FSIM: A Feature Similarity Index for Image Quality Assessment

Lin Zhang[a], *Student Member, IEEE,* Lei Zhang[a1], *Member, IEEE*
Xuanqin Mou[b], *Member, IEEE,* and Daivd Zhang[a], *Fellow, IEEE*

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

Abstract: Image quality assessment (IQA) aims to use computational models to measure the image quality consistently with subjective evaluations. The well-known structural-similarity (SSIM) index brings IQA from the pixel based stage to the structure based stage. In this paper, a novel feature-similarity (FSIM) index for IQA is proposed based on the fact that human visual system (HVS) understands an image mainly according to its low-level features. Specifically, the phase congruency (PC), which is a dimensionless measure of the significance of a local structure, is used as the primary feature in FSIM. Considering that PC is contrast invariant while the contrast information does affect HVS' perception of image quality, the image gradient magnitude (GM) is employed as the secondary feature in FSIM. PC and GM play complementary roles in characterizing the image contents. Experimental results on benchmark databases show that FSIM can achieve much higher consistency with the subjective evaluations than all the state-of-the-art IQA metrics used in comparison.

Index Terms: Image quality assessment, phase congruency, gradient, low-level feature

I. INTRODUCTION

With the rapid proliferation of digital imaging and communication technologies, image quality assessment (IQA) has been becoming an important issue in numerous applications such as image acquisition, transmission, compression, restoration and enhancement, etc. Since the subjective IQA methods cannot be readily and routinely used for many scenarios, e.g. real-time and automated systems, it is necessary to develop objective IQA metrics to automatically and robustly predict the perceived image quality. Meanwhile, it is anticipated that the evaluation results should be statistically consistent with those of the human observers. To this end, the scientific community has developed various IQA methods over the past decades.

[1] Corresponding author. Email: cslzhang@comp.polyu.edu.hk. This project is supported by the Hong Kong RGC General Research Fund (PolyU 5330/07E) and the Ho Tung Fund (5-ZH25).

According to the availability of a reference image, objective IQA metrics can be classified as full reference (FR), no-reference (NR) and reduced-reference (RR) methods [1]. In this paper, the discussion is confined to FR methods, where the original "distortion free" image is known as the reference image.

The conventional metrics such as the peak signal-to-noise ratio (PSNR) and the mean squared error (MSE) operate directly on the intensity of the image and they do not correlate well with the subjective fidelity ratings. Thus many efforts have been made on designing human visual system (HVS) based IQA metrics. Such kinds of models emphasize the importance of HVS's sensitivity to different visual signals, such as the luminance, the contrast, the frequency content, and the interaction between different signal components [2-4]. The noise quality measure (NQM) [2] and the visual signal-to-noise ratio (VSNR) [3] are two representatives. Instead of building on the assumptions about HVS models, methods such as the structural similarity (SSIM) index [1] are motivated by the need to capture the loss of structure in the image. SSIM is based on the hypothesis that HVS is highly adapted to extract the structural information from the visual scene; therefore, a measurement of structural similarity should provide a good approximation of perceived image quality. The multi-scale extension of SSIM, called MS-SSIM [5], produces better results than its single-scale counterpart. In [6], Sheikh *et al.* introduced the information theory into image fidelity measurement, and proposed the information fidelity criterion (IFC) for IQA by quantifying the information shared between the distorted and the reference images. IFC was later extended to the visual information fidelity (VIF) metric in [4]. 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.

The great success of SSIM and its extensions owes to the fact that HVS is adapted to the structural information in images. The visual information in an image, however, is often very redundant, while the HVS understands an image mainly based on its low-level features, such as edges and zero-crossings [8-10]. In other words, the salient low-level features convey crucial information for the HVS to interpret the scene. Accordingly, perceptible image degradations will lead to perceptible changes in image low-level features, and hence a good IQA metric could be devised by comparing the low-level feature sets between the reference image and the distorted image. Based on the above analysis, in this paper we propose a novel low-level feature similarity induced IQA metric, namely FSIM (Feature SIMilarity).

One key issue is then what kinds of features could be used in designing FSIM? Based on the physiological and psychophysical evidence, it is found that visually discernable features coincide with those

points where the Fourier waves at different frequencies have congruent phases [10-13]. That is, at points of high phase congruency (PC) we can extract highly informative features. Therefore, PC is used as the primary feature in computing FSIM. Meanwhile, considering that PC is contrast invariant but image local contrast does affect HVS's perception on the image quality, the image gradient magnitude (GM) is computed as the secondary feature to encode the contrast information. PC and GM are complementary and they reflect different aspects of the HVS in assessing the input image. Although FSIM is designed for grayscale image (or the luminance component of a color image), the chrominance information can be easily incorporated by means of a simple extension of FSIM, and we call this extension FSIM$_C$. FSIM and FSIM$_C$ are evaluated on six benchmark IQA databases in comparison with seven state-of-the-art IQA methods. The extensive results show that FSIM and FSIM$_C$ can achieve very high consistency with the human subjective evaluations, outperforming all the other competitors. Particularly, FSIM or FSIM$_C$ works consistently well across all the databases, while other methods may work well only on some specific databases.

