

Visual-Saliency-Enhanced Image Quality Assessment Indices

Joe Yuchieh Lin*, Tsung Jung Liu*, Weisi Lin[†] and C.-C. Jay Kuo*

*Ming Hsieh Department of Electrical Engineering, and Signal and Image Processing Institute,
University of Southern California, Los Angeles, CA 90089, USA

E-mail: yu-chieh.lin@usc.edu, liut@usc.edu, cckuo@sipi.usc.edu

[†]School of Computer Engineering, Nanyang Technological University, Singapore 639798

E-mail: wslin@ntu.edu.sg

Abstract—Modern image quality assessment (IQA) indices, *e.g.* SSIM and FSIM, are proved to be effective for some image distortion types. However, they do not exploit the characteristics of the human visual system (HVS) explicitly. In this work, we investigate a method to incorporate the human visual saliency (VS) model in these full-reference indices, and call the resulting indices $SSIM_{VS}$ and $FSIM_{VS}$, respectively. First, we decompose an image into non-overlapping patches, calculate visual saliency, and assign a parameter ranging from 0 and 1 to each patch. Then, the local SSIM or FSIM values of the patches are weighed by the said parameter. Finally, the weighed similarity of all patches are integrated into one single index for the whole image. Experimental results are given to demonstrate the improved performance of the proposed VS-enhanced indices.

I. INTRODUCTION

The mean-squared-errors (MSE) index has been widely used to assess the quality of distorted images or videos with respect to their original ones for a long history. Human visual experience is affected by several psychovisual factors [1], [2], but the MSE index does not take these factors into account. Quite a few full-reference image and video quality assessment indices have been proposed during the last decade to capture human visual experience more accurately [3], [4]. Such image quality assessment (IQA) indices use statistical features to measure perceived image quality. They characterize statistical properties of local image patches such as those related to luminance, contrast and structure (*e.g.* edges). The index of the whole image is the average of IQA values from all patches, where each patch in an image weighs equally.

Based on psychovisual experiments [5], the HVS tends to focus on one particular region of an image while neglecting others in its surroundings. This is known as visual saliency (VS) or visual attention. While the VS effect has been demonstrated, few feature-based IQA indices take it into account. Intuitively speaking, if the VS can be properly integrated

in IQA indices, we should obtain more accurate ones.

In this work, we investigate to incorporate the VS in exemplary IQA indices such as SSIM and FSIM, and call the resulting ones $SSIM_{VS}$ and $FSIM_{VS}$, respectively. To be specific, we first decompose an image into non-overlapping patches, calculate the visual saliency degree (a higher value implying higher saliency), and assign a parameter ranging from 0 and 1 to each small patch. Then, the similarity degree of the corresponding patches is weighed by the associated VS parameter. Finally, the similarity of all patches are integrated into one single index for the whole image.

An example is given in Fig. 1 to illustrate the above idea. Fig. 1 (a) is a blurred image from the TID2008 database [7]. Its visual saliency (VS) map, generated by a method in [8], is shown in Fig. 1 (b). The values of these maps are normalized in $[0, 1]$, where a brighter region has a larger value. As indicated by the VS map in Fig. 1 (b), the airplane is more visible to the HVS as compared to the background while the yellow propeller region receives the highest attention. Thus, we should give a different weight to the IQA value of a different region.

The rest of this work is organized as follows. The related previous work is reviewed in Section II. Then, the details of the proposed HVS-based IQA indices are presented in Section III. Experimental results are shown in Section IV and concluding remarks are given in Section V.

II. REVIEW OF PREVIOUS WORK

In this work, we will choose two representative full-reference IQA indices; namely, SSIM [9] and

