EFFECTIVE PREPROCESSING STAGE IN THE FOURIER TRANSFORM DOMAIN FOR IMAGE QUALITY ASSESSMENT

Min Liu, Guangtao Zhai, Ke Gu, and Xiaokang Yang

Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, China

ABSTRACT

Image quality assessment (IQA) is currently an important research topic. In fact, the varying viewing distance between audiences and the display seriously affect the IQA accuracy, which has been largely overlooked. To this end, in this paper we take into account the image size and viewing distance as well as the preferential activation of V1 cells by vertical and horizontal contours, and thereby propose an adaptive frequency selection (AFS) model to preprocess input visual signals before IQA metrics are used. Our algorithm works by first applying Fourier Transform (FT) to the reference and distorted images to approximate the low-pass behavior of the human visual system, then extracting proper amount of low frequency components and partial high frequency components corresponding to vertical and horizontal directions, and finally reconstructing spatial images with the inverse FT. We validate the performance of AFS based PSNR and SSIM on the related LIVE, Toyama and IVC databases with clearly specified viewing conditions. Experimental results and comparative studies show the effectiveness of the proposed model.

Index Terms— Image quality assessment (IQA), Fourier transform (FT), image size, viewing distance, human visual system (HVS), V1 cells

1. INTRODUCTION

With the booming of digital imaging and signal processing technologies, image quality assessment (IQA) has become increasingly important in many practical applications, such as image enhancement [1]-[2], compression [3], restoration [4] and etc. In general, IQA can be divided into two categories: subjective and objective assessments. Subjective assessment is often regarded as the ultimate quality criterion, but it is greatly expensive, time-consuming and impractical for real-time systems. As a consequence, objective metrics have become an intensely research topic during the last decade. Based on the availability of the original image, objective IQA can be further divided into three types, namely full-reference (FR), reduced-reference (RR), and no-reference (NR) IQA. In this work, we focus on FR IQA approaches.

The vast majority of IQA methods were designed for the FR scenario, and most of them try to predict the perceptual difference between an original image and its distorted counterpart to predict the commonly encountered distortion types, e.g. compression, noise and blurring. Classical mean squared error (MSE) and peak signal-to-noise ratio (PSNR) measure the difference between the original and distorted images, but unfortunately, it has been widely recognized that MSE and PSNR are not well correlated with human judgment of quality, i.e. the mean opinion score (MOS) [5]. Therefore, numerous FR IQA methods have been developed for better performance. Up to now, one of the most popular methods is perhaps the structural similarity (SSIM) index [6], which combines three factors which is separately used to measure the loss of correlation, luminance distortion and contrast distortion. SSIM turned out to be more effective than PSNR on some existing image quality databases [7]-[9].

In our early work, it has been revealed that the image size and viewing distance have considerable impacts on the IQA performance [10]. Generally, as the viewing distance increases, human eyes capture fewer image details. On this base, we recently proposed an effective self-adaptive scale transform (SAST) model [10] to estimate the optimal scale in the spatial domain. Realizing the fact that images in frequency domain are more suitable for post-filtering, so we expand the idea of scale transform into the frequency domain based on the reasonable hypothesis that more frequency loss occurs as the viewing distance becomes farther.

The filtering process of human eyes can be assumed as a low-pass filter due to limited number of rods and the optic abbrevation [11]. On the other hand, previous studies on natural scene statistics have revealed that natural images tend to have more structures in the horizontal and vertical directions [12]. This paper therefore proposes a simple yet effective filtering method by adaptively selecting proper amount of low frequency components in the Fourier Transform (FT) domain. In brief, we adopt an adaptive star-shaped 0-1 mask to extract the valid low-frequency components from the FT transformed images, and then calculate the quality score using mainstream IQA methods (PSNR and SSIM are used in this paper) on the inversely transformed images.

