COLOR FACE-TUNED SALIENT DETECTION FOR IMAGE QUALITY ASSESSMENT

Tong Yubing, Hubert Konik, and Alain Tremeau

Université de Lyon, Université Jean Monnet, Laboratoire Hubert Curien UMR 5516
42000 Saint-Etienne, France.

ABSTRACT

PSNRHVS and PSNRHVSM are two new emerging image quality assessment methods but they fail when assessing the quality of some distorted images called as "extreme" images. In this paper an algorithm is proposed to enhance their performance on extreme images while keeping their good performance on "normal" images unchanged. First, extreme images derived from PSNRHVS are labeled with an iterative algorithm. Then an SVM classifier is used to decide if current images are extreme images or not. Next, region saliency information is computed only for this kind of images. Then region saliency information is used instead of point saliency information in image quality assessment. We use color, intensity and orientation to compute the saliency of regions. We use also a face descriptor as faces play an important role in visual perception. The algorithm that we propose has been tested on the TID2008 database. The results that we have obtained show that the performance on extreme images is greatly enhanced compared with the original PSNRHVS.

Index Terms— image quality assessment, region saliency map, face detection, image classification.

1. INTRODUCTION

Since several years there is a trend to assess image quality with objective methods to decrease computing time, and to develop real-time algorithms. Many papers proposed image quality metrics, such as UQI, SSIM, LINLAB, PSNRHVS, PSNRHVSM, WSNR, IFC and VSNR, which give better results than PSNR and MSE [1-14]. But considering the wide range of possible distortion types, none of these metrics perform well enough. It has been shown in [15] that with the PSNRHVS and PSNRHVSM metrics we can nevertheless obtain excellent performance on Noise, Noise2, Safe and Hard subsets of TID2008 database, and that these two metrics could be recommended for evaluating efficiency of image filtering and lossy compressing. But both fail when assessing the quality of some distorted images, called "extreme" images. With such images the Mean Opinion Scores (MOS) of subjective assessments deviates much from objective values, meanwhile for other "normal" points, the subjective scores are consistent with the objective scores.

The aim of this paper was first to define, next to label, "extreme" images with an iterative algorithm. We have shown that effectively the lower quality on extreme images reduces much the performance of image quality metrics than those computed for other images. We propose to use a SVM classifier based on a RBF kernel to decide if an image, assessed with PSNRHVS, is extreme or not. Next, in order to enhance the performance of image quality metrics we propose to weight these metrics with saliency region information based on the presence of extreme images. Most of saliency maps, such as those proposed in [16-17] are computed, directly, at each point. Here we propose to consider additional information, based on image content, which characterizes the local saliency of regions. Our idea is to consider not only the saliency of every pixel but also the relative saliency of the current pixel in regards to its neighboring field and to the global image.

In section 2 we present briefly PSNRHVS and PSNRHVSM metrics, next in section 3 we define what we call "extreme" images and we present the classification algorithm that we propose to detect extreme images. Then, in section 4 we propose a new image quality assessment metrics based on region saliency. Lastly, in section 5 we present experimental results showing that the performance on extreme images is greatly enhanced with these new metrics compared with PSNRHVS. Experimental results from TID2008 database are discussed; next section 6 concludes the paper.

2. RELATED WORK AND ANALYSIS

Lacks of classical PSNR and MSE metrics are well-known. As example, Ponomarenko et al. showed in [15] that these metrics give the worst scores on TID2008 database according to Spearman's correlation and Kendall's correlation. The TID2008 database contains 1700 test images [18]. More than 800 image quality subjective assessments have been done from this database in order to compute reliable MOS. PSNRHVS and PSNRHVSM are two recent metrics which had been designed to improve the performance of PSNR and MSE.

PSNRHVS divides the image into 8x8 pixels non-overlapping blocks. Next, the difference between the original and the distorted blocks is weighted for every 8x8 block by the coefficients of the Contrast Sensitivity Function (CSF):

$$\delta_{PSNRHVS}(i,j) = [a(i,j) - \hat{a}(i,j)].CSF_{Coef}(i,j) \qquad (1)$$

PSNRHVSM is defined in a similar way, but the difference between the DCT coefficients is further multiplied by a Contrast Masking metric (CM) for every 8x8 block. The result is then weighted by the CSF_{Coef} as follows:

$$\delta_{PSNRHVSM}(i,j) = ([a(i,j) - \hat{a}(i,j)].CM(i,j)).CSF_{Coef}(i,j) \qquad (2)$$

