SR-SIM: A FAST AND HIGH PERFORMANCE IQA INDEX BASED ON SPECTRAL RESIDUAL

Lin Zhang and Hongyu Li[*]
School of Software Engineering, Tongji University, Shanghai, China

ABSTRACT

Automatic image quality assessment (IQA) attempts to use computational models to measure the image quality in consistency with subjective ratings. In the past decades, dozens of IQA models have been proposed. Though some of them can predict subjective image quality accurately, their computational costs are usually very high. To meet real-time requirements, in this paper, we propose a novel fast and effective IQA index, namely *spectral residual based similarity* (SR-SIM), based on a specific visual saliency model, spectral residual visual saliency. SR-SIM is designed based on the hypothesis that an image's visual saliency map is closely related to its perceived quality. Extensive experiments conducted on three large-scale IQA datasets indicate that SR-SIM could achieve better prediction performance than the other state-of-the-art IQA indices evaluated. Moreover, SR-SIM can have a quite low computational complexity. The Matlab source code of SR-SIM and the evaluation results are available online at http://sse.tongji.edu.cn/linzhang/IQA/SR-SIM/SR-SIM.htm.

***Index Terms*—** IQA, visual saliency, spectral residual

1. INTRODUCTION

Propelled by numerous potential applications, quantitative evaluation of image perceptual quality has become one of the most fundamental yet challenging problems in image processing and vision research. According to the availability of a reference image, objective IQA indices fall into three categories: full reference (FR) methods, no-reference (NR) methods, and reduced-reference (RR) methods. In this paper, the discussion is confined to FR methods.

Since the conventional pixel-based IQA indices, such as the peak signal-to-noise ratio (PSNR), do not correlate well with human beings' subjective fidelity ratings, in the past decade, several sophisticated IQA indices have been proposed. Representative ones include the noise quality measure (NQM) index [1], the universal quality index (UQI) [2], the structural similarity (SSIM) index [3], the multi-scale SSIM (MS-SSIM) index [4], the information fidelity criterion (IFC) index [5], the visual information fidelity (VIF) index [6], the Riesz transforms based feature similarity (RFSIM) index [7], the information content weighted SSIM (IW-SSIM) index [8], and the feature similarity (FSIM) index [9].

An IQA index is usually used in optimizing visual processing algorithms or systems. Thus, a perfect IQA index should perform well in two aspects. At first, objective quality scores predicted by an IQA index need to highly correlate with subjective evaluations. Secondly, to be suitable for real-time applications, an IQA index should have a low computational cost. Through our investigation, we find that the computational cost is often ignored in designing IQA indices. Therefore, though several recently proposed IQA indices, such as IW-SSIM [8] and FSIM [9], could achieve outstanding prediction performances, their computational costs are very high, which limits their applications in practice.

Based on these considerations, in this paper, we propose a novel IQA index having a high prediction performance and a low computational cost simultaneously, namely *spectral residual based similarity* (SR-SIM). SR-SIM is based on an effective and efficient visual saliency model, *spectral residual visual saliency* (SRVS) [10]. In SR-SIM, SRVS map acts as two roles, a feature map characterizing the image's local quality, and a weighting function indicating the importance of a local region to the human visual system (HVS) when pooling the final quality score. The performance of SR-SIM is examined on three large-scale IQA image datasets and is compared with other nine state-of-the-art IQA indices. Efficacy and efficiency of SR-SIM are corroborated by the experimental results.

The rest of this paper is organized as follows. Section 2 states the relationship between visual saliency and the perceived quality. Section 3 presents our proposed SR-SIM IQA index. Section 4 gives the experimental results and related discussions. Finally, Section 5 concludes the paper.

2. VISUAL SALIENCY AND PERCEIVED QUALITY

Building effective computational models to simulate human visual attention has been studied by scholars for a long time. Most of the existing visual attention models are bottom-up visual saliency (VS) models since bottom-up attention

[*] Corresponding author. Email: hyli@tongji.edu.cn.

mechanisms is more thoroughly studied than top-down mechanisms [11]. In the literature, dozens of various VS models have been proposed and several of them could predict the human visual attention accurately [11].

