality assessment is performed on multiple scales of the reference and the distorted images. Low-pass filtering and dyadic down-sampling is applied iteratively, and elements of the SSIM index applied at each scale, indexed from 1 (original image) through and the finest scale M obtained after $M-1$ iterations.

At each scale i, the contrast and structure terms are calculated: $c_j(x, y)$ and $s_j(x, y)$ respectively. The luminance term is computed only at scale M and represented as $l_M(x, y)$. The overall quality evaluation is obtained by combining the measurement over scales:

$$MS\text{-}SSIM\,(x,y) = [l_M(x,y)]^{\alpha_M} \prod_{j=1}^{M} [c_j(x,y)]^{\beta_j} [s_j(x,y)]^{\gamma_j}$$

where, typically M=5, and the exponents $\sigma_M, \beta_j, \gamma_j$ are selected such that $\sigma_M = \beta_j = \gamma_j$ and $\sum_{j=1}^{M} \gamma_j = 1$ [3].

3. LUMINANCE TERM

As noted by Rouse and Hemami [8], the luminance term of the SSIM index often plays a less significant perceptual role in predicting visual quality that the other terms. They propose eliminating it to reduce complexity. We choose to preserve the luminance term since images may suffer from a luminance bias, even if image quality databases do not explicitly include such distortions. Nevertheless, we have sought to expend as little computation as possible on the luminance term.

The luminance term in Fast SSIM utilizes an 8x8 square window, and an integral image technique [9] to compute the luminance similarity between the reference and test images.

By utilizing the so-called integral image, extracting the mean value of the pixels within a square window can be made quite efficient. As shown in **Fig. 1**, the value of the integral image at (x, y) is the sum of the pixels values above and to the left of (x, y), and including the value at (x, y).

Computing the sum over any rectangular area can be achieved with only two additions and one subtraction. As shown in **Fig. 1**, the sum of the pixel values within the rectangle D can be computed using four array references. The value of the integral image at location 1 is the sum of the pixels in rectangle A. The value at location 2 is $A+B$, at location 3 is $A+C$, and at location 4 is $A+B+C+D$. The sum over region D can be computed as '4'+'1'-('2'+'3') where 'i' is the value of the integral image at location i.

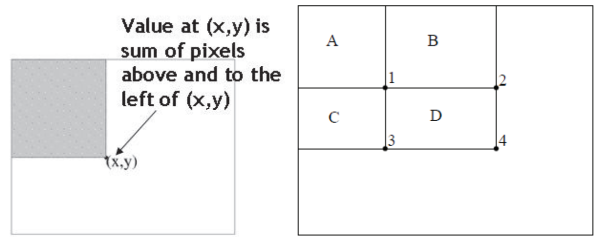

Fig. 1. Left: Integral image. Right: How to compute sum value over region D in integral image domain.

Using the integral image [9] and a square window, the complexity of computing the luminance term is reduced considerably. Assuming the window size is nxn, the standard SSIM index algorithm (using a Gaussian weighted window) requires n^2 multiplies and (n^2-1) additions to calculate the mean value, while the proposed Fast SSIM algorithm only requires 3 additions and 1 subtraction.

4. CONTRAST AND STRUCTURE TERMS

The computation of the variance term is the most time consuming part of the SSIM algorithm. In order to lower the complexity, we substitute a gradient value in Fast SSIM. Following Field [10], while images of real-world scenes vary greatly in their absolute luma and chroma distributions, the gradient magnitudes of natural images generally obey heavy tailed distribution laws. Indeed, some no-reference image quality assessment algorithms [11], [12] use the gradient image to assess blur severity. Similarly, the performance of the Gradient-based SSIM index [6] suggests that applying SSIM on the gradient magnitude may yield higher performance. The gradient is certainly responsive to image variation. Moreover, the gradient magnitude has low complexity and is amenable to integer-only implementation.

We generate the gradient image using the Roberts gradient templates depicted in **Fig. 2**.

-1	0
0	1

0	-1
1	0

Fig. 2. Roberts gradient templates.