Next, a MSE metric for both PSNRHVS and PSNRHVSM can be defined by summing differences between the original image and its distorted version inside blocks. The idea is therefore to divide the image into 8x8 blocks and inside each of them to compute CSF or CM metric to improve the absolute difference in MSE and PSNR. So all 8x8 blocks are treated independently from each other and contribute equally to the image quality. But the problem is that some blocks or even large regions are more salient than others and therefore contribute more to the image quality than other ones. So PSNRHVS and PSNRHVSM fail to assess image quality of some images. Figure 1 illustrates this drawback:
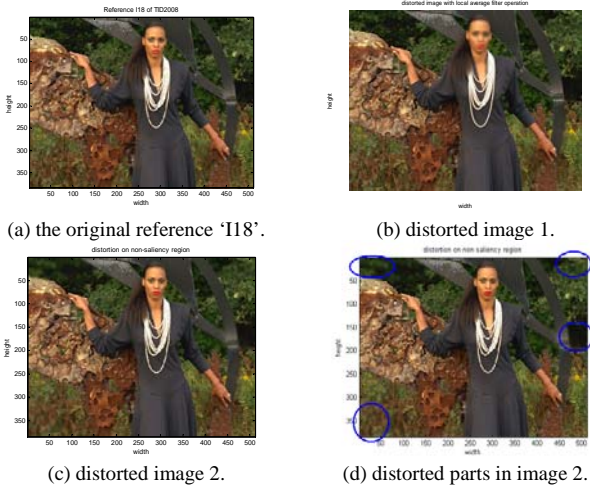

| (a) the original reference 'I18'. | (b) distorted image 1. |

| (c) distorted image 2. | (d) distorted parts in image 2. |

Figure 1. Examples of distorted images of 'I18' from TID2008 database.

In Figure 1 (b) the face, neck and breast part are distorted by noise. The objective image quality of sub-figure (b) is: 46.3db with PSNR, 33.74db with PSNRHVS and 36.3 db with PSNRHVSM. The objective quality of sub-figure (c) is: 41.6 db with PSNR, 32.4 db with PSNRHVS and 35.8 db with PSNRHVSM. These values show that the quality of sub-figure (b) is better than (c). But visually speaking (c) is better than (b) as less salient regions are distorted.

3. EXTREME DISTORTED IMAGES ANALYSIS

3.1. Extreme Images Definition and Labeling

In theory, the output of an ideal image quality assessment model should be consistent with subjective score; MOS. MOS can be divided into Low, Middle and High levels. Likewise, objective assessing scores can be divided into small, middle and large levels (as example see Figure 2 (a)). The consistence of subjective assessments and objective scores is established when their correlation coefficient is high, i.e. when most of scatter points are in the areas A_1, A_2 and A_3. In this paper, the points in the areas A_8 and A_9 whose subjective score deviates much from objective score are called "extreme" images. Here only A_8 and A_9 areas are considered as sets of extreme images meanwhile others are considered as "normal". As example let us consider the scatter plot of TID2008 images database, see Figure 2 (b). According to the above description, extreme images are underlined by red circles. Likewise, in the following section we will show that the Spearman and Kendall correlations also lower for these areas and that they pull down the performance of all areas. Spearman correlation and Kendall correlation are two important indexes used to describe the correlation between MOS and objective score.
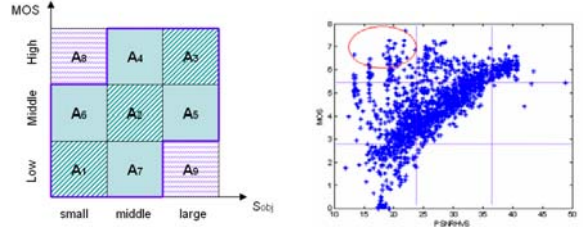

| (a) Common scatter plot | (b) Example with PSNRHVS-MOS. |

Figure 2. Set of extreme images computed from the scatter plot.

Extreme images can be labeled according to their belonging to areas A_8 and A_9 but there may be images at the boundary of these areas, such as images between A_8 and A_4, for which there is an ambiguity in using this delimitation. To avoid that case and make the extreme image selection more robust, we propose to detect extreme images with Spearman and Kendall correlations as follows:

Extreme Points Definition. *For the set of images considered, S_{all}, at each image i corresponds a point X_i (mos_i, obj_i) in the scatter plot of subjective scores and objective values, which can be defined as extreme image or normal image according to the following equation:*

$$\begin{cases} X_i(mos_i, obj_i) \in S_e & if \quad \Gamma(mos_i, obj_i) \leq \Gamma_T \\ X_i(mos_i, obj_i) \in S_n & else \end{cases} \qquad (3)$$

$$S_{all} = S_e \bigcup S_n \qquad (4)$$

where S_e is the set of extreme images and S_n is the set of normal images, obj_i is the objective score derived from image quality assessing method, (e.g. PSNRHVS). $\Gamma()$ is a distance function describing the correlation between MOS and objective score.

