RFSIM: A FEATURE BASED IMAGE QUALITY ASSESSMENT METRIC USING RIESZ TRANSFORMS

Lin Zhang[a], Lei Zhang[a][1], and Xuanqin Mou[b]

[a] Dept. of Computing, The Hong Kong Polytechnic University, Hong Kong
[b] Institute of Image Processing and Recognition, Xi'an Jiaotong University, China

ABSTRACT

Image quality assessment (IQA) aims to provide computational models to measure the image quality in a perceptually consistent manner. In this paper, a novel feature based IQA model, namely Riesz-transform based Feature SIMilarity metric (RFSIM), is proposed based on the fact that the human vision system (HVS) perceives an image mainly according to its low-level features. The 1st-order and 2nd-order Riesz transform coefficients of the image are taken as image features, while a feature mask is defined as the edge locations of the image. The similarity index between the reference and distorted images is measured by comparing the two feature maps at key locations marked by the feature mask. Extensive experiments on the comprehensive TID2008 database indicate that the proposed RFSIM metric is more consistent with the subjective evaluation than all the other competing methods evaluated.

Index Terms— Image quality assessment, monogenic signal, Riesz transform

1. INTRODUCTION

Image quality assessment (IQA) is of paramount importance for numerous image/video processing and computer vision applications. A number of full reference IQA metrics have been developed. In the early stage, some mathematical statistics based metrics were employed as objective IQA metrics. Some later developed models emphasize the importance of human visual system's (HVS) sensitivity to different visual signals, such as the luminance, contrast, the contrast interaction between spatial frequencies, and the visual attention, etc. Some representative methods of this kind include NQM [1] and VSNR [2].

The Structural-SIMilarity (SSIM) index proposed in [3] can be considered as a milestone of the development of IQA metrics. The fundamental principle of SSIM is that the HVS is highly adapted to extract structural information from the visual scene, and therefore, a measurement of structural similarity should provide a good approximation of perceived image quality. The multi-scale extension of SSIM, called MS-SSIM [4], produces much better results than its single-scale counterpart. In [5], Sheikh *et al.* introduced information theory into image fidelity measurement, and they proposed a novel information fidelity criterion (IFC) for IQA by using natural scene statistics models. Later, IFC was extended to the VIF (Visual Information Fidelity) by involving an image information measure [6]. In [7], Sampat *et al.* made use of the steerable complex wavelet transform to measure the structural similarity of the two images and proposed the CW-SSIM index for IQA.

In some sense, metrics such as SSIM and VIF aim to fully use the visual information inside an image to measure its quality. However, the visual information in an image is often very redundant, while the HVS perceives an image mainly based on its low-level features at key locations, such as edges, zero-crossings, corners and lines [8]. In other words, a small amount of key image points with salient features convey most of the crucial information for the HVS to interpret the scene. Thus, we believe that the perceptible image degradations will induce corresponding changes in image low level features at key locations. Based this belief, in this paper we propose a novel IQA metric, namely Riesz-transform based Feature SIMilarity (RFSIM) index. RFSIM is computed by comparing Riesz transform features at key locations between the reference image and a distorted image. Considering the fact that HVS is sensitive to image edges, key locations are marked by a mask formed by the Canny operator (without thinning operation). With respect to the feature extraction, the 1st-order and 2nd-order Riesz transforms [9, 10] are used because they can extract several types of image low-level features effectively and efficiently in a unified theoretic framework. Finally, only those Riesz transform coefficients within the feature mask are used in the RFSIM index calculation. Experiments on the recently established TID2008 database [11] show that RFSIM achieves higher consistency with the subjective evaluation than the other state-of-the-art IQA models.

---

[1] Corresponding author: cslzhang@comp.polyu.edu.hk.

The rest of this paper is organized as follows. Section 2 introduces the Riesz transforms. Section 3 describes the proposed RFSIM metric. Section 4 reports the experimental results and Section 5 concludes the paper.