The gradient magnitude is approximated by[1]

$$|\nabla I| = \max\{|\nabla i|, |\nabla j|\} + (1/4)\min\{|\nabla i|, |\nabla j|\} \qquad (1)$$

where ∇i and ∇j are the Roberts template responses in the two orthogonal directions. This approximation is based upon a simple expansion of the gradient. The contrast $c(x, y)$ and

[1] An accurate truncated expansion approximation learned by author ACB whilst lecturing at Texas Instruments in the 1990s.

structure $s(x,\ y)$ terms of the Fast SSIM index algorithm are then defined:

$$c(x,y) = \frac{(2\mu_{Gx}\mu_{Gy} + C_2)}{\left(\mu_{Gx}^2 + \mu_{Gy}^2 + C_2\right)}$$

$$s(x,y) = \frac{\left(\mu_{GxGy} + C_3\right)}{\left(\mu_{Gx}\mu_{Gy} + C_3\right)}$$

where $C_3 = C_2/2$, and

$$\mu_{Gx} = \frac{1}{N}\sum_{i=1}^{N} \left|\nabla x_i\right|$$

$$\mu_{GxGy} = \frac{1}{N}\sum_{i=1}^{N} \left|\nabla x_i\right|\left|\nabla y_i\right|$$

where $\left|\nabla x_i\right|$ and $\left|\nabla y_i\right|$ are the gradient magnitude values of the images x and y at spatial coordinate i, estimated using the approximation (1).

The Fast SSIM index between signals x and y is then:

$$\textit{Fast-SSIM}\,(x,y) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\mu_{GxGy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\mu_{Gx}^2 + \mu_{Gy}^2 + C_2\right)}.$$

In MS-SSIM, the contrast and structural terms are calculated over multiple scales. Therefore, the Fast MS-SSIM index between signal x and y is defined as:

$$\textit{Fast-MS-SSIM}\,(x,y) = [l_M(x,y)]^{\alpha_M}\prod_{j=1}^{M}[c_j(x,y)]^{\beta_j}[s_j(x,y)]^{\gamma_j}$$

where $M = 5$, the exponents $\sigma_M, \beta_j, \gamma_j$ are selected such that $\sigma_M = \beta_j = \gamma_j$ and $\sum_{j=1}^{M}\gamma_j = 1$.

Another modification that we make is to use an integer approximation to the Gaussian weighting window. In this way, the computation of the contrast and structural terms uses only integer operations. **Fig. 3** shows the window.

The simplifications include: all computations reduced to integer operations, with square roots eliminated.

0	0	0	1	1	0	0	0
0	0	1	2	2	1	0	0
0	1	2	4	4	2	1	0
1	2	4	8	8	4	2	1
1	2	4	8	8	4	2	1
0	1	2	4	4	2	1	0
0	0	1	2	2	1	0	0
0	0	0	1	1	0	0	0

Fig. 3. 8x8 integer approximation to Gaussian window.

5. OPTIMIZATION

Optimization is an essential process when implementing an algorithm for industrial applications. We propose to apply parallel computing and sub-sampling on the Fast SSIM index algorithm in order to achieve the best performance.

Data-level parallelization and frame–level parallelization are adopted in this study to optimize the parallel computing. After optimizing computation of the luminance part, about 80% of the computation is consumed on the contrast and structure terms. Since most operations in Fast SSIM are integer-only, Fast SSIM is amenable to Single Instruction Multiple Data (SIMD) optimization. Also, since Fast SSIM does not currently use any dependency between frames, it is natural to conduct frame level parallelization.

Regarding sub-sampling, we suggest that the contrast and structure terms need not be computed at the original scale in Fast MS-SSIM. Since humans are less sensitive to higher spatial frequencies, skipping computation of the contrast and structure terms at the first scale appears to not lower performance, but it does increase the computation speed dramatically. The experiment results shown in the next section support this assumption.

6. EXPERIMENTAL RESULTS

The LIVE database of images [13] was used in the following experiment. The database includes DMOS subjective scores for each image and 6 types of distortions. The distortions include JPEG2000 compression distortion (227 images), JPEG compression distortion (233 images), white noise (174 images), gaussian blur (174 images), and fast fading channel noise (174 images).

Fast-SSIM was evaluated against the LIVE DMOS scores using the Spearman Rank Order Correlation Coefficient (SROCC). The performance numbers on speed were tested on a 768x432 video with 250 frames. All experiments were conducted on a Intel Core2Duo 2.2 GHz platform, except the experiment result on multi-threading optimization, which was run on an Intel Core2 Quad Q6600 platform.

Algorithm	SROCC	Speed(fps)
SSIM	0.9244	3.42
Fast SSIM (*Luminance term optimized only*)	0.9204	5.09
Fast SSIM	0.9214	9.17
Fast SSIM (*SIMD*)	0.9214	16.6
Fast SSIM (*SIMD+multi-threading*)	0.9214	57.83

Table 1. SROCC and speeds of SSIM algorithms.