Γ_T is an adaptive threshold which depends on the results of the previous extreme images selection. In practice, the first threshold depends of the Spearman correlation SP_0

of the subset S_0. Next, the following thresholds Γ_T are adjusted adaptively by the following iterative algorithm. The selection of the next set of extreme images is based on the computation of absolute changes of Spearman and Kendall correlation coefficients for the current image. It is based also on the computation of contrast changes (relative change) between the current selection and the previous selection.
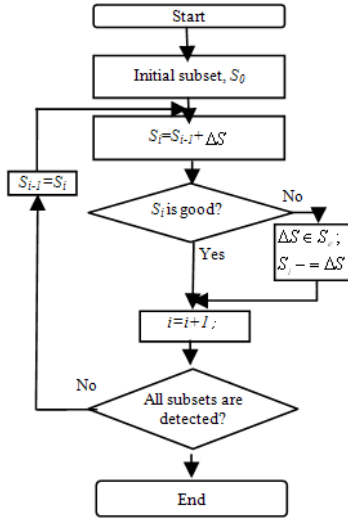

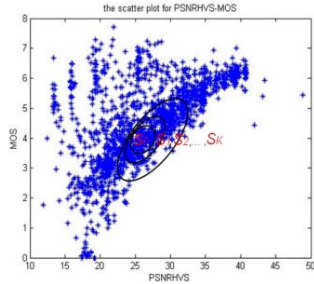*Figure 3. Flowchart of the extreme images detection process.*

Figure 4. Example of extreme images searching.

According to the flowchart of Figure 3, the searching algorithm of extreme images can be described by the following pseudo code:

STEP1. *select the initial subset S_0, next check the Spearman correlation SP_0 and Kendall correlation Ken_0 of the subset S_0.*

if $(SP_0 > Tsp_0 \ \& \ Ken_0 > T_{Ken0})$ chose the current subset S_0 as the initial S_0
$S_0 \in S_e$;
else reduce the member in subset S_0, and re-check S_0;
end

% Usually SP_0 and Ken_0 can be settled as a target correlation with range from 0 to 1.

STEP2. *create the current subset by extending last subset S_{i-1} with ΔS as follow: $S_i = S_{i-1} + \Delta S$*

STEP3. *check the performance of the current subset S_i by calculating the Spearman correlation SP_i and Kendall correlation Ken_i,*

if $(SP_i > SP_{i-1} \ \& \ Ken_i > Ken_{i-1})$ then ΔS is good and $\Delta S \in S_n$;
else
    if $(SP_i < SP_{i-1} \ \& \ Ken_i < Ken_{i-1} \ \& \ (SP_{i-1} - SP_i)/ \ SP_{i-1} < T_0 \ \& \ (Ken_{i-1} - Ken_i)/Ken_{i-1} < T_0)$ then $\Delta S \in S_n$;
        else $\Delta S \in S_e$;
        end
end

STEP4. *check if all subsets are detected,*

if all subsets have been detected go to step 5;
else $S_{i-1} = S_i$; go to step 2;
end

STEP5. *end*

The process is repeated until all the subsets of extreme images have been detected. As illustration see Figure 4. T_0 is an experiential threshold which describes contrast changes (relative change) between the current selection and the previous selection. Thanks to the above iterative searching algorithm the extreme images in the scatter plot of PSNRHVS can be labeled. As example, extreme images of TID2008 images database are labeled in Figure 5 by red points. This figure shows that most of the TID2008 images can be considered as normal images except only few of them. That means that PSNRHVS work well for most of images except for images on which the distance between MOS and objective score is higher than those of normal images (i.e. extreme images). Moreover, the Spearman and Kendall correlation coefficients are much lower for these images than for normal images, as shown in Table 1.
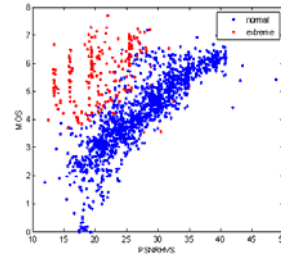

Figure 5. Scatter plot of PSNRHVS-MOS.

Table 1. Correlation coefficient on the TID2008 images database.

Full distortion subsets		
Subset Coefficient	Spearman	Kendall
PSNR	0.525	0.369
PSNRHVS-FULL	0.594	0.476
PSNRHVS-NORMAL	0.88	0.7033
PSNRHVS-EXTREME	0.3896	0.2698
PSNRHVSM-FULL	0.559	0.449
PSNRHVSM-NORMAL	0.8797	0.7015
PSNRHVSM-EXTREME	0.2586	0.1843

No matter PSNRHVS or PSNRHVSM, Spearman and Kendall correlation coefficients are high on normal subsets and low on extreme subset. Furthermore, on full subset the Spearman and Kendall correlation coefficients are not very high. That means that the lower values of extreme images

pull down the average value of all images. If we can identify extreme images then correspondingly we could enhance their correlation coefficients, and thus the overall correlation coefficient could be definitely raised. In this paper, we propose to use a SVM classifier to predict which images are extreme or normal.

