A NOVEL APPROACH FOR COMPUTING AND POOLING STRUCTURAL SIMILARITY INDEX IN THE DISCRETE WAVELET DOMAIN

Soroosh Rezazadeh , Stéphane Coulombe

École de technologie supérieure, Université du Québec, Montréal, Canada
soroosh.rezazadeh.1@ens.etsmtl.ca , stephane.coulombe@etsmtl.ca

ABSTRACT

The Structural SIMilarity (SSIM) index is an objective metric that gives relatively accurate similarity prediction scores with reasonable complexity. In this paper, an excellent trade-off between accuracy and complexity is presented in the form of a wavelet structural similarity index (WSSI), which is more accurate and less complex than the spatial SSIM index. Like the spatial SSIM index, the WSSI has the feature of boundedness. It computes an edge structural similarity map and an approximation structural similarity map to obtain the final similarity score. A contrast map is introduced in the wavelet domain for pooling structural similarity maps. Experimental results show that the low-complexity WSSI gives a correlation coefficient of 0.9548 between objective and subjective scores, and competes with visual information fidelity (VIF) performance.

Index Terms— Structural similarity, image quality assessment, wavelet transform

1. INTRODUCTION

Generally speaking, the full-reference (FR) quality assessment of image signals involves two categories of approach: bottom-up and top-down [1]. In the bottom-up approaches, the perceptual quality scores are best estimated by quantifying the visibility of errors. These methods have several important limitations, which can be studied in [1]. In the top-down approaches, the whole human visual system (HVS) is considered as a black box, and the input/output relationship is computed.

One of the main methods in the top-down category is the Structural SIMilarity (SSIM) index [2], which gives an accurate score with acceptable computational complexity compared to other quality metrics [3]. SSIM has attracted a great deal of attention in recent years, and has been considered for a range of applications. The principal idea underlying the SSIM approach is that the HVS is highly adapted to extracting structural information from visual scenes, and, therefore, a measurement of structural similarity (or distortion) should provide a good approximation of perceptual image quality. Some approaches have tried to improve the SSIM index. The Multi-scale SSIM [4] attempts to increase SSIM assessment accuracy by incorporating image details at different resolutions in the pixel domain. In [5], the authors investigate ways to simplify SSIM in the pixel domain. The authors in [6] propose to compute it using subbands at different levels in the discrete wavelet domain. Five-level decomposition using the Daubechies 9/7 wavelet is applied to both original and distorted images, and then SSIM is computed between corresponding subbands. Finally, the similarity score is obtained by the weighted mean of all SSIMs. To determine the weights, a large number of experiments have been performed to measure the sensitivity of the human eye to different frequency bands.

In this paper, we propose a novel approach to calculating SSIM more accurately, yet with less complexity, in the discrete wavelet domain. This method computes two image adaptive structural similarity scores and then linearly combines them. We developed the new approach mainly because of the following shortcomings of the current one. First, an SSIM map [2] computes local statistics within a local square window in the pixel domain, even though the statistics of blocks in the wavelet domain are more accurate. Second, in multi-scale and multi-level SSIMs [4,6], determining the sensitivity of the HVS to different scales or subbands requires many experiments. Moreover, if we change the wavelet or filter, the computed weights and parameters are no longer optimum and may not even be valid. Our new approach does not require such heavy experiments to determine parameters, and it is adaptive to different wavelet filters. Third, the five-level decomposition of images, as in [6], would make the size of the approximation subband very small, so it would no longer be able to help in the effective extraction of image statistics. In contrast, the approximation subband contains main image contents, and we have observed that this subband has a major impact on improving SSIM accuracy. Fourth, previous methods use the mean of the SSIM map to give the overall image quality score. However, distortions in various image areas have different impacts on the HVS. In our approach, we introduce a contrast map in the wavelet domain for pooling SSIM maps. This map is computed based on statistics previously calculated for SSIM maps.

2. STRUCTURAL SIMILARITY COMPUTATION

2.1. Description of the Proposed Method

Let **X** and **Y** denote the original and distorted images respectively. The procedure for calculating the proposed SSIM is described, and explained, in the following steps.

Step 1. We perform one-level discrete wavelet decomposition on both the original and the distorted images. With one-level decomposition, the approximation subbands are still large enough compared to the original images to provide accurate image statistics.