The remainder of this paper is organized as follows. Section II discusses the extraction of PC and GM. Section III presents in detail the computation of the FSIM and FSIM$_C$ indices. Section IV reports the experimental results. Finally, Section V concludes the paper.

II. EXTRACTION OF PHASE CONGRUENCY AND GRADIENT MAGNITUDE

A. Phase congruency (PC)

Rather than define features directly as points with sharp changes in intensity, the PC model postulates that features are perceived at points where the Fourier components are maximal in phase. Based on the physiological and psychophysical evidences, the PC model is a simple but biologically plausible model of how mammalian visual systems could detect and identify features in an image [10-13]. PC can be considered as a dimensionless measure for the significance of a local structure.

Under the definition of PC in [11], there can be different implementations to compute the PC map of a given image. In this paper we adopt the method developed by Kovesi in [13], which is widely used in literature. We start from the 1D signal $g(x)$. Denote by M_n^e and M_n^o the even-symmetric and odd-symmetric filters at scale n and they form a quadrature pair. Responses of each quadrature pair to the signal will form a response vector at position x and scale n: $[e_n(x), o_n(x)] = [g(x)* M_n^e, g(x)* M_n^o]$, and the local amplitude at

scale n is $A_n(x) = \sqrt{e_n(x)^2 + o_n(x)^2}$. Let $F(x)=\sum_n e_n(x)$ and $H(x)=\sum_n o_n(x)$. The 1D PC can be computed as

$$PC(x) = E(x) / \left(\varepsilon + \sum\nolimits_n A_n(x)\right) \tag{1}$$

where $E(x) = \sqrt{F^2(x) + H^2(x)}$ and ε is a small positive constant.

With respect to the quadrature pair of filters, i.e. M_n^e and M_n^o, we adopt the log-Gabor filter because its transfer function has an extended tail at the high frequency end, which makes it more capable to encode natural images and it is consistent with measurements on mammalian visual systems [13, 14]. The log-Gabor filter has a transfer function in the frequency domain of the form $G(\omega)=\exp(-(\log(\omega/\omega_0))^2/2\sigma_r^2)$, where ω_0 is the filter's center frequency and σ_r controls the filter's bandwidth.

To compute the PC of 2D grayscale images, we can apply the 1D analysis over several orientations and then combine the results using some rule. The 1D log-Gabor filters described above can be extended to 2D ones by simply applying some spreading function across the filter perpendicular to its orientation. By using Gaussian as the spreading function, the 2D log-Gabor function has the following transfer function

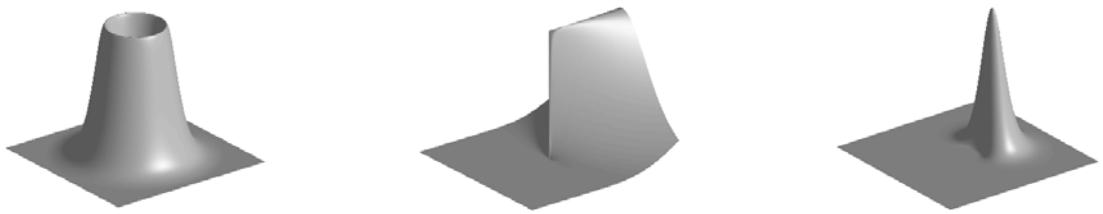
$$G_2(\omega, \theta_j) = \exp\left(-\frac{\left(\log(\omega / \omega_0)\right)^2}{2\sigma_r^2}\right) \cdot \exp\left(-\frac{\left(\theta - \theta_j\right)^2}{2\sigma_\theta{}^2}\right) \tag{2}$$

where $\theta_j = j\pi / J$, $j = \{0,1,\ldots, J\text{-}1\}$ is the orientation angle of the filter, J is the number of orientations and σ_θ determines the filter's angular bandwidth. An example of the 2D log-Gabor filter in the frequency domain, with $\omega_0 = 1/6$, $\theta_j = 0$, $\sigma_r = 0.3$, and $\sigma_\theta = 0.4$, is shown in Fig. 1.