2. RIESZ TRANSFORMS

The Hilbert transform of a 1-D function has been widely used in signal processing since Gabor proposed the analytic signal. It renders the possibility to estimate the local phase and the local amplitude of a 1-D signal simultaneously. However, when the 1-D signal is embedded into the 2-D ambient Euclidean space, the analytic signal cannot be directly used. To solve such a problem, in [9], Felsberg proposed the monogenic signal, which is an isotropic 2-D extension of the analytic signal. The monogenic signal is built upon the 1st-order Riesz transform which is a vector valued extension of the Hilbert transform in the n-D Euclidean space. If we restrict to the case when $n = 2$, the spatial representation of the Riesz kernel is

$$\left(R_x(\mathbf{x}), R_y(\mathbf{x})\right) = \left(\frac{x}{2\pi|\mathbf{x}|^3}, \frac{y}{2\pi|\mathbf{x}|^3}\right), \mathbf{x} = (x, y) \in \mathbb{R}^2 \quad (1)$$

and its transfer function in the Fourier domain is

$$\left(H_u(\mathbf{u}), H_v(\mathbf{u})\right) = \left(-i\frac{u}{|\mathbf{u}|}, -i\frac{v}{|\mathbf{u}|}\right), \mathbf{u} = (u, v) \in \mathbb{R}^2 \quad (2)$$

For an image $f(\mathbf{x})$, the monogenic signal is defined as the combination of f and its Riesz transform

$$\begin{aligned} f_M(\mathbf{x}) &= \left(f(\mathbf{x}), R_x\{f\}(\mathbf{x}), R_y\{f\}(\mathbf{x})\right) \\ &= \left(f, R_x * f, R_y * f\right) \end{aligned} \quad (3)$$

where * stands for the convolution. For i1D (intrinsic 1 dimension) signals, the local orientation can be calculated as [9, 10] (see Fig. 1 for an example)

$$\theta = \arctan\frac{R_y\{f\}}{R_x\{f\}}, \theta \in [0, \pi) \quad (4)$$

Moreover, for i1D signal $f(\mathbf{x})$, it can be proved that [10]

$$\begin{aligned} &\sqrt{R_x^2\{f\}(0,0) + R_y^2\{f\}(0,0)} \\ &= |(h_1 * f_\theta)(0)| = \left|-\frac{1}{\pi}\int_{t\in\mathbb{R}}\frac{f(t\cos\theta, t\sin\theta)}{t}dt\right| \end{aligned} \quad (5)$$

where $h_1 * f_\theta$ is actually the partial Hilbert transform of f along its main orientation θ and $h_1 = 1/\pi x$ is the Hilbert transform kernel. Thus, the local phase of the i1D signal $f(\mathbf{x})$ can be defined analogously to the 1-D analytic signal as

$$\varphi = \operatorname{atan2}\left(\sqrt{R_x^2\{f\} + R_y^2\{f\}}, f\right), \varphi \in [0, \pi) \quad (6)$$

The local amplitude is defined as

$$A = |f_M(\mathbf{x})| = \sqrt{f^2 + R_x^2\{f\} + R_y^2\{f\}} \quad (7)$$

It can be seen that the local orientation, the local phase, and the local amplitude of the i1D signal can be conveniently represented via the 1st-order Riesz transform based monogenic signal.

In order to analyze i2D image structures, higher order Riesz transforms have been proposed [10]. For example, the 2nd order Riesz transform $R_x\{R_x\{f\}\}$ can be obtained by applying Eq. (1) to $R_x\{f\}$. For our needs, we make use of three 2nd-order Riesz transforms, including $R_x\{R_x\{f\}\}$, $R_x\{R_y\{f\}\}$, and $R_y\{R_y\{f\}\}$. Some i2D features, such as the local image surface type and the apex angle, can be derived based on the 2nd-order Riesz transforms. Details can be found in [10].
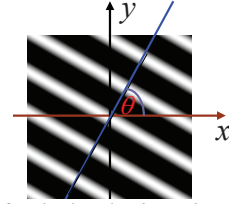

Fig. 1: Illustration for the local orientation of an i1D signal.

