Journal of Imaging Science and Technology® 54(3): 030503–030503-14, 2010.

Full Reference Image Quality Assessment Based on Saliency Map Analysis

Yubing Tong and Hubert Konik
Laboratoire Hubert Crurien UMR 5516, Université Jean Monnet-Saint-Etienne, Université de Lyon, 42000 Saint-Etienne, France
E-mail: yubing.tong@univ-st-etienne.fr

Faouzi A. Cheikh
Computer Science and Media Technology, Gjϕvik University College, P.O. Box 191, N-2802 Gjϕvik, Norway

Alain Tremeau^
Laboratoire Hubert Crurien UMR 5516, Université Jean Monnet-Saint-Etienne, Université de Lyon, 42000 Saint-Etienne, France

Abstract. *Region saliency has not been fully considered in most previous image quality assessment models. In this article, the contribution of any region to the global quality measure of an image is weighted with variable weights computed as a function of its saliency. In salient regions, the differences between distorted and original images are emphasized as if the authors are observing the difference image with a magnifying glass. Here a mixed saliency map model based on Itti's model and face detection is proposed. Both low-level features including intensity, color, orientation, and high-level features such as face are used in the mixed model. Differences in salient regions are then given more importance and thus contribute more to the image quality score. The experiments done on the 1700 distorted images of the TID2008 database show that the performance of the image quality assessment on full subsets is enhanced. © 2010 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2010.54.3.030503]

INTRODUCTION

Subjective image quality assessment procedure is a costly process which requires a large number of observers and takes lots of time. Therefore, it cannot be used in automatic evaluation programs or in real time applications. Hence there is a trend to assess image quality with objective methods.[1] Usually image quality assessment models are set up to approximate the subjective score on image quality. Some referenced models had been proposed such as Video Quality Experts Group (VQEG).[2] Some methods have gotten better results than peak signal-to-noise ratio (PSNR) and mean squared error (MSE), including Univeral Quality Index (UQI), Structural Similarity Index (SSIM), LINLAB, peak signal-to-noise ratio based on human visual system (PSNRHVS), modified metric based on PSNRHVSM, noise quality measure (NQM), weighted signal-to-noise ratio (WSNR), visual signal-to-noise ratio (VSNR), etc.[3–16] But it has been demonstrated that with respect to the wide range of possible distortion types no existing performance metric will be good enough. PSNRHVS and PSNRHVSM are two new methods with high performance on noise, noise2, safe, simple, and hard subsets of TID2008, which makes them appropriate for evaluating the efficiency of image filtering and lossy image compression.[1] But PSNRHVS and PSNRHVSM show very low performance on Exotic and Exotic2 subset of TID2008 database. With PSNRHVS and PSNRHVSM, images are divided into fixed size blocks. Moreover, every block is processed independently in the same way with the same weights.

Such a way of comparing images is contradictory with the way our human visual system (HVS) operates. Dividing an image into blocks of equal size irrespective of its content is definitely counterproductive since it breaks large objects and structures of the image into semantically nonmeaningful small fragments. Additionally it introduces strong discontinuities that were not present in the original image. Furthermore, it is proven that our HVS is selective in its handling/processing of the visual stimuli. Thanks to this selectivity of our visual attention mechanism, human observers usually focus more on some regions than another irrespective of their size. Therefore, it is intuitive to think that an approach that treats the image regions in the same way, disregarding the variation of their contents will never be able to faithfully estimate the perceived quality of the visual media. Therefore, we propose to use the saliency information to mimic the selectivity of the HVS and integrate it into existing objective image quality metrics to give more importance to the contribution of salient regions over those of nonsalient regions.

An image saliency map could be used to provide weights on the results of SSIM, VIF, etc.,[17] but the saliency map used in this study was, in fact, the image reconstructed from the phase spectrum and inverse Fourier transform which could reflect the presence of contours. This may not

^IS&T Member.

Received Sep. 16, 2009; accepted for publication Mar. 3, 2010; published online Apr. 27, 2010.
1062-3701/2010/54(3)/030503/14/$20.00.

be enough since the contour of an image is far from containing all information in the image. The detection order of region saliency was used to weight the difference between reference and distorted images.[18] For every image, there are 20 time steps to find the saliency region. If a salient region is found first, it is assigned the largest weight and vice versa. For pixels in the detected salient region, equal weighting and simple linear weighting were used. In this article, we propose to consider additional information computed from the image contents that affects region saliency. We will consider not only the saliency value of every pixel but also the saliency degree of the current pixel relative to its neighboring field and to the global image. Furthermore, the contribution of nonsalient regions to image quality score will be reduced by assigning lower weights to them.

Face plays an important role in recognition and can focus much of our attention.[19] Face should thus be used as a high-level feature for the saliency map analysis in addition to low-level features such as those used in Itti's model[20] based on color, intensity, and orientations. In this article, we propose a mixed saliency map model based on Itti's model and a face detection model.

ANALYSIS OF PREVIOUS WORK AND PRIMARY CONCLUSIONS

PSNR and MSE are two common methods used to assess the quality of the distorted image defined by

$$\mathrm{PSNR} = \log_{10}\left(\frac{255^2}{\mathrm{MSE}}\right), \tag{1}$$

where

$$\mathrm{MSE} = \frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N}[\delta(i,j)]^2 \tag{2}$$

and

$$\delta(i,j) = [a(i,j) - \hat{a}(i,j)], \tag{3}$$

where (i,j) is the current pixel position, $a(i,j)$ and $\hat{a}(i,j)$ are the original image and the distorted image respectively, and M and N are the height and width of the image. Neither image content information nor HVS characteristics are taken into account by PSNR and MSE when they are used to assess image quality. Consequently PSNR and MSE cannot achieve good results when compared to subjective quality scores, especially for images such as those in noise, noise2, Exotic, and Exotic2 subsets which include images corrupted with additive Gaussian noise, high frequency noise, impulse noise, Gaussian blur, etc. Since PSNR is only dependent on the absolute difference between the original image and the distorted image, there is no additional factor, such as saliency information, that might affect our visual perception. Some distorted images with the same PSNR look much different in image quality.[6] On the TID2008 database, PSNR gives the worst results according to Spearman's correlation and Kendall's correlation.[1]

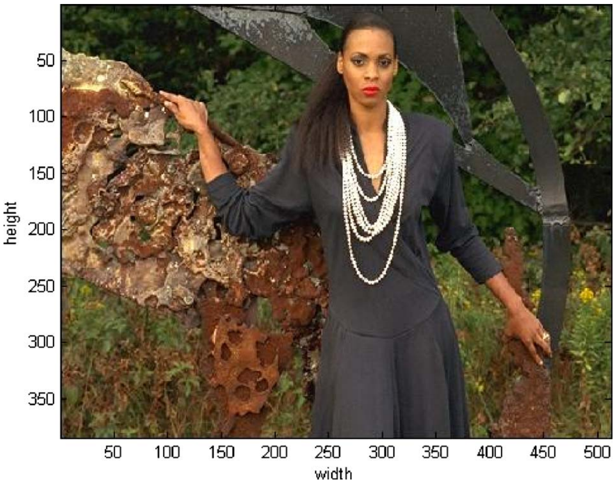


Figure 1. Reference image I18.

PSNRHVS and PSNRHVSM are two models which had been designed to improve the performance of PSNR and MSE. The PSNRHVS divides the image into 8×8 pixel nonoverlapping blocks. Then the $\delta(i,j)$ difference between the original and the distorted blocks is weighted for every 8×8 block by the coefficients of the contrast sensitivity function (CSF). So Eq. (3) can be rewritten as follows:

$$\delta_{\mathrm{PSNRHVS}}(i,j) = \delta(i,j) \cdot \mathrm{CSF}_{\mathrm{cof}}(i,j), \tag{4}$$

where $\delta(i,j)$ is calculated using DCT coefficients.

PSNRHVSM is defined in similar way to PSNRHVS, but the difference between the discrete cosinus transform (DCT) coefficients is further multiplied by a contrast masking (CM) metric for every 8×8 block. The result is then weighted by the $\mathrm{CSF}_{\mathrm{Cof}}$ as follows:

$$\delta_{\mathrm{PSNRHVSM}}(i,j) = [\delta(i,j) \cdot \mathrm{CM}(i,j)] \cdot \mathrm{CSF}_{\mathrm{cof}}(i,j), \tag{5}$$

$$\mathrm{MSE}_{\mathrm{PSNRHVS}}(i,j,I,J) = \frac{1}{M \times N}\sum_{I=1}^{M/8}\sum_{J=1}^{N/8}\left\{\sum_{i=1}^{8}\sum_{j=1}^{8}[\delta_{\mathrm{PSNR\_HVS}}(i,j)]^2\right\}, \tag{6}$$

where (I,J) is the position of an 8×8 block in the image and (i,j) is the position of a pixel in the 8×8 block. $\mathrm{MSE}_{\mathrm{PSNRHVSM}}$ can be defined in the same way. Then PSNRHVS or PSNRHVSM can be computed by replacing the MSE in Eq. (1) with $\mathrm{MSE}_{\mathrm{PSNRHVS}}$ or $\mathrm{MSE}_{\mathrm{PSNRHVSM}}$.