3.2. Extreme Points Decision with SVM Classifier

With the above iterative algorithm, all images can be labeled into normal subset and extreme subset thanks to their MOS and objective scores. Then they can be used as training dataset for any learning machine algorithm. Usually, when we want to assess an image we have no MOS information. In order to enhance the relevance of PSNRHVS for such images, we propose to use a SVM classifier to decide if the current image belongs or not to the subset of extreme images. The idea is to improve image quality assessment metric in adding region saliency information to PSNRHVS, and this only for extreme images. The flowchart of the new image quality assessment metric based on a SVM classifier that we propose is illustrated by Figure 6.
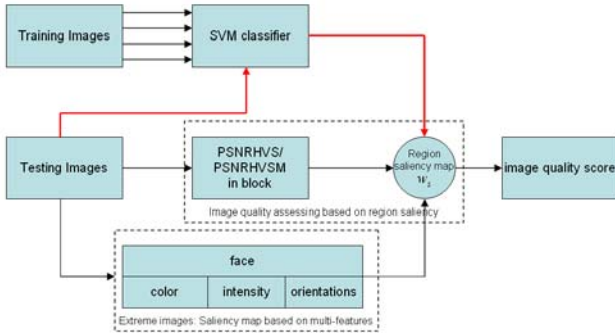

Figure 6. The flowchart of new image quality assessment metric proposed.

A two-label SVM classifier has been used with four inputs including the PSNRHVS and the three other features defined below. These three features have been used for SSIM [2, 3].

$$l(x,y) = \frac{2\ \mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad C_1 = (K_1 L)^2 \quad (5)$$

$$c(x,y) = \frac{2\ \sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad C_2 = (K_2 L)^2 \quad (6)$$

$$s(x,y) = \frac{2\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \quad (7)$$

With $\quad \sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)$

Where $\mu_x, \mu_y, \sigma_x, \sigma_y$ are the average value and standard square error of image x and image y. The SVM used is based on a RBF kernel and the LIBSVM toolbox has been used to train and predict images with this classifier [19].

3.3. Image Quality Analysis of Extreme Images

In order to show that the overall image quality deviate much from MOS subjective score, when the quality of salient regions of extreme images is not fully assessed, we used the following reference images. The first image used corresponds to the TID2008 reference image 'I14' and to two distorted versions of this image 'I14-17-2' and 'I14-17-3' (see Figure 7). This image has been used as 'I14-17-2' and 'I14-17-3' belong to the set of extreme images according to the SVM classifier.
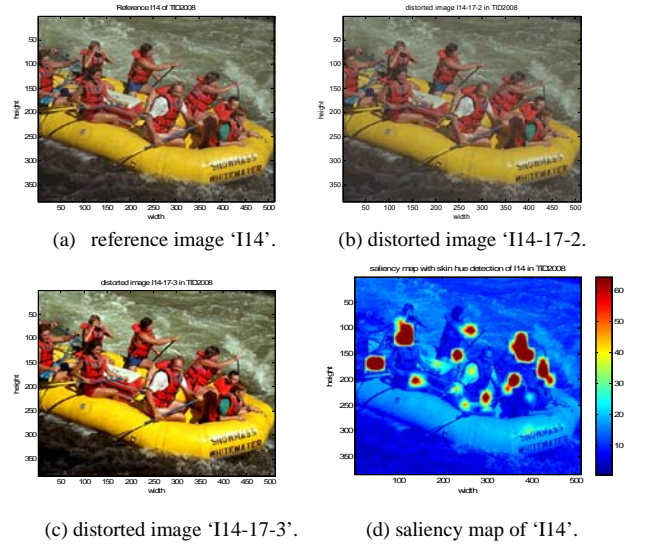

(a) reference image 'I14'. (b) distorted image 'I14-17-2.

(c) distorted image 'I14-17-3'. (d) saliency map of 'I14'.

Figure 7. reference image 'I14' and corresponding distorted images.

According to TID2008, the subjective score of 'I14-17-2' is lower than that of 'I14-17-3', meanwhile PSNRHVS and PSNRHVSM are higher for 'I14-17-2' than for 'I14-17-3'. For 'I14-17-2', PSNRHVS and PSNRHVSM are 23.3 db, 23.95db. For 'I14-17-3', PSNRHVS and PSNR-HVS_M are 19.3db and 19.87 db. But if we look at these images, our attention is much more focused by saliency regions, such as face, hands etc. than by other regions (see Figure 7 (d)). Consequently, these salient regions contribute more to subjective image quality than other ones. When the quality of these salient regions is acceptable, the overall image quality is also perceived as good. For extreme images, saliency perception influences significantly image quality assessment. That means that for extreme images we must more focus on saliency regions than for normal images as if these regions have been enlarged behind a magnifying glass. That also means that the distortions in salient regions should be much more taken into account with high weights than non-saliency regions, such as the "water" region in the background of image 'I14'.

4. IMAGE QUALITY ASSESSMENT BASED ON REGION SALIENCY

4.1. Saliency map with face detection

To compute the saliency we used the mixed saliency model introduced in [20]. This saliency model is based on three low level features (color, intensity and orientation), such as the Itti's saliency map model [21], and one high level features (faces and hands), such as the Koch's face detection model [22]. Cerf et al. showed in [23, 24] that faces are features which focus more attention than other features in many images. Psychological tests have proven that face, head or hands can be perceived by observers prior to any other details. For this reason we have considered faces as high level feature in our saliency map model.