This paper proceeds as follows. In Section 2, we first re-

978-1-4799-6536-6/14/$31.00 ©2014 IEEE

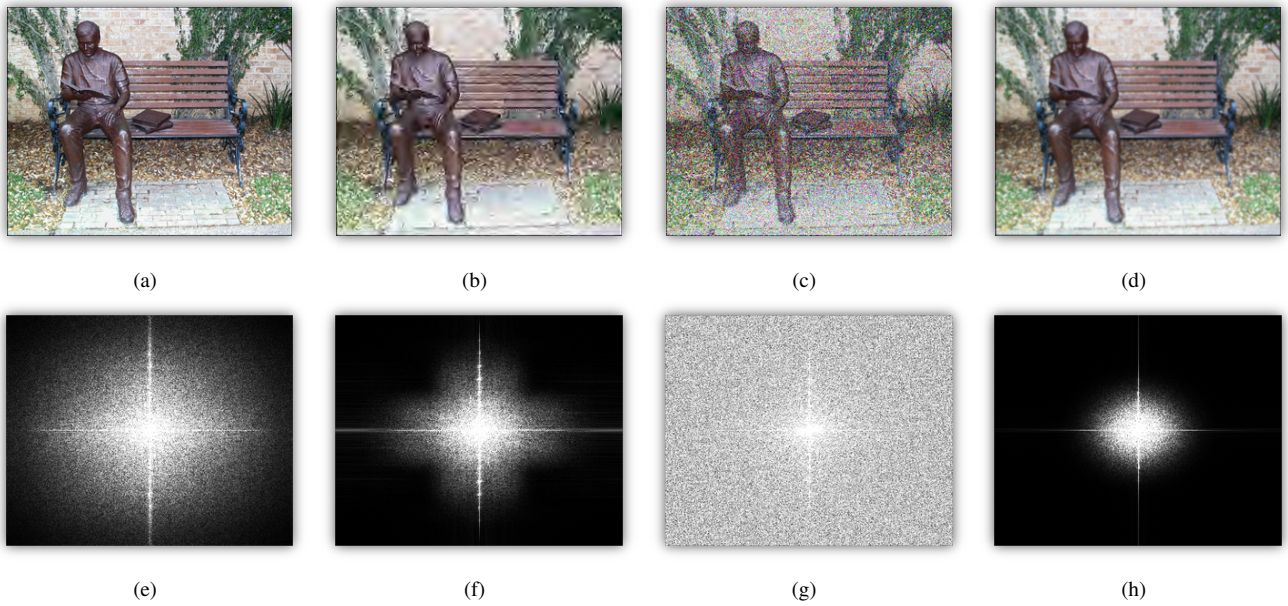

Fig. 1: (a)-(d) are the the original image, the JPEG, the noise and the blur distorted images of "sculpture", and (e)-(h) are respectively the spectrograms of (a)-(d).

view the existing scale transform models, and then put forward a new model by selecting suitable low and high frequency components from the input reference and distorted images according to the preferential activation of V1 cells and viewing conditions. Experimental results and comparative studies are given in Section 3. Section 4 concludes this paper.

2. THE PREPROCESSING STAGE

2.1. Introduction of PSNR/SSIM

A booming number of FR IQA algorithms have been put forward over the years, in this paper, we only adopt the most traditional PSNR and the benchmark SSIM. For the original image **x** and the distorted image **y**, PSNR is estimated by averaging the squared intensity differences of the reference and distorted images. It is simple to calculate, have clear physical meaning, i.e.,

$$\mathrm{PSNR} = 10 \times \log10((2^{\mathrm{n}}-1)^2/\mathrm{MSE}) \tag{1}$$

where n is the bit number of every sampling value. MSE is the mean square error between **x** and **y**.

SSIM index is a combination of luminance, contrast and structural similarity. The three components between two local image patches of the reference and distorted images are defined as

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

$$c(\mathbf{x},\mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{3}$$

$$s(\mathbf{x},\mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{4}$$

where C_1, C_2 and $C_3 = C_2/2$ are constant small numbers to avoid instability when $\mu_x^2 + \mu_y^2$, $\sigma_x^2 + \sigma_y^2$, $\sigma_x\sigma_y$ are very close to zero.

Finally, the aforementioned three comparisons are combined and the resulting similarity measure between the reference and distorted images is defined as

$$\begin{aligned}\mathrm{SSIM}(\mathbf{x},\mathbf{y}) &= l(\mathbf{x},\mathbf{y}) \cdot c(\mathbf{x},\mathbf{y}) \cdot s(\mathbf{x},\mathbf{y}) \\ &= \frac{(2\mu_x\mu_y + C_1)(2\sigma_x y + C_2)}{(\mu_x^2 + \mu_y^2 + C1)(\sigma_x^2 + \sigma_y^2 + C_2)}.\end{aligned} \tag{5}$$

The overall image quality is evaluated by averaging SSIM.