Analysis

For PSNRHVS and PSNRHVSM, images are processed with nonoverlapping 8×8 blocks. Every 8×8 block is considered to contribute equally to the image quality metric. From the point of view of human visual perception, an 8×8 block size is not optimal considering the variability of image content. In fact, the size of the salient region is not fixed. Independent blocks with fixed size might result in blockiness or sudden change that greatly affects the subjective quality perception. As an illustration the following figures show that different parts of an image contribute differently to the per-

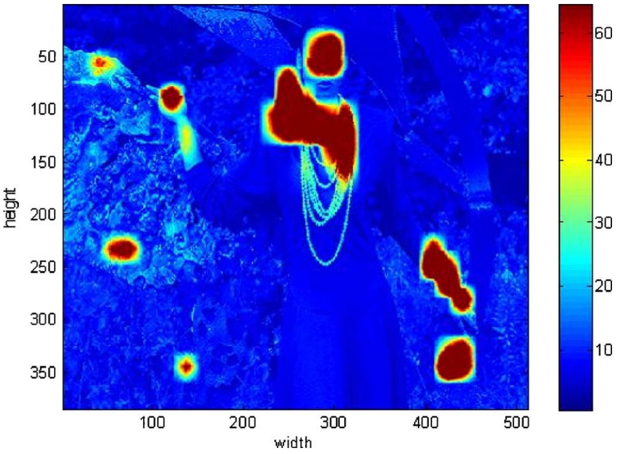


Figure 2. Saliency map of I18 with face detection.

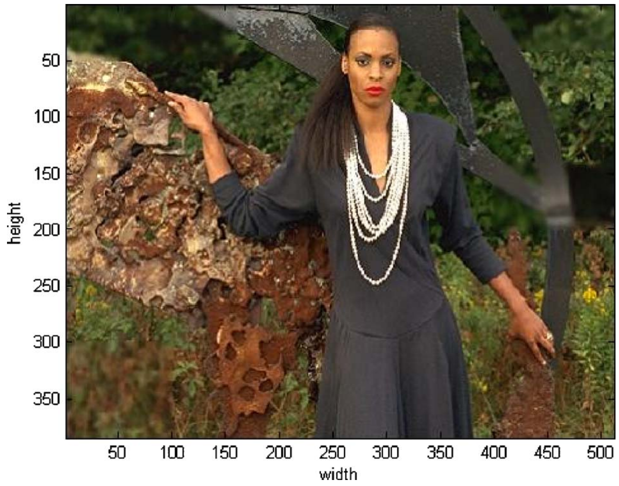


Figure 4. I18 with noise in four nonsalient regions.

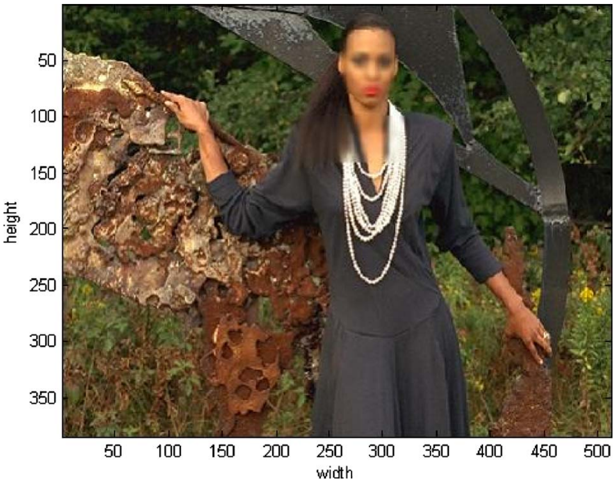


Figure 3. I18 with noise in one salient region.

ceived image quality and that degradation in salient regions may be more prominent and hence should contribute more to the final quality measure.

The image "I18" and its corresponding saliency map are illustrated in Figures 1 and 2, respectively. Figure 3 is a distorted image of I18 with noise on the saliency region including face, neck, and breast part. The objective image quality of this distorted image is equal to 46.3 dB with PSNR, 33.74 dB with PSNRHVS, and 36.3 dB with PSNRHVSM. Figure 4 is another distorted image of I18 with noise on the nonsaliency region. The objective image quality of this second distorted image is equal 41.6 dB with PSNR, 32.4 dB with PSNRHVS, and 35.8 dB with PSNRHVSM. Here a local smoothing filter was used to filter the corresponding parts in the saliency map with noise. The objective image quality metric values show that the quality of Fig. 3 is better than that of Fig. 4. However it is easy to see that the perceived quality of Fig. 4 is better than that of Fig. 3, as the filter operation was added on the nonsaliency region of Fig. 4. All the distorted parts in Fig. 4 are not perceptibly noticeable unless carefully observed pixel by pixel. In Figure 5, the nonsaliency regions with noise in Fig. 4 are marked out with blue circles.

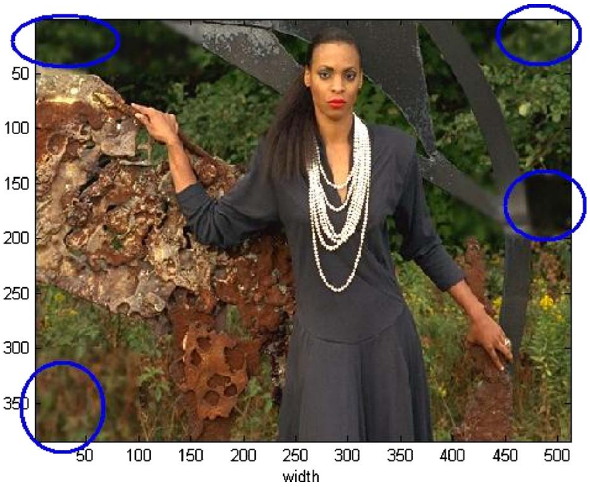
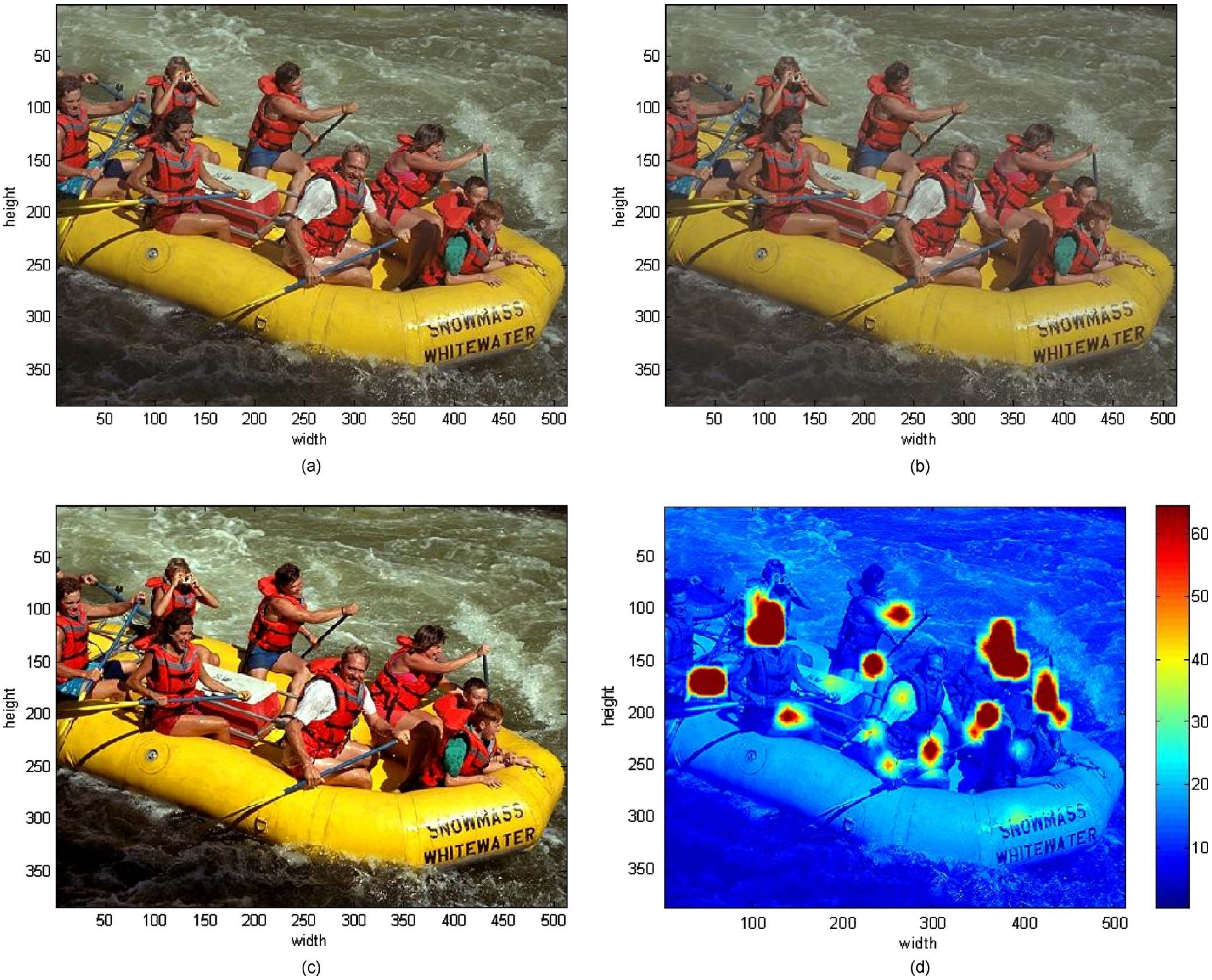


Figure 5. I18 with distortion in four nonsalient regions.