The mixed saliency model based on Itti's model and face detection model (based on skin hue detection) can be defined with linear weights as follow:

$$S_{MIX} = f(S_{Itti}, S_{face}) \qquad (8)$$

Here, the linear weights have been defined empirically from the TID2008 database as follow:

$$S_{MIX} = \alpha \cdot S_{Itti} + (1 - \alpha) \cdot S_{face} \qquad (9)$$

For most of images containing faces, heads or hands, we obtain better results with the mixed model (with $\alpha = 3/7$) than with the Itti's model, i.e. more accurate saliency maps.

4.2. Image Quality Assessment Based on Salient Regions

In order to analyze image quality, we propose here to take into account regions saliency maps instead of pixels saliency maps. First, a one-zero mark metric, $B_{i,j}$ is defined as follow:

$$B_{i,j} = \begin{cases} 0 & if \ S_{MIX}(i,j) < T_1 \\ 1 & else \end{cases} \qquad (10)$$

Where T_1 is an experimental threshold that can be adaptively adjusted accordingly to the average of S_{MIX} values, and $S_{MIX}(i,j)$ is the mixed saliency value of pixel (i,j).

Next, the image is divide is divided block by block and the saliency of the pixel $A(i,j)$ is analysed in function of its neighboring field $N(i,j)$ and relatively to the saliency of the corresponding block (I,J) (see Figure 8). The size of neighboring field is fixed to $k \times k$ pixels.

For each block (I,J) a saliency flag $\phi_{I,J}$ is thus defined as follow:

$$\phi_{I,J} = \begin{cases} false & if \ \sum_{i=1}^{8} \sum_{j=1}^{8} B_{Block(I,J)}(i,j) < T_2 \\ true & else \end{cases} \qquad (11)$$

Where T_2 is an experimental threshold that can be adaptively adjusted accordingly to the average of $B_{Block(I,J)}$ values, and (i,j) is the pixel position in the *Block(I,J)*.
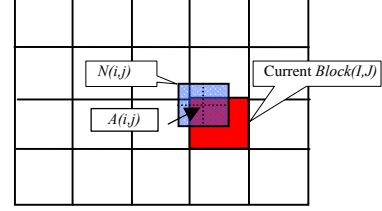

Figure 8. Current block, current pixel and corresponding neighboring field.

As salient regions focus more the observers' attention than non-salient regions, we propose to give less weight to pixels belonging to non-salient regions. This means that the saliency value of each pixel has to be weighted relatively to the saliency values of pixels belonging to its neighboring field. In this study, we have considered several weighting functions to compute the relative saliency of the current neighboring field, current block and current pixel.

Let us define $\rho_{Block}(I,J)$ and $\rho_{region}(i,j)$ the relative saliency degree of the current block and of the current neighboring field as follow:

$$\rho_{Block}(I,J) = \frac{1}{S_{Global}} \left(\frac{1}{64} \sum_{i=1}^{8} \sum_{j=1}^{8} S_{MIX}(i,j) \right) \qquad (12)$$

$$\rho_{region}(i,j) = \frac{\overline{S_{Local}}}{S_{Global}} \qquad (13)$$

with

$$\overline{S_{Local}} = \frac{1}{k \times k} \sum_{i=1}^{k} \sum_{j=1}^{k} S_{MIX}(i,j) \qquad (14)$$

$$\overline{S_{Global}} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} S_{MIX}(i,j) \qquad (15)$$

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

$$\rho_{pixel\_average}(i,j) = \max \left\{ \frac{S_{MIX}(i,j)}{S_{Local}}, \frac{S_{MIX}(i,j)}{S_{GLobal}} \right\} \qquad (16)$$

$$\rho_{pixel\_max}(i,j) = \frac{S_{MIX}(i,j)}{S_{Max\_Local}} \qquad (17)$$

with

$$S_{Max\_Local} = \max \{ S_{MIX}(i,j) | i \le k, j \le k \} \qquad (18)$$

Next, to decrease the influence of non-salient regions, we propose to compute a weighted saliency map $w_s(i,j)$ defined as follow:

$$w_s(i,j) = \left\{ \rho_{Block}(i,j) \ \middle| \ \rho_{region}(i,j) > T_3 \right\} \qquad (19)$$

Where T_3 is an experimental threshold that can be adjusted accordingly to the average of $\rho_{region}(i,j)$ values.