Recently, researchers have gradually found that VS is closely relevant to the image's perceptual quality. Various methods have been attempted to integrate VS information into IQA metrics [12]. These methods share some common characteristics. First, a VS map is only exploited as a weighting function to reflect the different importance of different regions in the local quality map. Second, for these methods, the motivation is actually not to design a new IQA index but to demonstrate that a VS-weighted pooling strategy could perform better than the simple "mean" scheme. Thus, for computing the local quality map, they all adopt some existing methods, such as PSNR, SSIM, and VIF, without discussing whether there could be more effective methods to characterize the local image quality.

3. SPECTRAL RESIDUAL BASED SIMILARITY

Here we propose a novel VS-based IQA index. With respect to the visual saliency model, we adopt the *spectral residual visual saliency* (SRVS) [10], which has been proved to be efficient and effective. Thus, the proposed IQA index is named as *spectral residual based similarity* (SR-SIM).

3.1. Spectral residual visual saliency

Spectral residual visual saliency (SRVS) model was proposed in [10]. In SRVS, the spectral residual is obtained at first from the log spectrum of the examined image. And then, the VS map is obtained by transforming the spectral residual to the spatial domain. Spectral residual actually approximately represents the innovation part of an image by removing the statistical redundant components.

Suppose f is the examined image. According to [10], SRVS can be computed as the following,

$$M(u,v) = abs\left(\mathcal{F}\{f(x,y)\}(u,v)\right) \tag{1}$$

$$A(u,v) = angle\left(\mathcal{F}\{f(x,y)\}(u,v)\right) \tag{2}$$

$$L(u,v) = \log\left(M(u,v)\right) \tag{3}$$

$$R(u,v) = L(u,v) - h_n(u,v) * L(u,v) \tag{4}$$

$$SRVS(x,y) = g(x,y) * \left(\mathcal{F}^{-1}\left\{\exp\left(R + jA\right)\right\}(x,y)\right)^2 \tag{5}$$

In the above equations, $\mathcal{F}$ ($\mathcal{F}^{-1}$) denotes the Fourier (inverse Fourier) transform, $abs(\cdot)$ returns the magnitude of a complex number, $angle(\cdot)$ returns the argument of a complex number, $h_n(u, v)$ is an $n\times n$ mean filter, $g(x, y)$ is a Gaussian function, and * represents the convolution.

3.2. SR-SIM: Spectral residual based similarity

Zhang *et al*. have shown that perceptible image quality degradation can lead to perceptible changes in image's low-level features [9]. Since bottom-up VS models are basically based on image's low level features, VS values themselves actually vary with the change of the image quality. Therefore, in our SR-SIM, we propose to use SRVS map as a feature to compute the local similarity map between the reference image and the distorted image.

However, SRVS value at a pixel actually is a measure reflecting its relative distinctiveness to its surroundings. Thus, it is weak to characterize image's absolute local contrast. Hence, we need to use an additional feature to compensate for the lack of contrast sensitivity of SRVS. The simplest feature of this kind may be the gradient modulus (GM). There are many operators to compute image gradient, and here we adopt the Scharr gradient operator, which has been proved very powerful in [9]. With Scharr gradient operator, partial derivatives $G_x(\mathbf{x})$ and $G_y(\mathbf{x})$ of an image $f(\mathbf{x})$ are calculated as

$$\begin{aligned} G_x(\mathbf{x}) &= \frac{1}{16}\begin{bmatrix} 3 & 0 & -3 \\ 10 & 0 & -10 \\ 3 & 0 & -3 \end{bmatrix} * f(\mathbf{x}) \\ G_y(\mathbf{x}) &= \frac{1}{16}\begin{bmatrix} 3 & 10 & 3 \\ 0 & 0 & 0 \\ -3 & -10 & -3 \end{bmatrix} * f(\mathbf{x}) \end{aligned} \tag{6}$$

The GM of $f(\mathbf{x})$ is then defined as $G(\mathbf{x}) = \sqrt{G_x^2(\mathbf{x}) + G_y^2(\mathbf{x})}$. VS and GM are complementary and they reflect different aspects of the HVS in assessing the local quality of the input image.