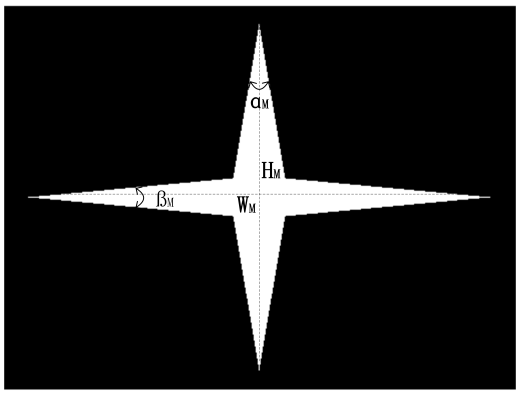

Fig. 2: The star-shaped binary mask for frequency selection.

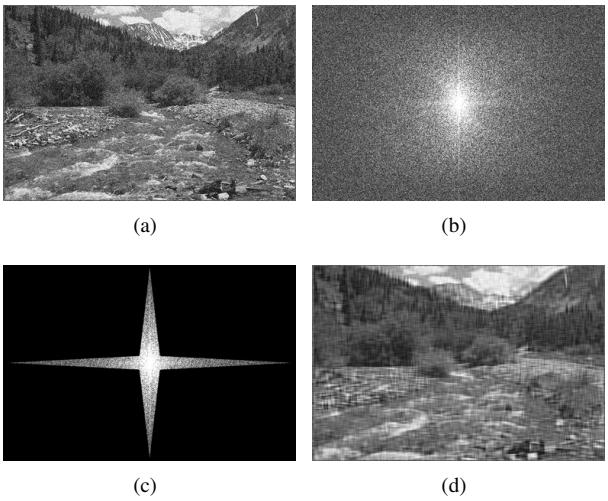

Fig. 3: (a) is the white noise contaminated image of "stream" from the LIVE database, (b) is the Fourier transform of (a), (c) is the frequency components selected by the proposed AFS model, (d) is the inverse transformed image from (c).

2.2. The previous preprocessing models

To approximate the real viewing conditions under different viewing distances, a down-sample model Z_α is provided for preprocessing images before using SSIM [13], so as to evaluate images viewed from a specific distance:

$$Z_\alpha = max(1, round(H_I/256)) \tag{6}$$

where H_I is the image height. However, due to the usage of the rounding operation, the resultant scale parameter is unstable with increased image sizes and therefore brings about only limited performance gain.

To this end, we recently proposed a simple and empirical self-adaptive down-sample scale Z_S [10] in the spatial domain using the concept of human visual angle:

$$\begin{aligned} Z_S &= \sqrt{\frac{H_I \cdot W_I}{H_V \cdot W_V}} \\ &= \sqrt{\frac{1}{4tan(\frac{\theta_{H_V}}{2}) \cdot tan(\frac{\theta_{W_V}}{2})} \cdot (\frac{H_I}{D})^2 \cdot \frac{W_I}{H_I}} \end{aligned} \tag{7}$$

where W_I is the image width. H_V and W_V are the visual height and width individually. θ_{H_V} ($\approx 40^o$) and θ_{W_V} ($\approx 50^o$) are the actual visual angle (i.e. angle of gaze) of human eyes, which is assumed to be one third of total visual angle.