As illustration, let us consider for example the reference image 'I18' (see Figure 1 (a)). Meanwhile the corresponding mixed saliency map is shown in Figure 9 (a), the weighted saliency map w_s is shown in Figure 9 (b). Comparing Figures 9 (a) and (b), we can see that w_s better 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 which are also salient with regards to the global image.
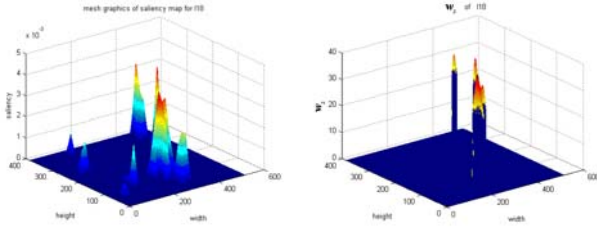
(a) surface plot of mixed saliency map.   (b) surface plot of W_s.

Figure 9. Surface plot of mixed saliency map and W_s metric.

For extreme images, in order to improve the efficiency of image quality metrics, we propose to weight image difference metrics by taking into account salient regions instead of salient pixels. Considering that human observers are unable to focus on several areas at the same time and that they assess the quality of an image firstly/mainly from the most salient areas, we propose to weight image difference metrics by the weighted saliency map w_s defined above. The corresponding improved PSNRHVS metric can be computed by the following pseudo code:

```
While (img_num)  {
    if (img_num ∈ Sₑ) {
    // for the current block of an image belonging to extreme set
       for i=1:8
          for j=1:8 {
             if (ϕI,J is false)
```
$$\delta_{PSNR\_HVS\_S}(i,j) = \delta(i,j) \cdot \left(\frac{CSF_{Coef}(i,j)}{CSF_{Coef}(i,j)+1} \right);$$
```
             else
                if (( ρpixel_max > T₄) & ( ρpixel_average > T₅))
```
$$\delta_{PSNRHVSS}(i,j) = \delta_{PSNRHVS}(i,j) \cdot \omega_s(i,j);$$
```
                else
```
$$\delta_{PSNRHVSS}(i,j) = \delta_{PSNRHVS}(i,j);$$
```
                end
             end
          }
       }
    else   PSNRHVS_img_num;
    img_num = img_num - 1;
}
```

Where (i,j) is the position of a pixel in an 8x8 block.

The thresholds T_3, T_4, T_5 have been empirically defined to 15, 0.5 and 40 respectively from the TID2008 database. In our experiments, parameters T_3, T_4, T_5 were selected via an exhaustive process in a 3D 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 were next separated into m sub-scales in order to get a data gird of m^3 grid points. Then we have chosen in the grid points set the best grid point (i.e. the values T_3, T_4, T_5) with the highest performance in regards to the dataset considered (cf. [20]).

5. EXPERIMENT RESULTS ANALYSIS

In this paper, the TID2008 database has been used to test our image quality assessment model. TID2008 is the largest database of distorted images intended for verification of full reference quality metrics [18]. The TID2008 database contains many distorted images, types of distortion and MOS computed from numerous subjective experiments. The TID2008 database contains 1700 distorted images (25 reference images x 17 types of distortions x 4 levels of distortions). The MOS (Mean Opinion Score) 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 is (0 - minimal, 9 - maximal, MSE of each score is 0.019), the higher the visual quality of the images is.

5.1. Experiment Results from saliency map and extreme images classification

Thanks to extreme images detection and region saliency map computation, the Spearman coefficients of PSNRHVS and PSNRHVSM are remarkably enhanced on all subsets. The two new metrics that we have proposed are called PSNRHVS-E and PSNRHVSM-E. In Tables 2 and 3, we can see that both PSNRHVS-E and PSNRHVSM-E metrics improve the image quality assessment. Spearman and Kendall correlation coefficients are enhanced by 20.2% and 15.97% on the full subset in regards to PSNRHVS.

Table 2. Spearman correlation coefficients.

Model / Subset	PSNR	PSNR HVS	PSNR HVS-E	Δ (%)	PSNR HVSM	PSNR HVSM-E	Δ (%)
Noise	0.704	0.917	0.92	0.32	0.918	0.923	0.54
Noise2	0.612	0.933	0.937	0.42	0.93	0.932	0.21
Safe	0.689	0.932	0.937	0.53	0.936	0.94	0.42
Hard	0.697	0.791	0.845	6.82	0.783	0.831	6.13
Simple	0.799	0.939	0.946	0.74	0.942	0.946	0.42
Exotic	0.248	0.275	0.6	118.1	0.274	0.56	104.3
Exotic2	0.308	0.324	0.621	91.66	0.287	0.548	90.94
Full	0.525	0.594	0.714	20.20	0.559	0.659	17.88

For PSNRHVSM, the Spearman and Kendall correlation coefficients are enhanced by 17.89% and 13.14% on the full subset. For the Exotic subset and Exotic2 subset, the performance is enhanced more than 85.23%. The performance with the original PSNRHVS on Noise, Noise2 and Simple subsets etc. are quite identical, meanwhile the performance on Exotic and Exotic2 subsets is much raised.

Table 3. Kendall correlation coefficients.