Suppose that we are going to calculate the similarity between images f_1 and f_2. Denote by R_1 and R_2 the SRVS maps extracted from images f_1 and f_2 using the SRVS model, and by G_1 and G_2 the GM maps extracted from f_1 and f_2. Similar to other IQA indices, the computation of SR-SIM consists of two stages. In the first stage, the local similarity map is computed, and in the second stage, we pool the similarity map into a single quality score.

We separate the SR-SIM measurement between $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ into two components, each for SRVS and GM. First, the similarity between $R_1(\mathbf{x})$ and $R_2(\mathbf{x})$ is defined as:

$$S_V(\mathbf{x}) = \frac{2R_1(\mathbf{x}) \cdot R_2(\mathbf{x}) + C_1}{R_1^2(\mathbf{x}) + R_2^2(\mathbf{x}) + C_1} \tag{7}$$

where C_1 is a positive constant to increase the stability of S_V. Similarly, the GM values $G_1(\mathbf{x})$ and $G_2(\mathbf{x})$ are compared as:

$$S_G(\mathbf{x}) = \frac{2G_1(\mathbf{x}) \cdot G_2(\mathbf{x}) + C_2}{G_1^2(\mathbf{x}) + G_2^2(\mathbf{x}) + C_2} \tag{8}$$

where C_2 is another positive constant. Then, $S_V(\mathbf{x})$ and $S_G(\mathbf{x})$ are combined to get the local similarity $S(\mathbf{x})$ of $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$. We define $S(\mathbf{x})$ as follows:

$$S(\mathbf{x}) = S_V(\mathbf{x}) \cdot [S_G(\mathbf{x})]^\alpha \tag{9}$$

where α is a constant used to adjust the relative importance of VS and GM features.

Having obtained the local similarity $S(\mathbf{x})$ at each location $\mathbf{x}$, the overall similarity between f_1 and f_2 can be calculated. It has been widely accepted that a good quality score pooling strategy should correlate well with human visual fixation. In our case, it is natural to use SRVS map to characterize the visual importance of a local region. Intuitively, for a given position $\mathbf{x}$, if anyone of $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ has a high SRVS value, it implies that this position $\mathbf{x}$ will have a high impact on HVS when evaluating the similarity between f_1 and f_2. Therefore, we use $R_m(\mathbf{x}) = \max(R_1(\mathbf{x}), R_2(\mathbf{x}))$ to weight the importance of $S(\mathbf{x})$ in the overall similarity. Thus, the SR-SIM between f_1 and f_2 is defined as:

$$\text{SR-SIM} = \frac{\sum_{\mathbf{x}\in\Omega} S(\mathbf{x}) \cdot R_m(\mathbf{x})}{\sum_{\mathbf{x}\in\Omega} R_m(\mathbf{x})} \tag{10}$$

where Ω means the whole image spatial domain. Fig. 1 illustrates the scheme for computing SR-SIM.

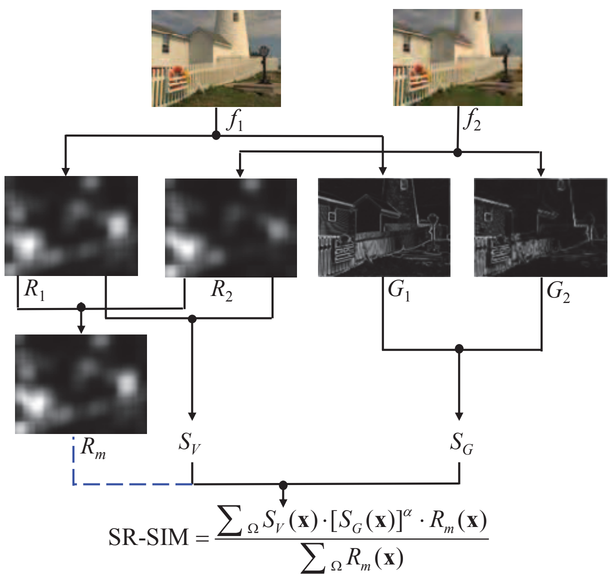

Fig. 1: Illustration for the computing scheme of SR-SIM. f_1 is the reference image and f_2 is a distorted version of f_1.

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

4.1. Test protocol

Experiments were conducted on three large-scale image datasets constructed for evaluating IQA indices, including TID2008 [13], CSIQ [14] and LIVE [15]. The important information of these three datasets, in terms of the number of reference images, the number of distorted images, the number of quality distortion types, and the number of subjects, is summarized in Table 1.