The above example might be considered as an artificially constructed case study. For this reason, we propose another image, the image "I14" of TID2008 [see Figure 6(a)], as another example where noise was added in equal quantity to different parts of the image. In Fig. 6, we have considered two distorted images "I14–17–2" and "I14–17–3" shown in Figs. 6(b) and 6(c). The saliency map of I14 is also illustrated in Fig. 6(d).

The subjective score of I14–17–2 is lower than that of I14–17–3, but PSNRHVS and PSNRHVSM are higher for I14–17–2 than that of I14–17–3; this result is consistent with data provided by TID2008. For I14–17–2, the value of PSNRHVS and PSNRHVSM are respectively 23.3 dB and 23.95 dB. For I14–17–3, the value of PSNRHVS and PSNRHVSM are respectively 19.3dB and 19.87 dB. In subjective experiments, the attention of observers is focused on saliency regions, such as face, hands, etc. [see Fig. 6(d)]. These parts can be considered as contributing more significantly to image quality. If the quality of these salient regions were acceptable, the final image quality should be considered as good. For each case study while PSNR scores were relatively close. the computed image quality scores were differ-



(a)

(b)

(c)

(d)

Figure 6. I14 and corresponding distorted image. (a) The reference I14. (b) The distorted image I14–17–2. (c) The distorted image I14–17–3. (d) The saliency map of I14.

ent. This result confirms our initial expectation, namely, that quantitatively equal distortions yield different image quality scores. Each part of an image contributes differently to the perceived image quality. Furthermore, distortions in salient regions image quality affect more profoundly than those in nonsalient regions.

IMAGE QUALITY ASSESSMENT BASED ON REGION SALIENCY

In this section, the saliency map of an image will be calculated using Itti's saliency map model or the following mixed saliency map model when faces are present in the image. First, a simple and fast face detection program in OPENCV based on Haar-like features was used to decide if the current image contains human faces.[21] Then according to that decision, Itti's model or the mixed model will be used to calculate the saliency map. The flowchart of the method that we propose is shown in Figure 7. The first step of the process is to compute the region saliency map of the input image; next the region saliency map is used to enhance the performance of the method used to assess the image quality (e.g., the PSNRHVS) of the original image.

Itti's Saliency Map Model

The saliency map model that we propose is based mainly on Itti's visual attention model. Considering that faces play an important role in our daily social interaction and thus easily focus our visual attention, we propose a mixed saliency map model based on Itti's visual attention model and face detection.

Itti's saliency map model is defined as a bottom-up visual attention mechanism, which is based on color, intensity, and orientation features. Each feature is analyzed using a Gaussian pyramid and multiscales. This model is based on seven feature maps including one intensity, four orientations (at 0°, 45°, 90°, and 135°) and two color opponencies (red/green and blue/yellow) maps. After a normalization step, all these feature maps are summed to three conspicuous maps including intensity conspicuous map C_i, color conspicuous map C_c, and orientation conspicuous map C_o. Finally the saliency maps are combined to get the saliency maps according to the following equation:

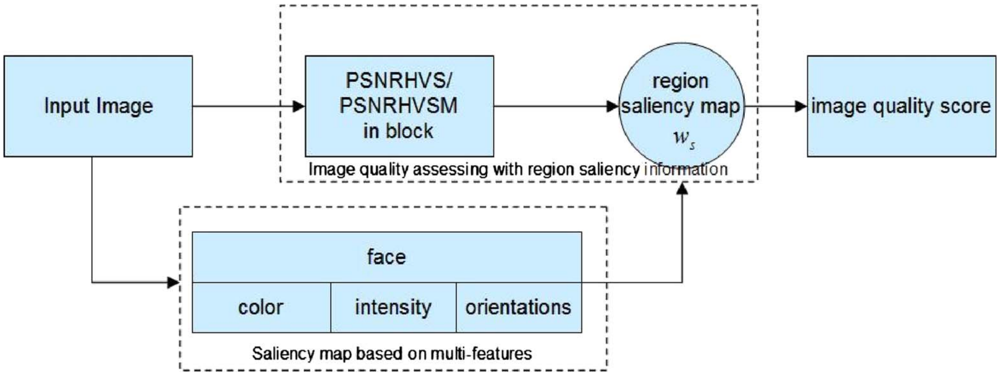
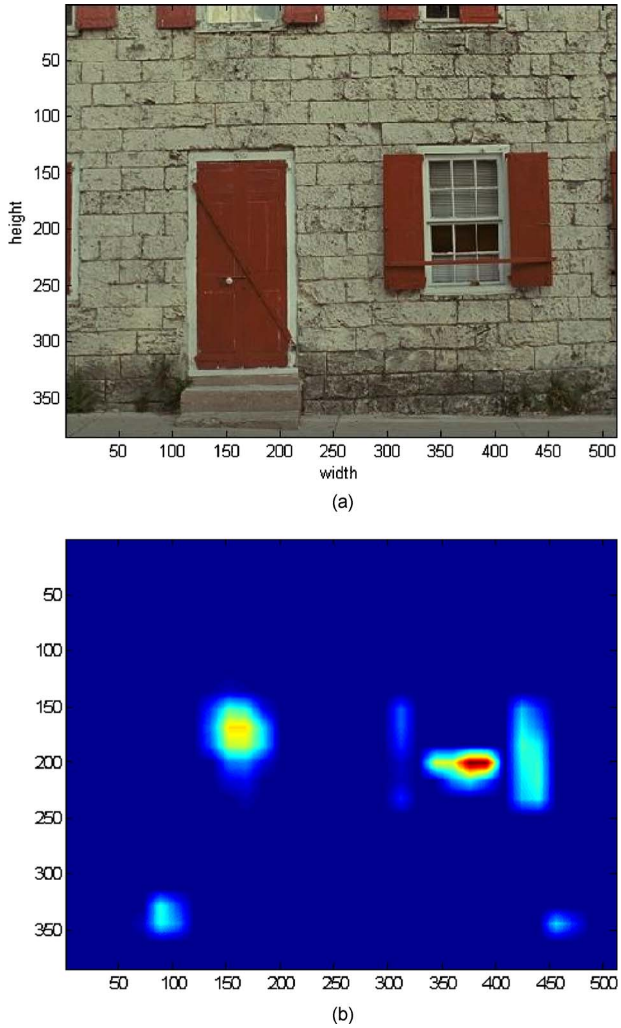


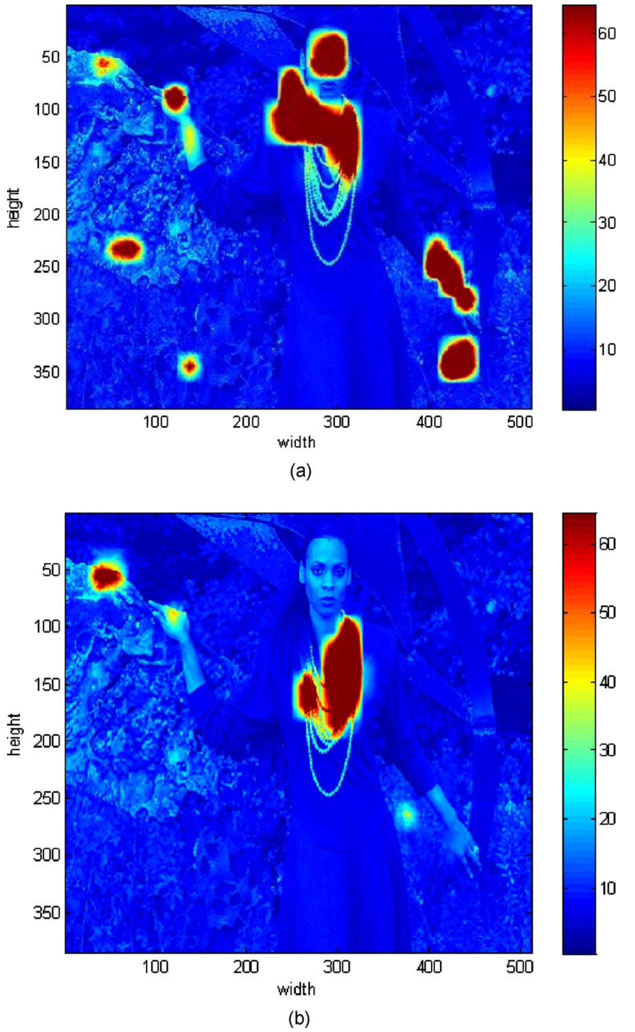
Figure 7. Flowchart of the method based on region saliency used to assess the image quality.



(a)

(b)

Figure 8. Image I01 with its saliency map and corresponding surface plot. (a) Reference image I01. (b) Saliency map of I01.