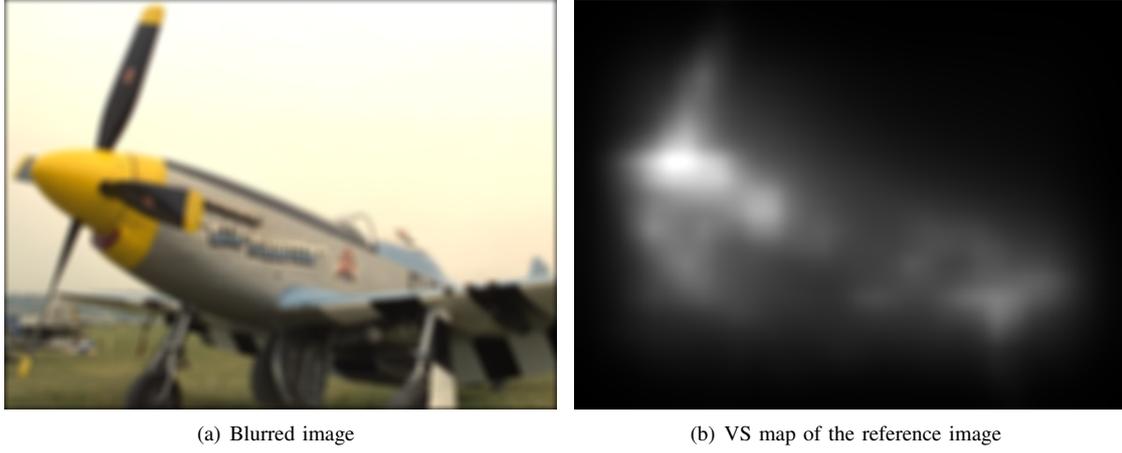


Fig. 1. An example of a blurred image and its VS map.

FSIM [10], and focus on their enhancement by incorporating the VS attribute. These two IQA indices are reviewed in this section.

The SSIM index[9] can be written as

$$SSIM = \overline{SSIM(\mathbf{x})_{\mathbf{x} \in \Omega}}, \quad (1)$$

where \mathbf{x} denotes a co-located block of the distorted and reference images, whose size is typically set to 11×11 , and $SSIM(\mathbf{x})$ is a function defined as

$$SSIM(\mathbf{x}) = \frac{(2\mu_d(\mathbf{x})\mu_r(\mathbf{x}) + c_1)(2\sigma_{d,r}(\mathbf{x}) + c_2)}{(\mu_d^2(\mathbf{x}) + \mu_r^2(\mathbf{x}) + c_1)(\sigma_d^2(\mathbf{x}) + \sigma_r^2(\mathbf{x}) + c_2)}, \quad (2)$$

and where μ_d and μ_r are block means, σ_d and σ_r are block standard deviation, and $\sigma_{d,r}$ are block cross-standard-deviation of block \mathbf{x} in distorted and reference images, respectively. The SSIM index is developed to measure the local change in luminance, contrast, and structure. As shown in Eq. (1), it is the mean of local SSIM values over the entire image domain denoted by Ω .

There are two FSIM indices[10] one for gray-level images and the other for color images. The gray-level FSIM index can be expressed as

$$FSIM = \frac{\sum_{\mathbf{x} \in \Omega} [S_{pc}(\mathbf{x})] [S_G(\mathbf{x})] [PC_m(\mathbf{x})]}{\sum_{\mathbf{x} \in \Omega} [PC_m(\mathbf{x})]}, \quad (3)$$

where S_{pc} and S_G measure the similarities in phase congruency (PC) and gradient magnitude (GM), respectively, and PC_m denotes the maximum phase congruency of co-located block \mathbf{x} between the distorted and the reference images. The color FSIM

index is of the following form:

$$FSIMC = \frac{\sum_{\mathbf{x} \in \Omega} [S_{pc}(\mathbf{x})] [S_G(\mathbf{x})] [S_C(\mathbf{x})]^\lambda [PC_m(\mathbf{x})]}{\sum_{\mathbf{x} \in \Omega} [PC_m(\mathbf{x})]}, \quad (4)$$

where S_C measures similarities of the color components in the YIQ color space.

As discussed earlier, both SSIM and FSIM do not take the VS into account. In the next section, we investigate a way to incorporate it to get their enhanced versions.

III. VS-ENHANCED SSIM AND FSIM

There are several methods to improve the performance of existing IQA indices. One is to fuse them together. A machine learning method was used in [11] to fuse multiple IQA methods for performance improvement. In this work, we would like to improve a single IQA index by adaptively changing the weighting of contribution from its spatial components.

We adopt the method in [8] to generate the VS map for an input image. It consists of three steps: 1) extracting certain features over the image, 2) forming activation maps based on the extracted features, and then 3) normalizing them and highlighting key locations. An input image Boat and its VS map normalized to $[0, 1]$ are shown in Fig. 2. As shown in Fig. 2(a), the boat and the shore attract most of the HVS attention while the ocean is less important. This experience is consistent with the generated VS map in Fig. 2(b).