Model \ Subset	PSNR	PSNR HVS	PSNR HVS-E	Δ (%)	PSNR HVSM	PSNR HVSM-E	Δ (%)
Noise	0.501	0.751	0.752	0.13	0.752	0.757	0.66
Noise2	0.424	0.78	0.783	0.384	0.771	0.774	0.38
Safe	0.486	0.772	0.775	0.38	0.778	0.782	0.51
Hard	0.516	0.614	0.659	7.32	0.606	0.643	6.10
Simple	0.598	0.785	0.79	0.63	0.789	0.794	0.63
Exotic	0.178	0.195	0.424	117.4	0.194	0.391	101.5
Exotic 2	0.225	0.238	0.448	88.2	0.21	0.389	85.23
Full	0.369	0.476	0.552	15.96	0.449	0.508	13.14

We have computed also Spearman and Kendall correlation coefficients on extreme images. We can see in Table 4 and 5 that these coefficients have been even enhanced by 31.5% and 25.07% for PSNRHVSM.

Table 4. Spearman Correlation on extreme points.

	Extreme subset original	Extreme subset new	Δ (%)
PSNRHVS	0.3896	0.4353	11.73
PSNRHVSM	0.2586	0.34	31.48

Table 5. Kendall Correlation on extreme points.

	Extreme subset original	Extreme subset new	Δ (%)
PSNRHVS	0.2698	0.2982	10.53
PSNRHVSM	0.1843	0.2305	25.07

As we can see on Figure 10, the extreme images detected in the scatter plot of modified PSNRHVS or PSNRHVSM, (underlined by blue points) are more clustered than the original extreme images detected (underlined by red points). Moreover more images belong to the "ideal" region shown in the Figure 2 (a).
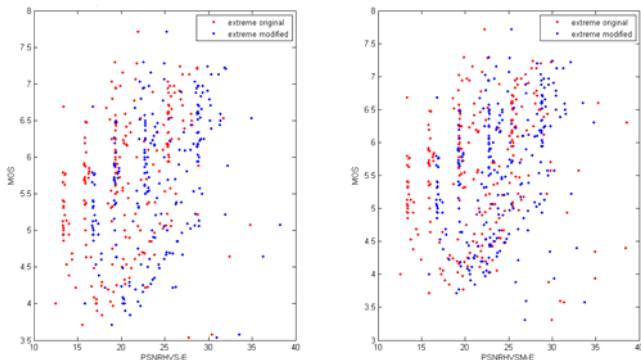

Figure 10. Scatter plot for extreme images.

5.2. Experimental Results from saliency without extreme images classification

To analyze the influence of extreme images classification on image quality assessment, we have also computed the correlation coefficients of the PSNRHVS and PSNRHVSM metrics enhanced by only region saliency information, without taking into account of extreme images classification. The corresponding Spearman and Kendall correlation coefficients are shown in Tables 6 and 7.

Table 6. Spearman correlation coefficients on all subsets.

Model \ Subset	PSNR	PSNR HVS	PSNR HVS_S	Δ(%)	PSNR HVSM	PSNR HVSM_S	Δ(%)
Noise	0.704	0.917	0.914	-0.32	0.918	0.92	0.21
Noise2	0.612	0.933	0.863	-7.5	0.93	0.871	-6.34
Safe	0.689	0.932	0.92	-1.28	0.936	0.924	-1.28
Hard	0.697	0.791	0.814	2.90	0.783	0.816	4.21
Simple	0.799	0.939	0.933	-0.63	0.942	0.935	-0.74
Exotic	0.248	0.275	0.465	69.09	0.274	0.442	61.31
Exotic2	0.308	0.324	0.377	16.36	0.287	0.331	15.33
Full	0.525	0.594	0.622	4.71	0.559	0.595	6.44

Table 7. Kendall correlation coefficients.

Model \ Subset	PSNR	PSNR HVS	PSNR HVS_S	Δ(%)	PSNR HVSM	PSNR HVSM_S	Δ(%)
Noise	0.501	0.751	0.745	-0.79	0.752	0.752	0
Noise2	0.424	0.78	0.68	-12.82	0.771	0.689	-10.63
Safe	0.486	0.772	0.752	-2.59	0.778	0.757	-2.69
Hard	0.516	0.614	0.634	3.25	0.606	0.637	5.11
Simple	0.598	0.785	0.773	-1.52	0.789	0.777	-1.52
Exotic	0.178	0.195	0.313	60.51	0.194	0.294	51.55
Exotic2	0.225	0.238	0.254	6.72	0.21	0.22	4.76
Full	0.369	0.476	0.472	-0.8	0.449	0.455	1.34

The corresponding metrics based on weighted saliency map are called PSNRHVS-S and PSNRHVSM-S. In Tables 6 and 7, $\Delta(\%)$ describes the enhancement of performance in regards to PSNRHVS and PSNRHVSM metrics. Considering Spearman correlation coefficients, PSNRHVS and PSNRHVSM perform well on Noise, Noise2, Safe, Hard and Simple subsets of TID2008. But they don't perform well on Exotic and Exotic2 subsets. With the weighted saliency map, the Spearman coefficients of PSNRHVS and PSNRHVSM are enhanced on full subsets but there is a little reduction on Noise2 subset. On Exotic and Exotic2 distorted subsets, the performance of the modified PSNRHVS and PSNRHVSM based on saliency map are remarkably enhanced. For PSNRHVS, the Spearman correlation coefficient on Exotic is enhanced nearly 69.1% and the Kendall correlation coefficient nearly 60.5%. Exotic and exotic2 are two subsets with contrast changes, i.e. shift distortions. The problem with PSNRHVS and PSNRHVSM metrics is that they are only based on intensity information while the improved metrics that we propose, based on salient information, includes color contrast, intensity and other information.