Step 2. We calculate the SSIM map between the approximation subbands of **X** and **Y**, and call it an approximation structural similarity map, $SSIM_A$. For each image patch $\mathbf{x}_A$ and $\mathbf{y}_A$ (of N pixels) within the approximation subbands of **X** and **Y**, $SSIM_A$ is computed as:

$$SSIM_A(\mathbf{x}_A, \mathbf{y}_A) = SSIM(\mathbf{x}_A, \mathbf{y}_A) \tag{1}$$

The SSIM map is calculated according to the method in [2].

Step 3. In this step, an edge-map function is defined for each image using the mean square of detail subbands.

$$\mathbf{X}_E(m,n) = \frac{1}{3}\left(\mathrm{X}_H^2(m,n) + \mathrm{X}_V^2(m,n) + \mathrm{X}_D^2(m,n)\right) \tag{2}$$

$$\mathbf{Y}_E(m,n) = \frac{1}{3}\left(\mathrm{Y}_H^2(m,n) + \mathrm{Y}_V^2(m,n) + \mathrm{Y}_D^2(m,n)\right) \tag{3}$$

where X_E and Y_E represent the edge maps of **X** and **Y** respectively; (m,n) shows the sample position within the wavelet subbands; X_H, X_V, and X_D denote horizontal, vertical, and diagonal detail subbands of image **X**; Y_H, Y_V, and Y_D are detail subbands of image **Y**. To simplify calculation of the edge maps, we have assumed that all detail subbands in the first level have the same sensitivity to the HVS; however, it is possible to calculate edge maps using a weighted squared sum. It is notable that the edge maps only reflect the fine-edge structures of images.

Step 4. The edge structural similarity map $SSIM_E$ is calculated between two images using the following formula:

$$SSIM_E(\mathbf{x}_E, \mathbf{y}_E) = \frac{2\sigma_{\mathrm{x}_E,\mathrm{y}_E} + c}{\sigma_{\mathrm{x}_E}^2 + \sigma_{\mathrm{y}_E}^2 + c} \tag{4}$$

$$c = (kL)^2\,, \quad k \ll 1 \tag{5}$$

where $\sigma_{\mathrm{x}_E,\mathrm{y}_E}$ is the cross correlation between image patches $\mathbf{x}_E$ and $\mathbf{y}_E$ (of X_E and Y_E); parameters $\sigma_{\mathrm{x}_E}^2$ and $\sigma_{\mathrm{y}_E}^2$ are variances of $\mathbf{x}_E$ and $\mathbf{y}_E$ respectively; k is a small constant; and L is a dynamic range of pixels (255 for gray-scale images). The correlation coefficient and variances are computed in the same manner as presented in [2]. In fact, as the edge map only forms image-edge structures and contains no luminance information, the luminance comparison part of the SSIM map in [2] is omitted for the edge structural similarity map.

Step 5. In this step, we form a contrast map function for pooling the approximation and edge SSIM maps. It is well known that the HVS is more sensitive to areas near the edges [1]. Therefore, the pixels in the SSIM map near the edges should be given more importance. On the other hand, high-energy (or high-variance) image regions are likely to contain more information to attract the HVS [7]. Thus, the pixels of an SSIM map within high-energy regions must also receive higher weights (more importance). Based on these facts, we can combine our edge map with the computed variance to form a contrast map function. Like approximation and edge SSIM maps, the contrast map is computed within a local Gaussian square window, which moves (pixel-by-pixel) over the entire edge map X_E and approximation subband X_A . As in [2], we define a Gaussian sliding window $\mathrm{W} = \{w_k | k = 1,2,\cdots,N\}$, with a standard deviation of 1.5 samples, normalized to unit sum. The contrast map is defined as follows:

$$Contrast(\mathrm{x}_E, \mathrm{x}_A) = (\mu_{x_E}\sigma_{x_A}^2)^{0.1} \tag{6}$$

$$\sigma_{x_A}^2 = \sum_{k=1}^{N} w_k (x_{A,k} - \mu_{x_A})^2 \tag{7}$$

$$\mu_{x_E} = \sum_{k=1}^{N} w_k x_{E,k} \quad , \quad \mu_{x_A} = \sum_{k=1}^{N} w_k x_{A,k} \tag{8}$$

It is notable that the contrast map just exploits the original image statistics to form the weighted function for SSIM map pooling. Fig. 1(b) demonstrates the resized contrast map, obtained by eq. (6), for a typical image in Fig. 1(a). As can be seen in Fig.1, the contrast map nicely shows the edges and important image structures to the HVS. Brighter (higher) sample values in the contrast map indicate image structures which are more important to the HVS and play an important role in judging image quality.