3. RFSIM: RIESZ-TRANSFORM BASED FEATURE SIMILARITY METRIC

With the belief that the perceptible image degradations will induce perceived changes in image features at key locations, we present a novel IQA model, namely Riesz-transform based Feature SIMilarity index (RFSIM). It is computed by comparing the feature maps at key locations marked by a feature mask between two images.

Considering the fact that HVS is sensitive to image edges, key locations are marked by a feature mask which is simply generated by an edge operator with a dilation operation. Suppose that we are going to calculate the similarity between two images f and g. Denote by M_1 the result of edge detection performed on f, and M_2 the result of edge detection on g. Then, the feature mask is defined as

$$M = M_1 \oplus M_2 \quad (8)$$

where $\oplus$ is the logical "OR" operation.

As introduced in Section 2, the 1st-order and the 2nd-order Riesz transforms can extract some low-level image features effectively and efficiently in a unified theoretic framework. Thus, in this paper, we utilize them for the feature extraction. However, we do not extract the features such as the local phase, local orientation and local surface type, etc., explicitly. Instead, we take the coefficients of the Riesz transforms as features directly. Let f_1, f_2, f_3, f_4, f_5 represent $R_x\{f\}$, $R_y\{f\}$, $R_x\{R_x\{f\}\}$, $R_x\{R_y\{f\}\}$, and $R_y\{R_y\{f\}\}$, respectively (the first two are 1st-order Riesz transform coefficients and the last three are 2nd-order coefficients). Similarly, let g_1, g_2, g_3, g_4, g_5 represent $R_x\{g\}$, $R_y\{g\}$, $R_x\{R_x\{g\}\}$, $R_x\{R_y\{g\}\}$, and $R_y\{R_y\{g\}\}$, respectively. The similarity between two feature maps f_i ($i = 1\sim5$) and g_i at the corresponding location (x, y) is defined as

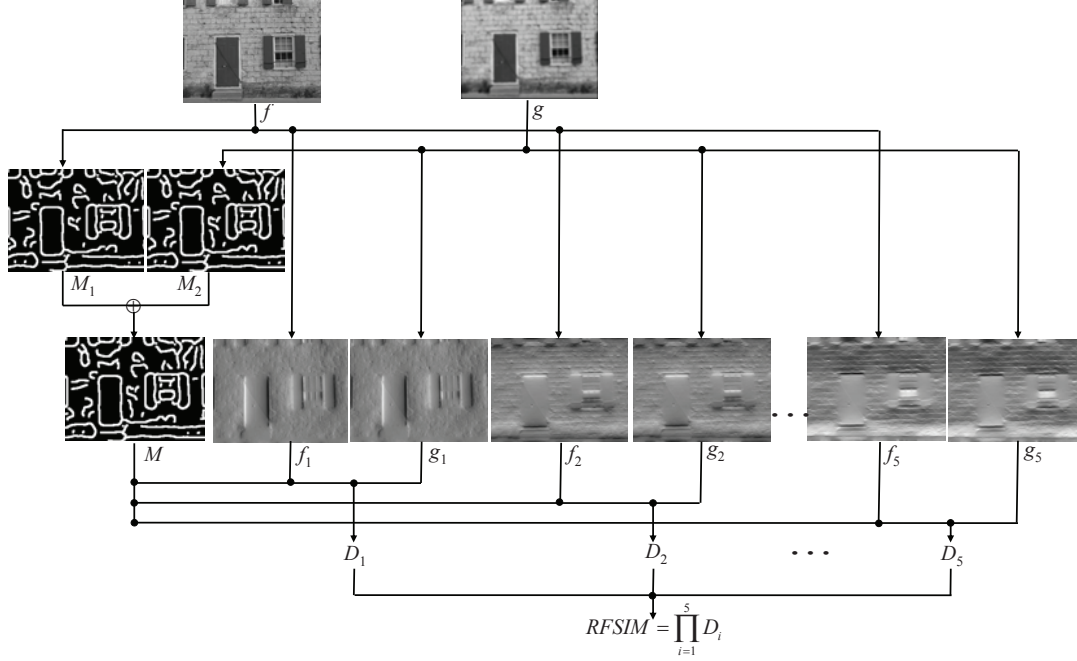$$d_i(x, y) = \frac{2f_i(x, y) \cdot g_i(x, y) + c}{f_i^2(x, y) + g_i^2(x, y) + c} \quad (9)$$