By modulating ω_0 and θ_j and convolving G_2 with the 2D image, we get a set of responses at each point $\mathbf{x}$ as $\left[e_{n,\theta_j}(\mathbf{x}), o_{n,\theta_j}(\mathbf{x})\right]$. The local amplitude at scale n, orientation θ_j is $A_{n,\theta_j}(\mathbf{x}) = \sqrt{e_{n,\theta_j}(\mathbf{x})^2 + o_{n,\theta_j}(\mathbf{x})^2}$, and the local energy along orientation θ_j is $E_{\theta_j}(\mathbf{x}) = \sqrt{F_{\theta_j}(\mathbf{x})^2 + H_{\theta_j}(\mathbf{x})^2}$, where $F_{\theta_j}(\mathbf{x}) = \sum_n e_{n,\theta_j}(\mathbf{x})$ and $H_{\theta_j}(\mathbf{x}) = \sum_n o_{n,\theta_j}(\mathbf{x})$. The 2D PC at $\mathbf{x}$ is defined as

$$PC_{2D}(\mathbf{x}) = \frac{\sum_j E_{\theta_j}(\mathbf{x})}{\varepsilon + \sum_n \sum_j A_{n,\theta_j}(\mathbf{x})} \tag{3}$$

It should be noted that $PC_{2D}(\mathbf{x})$ is a real number within 0 ~ 1. Examples of the PC maps of 2D images can be found in Fig. 2.



(a) (b) (c)

Fig. 1: An example of the log-Gabor filter in the frequency domain, with $\omega_0 = 1/6$, $\theta_j = 0$, $\sigma_r = 0.3$, and $\sigma_\theta = 0.4$. (a) The radial component of the filter. (b) The angular component of the filter. (c) The log-Gabor filter, which is the product of the radial component and the angular component.

B. Gradient Magnitude (GM)

Image gradient computation is a traditional topic in image processing. Gradient operators can be expressed by convolution masks. Some commonly used gradient operators are Robert operator, Laplace operator, Prewitt operator, Sobel operator, etc. In this paper, we simply use the Sobel operator to compute the gradient of an image. The partial derivatives of image $f(\mathbf{x})$ along horizontal and vertical directions are

$$G_x = \frac{1}{4}\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} * f(\mathbf{x}), \quad G_y = \frac{1}{4}\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * f(\mathbf{x}) \tag{4}$$

Then, the gradient magnitude (GM) of $f(\mathbf{x})$ can be defined as $G = \sqrt{G_x^2 + G_y^2}$.

III. THE FEATURE SIMILARITY (FSIM) INDEX FOR IQA

With the extracted PC and GM features, in this section we present a novel Feature SIMilarity (FSIM) index for IQA. Suppose that we are going to calculate the similarity between images f_1 and f_2. Denote by PC_1 and PC_2 the PC maps extracted from f_1 and f_2, and G_1 and G_2 the GM maps extracted from them. It should be noted that for color images, PC and GM features are extracted from their luminance channels. FSIM will be defined and computed based on PC_1, PC_2, G_1 and G_2. Furthermore, by incorporating the image chrominance information into FSIM, an IQA index for color images, denoted by FSIM_C, can be obtained.

A. The FSIM Index

We separate the feature similarity measurement between $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ into two components, each for PC or GM. First, the similarity measure for PC values $PC_1(\mathbf{x})$ and $PC_2(\mathbf{x})$ is defined as

$$S_{PC}(\mathbf{x}) = \frac{2PC_1(\mathbf{x}) \cdot PC_2(\mathbf{x}) + T_1}{PC_1^2(\mathbf{x}) + PC_2^2(\mathbf{x}) + T_1} \tag{5}$$

where the constant T_1 is introduced to increase the stability of S_{PC} (such a consideration was also included in SSIM [1]). In practice, the determination of T_1 depends on the dynamic range of PC values. Similarly, the GM values $G_1(\mathbf{x})$ and $G_2(\mathbf{x})$ are compared and the similarity measure is defined as

$$S_G(\mathbf{x}) = \frac{2G_1(\mathbf{x}) \cdot G_2(\mathbf{x}) + T_2}{G_1^2(\mathbf{x}) + G_2^2(\mathbf{x}) + T_2} \tag{6}$$

where T_2 is a constant depending on the dynamic range of GM values. In the experiments, both T_1 and T_2 will be fixed to all databases so that the proposed FSIM can be conveniently used. $S_{PC}(\mathbf{x})$ and $S_G(\mathbf{x})$ are then combined as follows to get the similarity measure $S_L(\mathbf{x})$ of $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$:

$$S_L(\mathbf{x}) = S_{PC}(\mathbf{x}) \cdot S_G(\mathbf{x}) \tag{7}$$

Having obtained the similarity $S_L(\mathbf{x})$ at each location $\mathbf{x}$, the overall similarity between f_1 and f_2 can be obtained. However, different locations will have different contributions to HVS's perception of the image. For example, edge locations convey more crucial visual information than the locations within a smooth area. Since PC is a dimensionless metric to measure the significance of a local structure [10-13], the PC value at a location can reflect how likely it is a perceptibly significant structure point. Intuitively, for a given location $\mathbf{x}$, if anyone of $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ has a significant PC value, it implies that this position $\mathbf{x}$ will have a high impact on HVS in evaluating the similarity between f_1 and f_2. Therefore, we use $PC_m(\mathbf{x}) = \max(PC_1(\mathbf{x}), PC_2(\mathbf{x}))$ to weight the importance of $S_L(\mathbf{x})$ in the overall similarity between f_1 and f_2, and accordingly the FSIM index between f_1 and f_2 is defined as

$$\text{FSIM} = \frac{\sum_{\mathbf{x} \in \Omega} S_L(\mathbf{x}) \cdot PC_m(\mathbf{x})}{\sum_{\mathbf{x} \in \Omega} PC_m(\mathbf{x})} \tag{8}$$

where Ω means the whole image spatial domain.

B. Extension to Color Image Quality Assessment

FSIM index is designed for grayscale images or the luminance components of color images. Since the chrominance information will also affect the HVS in understanding the images, better performance can be expected if the chrominance information is incorporated in FSIM for color IQA. Such a goal can be achieved by applying a straightforward extension to the FSIM framework.

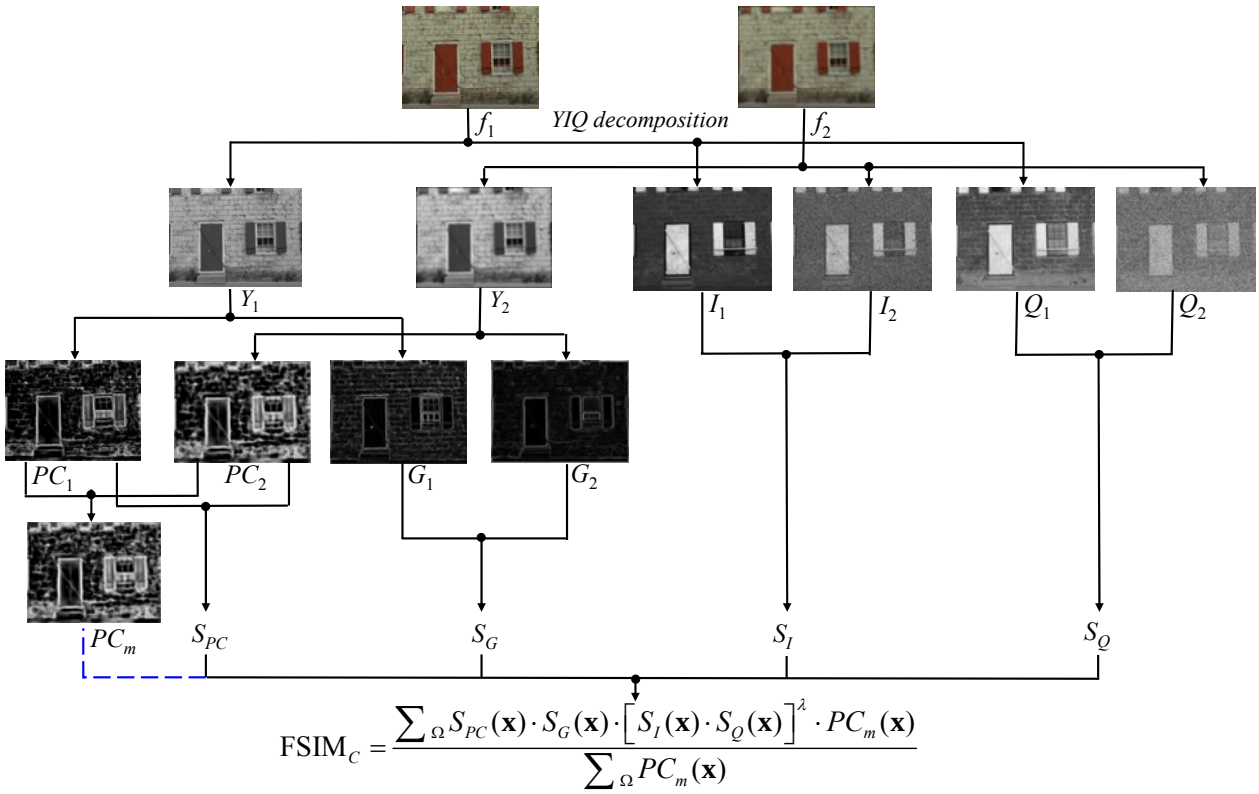


Fig. 2: Illustration for the FSIM/FSIM$_C$ index computation. f_1 is the reference image and f_2 is a distorted version of f_1.

