# Visual Horizontal Effect for Image Quality Assessment

Lin Ma, *Student Member, IEEE*, Songnan Li, *Student Member, IEEE*, and King Ngi Ngan, *Fellow, IEEE*

***Abstract*—In this paper, an image quality metric is proposed by modeling the visual Horizontal Effect (HE) and saliency property over structural distortions. Specifically, Structrue SIMilarity (SSIM) is firstly performed to obtain the structural distortion map. Subsequently, the obtained distortion map is refined by the visual HE model, which depicts visual sensitivities of oriented stimuli over different oriented contents. Finally, in order to describe the local Human Visual System (HVS) conspicuities, a saliency pooling strategy is proposed to generate the resulting image quality index. The experimental results have demonstrated that the proposed method outperforms SSIM and Visual Information Fidelity (VIF), which indicates that the obtained similarity index is more consistent with the perceptual evaluation of image quality.**

***Index Terms*—Horizontal effect (HE), human visual system (HVS), image quality assessment (IQA), structure SIMilarity (SSIM).**

## I. Introduction

THE measurement of image quality plays a very important role in many image processing tasks, such as image compression and enhancement, etc. As humans are the end-users of images and videos, one straightforward way for evaluating image quality is subjective testing. However, it is very expensive and time-consuming, which makes it impractical for the image processing applications. These drawbacks lead to the development of image quality metrics that can automatically evaluate the image perceptual quality.

The most popularly used image quality metrics are the Mean Squared Error (MSE) and the related Peak Signal-to-Noise Ratio (PSNR). These measurements are appealing for their simple formulation. However, they do not correlate well with the HVS [1], [2], because they just focus on the pixel value differences but ignore the image content and human perception properties. Recently, the psychophysical HVS features [4] have been discussed for assessing the image quality [3], of which the light adaptation, contrast sensitivity function, and masking have been considered in the Just-Noticeable Difference (JND) model [5]–[8]. However, the JND only models the HVS error tolerance property, which masks the distortions below the threshold. As to the distortions above the threshold, it cannot efficiently model the HVS sensitivity. Channel decomposition [9] is also employed to assess the image quality by modeling the HVS sensitivities over different frequencies. Recently, the Image Quality Assessment (IQA) methods attempt to characterize the features which HVS may associate with loss of quality, such as blurring, blocking, and so on. The IQAs that embody this approach include SSIM [2], [10], [12] and VIF [13]. SSIM is derived by capturing the information loss of image structures, while VIF employs the mutual information between the original and test image to evaluate the image quality. In [14], it has been demonstrated that SSIM and VIF have similar performances. And SSIM treats different oriented distortions and different located distortions equally. However, as the HVS perceives images with local varying saliencies [20], the pooling HVS feature [4] needs to be considered to evaluate the image quality. Also, the HVS HE property [15]–[18] of natural scenes has been researched on for modeling the visual sensitivities of different distortions over image contents with different orientations, in comparison with the HVS oblique effect property for the simple patterns, such as isolated gratings [16]. Therefore, the HVS HE property needs to be taken into account when evaluating the image quality.

In this paper, the HVS properties over structural distortions are considered to improve the IQA performance. Firstly, SSIM is employed to obtain the structural distortion map. Secondly, the distortion map is refined by the HVS orientation sensitivity modeled by the HE. Finally, the image quality index is obtained by a saliency pooling strategy over the distortion map.

Manuscript received March 05, 2010; revised April 14, 2010; accepted April 14, 2010. Date of publication April 22, 2010; date of current version May 21, 2010. This work was supported in part by a grant from the Chinese University of Hong Kong under the Focused Investment Scheme (Project 1903003). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. H. Vicky Zhao.

The authors are with the Department of Electronic Engineering, the Chinese University of Hong Kong, Hong Kong (e-mail: lma@ee.cuhk.edu.hk; snli@ee.cuhk.edu.hk; knngan@ee.cuhk.edu.hk).