Fig. 2: Illustration of the similarity computation process between two images by the proposed RFSIM IQA model. *f* is the reference image and *g* is a distorted version of *f*. Example images are taken from TID2008 database.

where c is a small constant value. This naturally leads to the following formula to define the similarity between the feature maps f_i and g_i by considering only the key locations marked by mask M,

$$D_i = \frac{\sum\sum d_i(x,y) \cdot M(x,y)}{\sum\sum M(x,y)} \qquad (10)$$

Then, we compute the RFSIM index between f and g as

$$RFSIM = \prod_{i=1}^{5} D_i \qquad (11)$$

The process to compute the RFSIM index is illustrated by a pair of images in Fig. 2.

4. EXPERIMENTS AND DISCUSSIONS

We conduct experiments on the TID2008 database [11], which is the largest database for the evaluation of full reference IQA models. (Our experimental results on other databases lead to the same conclusion.) TID2008 contains 25 reference images and 1,700 distorted images. Subjective scores are collected from 838 observers. In fact, there are several databases available for the evaluation of IQA models but they are quite different in several aspects, including the number of reference images, the number of distorted images, the number of distortion types, the image type, and the number of observers. The experimental configuration and the methodology to collect subjective scores also vary. Taking these factors into consideration, TID2008 is among the best currently available databases for the evaluation of IQA models.

We compare the performance of the proposed RFSIM metric with seven state-of-the-art IQA metrics, including SSIM [3], MS-SSIM [4], VIF [6], VSNR [2], IFC [5], VIFp [6], and NQM [1]. For the implementation of the compared IQA metrics, we use the publicly available software in [12]. The code of the proposed RFSIM method is available at http://www4.comp.polyu.edu.hk/~cslzhang/IQA.htm.

In the implementation of the RFSIM, several parameters are empirically set. The low threshold, high threshold, and the standard deviation of the Gaussian filter used in Canny operator are set as 0.08, 0.13, and 3.6, and c is set as 1.2. In addition, RFSIM will be most effective if used at the appropriate image scale. In our implementation, we use the following empirical steps proposed by Wang [13] to determine the scale for images viewed from a typical distance: 1) Let $F = \max(1, \text{round}(N / 256))$, where N is the number of pixels in image height or width; 2) average local $F \times F$ pixels and down-sample the image by a factor of F.

In order to evaluate the IQA models, four commonly used performance metrics are employed. The first two metrics are the Spearman rank-order correlation coefficient (SROCC) and the Kendall rank-order correlation coefficient (KROCC), which can measure the prediction monotonicity of an IQA model. To compute the third and the fourth metrics we need to apply a regression analysis, as suggested by the video quality experts group (VQEG), to provide a nonlinear mapping between the objective scores and the subjective mean opinion scores (MOS) [14]. The third metric is the Pearson linear correlation coefficient (CC) between MOS and the objective scores after nonlinear

regression. The fourth metric is the root mean square error (RMSE) between MOS and the objective scores after nonlinear regression. For the nonlinear regression analysis, we use the following mapping function [15]:

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(x - \beta_3)}} \right) + \beta_4 x + \beta_5 \qquad (12)$$

Table 1: Evaluation of IQA models on TID 2008

Model	SROCC	KROCC	CC	RMSE
RFSIM	**0.8632**	**0.6772**	**0.8618**	**0.6806**
MS-SSIM	0.8528	0.6543	0.8425	0.7229
SSIM	0.7749	0.5768	0.7732	0.8511
VIF	0.7496	0.5863	0.8090	0.7888
VSNR	0.7046	0.5340	0.6820	0.9815
IFC	0.5692	0.4261	0.7359	0.9086
VIFp	0.6546	0.4952	0.7702	0.8559
NQM	0.6243	0.4608	0.6135	1.0598

The performance comparison of the eight IQA models on the TID2008 database is given in Table 1. From Table 1, we can clearly see that the proposed RFSIM model performs much better than all the other seven methods evaluated in terms of all the four evaluation metrics. Fig. 3 shows the scatter plots of subjective MOS versus the predicted scores by the four IQA metrics which can achieve good results on the TID2008 database. Curves shown in Fig. 3 are obtained by a nonlinear fitting according to the model Eq. (12).