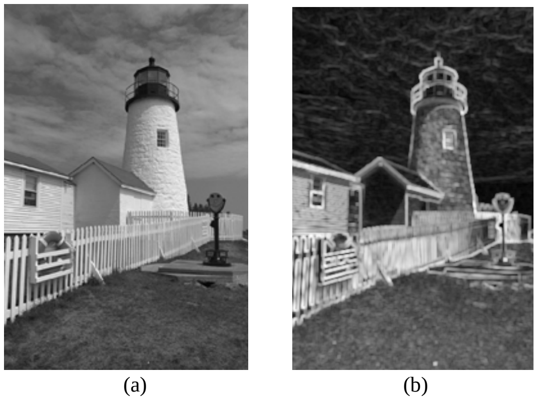

(a) (b)

Fig.1. (a) Original image; (b) Contrast map computed using eq. (6). The sample values of the contrast map scaled between [0,255] for easy observation.

Step 6. The contrast map in (6) is used for weighted pooling of the approximation SSIM map in (1) and the edge SSIM map in (4).

$$S_A = \frac{\sum_{j=1}^{M} Contrast(\mathrm{x}_{E,j}, \mathrm{x}_{A,j}) \cdot SSIM_A(\mathrm{x}_{A,j}, \mathrm{y}_{A,j})}{\sum_{j=1}^{M} Contrast(\mathrm{x}_{E,j}, \mathrm{x}_{A,j})} \tag{9}$$

$$S_E = \frac{\sum_{j=1}^{M} Contrast(\mathrm{x}_{E,j}, \mathrm{x}_{A,j}) \cdot SSIM_E(\mathrm{x}_{E,j}, \mathrm{y}_{E,j})}{\sum_{j=1}^{M} Contrast(\mathrm{x}_{E,j}, \mathrm{x}_{A,j})} \tag{10}$$

where $\mathbf{x}_{E,j}$, $\mathbf{y}_{E,j}$, $\mathbf{x}_{A,j}$, and $\mathbf{y}_{A,j}$ are image patches in the j-th local window; M is the number of samples in the SSIM maps; S_A and S_E represent the approximation and edge similarity scores respectively.

Step 7. Finally, we combine the approximation and edge similarity scores to obtain the overall quality measure between images $\mathbf{X}$ and $\mathbf{Y}$. A linear relationship is used to reach the final similarity score.

$$WSSI(\mathrm{X,Y}) = \alpha S_A + (1-\alpha) S_E \qquad (11)$$
$$0 < \alpha \leq 1$$

where *WSSI* gives the final Wavelet Structural Similarity Index score in the range [0,1], and α is a constant. As the approximation subband contains main image contents, α should be close to one to give the approximation similarity score much more weight. We set α = 0.94 in this paper.

2.2. Computational Complexity and Simplification

In spite of the number of steps required to calculate the *WSSI*, the computational complexity of the proposed algorithm is less than that of the SSIM presented in [2]. As in [9], we used MATLAB (v7.5.0 R2007b) for performance evaluation. We have observed that the running time for calculating the *WSSI* is, on average, about 0.65 of running time for SSIM calculation in the spatial domain. This test was conducted for a database of 1000 images. We discuss various different aspects of the complexity of the *WSSI*.

The resolution of the approximation subband and edge map is a quarter of that of the original image. Lower resolutions mean that fewer computations are required to obtain SSIM maps for the *WSSI*.

Because of the smaller resolution of the subbands in the wavelet domain, we can extract accurate local statistics with a smaller sliding window size. The spatial SSIM in [2] uses a window of size of 11×11 by default, while we show in the next section that the *WSSI* can provide accurate scores with a window of 4×4. A smaller window size reduces the number of computations required to obtain local statistics.