2.3. Proposed AFS model

The scale transforms proposed in equations (6) and (7) both work in the spatial domain. On the other hand, frequency domain processing has been an important tool for image processing due to its efficiency and effectiveness. So, in this paper, we will propose an adaptive scale transform in the frequency domain for image processing. Particularly, we use the fast Fourier transform (FFT) to firstly turn the images into the frequency domain:

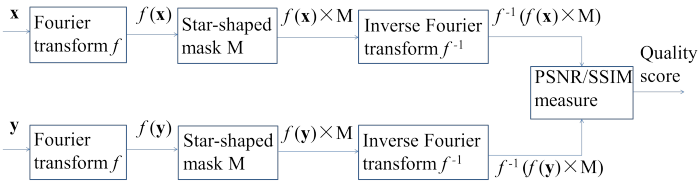

Fig. 4: The block of the AFS model.

$$\begin{aligned} f(\mathbf{x}) &= \sum_{i=0}^{I-1}\sum_{j=0}^{J-1} \mathbf{x}(i,j) W_I^{k_1 i} W_J^{k_2 j} \\ 0 \le k_1 \le I-1 &\quad 0 \le k_2 \le J-1. \end{aligned} \tag{8}$$

As shown in Fig. 1, most energies are concentrated on low frequencies in those transformed images. One exception is the white noise contaminated images which have spreading frequency components (but as can be seen, the image energy is still well concentrated).

To mimic the low-pass filtering property of human eyes, we use a 2D binary mask to extract the central section of the 2D frequency plane. As discussed, to better protect those vertical and horizontal frequencies, which correspond to important image structures as manifested in Fig. 1 (e)-(h), we use a star-shaped pattern in the mask as shown in Fig. 2. The frequencies located in the white part are extracted.

Considering that as viewing distance increases, the visible high frequency image structures reduce, we design the adaptive binary mask as:

$$H_M = max(1, \frac{a_1}{D^{a_2} + a_3}) - \frac{1}{2} \times H_I \tag{9}$$

where D is the viewing distance, and $\mathbf{a} = \{a_1, a_2, a_3\}$ is a set of tuning parameters. The width of the star is W_M, and W_M/H_M is set to be the same as the image aspect ratio, i.e. W_I/H_I. In addition, the angles of the mask are α_M and β_M respectively.

After extracting suitable amount of low and partial high frequency components from the Fourier domain, we transform the image back to the spatial domain. And this AFS postfiltering process is illustrated in Fig. 3. (a) and (b) are the original and its Fourier transformed images respectively. Then the star-shaped mask is utilized in (c) to extract corresponding frequency, and (d) is the inverse transformed image of (c).

Table 1: PLCC, SROCC and RMSE results (after nonlinear regression) of PSNR, $PSNR_\alpha$, MS-PSNR, $PSNR_S$, $PSNR_{FS}$, SSIM, $SSIM_\alpha$, MS-SSIM, $SSIM_S$, $SSIM_{FS}$, IGM and GMSD on the LIVE, IVC and Toyama databases

	LIVE database [7]			IVC database [8]			Toyama database [9]		
Metrics	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
PSNR	0.8701	0.8756	13.4685	0.7192	0.6886	0.8465	0.6355	0.6132	0.9662
$PSNR_\alpha$ [13]	0.8995	0.9031	11.9398	0.8791	0.8721	0.5808	0.7654	0.7583	0.8053
MS-PSNR [15]	0.9071	0.9110	11.5030	0.8388	0.8340	0.6634	0.7522	0.7411	0.8246
$PSNR_S$ [10]	0.9134	0.9160	11.1209	0.8953	0.8889	0.5428	0.8343	0.8272	0.6898
$PSNR_{FS}$	**0.9189**	**0.9208**	**10.7780**	**0.9035**	**0.8974**	**0.5221**	**0.8280**	**0.8237**	**0.7018**
SSIM [6]	0.9014	0.9104	11.8323	0.7924	0.7788	0.7431	0.7978	0.7870	0.7545
$SSIM_\alpha$ [13]	0.9300	0.9391	10.0439	0.9117	0.9017	0.5007	0.8877	0.8794	0.5762
MS-SSIM [15]	0.9338	0.9448	9.7788	0.8931	0.8846	0.5480	0.8926	0.8870	0.5641
$SSIM_S$ [10]	0.9306	0.9446	10.0020	0.9042	0.8905	0.5203	0.9072	0.9048	0.5265
$SSIM_{FS}$	**0.9383**	**0.9578**	**9.4514**	**0.9195**	**0.9118**	**0.4790**	**0.9011**	**0.8977**	**0.5425**
IGM [16]	0.9565	0.9581	7.9686	0.9128	0.9025	0.4976	0.8708	0.8654	0.6152
GMSD [17]	0.9568	0.9603	7.9447	0.8971	0.9148	0.0209	0.8576	0.8528	0.6437