5. CONCLUSION

In this paper, a novel IQA model, namely Riesz-transform based Feature SIMilarity index (RFSIM), was proposed. It is based on the belief that the image degradation will induce perceived changes in image features at key locations. RFSIM is computed via comparing the features at key locations of two images. Key locations are indicated by a feature mask which is generated by the Canny edge detection without thinning. We take the coefficients of the 1st-order and the 2nd-order Riesz transforms as features, and only the coefficients within the feature mask are used in the RFSIM index calculation. Experimental results indicate that the RFSIM index outperforms all the other IQA models evaluated.

ACKNOWLEDGEMENT

6. REFERENCES

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

[2] D.M. Chandler and S.S. Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images", *IEEE Trans. IP*, vol. 16, pp. 2284-2298, 2007.

[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*, 2003.

[5] H.R. Sheikh, A.C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics", *IEEE Trans. IP*, vol. 14, pp. 2117-2128, 2005.

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

[7] M.P. Sampat, Z. Wang, S. Gupta, A.C. Bovik and M.K. Markey, "Complex wavelet structural similarity: a new image similarity index", *IEEE Trans. IP*, vol. 18, pp. 2385-2401, 2009.

[8] D. Marr. Vision. W. H. Freeman and Company, 1980.

[9] M. Felsberg and G. Sommer, "The monogenic signal", *IEEE Trans. SP*, vol. 49, pp. 3136-3144, 2001.

[10] O. Fleischmann, "2D signal analysis by generalized Hilbert transforms", Thesis, University of Kiel, 2008.

[11] N. Ponomarenko, M. Carli, V. Lukin, K. Egiazarian, J. Astola, and F. Battisti, "Color image database for evaluation of image quality metrics", *IWMSP' 08*, 2008.

[12] M. Gaubatz, "Metrix MUX Visual Quality Assessment Package", http://foulard.ece.cornell.edu/gaubatz/metrix_mux.

[13] Z. Wang, "The SSIM Index for Image Quality Assessment", http://www.ece.uwaterloo.ca/~z70wang/research/ssim/.

[14] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment", http://www.vqeg.org, 2000.

[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.
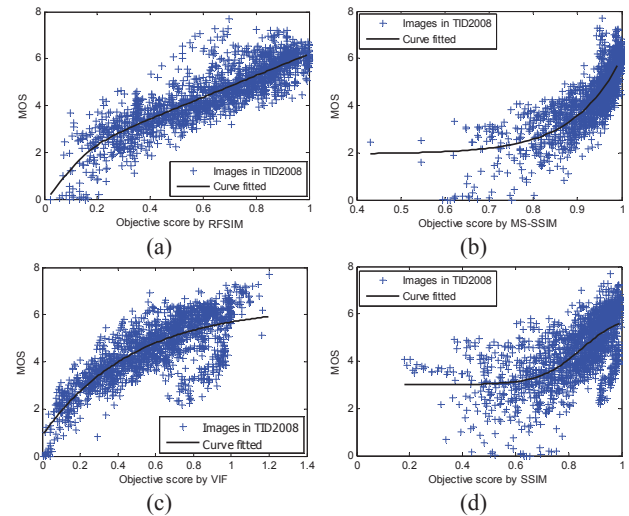
Fig. 3: Scatter plots of subjective MOS versus scores obtained by model prediction on the TID2008 database. (a) RFSIM; (b) MS-SSIM; (c) VIF; (d) SSIM.