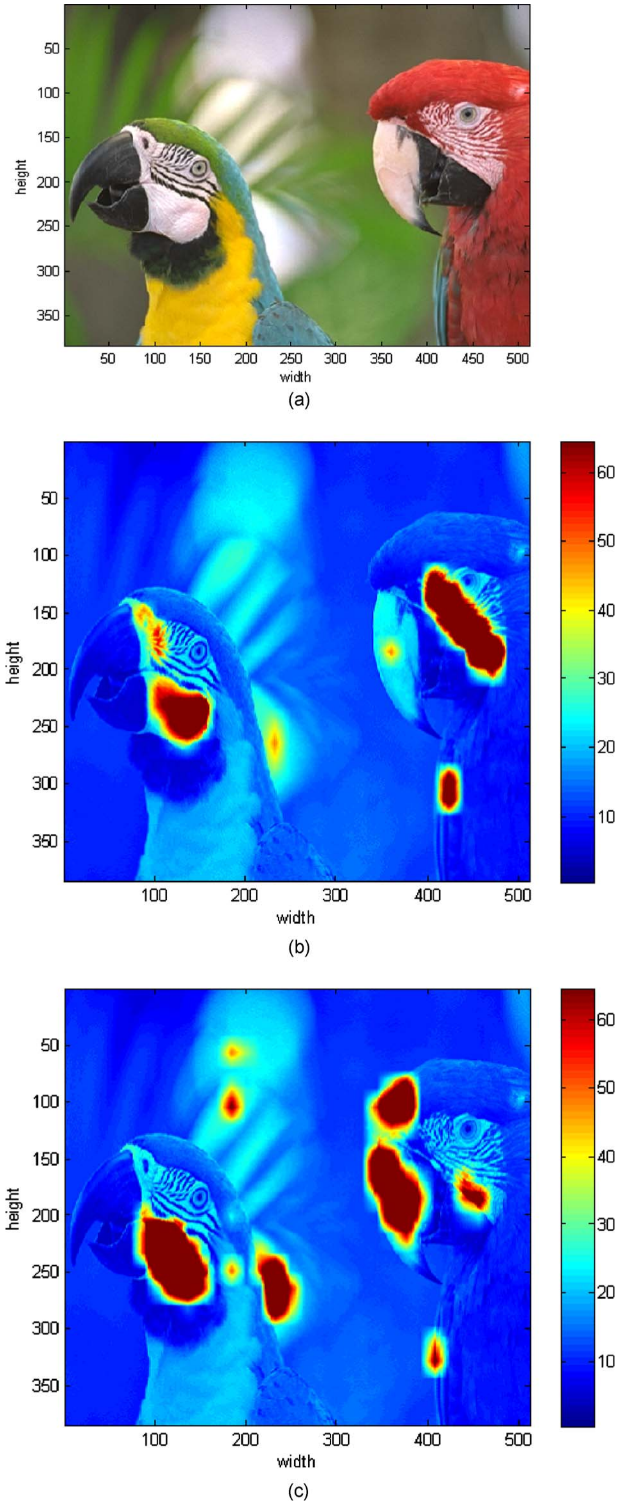
(a)

(b)

Figure 9. Saliency maps for mixed model and Itti's model on I18 reference image. (a) Saliency map from mixed model. (b) Saliency map from Itti's model.

$$S_{\text{Itti}} = \frac{1}{3}\sum_{k\in i,c,o} C_k. \qquad (7)$$

As an example, let us consider the image "I01" in TID2008 [see Figure 8(a)]; its saliency map [Fig. 8(b)] is computed using Itti's model. The more reddish a region of the saliency map, the more salient its corresponding image region. This concords with the selectivity of the HVS which focuses only on some parts of the image instead of the whole content.



(a)

(b)

(c)

Figure 10. Saliency maps from mixed model and Itti's model for I23 reference image. (a) I23 reference image. (b) Saliency map from mixed model. (c) Saliency map from Itti's model.

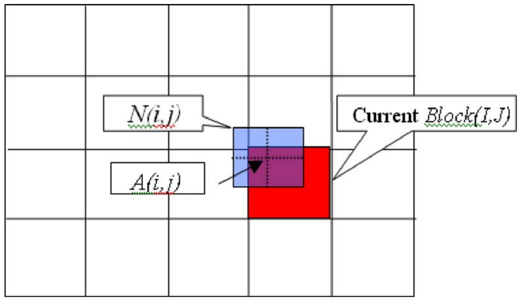
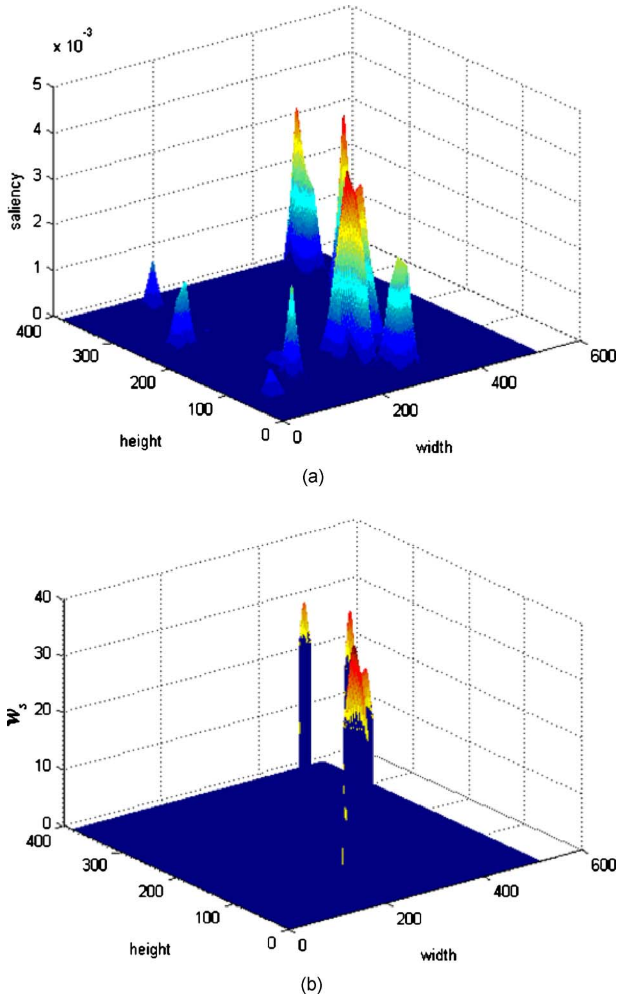


Figure 11. Current block, current pixel, and its neighboring field.



(a)

(b)

Figure 12. Surface plot of saliency map and weighted saliency map w_s. (a) Surface plot of saliency map. (b) Surface plot of weighted saliency map w_s.

Saliency Map Model Based on Face Detection

Faces are features which focus more attention than other features in many images. Psychological tests have proven that face, head or hands may be perceived prior to any other details.[20] So faces can be used as high level features for defining a saliency map. One drawback of Itti's visual attention mechanism model is that its saliency map model is not well adapted for images with faces. Several studies in face recognition have shown that skin hue features could be used to extract the face information. To detect heads and hands in images, we have used the face recognition and location algorithm of Walther et al.[22] This algorithm is based on a Gaussian model of the skin hue distribution in the (r',g') color space as an independent feature. For a given color pixel

Table I. Distortion subsets in TID2008.

No.	Distortion type	Noise	Noise2	Safe	Hard	Simple	Exotic	Exotic2	Full
1	Additive Gaussian noise	+	+	+	–	+	–	–	+
2	Different additive noise in color	–	+	–	–	–	–	–	+
3	Spatially correlated noise	+	+	+	+	–	–	–	+
4	Masked noise	–	+	–	+	–	–	–	+
5	High frequency noise	+	+	+	–	–	–	–	+
6	Impulse noise	+	+	+	–	–	–	–	+
7	Quantization noise	+	+	–	+	–	–	–	+
8	Gaussian blur	+	+	+	+	+	–	–	+
9	Image denoising	+	–	–	+	–	–	–	+
10	JPEG compression	–	–	+	–	+	–	–	+
11	JPEG2000 compression	–	–	+	–	+	–	–	+
12	JPEG transmission errors	–	–	–	+	–	–	+	+
13	JPEG2000 transmission errors	–	–	–	+	–	–	+	+
14	Non eccentricity pattern noise	–	–	–	+	–	+	+	+
15	Local blockwise distortions of different intensity	–	–	–	–	–	+	+	+
16	Mean shift (intensity shift)	–	–	–	–	–	+	+	+
17	Contrast change	–	–	–	–	–	+	+	+