Four metrics are employed to evaluate IQA indices, including Spearman rank-order correlation coefficient (SROCC), Kendall rank-order correlation coefficient (KROCC), Pearson linear correlation coefficient (PLCC), and Root Mean Squared Error (RMSE). Definitions, explanations, and ways for calculating these four performance metrics can be found in [8].

The performance of the proposed SR-SIM IQA index was evaluated and compared with the other 9 state-of-the-art or representative IQA indices, including NQM [1], UQI [2], SSIM [3], MS-SSIM [4], IFC [5], VIF [6], IW-SSIM [8], RFSIM [7], and FSIM [9].

Table 1: Benchmark image datasets for IQA

Dataset	Reference Images No.	Distorted Images No.	Distortion Types No.	Subjects No.
TID2008	25	1700	17	838
CSIQ	30	866	6	35
LIVE	29	779	5	161

4.2. Performance evaluation

The prediction performance of each IQA index is given in Table 2. For each performance measure, the three IQA indices producing the best results are highlighted in boldface. In addition, in order to provide an evaluation of the overall performance of the evaluated IQA indices, in Table 3 we present their weighted-average SROCC, KROCC and PLCC results over three datasets and the weight assigned to each dataset linearly depends on the number of distorted images contained in that dataset.

The running speed of each selected IQA index was also evaluated. Experiments were performed on a Lenovo ThinkCenter M6300t PC. The software platform was Matlab R2010b. The time cost consumed by each IQA index for measuring the similarity of a pair of 384×512 color images (taken from TID2008) is listed in Table 4.

4.3. Discussions

From Table 2, it can be seen that with respect to the prediction performance, SR-SIM performs better than all the other IQA indices evaluated on the two largest IQA datasets, TID2008 and CSIQ. Even on LIVE, it performs only a slightly worse than VIF and FSIM. In Table 3, the statistical superiority of SR-SIM to the other competing IQA indices is clearly exhibited since no matter which performance criterion is used, SR-SIM always achieves the best overall results.

From Table 4 it can be seen that with respect to the running speed, SR-SIM ranks second and it performs only a little slower than SSIM. However, it should be noted that SR-SIM could achieve greatly better prediction performance than SSIM. FSIM, RFSIM, and IW-SSIM, indices that could achieve comparable prediction performances with SR-SIM, all have much higher computational complexities than SR-SIM.

Thus, we conclude that the proposed SR-SIM could achieve the best overall prediction performance while it has a quite low computational complexity, similar as SSIM.

Table 2: Performance comparison of 10 IQA indices on three benchmark datasets

		NQM	UQI	SSIM	MS-SSIM	IFC	VIF	IW-SSIM	RFSIM	FSIM	SR-SIM
TID 2008	SROCC	0.6243	0.5851	0.7749	0.8542	0.5675	0.7491	0.8559	**0.8680**	**0.8805**	**0.8913**
	KROCC	0.4608	0.4255	0.5768	0.6568	0.4236	0.5860	0.6636	**0.6780**	**0.6946**	**0.7149**
	PLCC	0.6142	0.6643	0.7732	0.8451	0.7340	0.8084	0.8579	**0.8645**	**0.8738**	**0.8866**
	RMSE	1.0590	1.0031	0.8511	0.7173	0.9113	0.7899	0.6895	**0.6746**	**0.6525**	**0.6206**
CSIQ	SROCC	0.7402	0.8098	0.8756	0.9133	0.7671	0.9195	0.9213	**0.9295**	**0.9242**	**0.9319**
	KROCC	0.5638	0.6188	0.6907	0.7393	0.5897	0.7537	0.7529	**0.7645**	**0.7567**	**0.7725**
	PLCC	0.7433	0.8312	0.8613	0.8991	0.8384	**0.9277**	0.9144	**0.9179**	0.9120	**0.9250**
	RMSE	0.1756	0.1460	0.1334	0.1149	0.1431	**0.0980**	0.1063	**0.1042**	0.1077	**0.0997**
LIVE	SROCC	0.9086	0.8941	0.9479	0.9513	0.9259	**0.9636**	0.9567	0.9401	**0.9634**	**0.9618**
	KROCC	0.7413	0.7100	0.7963	0.8045	0.7579	**0.8282**	0.8175	0.7816	**0.8337**	**0.8299**
	PLCC	0.9122	0.8987	0.9449	0.9489	0.9268	**0.9604**	0.9522	0.9354	**0.9597**	**0.9553**
	RMSE	11.1926	11.9823	8.9455	8.6188	10.2643	**7.6137**	8.3473	9.6642	**7.6780**	**8.0811**