At first, the original RGB color images are converted to another color space, where the luminance can be separated from the chrominance. To this end, we adopt the widely used YIQ color space [15], in which Y represents the luminance information and I and Q convey the chrominance information. The transform from the RGB space to the YIQ space can be accomplished via [15]:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{9}$$

Let I_1 (I_2) and Q_1 (Q_2) be the I and Q chromatic channels of the image f_1 (f_2), respectively. Similar to the definitions of $S_{PC}(\mathbf{x})$ and $S_G(\mathbf{x})$, we define the similarity between chromatic features as

$$S_I(\mathbf{x}) = \frac{2I_1(\mathbf{x}) \cdot I_2(\mathbf{x}) + T_3}{I_1^2(\mathbf{x}) + I_2^2(\mathbf{x}) + T_3}, \quad S_Q(\mathbf{x}) = \frac{2Q_1(\mathbf{x}) \cdot Q_2(\mathbf{x}) + T_4}{Q_1^2(\mathbf{x}) + Q_2^2(\mathbf{x}) + T_4} \tag{10}$$

where T_3 and T_4 are constants. Since I and Q components have nearly the same dynamic range, in this paper we set $T_3 = T_4$ for simplicity. $S_I(\mathbf{x})$ and $S_Q(\mathbf{x})$ can then be combined as follows to get the chrominance similarity measure, denoted by $S_C(\mathbf{x})$, of $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$:

$$S_C(\mathbf{x}) = S_I(\mathbf{x}) \cdot S_Q(\mathbf{x}) \tag{11}$$

Then, the FSIM index can be extended to FSIM_C by incorporating the chromatic information in a straightforward manner

$$\text{FSIM}_C = \frac{\sum_{\mathbf{x}\in\Omega} S_L(\mathbf{x}) \cdot [S_C(\mathbf{x})]^{\lambda} \cdot PC_m(\mathbf{x})}{\sum_{\mathbf{x}\in\Omega} PC_m(\mathbf{x})} \tag{12}$$

where $\lambda > 0$ is the parameter used to adjust the importance of the chromatic components. The procedures to calculate the FSIM_C index is illustrated in Fig. 2. If the chromatic information is ignored in Fig. 2, the FSIM_C index is reduced to the FSIM index.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

In order to validate the efficacy of the proposed FSIM and FSIM_C indices, we evaluated their performance in comparison with seven IQA metrics, including six state-of-the-arts (SSIM [1], MS-SSIM [5], VIF [4], VSNR [3], IFC [6], NQM [2]) and the classical PSNR, on six publicly available test databases: TID2008 [16], CSIQ (http://vision.okstate.edu/csiq/), LIVE (http://live.ece.utexas.edu/research/quality), IVC (http://www.irccyn.ec-nantes.fr/ivcdb/), MICT (http://mict.eng.u-toyama.ac.jp/mict/index2.html) and A57 (http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html). The characteristics of the six databases are summarized in Table I. We used the public software MeTriX MuX (http://foulard.ece.cornell.edu/gaubatz/metrix_mux) for the implementation of the competing IQA metrics except for SSIM, whose implementation is available at [17].

TABLE I. BENCHMARK TEST DATABASES FOR IQA

Database	Source Images	Distorted Images	Distortion Types	Image Type	Observers
TID2008	25	1700	17	color	838
CSIQ	30	866	6	color	35
LIVE	29	779	5	color	161
IVC	10	185	4	color	15
MICT	14	168	2	color	16
A57	3	54	6	gray	unknown

The parameters required in the proposed methods were experimentally tuned and then fixed to all the six databases: $n = 4$, $J = 4$, $\sigma_r = 0.5978$, $\sigma_\theta = 0.6545$, $T_1 = 0.85$, $T_2 = 180$, $T_3 = T_4 = 200$, and $\lambda = 0.03$. Besides, the center frequencies of the log-Gabor filters at four scales were set as: 1/6, 1/12, 1/24 and 1/48. It should be noted that the FSIM/FSIM_C index will be most effective if used at the appropriate scale. The precisely "right" scale depends on both the image resolution and the viewing distance and hence is difficult to be obtained. In practice, we used the following empirical steps proposed by Wang [17] to determine the scale

for images viewed from a typical distance. 1) Let F=max(1, round(N / 256)), where N is the number of pixels in image height or width; 2) average local $F \times F$ pixels and then down-sample the image by a factor of F.