The AFS model processed images are then fed into the IQA metrics of PSNR/SSIM as:

$$\mathrm{PSNR}_{FS}(\mathbf{x},\mathbf{y}) = \mathrm{PSNR}(\mathbf{x}',\mathbf{y}') \tag{10}$$

$$\mathrm{SSIM}_{FS}(\mathbf{x},\mathbf{y}) = \mathrm{SSIM}(\mathbf{x}',\mathbf{y}') \tag{11}$$

where

$$\mathbf{x}' = f^{-1}(f(\mathbf{x}) \times M) \tag{12}$$

$$\mathbf{y}' = f^{-1}(f(\mathbf{y}) \times M) \tag{13}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the original and distorted images respectively. M is the star-shaped binary mask with the same size of $\mathbf{x}$ and $\mathbf{y}$. f represents the Fourier transform function, and f^{-1} is the inverse Fourier transform function. The detail of the AFS model is displayed in Fig. 4.

Table 2: Specifications of LIVE, IVC and Toyama databases.

Dataset	LIVE	IVC	Toyama
Image size ($W_I \times H_I$)	768×512, 480×720, 640×512, 632×505, 634×505, 618×453, 610×488, 627×482, 634×438	512×512	768×512
D / H_I	3∼3.75	4	6
No.	779	185	168

3. EXPERIMENTAL RESULTS

Three image databases, including LIVE [7], IVC [8] and Toyama [9], are utilized in this paper as testing beds. We chose those databases among many candidates because of their specific image sizes, viewing distances/image heights. A four-parameter logistic function is chosen to fit the scores of our method to subjective scores [14]

$$Quality(z) = \frac{\beta_1 - \beta_2}{1 + exp(-(z-\beta_3)/\beta_4)} + \beta_2 \tag{14}$$

where z is the input score, $Quality(z)$ is the mapped score, and β_1 to β_4 are free parameters to be determined during the curve fitting process. According to the suggestion provided by VQEG [14], we then use three evaluation metrics to compare the performance against the MOS/DMOS scores: 1) Pearson linear correlation coefficient (PLCC), which is employed to assess the prediction accuracy; 2) Spearman rank-ordered correlation coefficient (SROCC), which aims to evaluate prediction monotonicity; 3) root mean-squared error (RMSE), which measures how well an algorithm's prediction correlates with the raw opinion scores.

The optimal model parameters of the proposed AFS model were obtained on LIVE database, and validated on IVC and Toyama databases. To carefully compare the performances of our algorithm and some existing related metrics (based on benchmark algorithms PSNR/SSIM): $PSNR_\alpha$/MS-PSNR/$PSNR_S$, $SSIM_\alpha$/MS-SSIM/$SSIM_S$ as tabulated in Table 1, we can easily find that the proposed $PSNR_{FS}$ and $SSIM_{FS}$ methods have achieved promising performance on the three databases. Furthermore, our results can also be comparable to the state-of-the-art internal generative mechanism (IGM)