Probably the most complex part of the *WSSI* method is wavelet decomposition. Since the sensitivity of the *WSSI* to different wavelets is negligible, a simple wavelet can be used to reduce complexity. We used the Haar wavelet for image decomposition. As this wavelet has the shortest filter length, it makes the filtering process simpler.

As can be seen from eq. (6), the local statistics calculated for eq. (1) and eq. (4) are used to form the contrast map. Therefore, computing this map does not impose a large computational burden.

Further simplification of the *WSSI* is also possible if the luminance comparison part of SSIM in eq. (1) is ignored and an approximation SSIM map similar to eq. (4) is calculated. Based on our experiments, such simplification reduces the accuracy of the *WSSI* by just 0.03%, which is less than the 1% effect of spatial SSIM simplification [5].

3. SIMULATION RESULTS AND ANALYSIS

The performance evaluation of the proposed *WSSI* is carried out on *LIVE Image Quality Assessment Database Release 2* [8]. This database consists of 779 distorted images derived from 29 original color images using five types of distortion. Distortion types are JPEG compression, JPEG2000 compression, Gaussian white noise, Gaussian blurring, and the Rayleigh fast fading channel model. The realigned subjective quality data for the database are used in all experiments [8].

In this paper, three performance metrics are adopted to measure the performance of objective models. The first metric is the correlation coefficient (CC) between the Difference Mean Opinion Score (DMOS) and the objective model outputs after nonlinear regression. The correlation coefficient gives an evaluation of *prediction accuracy*. We use the five-parameter logistical function defined in [3] for nonlinear regression. The second metric is the root mean square error (RMSE) between DMOS and the objective model outputs after nonlinear regression. The RMSE is considered as a measure of *prediction consistency*. The third metric is Spearman rank order correlation coefficient (ROCC), which provides a measure of *prediction monotonicity*.

In order to put the performance evaluation of our method in the proper context, we compared the proposed *WSSI* against other quality metrics, including PSNR, Mean SSIM [2], DWT-SSIM [6], and Visual Information Fidelity (VIF) [9]. In the *WSSI* simulation, we used the Haar wavelet, $k = 0.03$, and a Gaussian window size of 4×4. The other metrics, except for VIF, were implemented and simulated with the default parameters described in their reference papers. In our simulations, we used an enhanced version of VIF implementation, which is available in [8].

When we calculate the RMSE for different α values in eq. (11), it reaches its minimum (global) for α = 0.94. This value of α meets our expectation that α should be close to one. We must note that CC has very low sensitivity to small variations in α, that is, the proposed α does not affect *WSSI* performance for the quality prediction of a different image database.

To better understand the effect of the wavelet transform in quality assessment, we considered a $SSIM_A$ mean as a separate objective quality assessment model. Obviously, a mean $SSIM_A$ has even lower complexity than the *WSSI*. Table 1 lists values of performance metrics for each objective model. It can be seen that the CC value for the mean $SSIM_A$ (0.9412) is higher than the CC value for DWT-SSIM (0.9346). This shows that we can calculate the similarity of images with very good precision by just considering their first-level approximation subband. The reason is that most of the useful image information is concentrated in the first-level approximation subband. As mentioned earlier, neglecting a luminance comparison in calculating the mean $SSIM_A$ has a negligible effect on performance (just 0.03% in CC). This makes a very low complexity metric with very good performance possible. Following other simple steps in our algorithm can raise the performance to reach the correlation coefficient of 0.9548 for the *WSSI*. While the complexity of the *WSSI* is much less than that of the VIF, its performance is very close to that of the enhanced VIF implementation.

Table 1. Performance comparison of image quality assessment models (all 779 distorted images included)

Model	CC	RMSE	ROCC
PSNR	0.8701	13.4685	0.8756
Mean SSIM [2]	0.9041	11.6736	0.9104
DWT-SSIM [6]	0.9346	9.7201	0.9346
VIF [8]	0.9593	7.7122	0.9635
Mean $SSIM_A$	0.9412	9.2270	0.9441
WSSI	**0.9548**	**8.1176**	**0.9586**

Fig.2 shows the scatter plots of DMOS versus mean SSIM and *WSSI* predictions for all the distorted images. It is evident that *WSSI* prediction is more consistent with the subjective scores than the mean SSIM.