TABLE II: PERFORMANCE COMPARISON OF IQA MODELS ON 6 BENCHMARK DATABASES

		FSIM	FSIM$_C$	MS-SSIM	VIF	SSIM	IFC	VSNR	NQM	PSNR
TID 2008	SROCC	**0.8791**	**0.8846**	**0.8528**	0.7496	0.7749	0.5692	0.7046	0.6243	0.5245
	KROCC	**0.6946**	**0.7016**	**0.6543**	0.5863	0.5768	0.4261	0.5340	0.4608	0.3696
	CC	**0.8732**	**0.8770**	**0.8425**	0.8090	0.7732	0.7359	0.6820	0.6135	0.5309
	RMSE	**0.6541**	**0.6448**	**0.7299**	0.7888	0.8511	0.9086	0.9815	1.0598	1.1372
CSIQ	SROCC	**0.9141**	**0.9218**	0.9072	**0.9108**	0.8719	0.7402	0.8075	0.7311	0.8053
	KROCC	**0.7442**	**0.7573**	0.7316	**0.7429**	0.6565	0.5644	0.6217	0.5532	0.6083
	CC	**0.9009**	**0.9084**	0.8918	**0.9153**	0.8569	0.8219	0.7963	0.7318	0.8002
	RMSE	**0.1140**	**0.1097**	0.1188	**0.1057**	0.1353	0.1496	0.1588	0.1789	0.1574
LIVE	SROCC	**0.9632**	**0.9647**	0.9445	**0.9631**	0.9479	0.9234	0.9274	0.9086	0.8755
	KROCC	**0.8340**	**0.8373**	0.7922	**0.8270**	0.7963	0.7540	0.7616	0.7413	0.6864
	CC	**0.9591**	**0.9611**	0.9430	**0.9598**	0.9449	0.9248	0.9213	0.9067	0.8721
	RMSE	**7.7324**	**7.5472**	9.0956	**7.6734**	8.9455	10.392	10.623	11.798	13.368
IVC	SROCC	**0.9272**	**0.9301**	0.8847	0.8966	**0.9018**	0.8978	0.7983	0.8347	0.6885
	KROCC	**0.7566**	**0.7641**	0.7012	0.7165	**0.7223**	0.7192	0.6036	0.6342	0.5220
	CC	**0.9375**	**0.9393**	0.8934	0.9028	**0.9119**	0.9080	0.8032	0.8498	0.7199
	RMSE	**0.4240**	**0.4179**	0.5474	0.5239	**0.4999**	0.5105	0.7258	0.6421	0.8456
MICT	SROCC	**0.9050**	**0.9056**	0.8864	**0.9086**	0.8794	0.8387	0.8614	0.8911	0.6130
	KROCC	**0.7282**	**0.7296**	0.7029	**0.7329**	0.6939	0.6413	0.6762	0.7129	0.4447
	CC	**0.9071**	**0.9072**	0.8935	**0.9144**	0.8887	0.8434	0.8710	0.8955	0.6426
	RMSE	**0.5268**	**0.5264**	0.5621	**0.5066**	0.5738	0.6723	0.6147	0.5569	0.9588
A57	SROCC	**0.9226**	-	**0.8394**	0.6223	0.8066	0.3185	**0.9355**	0.7981	0.6189
	KROCC	**0.7723**	-	**0.6478**	0.4589	0.6058	0.2378	**0.8031**	0.5932	0.4309
	CC	**0.9297**	-	**0.8504**	0.6158	0.8017	0.4548	**0.9472**	0.8020	0.6587
	RMSE	**0.0905**	-	**0.1293**	0.1936	0.1469	0.2189	**0.0781**	0.1468	0.1849

In order to evaluate the IQA models, four commonly used performance metrics are employed. The first two 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. These two metrics operate only on the rank of the data points and ignore the relative distance between data points. To compute the other two metrics we need to apply a regression analysis, as suggested by the video quality experts group (VQEG) [18], to provide a nonlinear mapping between the objective scores and the subjective mean opinion scores (MOS). 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 squared error (RMSE) between MOS and the objective scores after nonlinear regression. For the nonlinear regression, we used the following mapping function [19]:

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2 (x - \beta_3)}} \right) + \beta_4 x + \beta_5 \tag{13}$$

where β_i, i =1, 2, …, 5, are the parameters to be fitted.

TABLE III: RANKING OF THE IQA MODELS EVALUATED (EXCEPT FOR $FSIM_C$) ON THE 6 DATABASES

	TID2008	CSIQ	LIVE	IVC	MICT	A57
FSIM	**1**	**1**	**1**	**1**	2	2
MS-SSIM	2	3	4	5	4	3
VIF	4	2	2	4	**1**	6
SSIM	3	4	3	2	5	4
IFC	7	7	6	3	7	8
VSNR	5	5	5	7	6	**1**
NQM	6	8	7	6	3	5
PSNR	8	6	8	8	8	7