Digital Object Identifier 10.1109/LSP.2010.2048726

## II. Proposed Image Quality Assessment Framework

To model the HVS properties over the image structural distortions for depicting the image quality, SSIM is performed on the test image to obtain the structural distortion map SD:

$$\mathrm{SD}(i,j) = (l(I(i,j),T(i,j)))^{\alpha} \cdot (c(I(i,j),T(i,j)))^{\beta} \cdot (s(I(i,j),T(i,j)))^{\gamma} \quad (1)$$

where $I$ is the reference image, $T$ is the test image, $(i,j)$ is the pixel location, $l(I(i,j),T(i,j))$ indicates the luminance similarity, $c(I(i,j),T(i,j))$ denotes the contrast similarity, and $s(I(i,j),\mathrm{T}(i,j))$ is the structure similarity, $\alpha, \beta$, and $\gamma$ are three parameters used to balance the relative importance of the three components, which are set as 1 to simplify the expression. The definitions can be referred to [2], [10], and [12].

### A. Visual Horizontal Effect Modeling

According to Hansen *et al.*'s researches on human vision [15]–[18], the oblique content is perceived to be the best, whereas the horizontal content is the worst for natural images;

also the oblique stimuli are perceived to be the best in naturalistic broadband stimuli. The phenomenon is known as HE. Hence, we want to model the visual HE sensitivity, which the HVS may associate with the image quality.

In order to model the visual HE sensitivity, firstly we need to obtain the orientation and energy information for both the content and stimulus. In our approach, the reference image is regarded as the original content, while the difference between the test and reference images is denoted as the stimulus superposed onto the content. As we all know, the kernels of Gabor filters are similar to the 2-D receptive field profiles of the mammal cortical simple cells, which exhibit desirable characteristics of spatial locality and orientation selectivity [19]. Therefore, different oriented Gabor filters are employed to filter the original content and stimulus to generate different oriented responses. According to the maximum response, the orientation and energy are determined for depicting the local features of the visual inputs. However, in some local smooth regions, all the filtering responses may appear very small, which are regarded as isotropic for its weak influences over all the orientations.

Four representative images and their distorted versions from the LIVE database [21] are selected to train the visual HE sensitivities over the image structural distortions. The orientation information of the images and their distorted versions are shown in Fig. 1. It can be observed that certain oriented contents dominate each selected image, such as: the isotropic contents dominate PARROTS; 135 degrees and horizontal contents dominate BIKES and so on. Moreover, we can see that different oriented stimuli are superposed onto different oriented contents with different probabilities. Therefore, we can employ the four representative images (with different dominant oriented contents) and their distorted versions (with different oriented stimuli) to train the visual HE sensitivities. During the training process, the following three aspects should be considered.

a) The content orientation is isotropic. The stimuli presented in these regions are easy for HVS perception, which is modeled by contrast masking in JND models [5]–[8]. Therefore, HVS is highly sensitive.
b) The stimulus orientation is the same as the content orientation. It can be viewed as a signal enhancement rather than distortion. Therefore, the lower HVS sensitivity is expected and the distortion is difficult to detect.
c) The stimulus orientation is perpendicular to the content orientation. The HVS is extremely sensitive and the distortion is very easy to perceive.

The initial HE sensitivity values in [16] are firstly slightly modified (increased or decreased) by referring to the aforementioned three aspects. Based on the structural distortion map SD, if the HE refined SSIM values correlate better with the subjective Differential Mean Opinion Score (DMOS) values, which are provided by the database, the HE sensitivity values are tuned by following the same direction. Otherwise, the HE sensitivity values are tuned by following the opposite direction. After several iterations, the optimized HE sensitivity values are obtained. The optimized HE sensitivity values of four dominant oriented stimuli over five prevailing biased contents are indicated by the spots in Fig. 2. Based on these sensitivity values, the cubic polynomial functions are fitted to model the sensitivities of oriented stimuli over the same content, illustrated by the curves in Fig. 2.