Algorithm	SROCC	Speed(fps)
MS-SSIM	0.9429	2.54
Fast MS-SSIM (*Luminance term optimized only*)	0.9409	3.74
Fast MS-SSIM	0.9409	6.4
Fast MS-SSIM (***Proposed algorithm*** *using sub-sampling*)	0.9409	25.31
Fast MS-SSIM (*SIMD+sub-sampling*)	0.9409	35.34
Fast MS-SSIM (*SIMD+sub-sampling +multi-threading*)	0.9409	121.97

Table 2. SROCC and speeds of MS-SSIM algorithms.

Table 1 makes it clear that the Fast SSIM and Fast MS-SSIM algorithms suffer no performance loss in terms of their subjective quality prediction capability; the SROCC scores are very close. However, looking at the performance improvement on speed, the improvement from SSIM to Fast SSIM is 168% (from 3.42 fps to 9.17 fps). Thus Fast SSIM is 2.68 times faster than SSIM. For optimization, Intel SSE2 instructions were implemented to calculate the mean and correlation of the gradient images, to demonstrate the improvement on applying data-level parallelization. As shown in **Table 1**, Fast SSIM with SIMD enhances the performance from 9.7 fps to 16.6 fps. Finally, with multi-threading optimization, Fast SSIM reaches 57.83 fps on an Intel Core2 Quad platform, which qualifies the algorithm for real-time application.

The performance numbers for Fast MS-SSIM are shown in Table 2. The modifications in Fast MS-SSIM are the same as the modification in Fast SSIM, except that we propose to skip the analysis on contrast and structural terms on the original scales. Table 2 shows that the SROCC scores of Fast MS-SSIM and Fast MS-SSIM with sub-sampling are very close, but both are a little lower than MS-SSIM. However, if we compare the performance of Fast MS-SSIM with Fast MS-SSIM with sub-sampling, Fast MS-SSIM with sub-sampling yields better performance for assessing image quality, at speeds adequate for real-time application.

7. CONCLUDING REMARKS

In this paper we proposed Fast SSIM and Fast MS-SSIM index algorithms and verified their performance on the LIVE image database [13]. The experimental results show that the proposed algorithms not only have competitive performance with SSIM and MS-SSIM for assessing image quality, but have much lower computational complexity. Indeed, the proposed algorithms achieve real-time performance with simple optimization.

8. REFERENCES

[1] LC-Text: Recommendation J.144 (Rev.1) - Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference.

[2] Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans Image Process.,* vol. 13, no. 4, pp. 600-612, Apr. 2004.

[3] Z. Wang, E.P. Simoncelli, and A.C. Bovik, “Multi-scale structural similarity for image quality assessment,” *IEEE Asilomar Conf. Signals, Syst, Comput,* Nov. 2003.

[4] A.K. Moorthy and A.C. Bovik, “Visual Importance Pooling for Image Quality Assessment,” *IEEE J Selected Topics Signal Process.,* vol. 3, no. 2, pp. 193-201, April 2009.

[5] Z. Wang, E. P. Simoncelli, Translation insensitive image similarity in complex wavelet domain, *IEEE Int'l Conf. Acoustics, Speech, and Signal Process* 2, pp. 573-576, 2005.

[6] G.H. Chen, C.L. Yang and S.L. Xie, “Gradient-based structural similarity for image quality assessment,” *Proc. ICIP*, *2929-2932*, 2006.

[7] C. Li and A.C. Bovik**,** "Three-Component Weighted Structural Similarity Index"**,** *SPIE Conference on Image Quality and System Performance*, Jan 19-22, 2009, San Jose, California.

[8] D.M. Rouse and S.S. Hemami, “Understanding and simplifying the structural similarity metric,” *IEEE Int. Conf. on Image Process. (ICIP)*, Oct. 2008

[9] F. Porikli, “Integral histogram: A fast way to extract histograms in Cartesian spaces,” *IEEE CVPR,* 2005.

[10] D. Field, “What is the goal of sensory coding?” *Neural Computational,* vol. 6 pp. 559-601, 1994.

[11] M.-J. Chen and A.C. Bovik, “No-Reference Image Blur Assessment Using Multi-Scale Gradient,” *Int’l Wkshp on Quality of Multimedia Experience,* 2009.

[12] L. Renting, L. Zhaorong, J. Jiaya, “Image partial blur detection and classification,” *IEEE CVPR*, 2008.

[13] 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