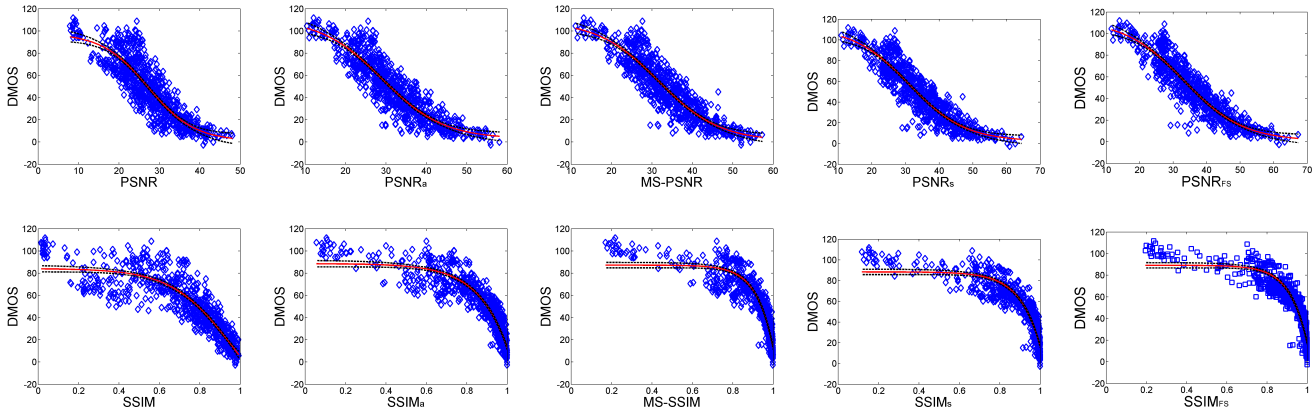

Fig. 5: Scatter plots of DMOS vs. PSNR/PSNR$_\alpha$/MS-PSNR/PSNR$_S$/PSNR$_{FS}$, SSIM/SSIM$_\alpha$/MS-SSIM/SSIM$_S$/SSIM$_{FS}$ on the LIVE database. The red lines are curves fitted with the logistic function and the back dash lines are 95% confidence intervals.

Table 3: Database size-weighted average performance of IQA metrics over three databases.

Metrics	PLCC	SROCC	RMSE
PSNR	0.8106	0.8061	9.5503
PSNR$_\alpha$	0.8763	0.8765	8.4310
MS-PSNR	0.8729	0.8732	8.1467
PSNR$_S$	0.8987	0.8984	7.8441
PSNR$_{FS}$	**0.9029**	**0.9026**	**7.6065**
SSIM	0.8682	0.8706	8.3760
SSIM$_\alpha$	0.9207	0.9241	7.0792
MS-SSIM	0.9210	0.9264	6.9027
SSIM$_S$	0.9228	0.9299	7.0462
SSIM$_{FS}$	**0.9297**	**0.9414**	**6.6629**
IGM	0.9366	0.9353	5.6563
GMSD	0.9323	0.9369	5.5662

Table 4: The rates of increase of PSNR$_\alpha$, MS-PSNR, PSNR$_S$, PSNR$_{FS}$ over PSNR, and SSIM$_\alpha$, MS-SSIM, SSIM$_S$, SSIM$_{FS}$ over SSIM on the database-weighted average results.

Metrics	PLCC	SROCC	RMSE
PSNR	-	-	-
PSNR$_\alpha$	8.11%	8.73%	11.72%
MS-PSNR	7.69%	8.32%	14.79%
PSNR$_S$	10.87%	11.45%	17.87%
PSNR$_{FS}$	**11.39%**	**11.97%**	**20.35%**
SSIM	-	-	-
SSIM$_\alpha$	6.05%	6.15%	15.48%
MS-SSIM	6.08%	6.41%	17.59%
SSIM$_S$	6.29%	6.81%	15.88%
SSIM$_{FS}$	**7.08%**	**8.13%**	**20.45%**

[16], and the newest gradient magnitude similarity deviation (GMSD) [17] metric. The database size-weighted average performance of IQA metrics over three databases and the rates of increase of the metrics over PSNR/SSIM are tabulated in Table 3 and Table 4 respectively. Figure. 5 shows the scatter plots of all metrics compared vs. DMOS on the largest LIVE database which present favorable convergency. Figure. 6 and 7 display the well performed PSNR/MS-PSNR/PSNR$_{FS}$, SSIM/MS-SSIM/SSIM$_{FS}$ on IVC and Toyama databases.

4. CONCLUSION

This paper proposes a simple yet effective adaptive frequency selection (AFS) model to improve performance of image quality metrics. The AFS model approximates the low-pass filtering process of human eyes with proper frequency protection for the vertical and horizontal directions. Experimental results on LIVE, IVC and Toyama databases, which have definite records of image sizes and viewing distances, are provided to confirm that both the AFS based methods PSNR$_{FS}$ and SSIM$_{FS}$ outperform existing scale transform models, such as the multi-scale model and SAST model, and is even matchable with that state of the art IGM and GMSD methods.

Acknowledgment

This work was supported in part by NSFC (61025005, 61371146, 61221001, 61390514), 973 Program (2010CB731401) and FANEDD (201339).