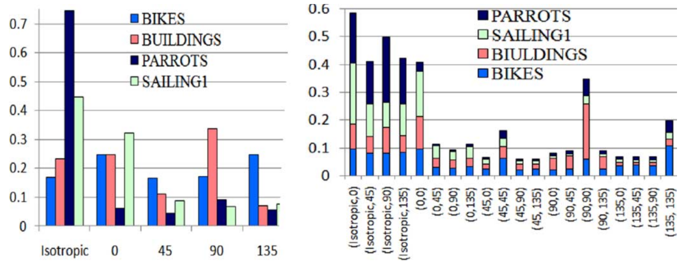


Fig. 1. Orientation information of the reference images and their distorted versions. Left: content orientation distribution (x-axis: content orientation; y-axis: pixel number probability). Right: content and stimulus orientation joint distribution (x-axis: (content orientation, stimulus orientation) pair; y-axis: pixel number probability).

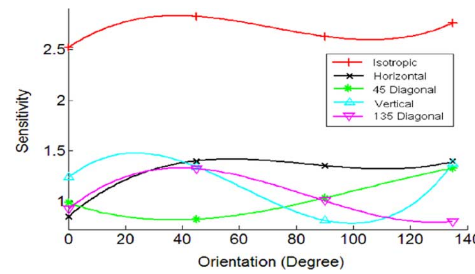


Fig. 2. HE sensitivity values of different orientated stimuli over different content bias (each color represents a biased content, and the horizontal axis indicates the stimulus orientation).

For the isotropic biased content, the visual sensitivity values are much larger than the other biased contents, which match the HVS contrast masking properties. For the isotropic and horizontal biased contents, the visual sensitivity values of oblique orientations (45 and 135 degrees) are higher than that of the vertical direction, while the horizontal sensitivity value appears the smallest, which matches the experimental results of HE. As for the 45- and 135-degree biased contents, sensitivity values of the orientations around its perpendicular direction are the largest, whereas the sensitivity value of the same orientation appears to be the smallest, which matches the aforementioned aspects. For the vertical biased content, the largest sensitivity value appears around 45 degrees according to the HVS HE property and around zero degrees for the perpendicular property. Therefore, the HVS appears to be the most sensitive between zero and 45 degrees by considering the HVS properties together. The cubic polynomial function for depicting the visual HE sensitivities of orientated stimuli over different oriented contents is expressed as:

$$S_{\mathrm{HE}} = \varphi(\theta_I, \theta_S) = a_{\theta_I}\theta_S^3 + b_{\theta_I}\theta_S^2 + c_{\theta_I}\theta_S + d_{\theta_I} \quad (2)$$

where $\varphi$ is the HE sensitivity function illustrated in Fig. 2, $\theta_I$ and $\theta_S$ denote the orientation information of the content and stimulus, respectively, which are determined by the maximum responses of the oriented Gabor filters. $a_{\theta I}, b_{\theta I}, c_{\theta I}$, and $d_{\theta I}$ are the parameters which relate to the content orientation $\theta_I$. Furthermore, the higher the stimulus energy, the worse is the visual quality of the test image. Therefore, a relationship between stimulus energy and image perceptual quality should be considered.

A stimulus energy adaptation factor $\alpha_{\mathrm{SE}}$ is used to refine the structural distortion value, which is defined as

$$\alpha_{\mathrm{SE}}(i,j) = a \cdot \mathrm{erf}(b \cdot E_S(i,j) + c) + d \tag{3}$$

where $E_S$ is the stimulus energy obtained from the Gabor filtering results, erf is the error function, $a = -0.175, b = 0.35, c = -2.5$, and $d = 0.825$ are set empirically for adjusting the stimulus energy adaptation factor. Then the refined structural distortion map $\mathrm{SM}_r$ is obtained by:

$$\mathrm{SM}_r(i,j) = \mathrm{SD}(i,j) \cdot \alpha_{\mathrm{SE}}(i,j) / S_{\mathrm{HE}}(\theta_I(i,j), \theta_S(i,j)). \tag{4}$$

Moreover, as we have mentioned before, when all the responses of Gabor filtering appear very small, the regions should be regarded as isotropic. It means that the signal has no inclined orientations. For the stimulus, it means that the distortion obtained is spread over all the orientations. As the stimulus energy is very small, it will have little effect on the image perceptual quality, which can be modeled by JND [5]–[8]. In this case, the HE sensitivity and stimulus energy adaptation factor should not be taken into consideration. Therefore, a signal-dependent JND model for the stimulus should be considered by neglecting the influence of the invisible distortion, the magnitude of which is smaller than a threshold *Thr*. However, according to our experiments, the performance will not be obviously affected as the threshold varies. Thereby, *Thr* is simply set as 2.2.