(r',g'), the model's hue response is then defined by the following equation:

$$h(r',g') = \exp\left\{-\frac{1}{2}\left[\frac{(r'-\mu_r)^2}{\sigma_r^2} + \frac{(g'-\mu_g)^2}{\sigma_g^2} + \frac{\rho(r'-\mu_r)(g'-\mu_g)}{\sigma_r\sigma_g}\right]\right\}, \tag{8}$$

where

$$r' = \frac{r}{r+g+b} \quad \text{and} \quad g' = \frac{g}{r+g+b}, \tag{9}$$

(μ_r,μ_g) is the average of the skin hue distributions, σ_r^2 and σ_g^2 are the variances of the r' and g' components, and ρ is the correlation between the components r' and g'. These parameters have been statistically estimated from 1153 photographs which contained faces. The function $h(r',g')$ can be considered as a color variability function around a given hue. Next a Gaussian pyramid (GP) based on a multiscale subsampling operation and a Gaussian smoothing was computed from $h(r',g')$. Then the center-surround (CS) map was calculated from the pyramid, in the same way as in Itti's model. Lastly, the results were normalized (Norm) to obtain the saliency map S_{face} defined as follows:

$$S_{\text{face}} = \text{Norm}(\text{CS}\{\text{GP}(h(r', \text{‘}g'))\}). \tag{10}$$

Mixed Saliency Map Model Based on Face Detection

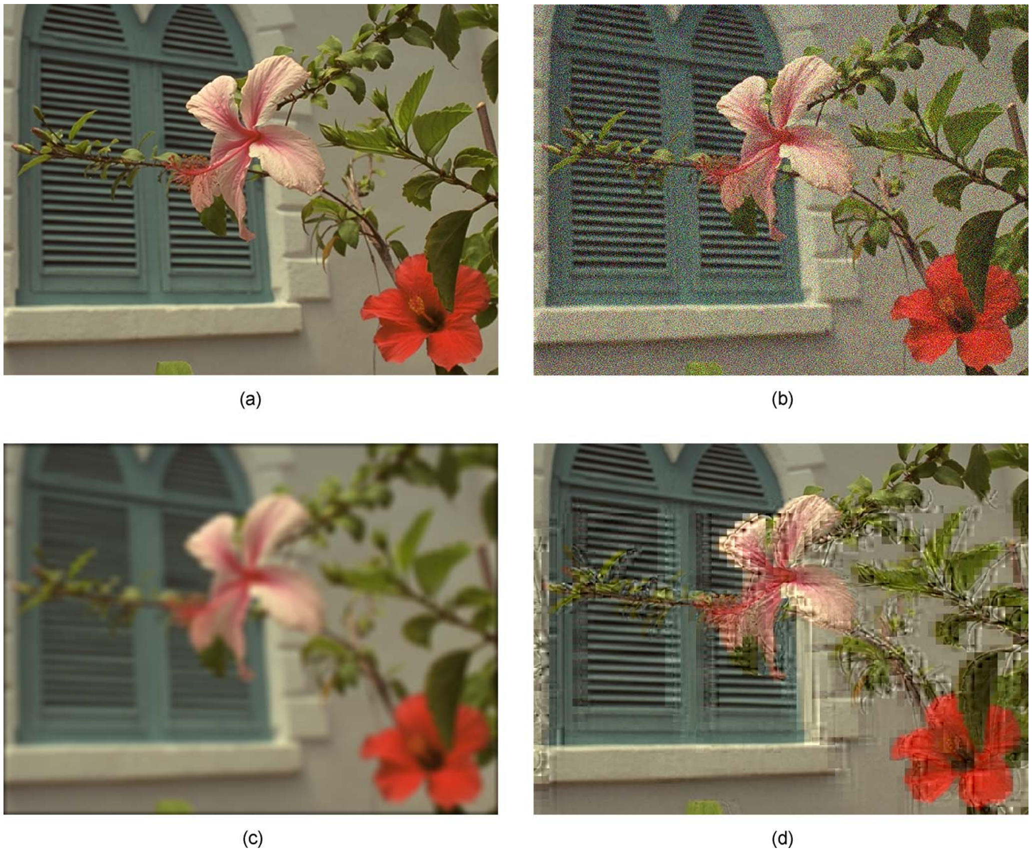
The mixed saliency analysis model that we propose is a linear combination model which combines both Itti's model and the Gaussian face detection model as follows:

$$S_{\text{MIX}} = \alpha S_{\text{Itti}} + (1-\alpha)S_{\text{Face}}, \tag{11}$$

where α is a constant. The best results that we obtained in our study have been achieved for $\alpha=3/7$.

For most images containing faces, heads, or hands, the mixed model with skin hue detection gives better results than Itti's model, i.e., more accurate saliency maps. The two examples given in this article show the difference between Itti's model and the mixed model for face images. The first example corresponds to the reference image “I18” in TID2008 which contains a face with eyes and hands. Figure 9(a) shows the saliency map computed from the mixed model. Figure 9(b) shows the saliency map computed from Itti's model. Fig. 1 shows that the most salient regions which attract the attention of observers are the face and the hands. Relative to the visual saliency map (i.e., Fig. 1) the mixed model appears more precise than Itti's model.

Another interesting example is the reference image “I23” which is a nonhuman face image as shown in Figure 10. The original reference image is shown in Fig. 10(a). The most salient regions which focus the attention are the heads of the parrots and in particular their eyes and their faces. Considering the hue of the faces of the parrots and in particular the hue of the region around the eyes, we computed the corresponding color variability function $h(r',g')$ and



(a) (b) (c) (d)

Figure 13. Examples of distortions in different subsets. (a) Original image. (b) Distortion 5: High frequency noise. (c) Distortion 8: Gaussian blur noise. (d) Distortion 12: JPEG transmission errors.

next the mixed model associated with this hue distribution. The saliency map computed from the mixed model is given in Fig. 10(b), and the one computed from Itti's model is given in Fig. 10(c). Comparison of Figs. 10(a) and 10(b) shows that the saliency map computed from the mixed model is more accurate than that computed from Itti's model. This second example shows that the mixed model could be extended to high level features other than human faces.

Mixed Saliency Map Model Based on Salient Region

We usually focus on the salient regions instead of salient points. That means that the saliency value of every pixel in the region should be a weighted function of the saliency value of pixels belonging to the neighboring field or of the saliency value of the region it belongs to. For each pixel belonging to a salient region, we propose to enlarge the area of the neighboring field as if we were using a magnifying glass. For each pixel belonging to a nonsalient region, we

Table II. Spearman correlation.

	PSNRHVS	PSNRHVS_S	Δ(%)	PSNRHVSM	PSNRHVSM_S	Δ(%)
Noise	0.917	0.914	−0.327	0.918	0.92	0.218
Noise2	0.933	0.863	−7.5	0.93	0.871	−6.344
Safe	0.932	0.92	−1.28	0.936	0.924	−1.282
Hard	0.791	0.814	2.908	0.783	0.816	4.215
Simple	0.939	0.933	−0.639	0.942	0.935	−0.743
Exotic	0.275	0.465	69.09	0.274	0.442	61.314
Exotic2	0.324	0.377	16.358	0.287	0.331	15.331
Full	0.594	0.622	4.71	0.559	0.595	6.44

Table III. Kendall correlation.

	PSNRHVS	PSNRHVS_S	Δ(%)	PSNRHVSM	PSNRHVSM_S	Δ(%)
Noise	0.751	0.745	−0.799	0.752	0.752	0
Noise2	0.78	0.68	−12.82	0.771	0.689	−10.63
Safe	0.772	0.752	−2.59	0.778	0.757	−2.69
Hard	0.614	0.634	3.257	0.606	0.637	5.11
Simple	0.785	0.773	−1.52	0.789	0.777	−1.52
Exotic	0.195	0.313	60.51	0.194	0.294	51.55
Exotic2	0.238	0.254	6.72	0.21	0.22	4.76
Full	0.476	0.472	−0.8	0.449	0.455	1.34

propose to give less weight to the neighboring field. We used a metric to define the salient regions and the neighboring field associated with a given pixel.

First we computed the binary mark metric, $B_{i,j}$ defined as follows:

$$B_{i,j} = \begin{cases} 0 & \text{if } S_{\mathrm{MIX}}(i,j) < T_1 \\ 1 & \text{otherwise,} \end{cases} \tag{12}$$

where T_1 is an experimental threshold that is adaptive to the average value of $S_{\mathrm{MIX}}(i,j)$ and $S_{\mathrm{MIX}}(i,j)$ is the saliency value computed from the saliency map model considered and (i,j) is the pixel position in the image.

Next we computed block by block the relative saliency degree of the current pixel as a function of its neighboring field. The current point $A(i,j)$, current block(I,J), and the overlapped neighboring field $N(i,j)$ with size $k \times k$ are illustrated in Figure 11. Accordingly $\phi_{I,J}$ was defined as a saliency flag of the current block as follows:

$$\phi_{I,J} = \begin{cases} \text{false} & \text{if } \sum_{i=1}^{8}\sum_{j=1}^{8} B_{\mathrm{Block}(I,J)}(i,j) < T_2 \\ \text{true} & \text{otherwise,} \end{cases} \tag{13}$$

where T_2 is an experimental threshold, and the average of the current block was used as T_2; (i,j) is the pixel position in the Block(I,J).

Then, as salient regions focus more the attention of the observers than nonsalient regions, we gave less weight to pixels belonging to nonsalient regions. This means that the saliency value of every pixel is weighted by a function of the saliency values of the pixels belonging to its neighboring area. We considered several variables to compute the relative saliency of the current neighboring area, current block, and current pixel.