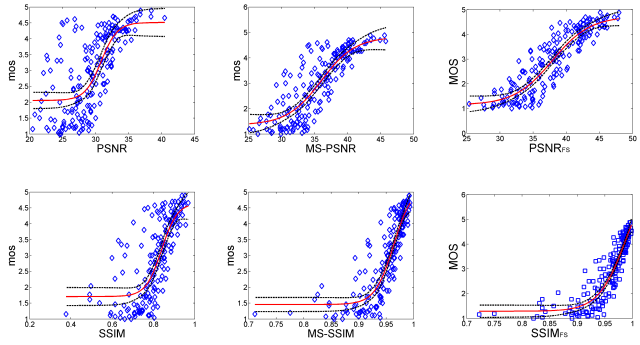

Fig. 6: Scatter plots of MOS vs. PSNR/MS-PSNR/PSNR$_{FS}$, SSIM/MS-SSIM/SSIM$_{FS}$ on IVC database.

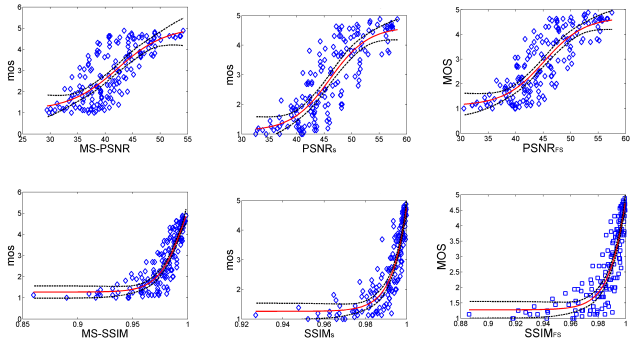

Fig. 7: Scatter plots of MOS vs. PSNR/MS-PSNR/PSNR$_{FS}$, SSIM/MS-SSIM/SSIM$_{FS}$ on Toyama database.

5. REFERENCES

[1] K. Panetta, S. S. Agaian, Y. Zhou, and E. J. Wharton, "Parameterized logarithmic framework for image enhancement," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 2, pp. 460-473, April 2011.

[2] G. Zhai, J. Cai, W. Lin, X. Yang, and W. Zhang, "Three dimensional scalable video adaptation via user-end perceptual quality assessment," *IEEE Trans. Broadcasting*, vol. 54, no. 3, pp. 719-727, September 2008.

[3] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang, and M. Etoh, "Cross-dimensional perceptual quality assessment for low bitrate videos," *IEEE Trans. Multimedia*, vol. 10, no. 7, pp. 1316-1324, November 2008.

[4] C. Liu, R. Szeliski, S. B. Kang, C. L. Zitnick, and W. T. Freeman, "Automatic estimation and removal of noise from a single image," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 30, no. 2, pp. 299-314, February 2008.

[5] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Processing Mag.*, vol. 26, pp.98-117, January 2009.

[6] 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*, 2004.

[7] H. R. Sheikh, K. Seshadrinathan, A. K. Moorthy, Z. Wang, A. C. Bovik, and L. K. Cormack, "Image and video quality assessment research at LIVE," [Online]. Available: http://live.ece.utexas.edu/research/quality/

[8] A. Ninassi, P. Le Callet, and F. Autrusseau, "Subjective quality assessment-IVC database," [Online]. Available: http://www2.irccyn.ec-nantes.fr/ivcdb

[9] Y. Horita, K. Shibata, Y. Kawayoke, and Z. M. P. Sazzad, "MICT image quality evaluation database," [Online]. Available: http://mict.eng.u-toyama.ac.jp/mict/index2.html

[10] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Self-adaptive scale transform for IQA metric," *IEEE International Symposium on Circuits and Systems*, 2012.

[11] [Online]. Available: http://en.wikipedia.org/wiki/Human-visual-system-model

[12] Bruce C. Hansen and Edward A. Essock, "A horizontal bias in human visual processing of orientation and its correspondence to the structural components of natural scenes," *Journal of Vision*, 2004.

[13] Weisi Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image R.*, 2011.

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

[15] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," *IEEE Asilomar Conference Signals, Systems and Computers*, November 2003.

[16] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism ," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 43-54, January 2013.

[17] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684-695, February 2014.