### *B. Saliency Pooling Strategy*

As HVS processes local regions of images with different visual acuities, artifacts that are present in the attended regions are better perceived than those present in the non-attended areas, which means that the observer's assessment of image quality is prejudiced by the perceived structural distortions in salient regions. Therefore, a relative measure of the importance of different regions, indicated by a saliency map, plays an important role in evaluating the image quality. In this paper, we employ the spectral residual model [20] to detect the saliency.

Given an image $I$, Fourier Transform (FT) $\zeta$ is firstly applied to obtain the amplitude spectrum $A(f)$ and phase spectrum $P(f)$. The spectral residual $R(f)$ can be generated based on the log-spectrum representation $L(f) = \log(A(f))$ of an image according to

$$R(f) = L(f) - L_a(f) \tag{5}$$

where $L_a(f)$ denotes the averaged spectrum, which is derived by convolving the log-spectrum $L(f)$ with an averaging filter. And it is claimed that the spectral residual contains some important information of an image related to the HVS perception [20]. The primary nontrivial part of the scene is constructed by inverse FT $\zeta^{-1}$, which could be interpreted as the unexpected portion of the image. The unexpected portion represents the saliency map $Sa_M$ in spatial domain, which indicates the different visual importances of different locations:

$$Sa_M = |\zeta^{-1}(\exp(R(f) + jP(f)))|^2. \tag{6}$$

Based on $Sa_M$, a saliency pooling strategy is proposed to generate the final similarity index for evaluating the image perceptual quality:

$$\mathrm{Index} = \sum Sa_M \cdot SM_r / \sum Sa_M. \tag{7}$$

## III. Experimental Results

In this section, we compare the performance of our proposed IQA method with other methods, i.e., PSNR, SSIM, and VIF. The IQA methods are evaluated on the LIVE [21] and A57 databases [24], which comprise the most prevailing distortions. The distorted images, excluding the ones generated from the four training images, are used for evaluating the IQA performances.

We follow the performance evaluation procedure employed in the Video Quality Experts Group (VQEG) HDTV test [22] and that in [11]. Firstly, a logarithmic function is employed to fit the objective and subjective scores through a nonlinear mapping process. Subsequently, the Correlation Coefficients (CC) between the subjective and the non-linearly mapped objective scores, which provides an evaluation of the prediction accuracy, and the Spearman Rank-Order Correlation Coefficients (SROCC), which measures the prediction monotonicity, are employed for evaluating the different IQA performances. Furthermore, the Mean Absolute prediction Error (MAE), and the Root Mean Square prediction Error (RMSE) of the fitting procedure are also utilized to measure the IQAs' efficiencies. On one hand, larger CC and SROCC values mean that the objective and subjective scores correlate better, that is to say, a better performance of the IQA. On the other hand, smaller RMSE and MAE values indicate smaller errors between the two scores, hence a better performance.

We compare the performance of the proposed IQA scheme with PSNR, S-SSIM (Single-scale SSIM) [2], LDW-SSIM (Local Distortion-Weighted SSIM) [10], ICW_SSIM (Information Content-Weighted SSIM) [12], SMW-SSIM (SMooth region-Weighted SSIM) [23], and VIF [13]. The results are listed in Table I. The performance of our proposed scheme outperforms the other IQAs on the provided two databases with larger SROCC and LCC, and smaller RMSE and MAE, which means that our method demonstrates better performance across a wide range of image distortions. The reason is that the SSIM methods just employ different weights for different locations of the image. However, they do not account for the orientation sensitivity and saliency property of HVS. VIF [13] employs the steerable pyramid to decompose the test image, which extracts the image features at different scales and different orientations. In this way, HVS orientation and saliency properties are included. That is why it can outperform the other IQAs. However, our method can more accurately model the HVS orientation and saliency sensitivities over structural distortions, which outperforms VIF. The scatter-plots of different IQAs are shown in Fig. 3. It can be observed that the results of our proposed method scatter more closely around the fitted line than other IQAs, which indicates a better performance. Furthermore, it can be observed that VIF performs well on LIVE, but poorly on A57 database. The reason may be that the distortion model embodied in VIF [13] cannot efficiently simulate the two new