(a) Boat Image

(b) The VS map of the boat image

Fig. 2. The input boat image and its VS map.

The proposed VS-enhanced SSIM ($SSIM_{VS}$) and VS-enhanced FSIM ($FSIM_{VS}$) are built based on the SSIM map and the FSIM map and weighed by the VS map. Mathematically, we have

$$SSIM_{VS} = \frac{\sum_{\mathbf{x} \in \Omega} [SSIM(\mathbf{x})]^\theta [VS(\mathbf{x})]^\nu}{\sum_{\mathbf{x} \in \Omega} [VS(\mathbf{x})]^\nu}, \quad (5)$$

where

$$\theta = K_{SSIM} \times \overline{SSIM_{\mathbf{x} \in \Omega}(\mathbf{x})}, \quad (6)$$

$$\nu = K_{VS} \times CORR(VS_r(\mathbf{x}), VS_d(\mathbf{x})), \quad (7)$$

As shown above, the $SSIM_{VS}$ includes local SSIM and VS, where the VS value is normalized to $[0, 1]$. The exponent parameter, θ , is used to adjust the contribution of local SSIM values. The exponent parameter, ν , depends on the correlation between the reference and the distorted images. If these two images have lower correlation, then the visual attention is changed by the distortions, and the contribution from the VS value would be reduced. Furthermore, these parameters are adaptive to image content since one can choose K_{SSIM} and K_{VS} to be adaptive to the input image. Similarly, we can define the VS-enhanced FSIM and FSIMC indices as follows:

$$FSIM_{VS} = \frac{\sum_{\mathbf{x} \in \Omega} [S_{pc}(\mathbf{x})]^\alpha [S_G(\mathbf{x})]^\beta [PC_m(\mathbf{x})]^\gamma [VS_c(\mathbf{x})]^\nu}{\sum_{\mathbf{x} \in \Omega} [PC_m(\mathbf{x})]^\gamma [VS_c(\mathbf{x})]^\nu}, \quad (8)$$

and

$$FSIMC_{VS} = \frac{\sum_{\mathbf{x} \in \Omega} [S_{pc}(\mathbf{x})]^\alpha [S_G(\mathbf{x})]^\beta [PC_m(\mathbf{x})]^\gamma [S_C(\mathbf{x})]^\lambda [VS(\mathbf{x})]^\nu}{\sum_{\mathbf{x} \in \Omega} [PC_m(\mathbf{x})]^\gamma [VS(\mathbf{x})]^\nu}, \quad (9)$$

where

$$\alpha = K_{pc} \times \overline{S_{pc, \mathbf{x} \in \Omega}(\mathbf{x}) + PC_{m, \mathbf{x} \in \Omega}(\mathbf{x})}, \quad (10)$$

$$\beta = K_G \times \overline{S_{G, \mathbf{x} \in \Omega}}, \quad (11)$$

$$\gamma = K_{PC_m} \times \overline{PC_{m, \mathbf{x} \in \Omega}(\mathbf{x})}, \quad (12)$$

$$\lambda = K_{Color} \times \overline{S_{I, \mathbf{x} \in \Omega} + S_{Q, \mathbf{x} \in \Omega}(\mathbf{x})}. \quad (13)$$

The parameters, K_{pc} , K_G , K_{PC_m} and K_{Color} , are determined by experiments. They should be kept in a range so that each component would not dominate in $FSIM_{VS}$ and $FSIMC_{VS}$. Also, parameter α is controlled by both the PC and the PC_m maps to amplify the PC effect in the resulting index.

IV. EXPERIMENTAL RESULTS

We conducted the performance evaluation on the proposed $SSIM_{VS}$, $FSIM_{VS}$ and $FSIMC_{VS}$ using the TID2008 database [7] since it contains a wide range of distortion types. We chose the following parameters empirically based on a subset of the TID2008, containing 10 reference images and their 680 distorted images:

$$\begin{aligned} K_{SSIM} &= 0.09, & K_{VS} &= 1.25, & K_{JND} &= 0.48, \\ K_{pc} &= 0.7, & K_G &= 0.41, & K_{PC_m} &= 1.96, \\ & & & & K_{Color} &= 0.02. \end{aligned}$$