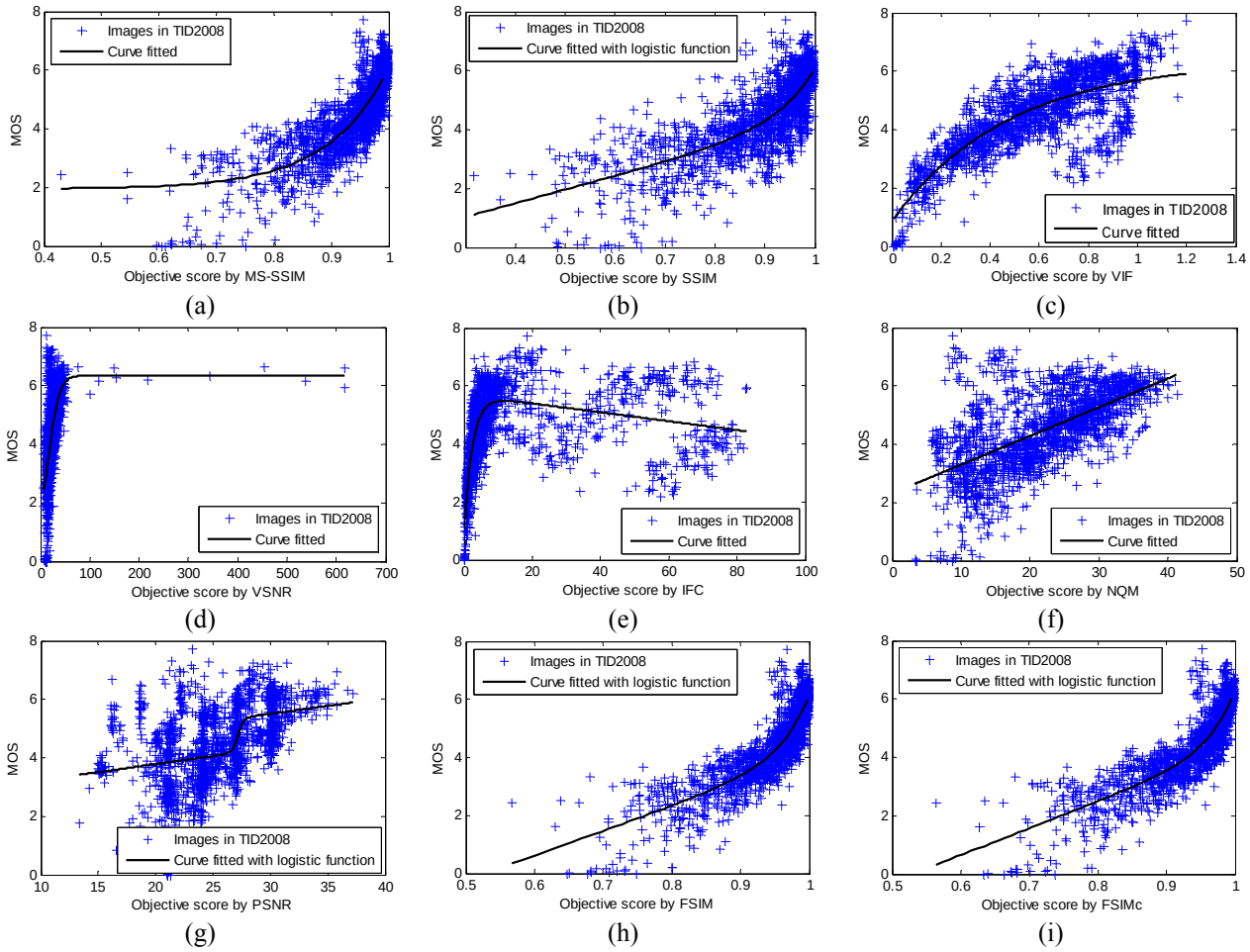


Fig. 3: Scatter plots of subjective MOS versus scores obtained by model prediction on the TID 2008 database. (a) MS-SSIM; (b) SSIM; (c) VIF; (d) VSNR; (e) IFC; (f) NQM; (g) PSNR; (h) FSIM and (i) $FSIM_C$.

Table II gives the performance comparison of the 9 IQA models on the 6 databases. The three models producing the greatest SROCC values for each database are highlighted in bold. It should be noted that except for $FSIM_C$, all the other eight IQA indices are based only on the luminance of the image. From Table II, we can see that the proposed feature-similarity based IQA model FSIM or $FSIM_C$ performs consistently well across all the databases. In order to demonstrate this consistency more clearly, in Table III we list the performance ranking of all the IQA models evaluated according to their SROCC values. For fairness, the

$FSIM_C$ index, which also exploits the chrominance besides luminance, is excluded in Table III.