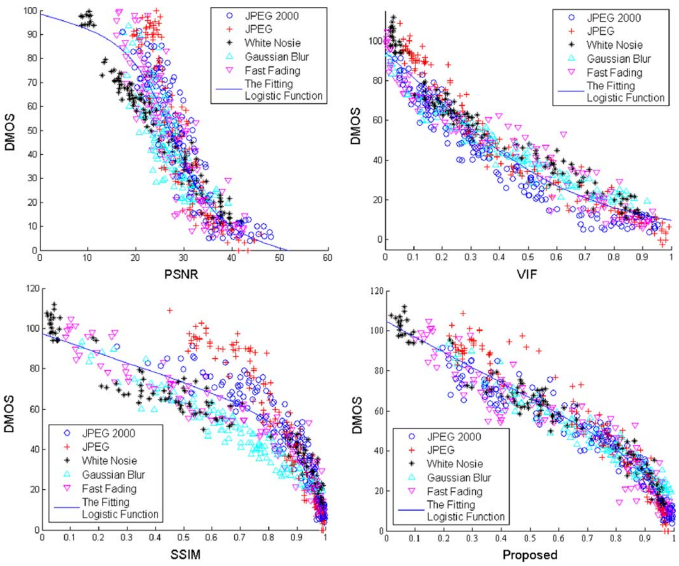


Fig. 3. Scatter plots of the DMOS values versus model predictions on the LIVE database. Each sample point represents one test image. Top left: PSNR; top right: VIF; bottom left: S-SSIM; bottom right: the proposed method.

TABLE I
PERFORMANCES OF DIFFERENT IQAS ON THE LIVE AND A57 DATABASE

| | | PSNR | S-SSIM | LDW-SSIM | ICW-SSIM | SMW-SSIM | VIF | Proposed |
|---|---|---|---|---|---|---|---|---|
| LIVE Database | CC | 0.891 | 0.914 | 0.915 | 0.936 | 0.947 | 0.961 | **0.966** |
| | SROCC | 0.897 | 0.922 | 0.919 | 0.942 | 0.953 | 0.966 | **0.971** |
| | RMSE | 12.425 | 11.060 | 11.051 | 9.641 | 8.769 | 7.523 | **7.057** |
| | MAE | 9.765 | 8.727 | 8.655 | 7.478 | 6.921 | 6.043 | **5.374** |
| A57 Database | CC | 0.644 | 0.415 | 0.545 | 0.518 | 0.607 | 0.614 | **0.848** |
| | SROCC | 0.570 | 0.407 | 0.495 | 0.455 | 0.557 | 0.622 | **0.857** |
| | RMSE | 0.192 | 0.224 | 0.206 | 0.210 | 0.195 | 0.194 | **0.130** |
| | MAE | 0.151 | 0.186 | 0.160 | 0.166 | 0.145 | 0.141 | **0.102** |

TABLE II
PERFORMANCE OF EACH PHASE OF THE PROPOSED SCHEME ON LIVE DATABASE

| | CC | SROCC | RMSE | MAE |
|---|---|---|---|---|
| HE Sensitivity | 0.9331 | 0.9405 | 9.820 | 7.593 |
| Saliency Pooling | 0.9443 | 0.9495 | 8.990 | 7.015 |

distortion types in A57, which are a) quantization of the LH subbands of the image with equal distortion contrast at each scale; b) JPEG 2000 with dynamic contrast-based quantization compression [24]. However, as the proposed method models the HE and saliency properties of the HVS, it can efficiently capture the distortions in the image which are sensitive to the HVS, no matter what the distortion type is. That is why the proposed metric performs well on both of the two databases.

Moreover, we demonstrate the efficiency of each phase (i.e., HE sensitivity and saliency pooling) of our proposed scheme individually by evaluating its performance on the LIVE database. The results are illustrated in Table II. Both the strategies improve the IQA performance. However, the saliency pooling strategy performs better than the HE sensitivity. Intuitively, the results are in accordance with the human perception of a visual input. While perceiving an image, we mainly focus on its interesting or salient portion. If the part appears really interesting and attractive, we will examine it more carefully. Therefore, the saliency pooling is important for image quality assessment. However, the visual HE sensitivity appears to play a lesser but nevertheless an important role in image quality assessment.