Let us define $\rho_{\mathrm{Block}}(I,J)$ and $\rho_{\mathrm{region}}(i,j)$, the relative saliency degree of the current block and the current neighboring field as functions of the average saliency of the global image;

$$\rho_{\mathrm{Block}}(I,J) = \frac{1}{\overline{S_{\mathrm{Global}}}}\left(\frac{1}{64}\sum_{i=1}^{8}\sum_{j=1}^{8} S_{\mathrm{MIX}}(i,j)\right), \tag{14}$$

$$\rho_{\mathrm{region}}(i,j) = \frac{S_{\mathrm{Local}}}{\overline{S_{\mathrm{Global}}}}, \tag{15}$$

with

$$\overline{S_{\mathrm{Local}}} = \frac{1}{k \times k}\sum_{i=1}^{k}\sum_{j=1}^{k} S_{\mathrm{MIX}}(i,j), \tag{16}$$

$$\overline{S_{\mathrm{Global}}} = \frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N} S_{\mathrm{MIX}}(i,j). \tag{17}$$

Let us define $\rho_{\mathrm{pixel\_average}}(i,j)$ and $\rho_{\mathrm{pixel\_max}}(i,j)$, the relative saliency degree of the current pixel as a function of its neighboring field and of the global image.

$$\rho_{\mathrm{pixel\_average}}(i,j) = \max\left\{\frac{S_{\mathrm{MIX}}(i,j)}{\overline{S_{\mathrm{Local}}}}, \frac{S_{\mathrm{MIX}}(i,j)}{\overline{S_{\mathrm{GLobal}}}}\right\}, \tag{18}$$

$$\rho_{\mathrm{pixel\_max}}(i,j) = \frac{S_{\mathrm{MIX}}(i,j)}{S_{\mathrm{Max\_Local}}}, \tag{19}$$

with

$$S_{\mathrm{Max\_Local}} = \max\{S_{\mathrm{MIX}}(i,j) | i \le k, j \le k\}. \tag{20}$$

Finally, to decrease the influence of nonsalient regions, we computed a weighted saliency map $w_s(i,j)$ as follows:

$$w_s(i,j) = \{\max\{\rho_{\mathrm{region}}(i,j), \rho_{\mathrm{Block}}(i,j)\} | \rho_{\mathrm{region}}(i,j) > T_3\}, \tag{21}$$

where T_3 is a threshold computed experimentally (see the Appendix).

Thus, if we consider, for example, the saliency map of reference I18 given by Fig. 9(a), we get the weighted saliency map w_s corresponding to Figure 12. Comparing Figs. 12(a) and 12(b), we can see that w_s reflects the fact that observers usually focus on the most salient parts instead of all locally salient parts. Most salient regions correspond to regions which are not only locally salient but also salient with regards to the global image.

Table IV. Spearman correlation comparison.

	WSNR	LINLAB	SNR	PSNR	PSNRHVSM	IFC	PSNRHVS	UQI	SSIM	PSNRHVS_S	PSNRHVSM_S
Noise	0.897	0.839	0.712	0.704	0.918	0.663	0.917	0.526	0.562	0.914	0.92
Noise2	0.908	0.853	0.687	0.612	0.93	0.743	0.933	0.599	0.637	0.863	0.871
Safe	0.921	0.859	0.699	0.689	0.936	0.775	0.932	0.638	0.632	0.92	0.924
Hard	0.776	0.761	0.646	0.697	0.783	0.736	0.791	0.759	0.812	0.814	0.816
Simple	0.931	0.877	0.794	0.799	0.942	0.817	0.939	0.784	0.769	0.933	0.935
Exotic	0.157	0.135	0.227	0.248	0.274	−0.269	0.275	0.292	0.385	0.465	0.442
Exotic2	0.059	0.033	0.29	0.308	0.287	0.276	0.324	0.546	0.594	0.377	0.331
Full	0.488	0.487	0.523	0.525	0.559	0.569	0.594	0.6	0.645	0.622	0.595

Table V. Kendall correlation comparison.

	PSNR	SNR	LINLAB	WSNR	IFC	UQI	PSNRHVSM	SSIM	PSNRHVS	PSNRHVS_S	PSNRHVSM_S
Noise	0.501	0.512	0.652	0.714	0.477	0.363	0.752	0.388	0.751	0.745	0.752
Noise2	0.424	0.492	0.671	0.736	0.547	0.42	0.771	0.45	0.78	0.68	0.689
Safe	0.486	0.497	0.682	0.753	0.581	0.454	0.778	0.437	0.772	0.752	0.757
Hard	0.516	0.464	0.569	0.586	0.552	0.565	0.606	0.618	0.614	0.634	0.637
Simple	0.598	0.593	0.715	0.766	0.624	0.587	0.789	0.564	0.785	0.773	0.777
Exotic	0.178	0.154	0.084	0.107	−0.156	0.196	0.194	0.266	0.195	0.313	0.294
Exotic2	0.225	0.205	0.026	0.047	0.208	0.389	0.21	0.431	0.238	0.254	0.22
Full	0.369	0.374	0.381	0.393	0.426	0.435	0.449	0.468	0.476	0.472	0.455

IMAGE QUALITY ASSESSMENT WEIGHTED BY SALIENT REGION

In order to improve the efficiency of image quality metrics taking into account the human visual attention mechanism, we propose to weight the image differences from the salient regions instead of salient points. Considering that human observers are unable to focus on several areas at the same time and that they assess the quality of an image first or mainly from the most salient areas, we propose to weight image difference metrics by the weighted saliency map w_s defined above. Thus the PSNRHVS metric can be computed with the following pseudo code:

```
//for the pixels in a target block with $8 \times 8$
for $i=1:8$
    for $j=1:8$
        if ($\phi_{I,J}$ is false)
            $\delta_{\mathrm{PSNRHVS\_S}}(i,j) = \delta(i,j) \cdot \left( \frac{\mathrm{CSF\ cof}(i,j)}{\mathrm{CSF\ cof}(i,j)+1} \right)$
;
        Else
            if $[(\rho_{\mathrm{pixel\_max}} > T_4)$ and $(\rho_{\mathrm{pixel\_average}} > T_5)]$
                $\delta_{\mathrm{PSNRHVS\_S}}(i,j) = \delta_{\mathrm{PSNRHVS}}(i,j) \cdot w_s(i,j)$;
            Else
                $\delta_{\mathrm{PSNRHVS\_S}}(i,j) = \delta_{\mathrm{PSNRHVS}}(i,j)$;
            end
        end
    End
end
```

In this algorithm (i,j) is the position of a pixel in an 8×8 block. The thresholds T_3, T_4, and T_5 have been empirically defined as 15, 0.5, and 40, respectively, for the TID2008 database. In our experiments, parameters T_3, T_4, and T_5 were selected via an exhaustive process in a three-dimensional search space $\{T_3, T_4, T_5\}$. In this space, every parameter T_3, T_4, T_5, was normalized to a scale which was next separated into m subscales in order to get a data grid of m^3 grid points. Then we have chosen in the grid points set the best grid point (i.e., the values for T_3, T_4, and T_5) with the highest performance in regards to the data set considered.

EXPERIMENTAL RESULTS AND ANALYSIS

In this article, the images in the TID2008 database were used to test our image quality assessment model. TID2008 is the largest database of distorted images intended for verification of full reference quality metrics.[23] We used the TID2008 database as it contains more distorted images, types of distortion and subjective experiments than the LIVE database.[24]

Table VI. Spearman correlation and Kendall correlation on LIVE database.

Correlation	SNR	PSNR	WSNR	UQI	IFC	SSIM	PSNRHVS_S	PSNRHVSM_S
Spearman	0.7811	0.8044	0.8479	0.802	0.8429	0.86	0.89	0.8963
Kendall	0.5922	0.6175	0.6883	0.6142	0.6677	0.7057	0.7179	0.7258

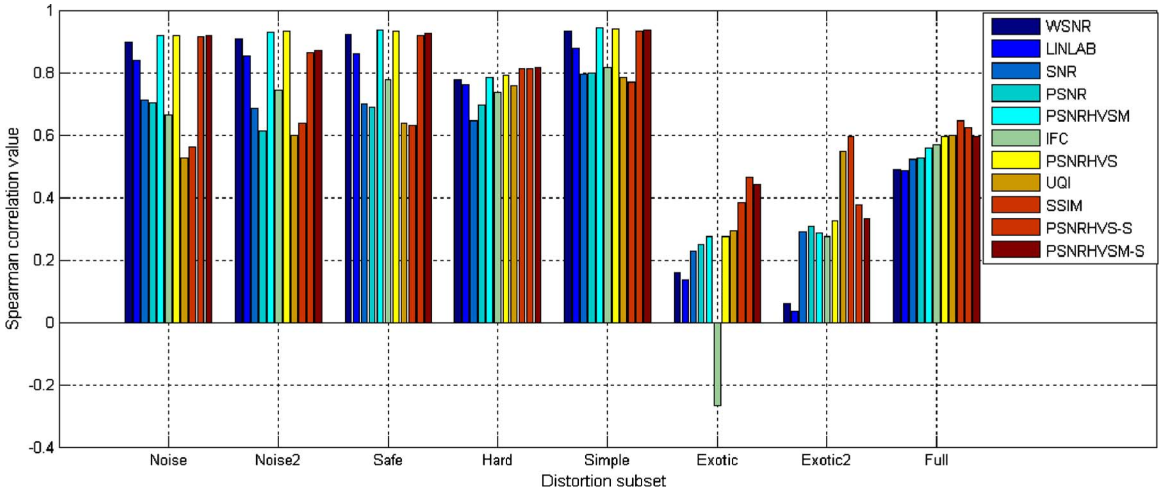


Figure 14. Spearman correlation comparison.