Next, we have compared the correlation coefficients computed for all images without extreme images classification with the results given in section 5.1. This comparison shows that the saliency information enhances

the performance of PSNRHVS and PSNRHVSM metrics for all images but this enhancement is not as significant as for extreme images.

6. CONCLUSION

A novel image quality assessing method based on saliency map and extreme images classification has been proposed in this paper. Thanks to the enhancement of the performance of image quality metrics for extreme images, the performance of the image quality assessing model proposed has been enhanced remarkably. Experiments done with the TID2008 database have shown that the new metrics proposed are effective. The next step of our study will be focused on the assumption that the saliency is scalable.

7. REFERENCES

[1] 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.

[2] Z. Wang, and A. Bovik, "A universal image quality index", IEEE Signal Proc. Letters, vol. 9, pp. 81-84, 2002.

[3] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity", *IEEE Transactions on Image Proc.*, vol. 13, issue 4, pp. 600-612, April, 2004.

[4] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," IEEE Asilomar Conference on Signals, Systems and Computers, Nov. 2003.

[5] B. Kolpatzik, and C. Bouman, "Optimized Error Diffusion for High Quality Image Display", *Journal Electronic Imaging*, pp. 277-292, 1992.

[6] B. W. Kolpatzik, and C. A. Bouman, "Optimized Universal Color Palette Design for Error Diffusion", *Journal Electronic Imaging*, vol. 4, pp. 131-143, 1995.

[7] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "New full-reference quality metrics based on HVS", Proceedings of the 2nd Int. Workshop on Video Proc. and Quality Metrics, Scottsdale, USA, 2006.

[8] 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 3rd Int. Workshop on Video Proc. and Quality Metrics, USA, 2007.

[9] H.R. Sheikh, and A.C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, Vol.15, no.2, 2006, pp. 430-444.

[10] N. Damera-Venkata, T. Kite, W. Geisler, B. Evans, and A. Bovik, "Image Quality Assessment Based on a Degradation Model", *IEEE Trans. on Image Processing*, Vol. 9, pp. 636-650, 2000.

[11] T. Mitsa, and K. Varkur, "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms", Proc. of IEEE, ICASSP, pp. 301-304, 1993.

[12] H.R. Sheikh, A.C. Bovik and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics", *IEEE Transactions on Image Processing*, vol.14 (12), pp. 2117-2128, 2005.

[13] D.M. Chandler, and S.S. Hemami, "VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images", *IEEE Transactions on Image Processing*, Vol. 16 (9), pp. 2284-2298, 2007.

[14] 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/.

[15] N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, and V. Lukin "Metrics performance comparison for color image database", Proc. of 5th Int. workshop on video processing and quality metrics for consumer electronics, Scottsdale, Arizona, USA, pp.14-16, 2009.

[16] Q. Ma, and L. Zhang, "Saliency-Based Image Quality Assessment Criterion", ICIC 2008, LNCS 5226, pp.1124–1133, 2008.

[17] X. Feng, T. Liu, D. Yang, and Y. Wang, "Saliency Based Objective Quality Assessment of Decoded Video Affected By Packet Losses", Proc. of IEEE, ICIP, pp.2560-2563, 2008.

[18] TID2008 image database, page: http://www.ponomarenko.info/tid2008.htm.

[19] C.C. Chang, and C.J. Lin, "LIBSVM : a library for support vector machines", 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[20] Y. Tong, H. Konik, F.A. Cheikh, A. Tremeau. "A Visual Saliency Model for Perception-based Video Surveillance", Journal of Imaging and Science, May 2010.

[21] L. Itti, C. Koch and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis" IEEE Trans. PAMI., vol. 20, No.11, pp.1254-1259, 1998.

[22] W. D., Koch, "Modeling Attention to Salient Proto-objects", Neural Networks 19, pp 1395–1407, 2006.

[23] M. Cerf, E.P. Frady, C. Koch, Faces and text attract gaze independent of the task: Experimental data and computer model. Journal of Vision, 9(12), 2009, pp 1-15.

[24] M. Cerf, J. Harel, W. Einhäuser and C. Koch. "Predicting human gaze using low-level saliency combined with face detection". In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems* (vol. 20), Cambridge, MA: MIT Press.