## IV. CONCLUSION

In this letter, an image quality assessment method is proposed by considering the visual HE sensitivity and saliency properties. The SSIM structural distortion map is refined by the visual HE model. The image quality index is generated by saliently pooling on the refined structural distortion map. Experimental results demonstrate that the proposed scheme outperforms the other IQAs, which means that the proposed metric correlates well with the human perception of image quality.

## REFERENCES

[1] B. Girod, "What's wrong with mean-squared error," in *Digital Images and Human Vision*. Cambridge, MA: MIT Press, 1993, pp. 207–220.

[2] 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.

[3] T. Pappas *et al.*, "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Processing*. New York: Academic, 2000.

[4] Z. Wang *et al.*, "Objective video quality assessment," in *Handbook of Video Databases: Design and Applications*. Boca Raton, FL: CRC, 2003.

[5] W. Lin *et al.*, "Visual distortion gauge based on discrimination of noticeable contrast changes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 7, pp. 900–909, Jul. 2005.

[6] W. Lin *et al.*, "Discriminative analysis of pixel difference towards picture quality prediction," *Proc. ICIP*, 2003.

[7] A. B. Watson, "DCTune: A technique for visual optimization of DCT quantization matrices for individual images," *Soc. Inform. Display Dig. Tech. Papers XXIV*, pp. 946–949, 1993.

[8] L. Ma and K. N. Ngan, "Adaptive block-size transform based just-noticeable difference profile for images," in *Pacific-Rim Conf. Multimedia*, 2009.

[9] A. P. Bradley, "A wavelet visible difference predictor," *IEEE Trans. Image Process.*, vol. 5, pp. 717–730, May 1999.

[10] Z. Wang *et al.*, "Stimulus synthesis for efficient evaluation and refinement of perceptual image quality metrics," *Proc. SPIE, Human Vision and Electronic Imaging*, vol. 5292, Jan. 2004.

[11] H. R. Sheikh *et al.*, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3441–3452, Nov. 2006.

[12] Z. Wang *et al.*, "Spatial pooling strategies for perceptual image quality assessment," *Proc. ICIP*, pp. 2945–2948, 2006.

[13] 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.

[14] K. Seshadrinathan *et al.*, "Unifying analysis of full reference image quality assessment," in *Proc. ICIP.*, 2008.

[15] E. A. Essock *et al.*, "Oblique stimuli are seen best (not worst!) in naturalistic broad-band stimuli: A horizontal effect," *Vision Research*, vol. 43, no. 12, pp. 1329–1335, Jun. 2003.

[16] B. C. Hansen *et al.*, "A horizontal bias in human visual processing of orientation and its correspondence to the structural components of natural scenes," *J. Vis.*, vol. 4, no. 12, pp. 1044–1060, 2004.

[17] B. C. Hansen and E. A. Essock, "Influence of scale and orientation on the visual perception of natural scenes," *Vis. Cogn.*, vol. 12, no. 6, pp. 1199–1234, 2005.

[18] B. C. Hansen and E. A. Essock, "Anisotropic local contrast normalization: The role of stimulus orientation and spatial frequency bandwidths in the oblique and horizontal effect perceptual anisotropies," *Vis. Res.*, vol. 46, no. 26, pp. 4398–4415, 2006.

[19] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vis. Res.*, vol. 20, no. 10, pp. 847–856, 1980.

[20] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," *Proc. CVPR*, 2007.

[21] H. R. Sheikh *et al.*, LIVE Image Quality Assessment Database [Online]. Available: http://live.ece.utexas.edu/research/quality [Online]. Available

[22] VQEG Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment 2000 [Online]. Available: http://www.vqeg.org

[23] S. Li *et al.*, "Influence of the smooth region on the structural similarity index," in *Pacific-Rim Conf. Multimedia*, 2009.

[24] A57 Database [Online]. Available: http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html