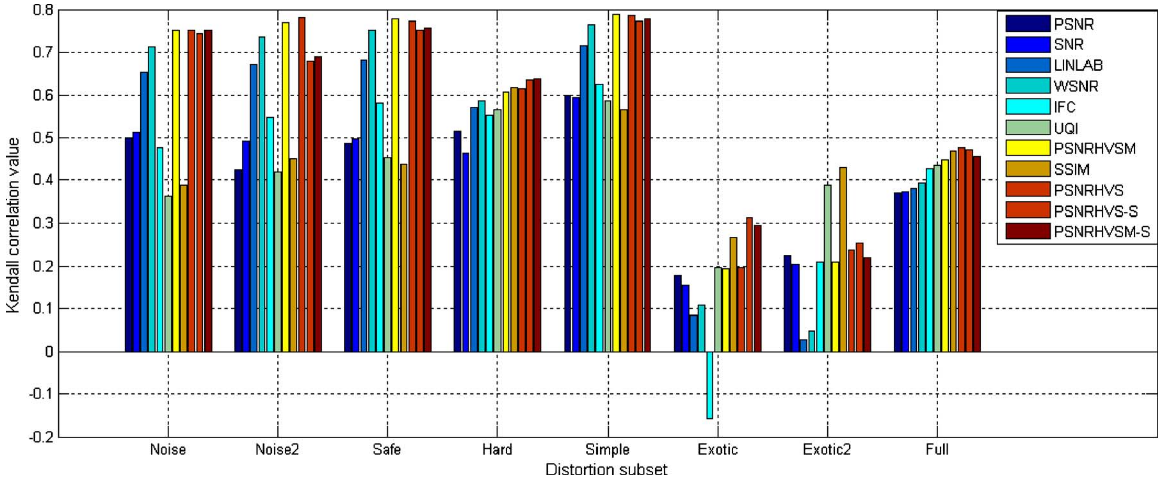


Figure 15. Kendall correlation comparison.

The TID2008 database contains 1700 distorted images (25 reference images × 17 types of distortions × 4 levels of distortions). LIVE on the other hand contains 779 distorted images with only five types of distortion and 161 subjective experiments. The mean opinion score (MOS) of image quality was computed from the results of 838 subjective experiments carried out by observers from Finland, Italy, and Ukraine. The higher the MOS (0-minimal, 9-maximal, MSE of each score is 0.019), the higher the visual quality of the images. In our experiments, both databases have been used to compare results from different image quality metrics.

All the distorted images are grouped together into a full subset or into different subsets including noise, noise2, safe, hard, simple, exotic, and exotic2 with different distortions. For example, in the Noise subset there are several types of distortions such as high frequency noise distortion, Gaussian blur, etc. Table I shows every subset and its corresponding distortion type. Distortion of types 12, 13, and 16, etc. are included in the exotic2 subset. Figures 13(b)–13(d) show, respectively, the distortion types 5, 8, and 12 in the noise and exotics subsets.

Experimental Results from TID2008

In order to compare the accuracy of the image quality metrics weighted by salient regions with those of nonweighted metrics, we compute the Spearman correlation and Kendall correlation coefficients. Spearman correlation and Kendall correlation coefficients are two indexes used in

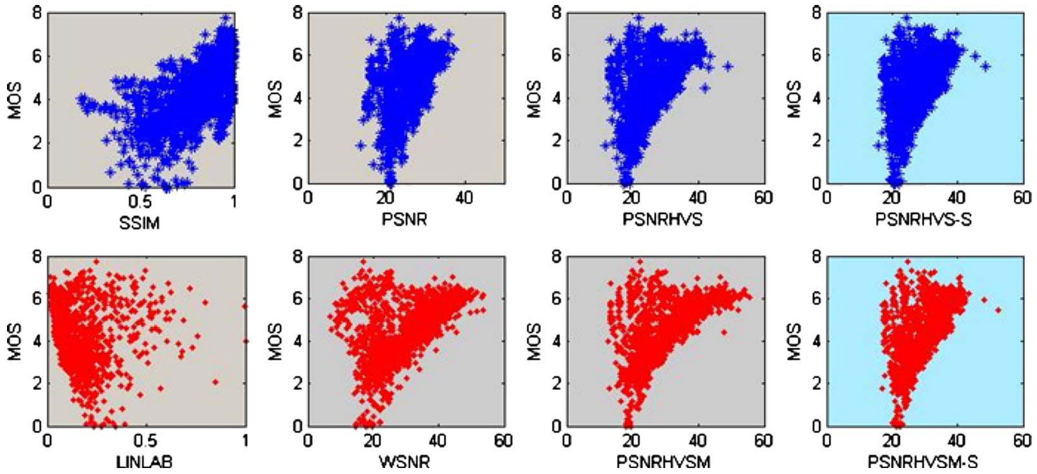


Figure 16. Scatter plots of the image quality assessment models, the plots with blue points are the results from the image quality assessment model based on weighted saliency map.

Table VII. Spearman correlation.

	PSNRHVS_s			PSNRHVSM_s		
Distortion type	$\rho_{\text{region}}(i,j)$	$\rho_{\text{Block}}(i,j)$	Max	$\rho_{\text{region}}(i,j)$	$\rho_{\text{Block}}(i,j)$	Max
Noise	0.913	0.913	0.914	0.920	0.920	0.92
Noise2	0.862	0.862	0.863	0.872	0.872	0.871
Safe	0.920	0.920	0.92	0.924	0.924	0.924
Hard	0.815	0.815	0.814	0.817	0.817	0.816
Simple	0.932	0.932	0.933	0.935	0.935	0.935
Exotic	0.463	0.463	0.465	0.440	0.440	0.442
Exotic2	0.377	0.377	0.377	0.331	0.331	0.331
Full	0.622	0.622	0.622	0.595	0.595	0.595

image quality assessment to compute the correlation of objective measures with human perception. Compared with the original PSNRHVS and PSNRHVSM metric, the method based on region saliency greatly enhances the performance on exotic and exotic2. In Tables II and III, PSNRHVS_S and PSNRHVSM_S are, respectively, the new modified PSNRHVS and PSNRHVSM based on the weighted saliency map. The original PSNRHVS and PSNRHVSM are based on image difference metrics which assess image quality by independent blocks without taking into account that salient regions contribute more in the image quality score. In this comparison Δ(%) is the enhancement of performance of PSNRHVS and PSNRHVSM.

From the point of view of Spearman correlation coefficients, PSNRHVS and PSNRHVSM perform well on noise, noise2, safe, hard, and simple subsets of TID2008. But they fail to perform well on exotic and exotic2 subsets. With the weighted saliency map, the Spearman coefficients of PSNRHVS and PSNRHVSM on full subsets are enhanced although there is reduction on the noise2 subset. On exotic and exotic2 distorted subsets, the performances of the modified PSNRHVS and PSNRHVSM based on saliency map are remarkably enhanced. For PSNRHVS, the Spearman correlations on exotic and exotics2 are enhanced 69.1% and 16.4%, respectively, and Kendall correlations are enhanced 60.5% and 6.7%, respectively. For PSNRHVSM, the Spearman correlations are enhanced 61.3% and 15.3%, respectively, and Kendall correlations are enhanced 51.55% and 4.8%, respectively. Exotic and exotic2 are two subsets with contrast change and mean shift distortions. PSNRHVS and PSNRHVSM only used the intensity information, but for our proposed method, color contrast, intensity and other information will be reflected in the image quality assessment. So our method can reflect the attributes of our visual attention more effectively than PSNRHVS or PSNRHVSM.

Furthermore besides the comparison between the algorithm that we propose and the original PSNRHVS, other image quality assessment metrics have been included to make the result more creditable. Nine other image quality assessment metrics, including SSIM UQI, SNR, PSNR, WSNR, LINLAB, PSNRHVS, PSNRHVSM, and IFC, had been also used for comparing results. The results computed from all the quality metrics considered are arranged in order of increasing value of the correlation coefficient on the full

Table VIII. Kendall correlation.

Distortion type	PSNRHVS_s			PSNRHVSM_s		
Distortion	$\rho_{\text{region}}(i,j)$	$\rho_{\text{Block}}(i,j)$	Max	$\rho_{\text{region}}(i,j)$	$\rho_{\text{Block}}(i,j)$	Max
Noise	0.743	0.743	0.745	0.752	0.752	0.752
Noise2	0.680	0.680	0.68	0.689	0.689	0.689
Safe	0.750	0.750	0.752	0.757	0.757	0.757
Hard	0.634	0.634	0.634	0.637	0.637	0.637
Simple	0.770	0.770	0.773	0.776	0.776	0.777
Exotic	0.313	0.313	0.313	0.293	0.293	0.294
Exotic2	0.255	0.255	0.254	0.220	0.220	0.22
Full	0.472	0.472	0.472	0.455	0.455	0.455

Table IX. PSNRHVS_S with different operator.