Table 3: Overall performance of IQA indices over 3 datasets

IQA Index	SROCC	KROCC	PLCC
NQM	0.7205	0.5528	0.7170
UQI	0.7152	0.5418	0.7621
SSIM	0.8413	0.6574	0.8360
MS-SSIM	0.8921	0.7126	0.8833
IFC	0.7026	0.5445	0.8059
VIF	0.8432	0.6858	0.8747
IW-SSIM	0.8963	0.7226	0.8945
RFSIM	0.9007	0.7245	0.8948
FSIM	0.9111	0.7431	0.9037
SR-SIM	**0.9182**	**0.7566**	**0.9125**

Table 4: Time cost of each IQA index

IQA Index	Time (milliseconds)
NQM	227.4
UQI	48.4
SSIM	25.1
MS-SSIM	78.9
IFC	685.4
VIF	705.2
IW-SSIM	386.3
RFSIM	85.7
FSIM	350.6
SR-SIM	29.1

5. CONCLUSION

In this paper, we proposed a novel efficient and effective IQA index, SR-SIM, based on a specific visual saliency model, spectral residual visual saliency. SR-SIM is designed based on the assumption that an image's visual saliency map has a close relationship with its perceptual quality. Experimental results indicate that SR-SIM could yield statistically better prediction performance than all the other competing methods evaluated. Moreover, SR-SIM has a very low computational complexity, similar as SSIM. Thus, SR-SIM can be the best candidate of IQA indices for real-time applications.

ACKNOWLEDGEMENT

This research is supported by the Fundamental Research Funds for the Central Universities (2100219033), the NSFC (60903120), and Shanghai NSF (09ZR1434400).

6. REFERENCES

[1] N. Damera-Venkata, T.D. Kite, W.S. Geisler, B.L. Evans, and A.C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. IP*, vol. 9, pp. 636-650, 2000.

[2] Z. Wang and A.C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, pp. 81-84, 2002.

[3] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. IP*, vol. 13, pp. 600-612, 2004.

[4] Z. Wang, E.P. Simoncelli, and A.C. Bovik, "Multi-scale structural similarity for image quality assessment," *ACSSC'03*, pp. 1398-1402, 2003.

[5] H.R. Sheikh, A.C. Bovik, and G. Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. IP*, vol. 14, 2005.

[6] H.R. Sheikh and A.C. Bovik, "Image information and visual quality," *IEEE Trans. IP*, vol. 15, pp. 430-444, 2006.

[7] L. Zhang, L. Zhang, and X. Mou, "RFSIM: a feature based image quality assessment metric using Riesz transforms," *ICIP'10*, pp. 321-324, 2010.

[8] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. IP*, vol. 20, pp. 1185-1198, 2011.

[9] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE Trans. IP*, vol. 20, pp. 2378-2386, 2011.

[10] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," *CVPR'07*, pp. 1-8, 2007.

[11] A. Toet, "Computational versus psychophysical bottom-up image saliency: a comparative evaluation study," *IEEE Trans. PAMI*, vol. 33, pp. 2131-2146, 2011.

[12] U. Engelke, H. Kaprykowsky, H. Zepernick, and P. Ndjiki-Nya, "Visual attention in quality assessment," *IEEE Signal Process. Magazine*, vol. 28, pp. 50-59, 2011

[13] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008 - A database for evaluation of full-reference visual quality assessment metrics," *Adv. Modern Radioelectron.*, vol. 10, pp. 30-45, 2009.

[14] E.C. Larson and D.M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *J. Electr. Imaging*, vol. 19, pp. 001006:1-21, 2010.

[15] H.R. Sheikh, M.F. Sabir, and A.C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. IP*, vol. 15, pp. 3440-3451, 2006.