From the experimental results summarized in Table II and Table III, we can see that our methods get the best results on almost all the databases, except for MCIT and A57. However, even on these two databases, the proposed FSIM (or $FSIM_C$) is only slightly worse than the best results. FSIM and $FSIM_C$ can get the most consistent and stable performance across all the 6 databases. By contrast, for the other methods, they may work well on a specific database but fail to provide good results on other databases. For example, although VIF can get rather pleasing results on LIVE, it performs very poorly on TID2008 and A57. Experimental results also demonstrate that the chromatic information of an image does affect its perceptible quality since $FSIM_C$ has better performance than FSIM on all color images databases.

Although 6 publicly available databases were used in evaluating the IQA models, they do not reflect equally well the performance of an IQA model because the 6 databases are very different in various 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 (refer to Table I please). In addition, the experimental configuration and the methodology to collect subjective scores also vary. Taking these factors into account, we think that TID 2008 [16] may be the most comprehensive database currently available for IQA model evaluation and thereby the evaluation results obtained on this database are the most convincing. Our proposed methods FSIM and $FSIM_C$ perform much better than the other IQA models evaluated on this database. Fig. 3 shows the scatter distributions of subjective MOS versus the predicted scores by the 9 IQA indices on the TID 2008 database. The curves shown in Fig. 3 were obtained by a nonlinear fitting according to Eq. (13). It can be seen that the objective scores predicted by FSIM and $FSIM_C$ correlate much more consistently with the subjective evaluations than the other methods.

V. CONCLUSIONS

In this paper, we proposed a novel human visual system (HVS) driven image quality assessment (IQA) metric, namely Feature-SIMilarity (FSIM) index. The underlying principle of FSIM is that HVS perceives an image mainly based on its salient low-level features. Two kinds of features, the phase congruency (PC) and the gradient magnitude (GM), are used in FSIM, and they represent complementary aspects of the image visual quality. The PC value is also used to weight the contribution of each point to the overall similarity of

two images. FSIM was then extended to $FSIM_C$ by incorporating the image chromatic features into consideration. The FSIM/$FSIM_C$ index reflects the similarity of two images at the feature level, which conforms well to the HVS. The FSIM/$FSIM_C$ index was compared with seven representative and prominent IQA models on six benchmark databases. The experimental results clearly demonstrated that FSIM and $FSIM_C$ outperform all the other state-of-the-art IQA models used in comparison. Particularly, they perform consistently well across all the test databases, validating that they are very robust IQA models.

REFERENCES

[1] 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.

[2] 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. Image Process.*, vol. 9, no. 4, pp. 636-650, Apr. 2000.

[3] D.M. Chandler and S.S. Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images", *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284-2298, Sep. 2007.

[4] H.R. Sheikh and A.C. Bovik, "Image information and visual quality", *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430-444, Feb. 2006.

[5] Z. Wang, E.P. Simoncelli, and A.C. Bovik, "Multi-scale structural similarity for image quality assessment", presented at the IEEE Asilomar Conf. Signals, Systems and Computers, Nov. 2003.

[6] H.R. Sheikh, A.C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics", *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117-2128, Dec. 2005.

[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. Image Process.*, vol. 18, no. 11, pp. 2385-2401, Nov. 2009.

[8] D. Marr, *Vision*. New York: W. H. Freeman and Company, 1980.

[9] D.Marr and E. Hildreth, "Theory of edge detection", *Proc. Roy. Soc. London. B.*, vol. 207, pp. 187-217, Feb. 1980.

[10] M.C. Morrone and D.C. Burr, "Feature detection in human vision: a phase-dependent energy model", *Proc. R. Soc. Lond. B*, vol. 235, no. 1280, pp. 221-245, Dec. 1988.

[11] M.C. Morrone, J. Ross, D.C. Burr, and R. Owens, "Mach bands are phase dependent", *Nature*, vol. 324, no. 6049, pp. 250-253, Nov. 1986.

[12] M.C. Morrone and R.A. Owens, "Feature detection from local energy", *Pattern Recognition Letters*, vol. 6, no. 5, pp. 303-313, Dec. 1987.

[13] P. Kovesi, "Image features from phase congruency", *Videre: J. Comp. Vis. Res.*, vol. 1, no. 3, pp. 1-26, 1999.

[14] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells", *J. Opt. Soc. Am. A*, vol. 4, no. 12, pp. 2379-2394, Dec. 1987.

[15] C. Yang and S.H. Kwok, "Efficient gamut clipping for color image processing using LHS and YIQ", *Optical Engineering*, vol. 42, no. 3, pp.701-711, Mar. 2003.

[16] 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", *Advances of Modern Radioelectronics*, vol. 10, pp. 30-45, 2009.

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

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

[19] H.R. Sheikh, M.F. Sabir, and A.C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms", *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440-3451, Nov. 2006.