Finally, we tested the *WSSI* with the previously defined parameters for various wavelet filters. We observed that the wavelet basis has very little effect on performance. The worst case is for the Daubechies 9/7 wavelet, which results in CC=0.9489, RMSE=8.6232, and ROCC=0.9529. These values are still quite acceptable, and so this model outperforms DWT-SSIM.

4. CONCLUSION

In this paper, we proposed a Wavelet Structural Similarity Index (*WSSI*) to improve the accuracy of spatial domain SSIM prediction, while keeping computational complexity as low as possible. To compute the *WSSI*, we defined a contrast map, which takes advantage of basic HVS characteristics, for discrete wavelet domain pooling of SSIM maps. Although the *WSSI* is much less complex than the VIF, its prediction scores are very close to VIF values and compete with them very well. Our results show that the first-level approximation subband of decomposed images has an important role to play in improving quality assessment performance and also complexity reduction. Since the ways of making these improvements that we have discussed here provide very good tradeoffs between accuracy and complexity, they can be used efficiently in wavelet-based image/video processing applications.

5. ACKNOWLEDGMENT

This work was funded by Vantrix and by the Natural Sciences and Engineering Research Council of Canada. The authors would like to thank Dr. Hamid R. Sheikh for providing the logistic function code used in [3].

6. REFERENCES

[1] Wang, Z., A.C. Bovik, *Modern Image Quality Assessment*, Morgan & Claypool, United States, 2006.

[2] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, April 2004.

[3] H.R. Sheikh, M.F. Sabir, and A.C. Bovik, “A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms,” *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3441-3452, November 2006.

[4] Z. Wang, E.P. Simoncelli, and A.C. Bovik, “Multi-Scale Structural Similarity for Image Quality Assessment,” *37th IEEE Asilomar Conference on Signals, Systems and Computers*, pp. 1398-1402, November 2003.

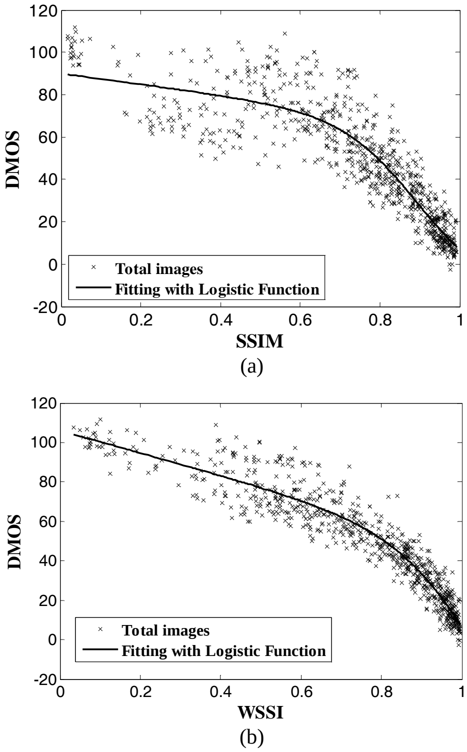
[5] D.M. Rouse, and S.S. Hemami, “Understanding and Simplifying the Structural Similarity Metric,” *IEEE International Conference on Image Processing*, San Diego, pp. 1188-1191, October 2008.

[6] C-L. Yang, W-R. Gao, and L-M. Po, “Discrete Wavelet Transform-based Structural Similarity for Image Quality Assessment,” *IEEE International Conference on Image Processing*, San Diego, pp. 377-380, October 2008.

[7] Z. Wang, X. Shang, “Spatial Pooling Strategies for Perceptual Image Quality Assessment,” *IEEE International Conference on Image Processing*, Atlanta, pp. 2945-2948, October 2006.

[8] H.R. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik, “LIVE Image Quality Assessment Database Release 2”, http://live.ece.utexas.edu/research/quality.

[9] H.R. Sheikh, A.C. Bovik, “Image Information and Visual Quality,” *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430-444, February 2006.



(a)

(b)

Fig.2. Scatter plots of DMOS versus model prediction for all 779 distorted images: (a) Mean SSIM model; (b) *WSSI* model.