	PSNRHVS_S				
	Spearman correlation			Kendall correlation	
Distortion type	$\rho_{\text{Block}}(i,j)$ without T_3	With T_3	Distortion type	$\rho_{\text{Block}}(i,j)$ Without T_3	With T_3
Noise	0.707	0.913	Noise	0.521	0.743
Noise2	0.657	0.862	Noise2	0.475	0.68
Safe	0.732	0.92	Safe	0.537	0.75
Hard	0.587	0.815	Hard	0.422	0.634
Simple	0.716	0.932	Simple	0.517	0.77
Exotic	0.228	0.463	Exotic	0.162	0.313
Exotic2	0.201	0.377	Exotic2	0.138	0.255
Full	0.446	0.622	Full	0.312	0.472

subset; the methods that we propose are also listed at the right of the Tables IV and V for comparison.

Figures 14 and 15 show the results obtained from different image quality metrics on different subsets of TID2008. SSIM showed nearly the best performance on the full subset in terms of Spearman correlation; however, according to Figs. 14 and 15 SSIM performance on noise, noise2, simple, etc., are much lower than that of the method that we propose. The high values of Spearman and Kendall correlations computed from the original methods PSNRHVS_S and PSNRHVSM_S are preserved by the modified PSNR-HVS and PSNR-HVS-M on noise, safe, hard and simple subsets, while the performance on Exotic and Exotic2 subsets is improved remarkably. The method PSNRHVS_S that we propose gets almost the highest values on every subset.

Figure 16 illustrates the scatter plots for the MOS for different models including PSNR, LINLAB, WNSR, PSNRHVS and PSNRHVS_S, etc. Usually we expect the scatter plot to define a cluster, which means that the subjective score and objective assessing value are tightly correlated since the ideal image quality metric should accurately reflect the subjective score, i.e., the MOS. The plots from the methods that we propose, PSNRHVS-S and PSNRHVSM-S are effectively better clustered than that of original models, PSNRHVS and PSNRHVSM, except for only few extreme points.

Experiment on LIVE Database

Besides the TID2008 database, LIVE database (release 1) used for image quality assessing from the University of Texas has also been used to test the methods that we propose. Since LIVE database was first set up with popular SSIM and UQI metrics, we also test the metrics that we propose on LIVE database and compare our results with them. Besides SSIM and UQI, we also compared our proposed methods with IFC, WSNR, SNR, PSNR, etc. Metrix'Mux toolbox was used in our experiments to compute image quality with SSIM and UQI.[25] The results show that the methods that we propose with region saliency, PSNRHVS_S and PSNRHVSM_S, get nearly the highest values of Spearman and Kendall correlation for the LIVE database (Table VI).

CONCLUSIONS AND FURTHER RESEARCH

In this article, a saliency map has been introduced to improve image quality assessment based on the observation that salient regions contribute more to the perceived image quality. The saliency map is defined by a mixed model based on Itti's model and a face detection model. Salient region information including local contrast saliency and local average saliency, etc. were used instead of salient pixel information as weights of the output of previous methods. The experimental results from TID2008 database show that the weighted saliency map can be used to enhance the performance of PSNRHVS, PSNRHVS-M on specific subsets remarkably.

Future research involves extending the test database and analyzing the extreme points in scatter plots for which the distance between objective metrics and MOS is large, i.e., images for which the image quality assessment models do not work accurately. The performance of image quality assessment models will be enhanced by reducing the number of these extreme points. Besides that, some machine learning methods, such as the neural network approach, might be employed to acquire well-chosen coefficients in the mixed saliency map and thresholds although much more complexity could be introduced thereby.

ACKNOWLEDGMENT

We thank the Rhône-Alpes region for its support through the LIMA project of cluster ISLE.

APPENDIX

Region Saliency Map w_s and Its Simplification

This part shows while the function defined by the maximum of the two parameters $\rho_{\text{region}}(i,j)$ and $\rho_{\text{Block}}(i,j)$ has been chosen among other tested functions for Eq. (21), which was defined above as

$$w_s(i,j) = \{\max\{\rho_{\text{region}}(i,j), \rho_{\text{Block}}(i,j)\} | \rho_{\text{region}}(i,j) > T_3\}.$$

We have tested different functions to calculate $w_s(i,j)$; for example, we have tried to use only $\rho_{\text{region}}(i,j)$ or $\rho_{\text{Block}}(i,j)$ instead of the function 'max' with the following results on TID2008 (see Tables VII and VIII).

From these tables we can see that the results obtained from only $\rho_{\text{region}}(i,j)$ or $\rho_{\text{Block}}(i,j)$ are almost similar to the max result although the results from *max* was slightly higher than from the others. For reduced computation, Eq. (21) could also be simplified as follows:

$$w_s(i,j) = \{\rho_{\text{Block}}(i,j) | \rho_{\text{region}}(i,j) > T_3\}. \tag{22}$$

The reason for using the threshold T_3 is that when the threshold is limited to a lower value, then $\rho_{\text{region}}(i,j)$ and $\rho_{\text{Block}}(i,j)$ are more effective. The following test results, illustrated in Table IX, computed from $\rho_{\text{Block}}(i,j)$, with and without T_3, show the influence of the T_3 threshold. The Spearman and Kendall correlation with the T_3 threshold are much higher than that without T_3.

REFERENCES

[1] N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, and V. Lukin, "Metrics performance comparison for color image database", *Proc. Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics* (Scottsdale, AZ, 2009) pp. 14–16.

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

[3] M. Gaubatz, "Metrix MUX Visual Quality Assessment Package: MSE, PSNR, SSIM, MSSIM, VSNR, VIF, VIFP, UQI, IFC, NQM, WSNR, SNR", http://foulard.ece.cornell.edu/gaubatz/metrix_mux/.

[4] A. B. Watson, "DCTune: A technique for visual optimization of DCT quantization matrices for individual images", *Soc. Inf. Display Digest,Technical Papers: XXIV* (SID, Santa Anaheim, CA, 1993), pp. 946–949.

[5] Z. Wang and A. Bovik, "A universal image quality index", IEEE Signal Process. Lett. **9**, 81–84 (2002).

[6] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity", IEEE Trans. Image Process. **13**, 600–612 (2004).

[7] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment", *Proc. 37th Asilomar Conference on Signals, Systems and Computers* (IEEE, Piscataway, NJ, 2003) pp. 1398–1402.

[8] B. Kolpatzik and C. Bouman, "Optimized error diffusion for high quality image display", J. Electron. Imaging **1**, 277–292 (1992).

[9] B. W. Kolpatzik and C. A. Bouman, "Optimized universal color palette design for error diffusion", J. Electron. Imaging **4**, 131–143 (1995).

[10] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions", *Proc. of the Third International Workshop on Video Processing and Quality Metrics, VPQM 2007* (Scottsdale, AZ, 2007) p. 4.

[11] H. R. Sheikh and A. C. Bovik, "Image information and visual quality", IEEE Trans. Image Process. **15**, 430–444 (2006).

[12] N. Damera-Venkata, T. Kite, W. Geisler, B. Evans, and A. Bovik, "Image quality assessment based on a degradation model", IEEE Trans. Image Process. **9**, 636–650 (2000).

[13] T. Mitsa and K. Varkur, "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms", *Proc. ICASSP (IEEE)* (IEEE, Piscataway, NJ, 1993) pp. 301–304.

[14] 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. **14**, 2117–2128 (2005).

[15] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images", IEEE Trans. Image Process. **16**, 2284–2298 (2007).

[16] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "New full-reference quality metrics based on HVS", *Proceedings of the Second International Workshop on Video Processing and Quality Metrics* (Scottsdale, AZ, 2006) p. 4.

[17] Q. Ma and L. Zhang, "Saliency-based image quality assessment criterion", *Proceedings of ICIC 2008, LNCS 5226* (Springer, Shanghai, 2008) pp. 1124–1133.

[18] X. Feng, T. Liu, D. Yang, and Y. Wang, "Saliency-based objective quality assessment of decoded video affected by packet losses", *Proceedings of ICIP (IEEE)* (IEEE, Piscataway, NJ, 2008) pp. 2560–2563.

[19] R. Desimone, T. D. Albright, C. G. Gross, and C. Bruce, "Stimulus selective properties of inferior temporal neurons in the macaque", J. Neurosci. **4**, 2051–2062 (1984).

[20] L. Itti and C. Koch, "A saliency-based search mechanism forovert and covert shifts of visual attention", Vision Res. **40**, 211–227 (2008).

[21] Face detection using OPENCV, http://opencv.willowgarage.com/wiki/FaceDetection, (accessed May 2009).

[22] W. D. Koch, "Modeling attention to salient proto-objects", Neural Networks **19**, 1395–1407 (2006).

[23] TID2008, http://www.ponomarenko.info/tid2008.htm (accessed May 2009).

[24] 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. **15**, 3440–3451 (2006).

[25] M. Gaubatz, "Metrix MUX Visual Quality Assessment Package", http://foulard.ece.cornell.edu/gaubatz/metrix_mux/ (accessed May 2009).