TABLE I
PERFORMANCE OF SSIM_{VS}

| | KROCC | PLCC | SROCC | RMSE |
|--------------------|--------|--------|--------|--------|
| SSIM | 0.5768 | 0.7732 | 0.7749 | 0.8511 |
| SSIM _{VS} | 0.6409 | 0.8258 | 0.8336 | 0.7567 |

TABLE II
PERFORMANCE OF FSIM_{VS}

| | KROCC | PLCC | SROCC | RMSE |
|--------------------|--------|--------|--------|--------|
| FSIM | 0.6946 | 0.8738 | 0.8805 | 0.6525 |
| FSIM _{VS} | 0.7133 | 0.8929 | 0.8941 | 0.5997 |

The performance measures in this work include: the Spearman rank-order correlation coefficient (SROCC), the Kendall rank-order correlation coefficient (KROCC), the Pearson linear correlation coefficient (PLCC), and the root mean squared error (RMSE). The first two are used to measure the prediction monotonicity of an IQA index. The third one, PLCC, computes the correlation coefficient between the MOS and the predicted MOS obtained by an IQA index to evaluate the prediction accuracy. The last one, RMSE, is used to measure the error between the predicted and the true MOS.

We compare the performance of SSIM_{VS}, FSIM_{VS} and FSIMC_{VS} with that of SSIM, FSIM and FSIMC in Tables I, II and III, respectively. We see consistent performance improvement of the proposed method with respect to all performance measures in these three tables.

V. CONCLUSION AND FUTURE WORK

In this work, we have incorporated the visual saliency (VS) model in SSIM, FSIM and FSIMC and showed that the resulting indices, SSIM_{VS}, FSIM_{VS} and FSIMC_{VS} outperform the original indices, respectively, when they have tested on the TID-2008 IQA database. It has been demonstrated that a VS model can facilitate better visual quality prediction. The proposed metric can be used as evaluation feedback for image enhancement algorithms.

We would like to extend the current work along two directions in the near future. First, it is interesting to consider other HVS attributes such as

TABLE III
IMPROVED PERFORMANCE OF FSIMC_{VS}

| | KROCC | PLCC | SROCC | RMSE |
|---------------------|--------|--------|--------|--------|
| FSIMC | 0.6991 | 0.8762 | 0.8840 | 0.6468 |
| FSIMC _{VS} | 0.7226 | 0.8997 | 0.9021 | 0.5871 |

the masking effect and see what they enhance IQA indices alongside with a VS model. Second, it is worthwhile to incorporate the HVS attributes to video quality assessment (VQA) indices, since VS is more meaningful in video; For example, a viewer may have time to view all details throughout the image given a reasonable time interval, while he or she definitely has no time to examine every detail in video given a typical frame rate. We attempt to understand the HVS in temporal saliency and explore these attributes to design VQA indices.

REFERENCES

- [1] B. Girod, "What's wrong with mean-squared error?" *Digital images and human vision*, MIT press, 1993.
- [2] L. Itti, C. Koch and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.20, no.11, pp.1254,1259, Nov. 1998.
- [3] W.S. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: a survey," *Journal of Visual Communication and Image Representation*, Vol. 22, No. 4, pp. 297-312, May 2011.
- [4] T.-J. Liu, Y.-C. Lin, W.S. Lin and C.-C. J. Kuo, "Visual quality assessment: recent developments and coding applications," *APSIPA Transactions on Signal and Information Processing*, vol. 3, 2013.
- [5] J. Jonides, "Further toward a model of the mind: eye's movement." *Bulletin of the Psychonomic Society* 21.4, pp. 247-250, 1983.
- [6] A. B. Watson, "DCTune: a technique for visual optimization of DCT quantization matrices for individual images," *Soc. Inf. Display Dig. Tech. Papers*, vol. XXIV, pp. 946-949, 1993.
- [7] 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.
- [8] J. Harel, C. Koch and P. Perona, "Graph-based visual saliency," *Proceedings of Neural Information Processing Systems (NIPS)*, 2006.
- [9] 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, 2004.
- [10] Lin Zhang, Lei Zhang, X. Mou and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE Trans. Image Processing*, vol. 20, no. 8, pp. 2378-2386, 2011.
- [11] T.-J. Liu; W.S. Lin and C.-C. J. Kuo, "Image quality assessment using multi-method fusion," *IEEE Transactions on Image Processing*, vol.22, no.5, pp.1793-1807, May 2013.