AN IMAGE QUALITY ASSESSMENT METHOD BASED ON PERCEPTION OF STRUCTURAL INFORMATION

Mathieu CARNEC, Patrick LE CALLET and Dominique BARBA

Ecole Polytechnique de l'Universite de Nantes, FRANCE
{Mathieu.Carnec, Patrick.LeCallet, Dominique.Barba}@polytech.univ-nantes.fr

ABSTRACT

This paper presents a new method to evaluate the quality of distorted images. This method is based on a comparison between the structural information extracted from the distorted image and from the original image. The interest of our method is that it uses reduced references containing perceptual structural information. First, a quick overview of image quality evaluation methods is given. Then the implementation of our Human Visual System (HVS) model is detailed. At last, results are given for quality evaluation of JPEG and JPEG2000 coded images. They show that our method provides results which are highly correlated with human judgments (Mean Opinion Score). This method has been implemented in an application available on the Internet.

1. INTRODUCTION

Image quality assessment is an active domain of investigation in the image community. The aim of these studies is to find automatic methods providing computed quality scores which are correlated with quality scores given by human observers. These methods can supply precise control of image compression schemes and more generally a relevant human-related Quality of Service (QoS) for images transmission. In literature, methods can be classified in three categories. The first category gathers full reference methods in which the original image and the distorted image are required to compute the quality score. Basically these methods often operate on an error image and can include simple or complex Human Visual System (HVS) models. In the second category, reduced reference methods need a description of the original image and the distorted image. The reduced reference is designed to contain relevant information with respect to the image quality to compute. In the third and last category, no reference methods only use the distorted image to compute the quality score. In a broadcasting purpose, only reduced reference (RR) and no reference methods are suitable. Since no reference methods focus on artifacts due to a given transmission scheme, they depend on the service (compression type, transmission channel, etc...). Thus, we are interested in RR methods since the reduced reference can be coded and embedded in the bitstream and therefore constitutes a practical approach to quality evaluation. This paper presents a new method to measure images quality using a reduced reference. This reduced reference represents the structural information of the image in a perceptual space. The whole method is based on the implementation of a rather elaborated model of the Human Visual System (HVS) concerning the stages needed to project the image into a perceptual space but also the type of structural information which is extracted. This model has been established from descriptions given in papers from neurophysiology literature [1] [2] [3]. It contains the two visual pathways running from the eye to the visual parts of the human brain: the ventral pathway (ending in the V4 area) and the dorsal pathway (running to the V5 area). The HVS' superior colliculi are also included. The model is detailed in [4].

The ventral pathway and the superior colliculi are implemented. On the opposite, we haven't implemented the dorsal pathway as it processes only moving stimuli and, in this paper, we only deal with still images. The basic idea used in our method is that image quality is mainly based on annoyance. This one is directly related to the difficulty the HVS has to extract patterns and recognize them when viewing distorted images. By simulating the HVS, we want to extract the information which is relevant for recognition. The full process can be decomposed in two stages. During the first stage, perceptual representation are built for the original and the distorted images. Then during the second stage, representations are compared in order to compute a quality score.

2. PERCEPTUAL REPRESENTATION OF IMAGES

The whole process is shown on figure 1. At first, RGB images are converted into a perceptual space. Then perceptual features are extracted.

0-7803-7750-8/03/$17.00 ©2003 IEEE

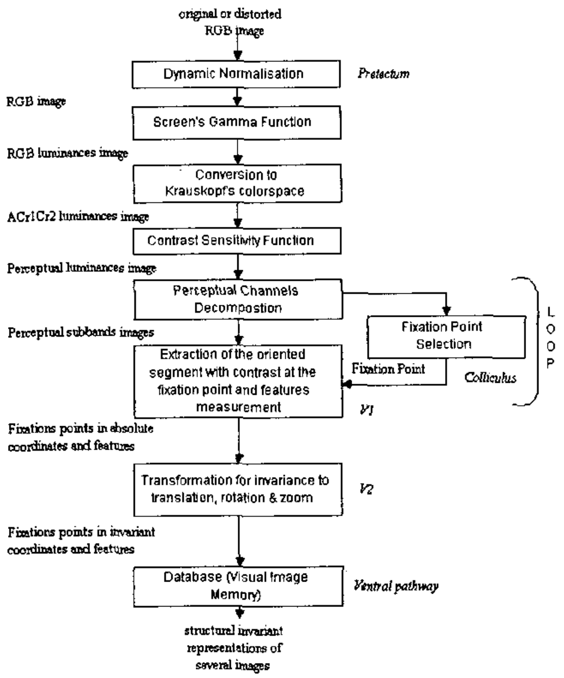

Fig. 1. Implementation Diagram

2.1. Perceptual Representation

The perceptual representation results from two main stages. The first stage is a set of low level processings:
- color normalization of the image range emulating the light adaptation achieved by the pretectum in the HVS,
- transformation of RGB values into luminances using the monitor gamma function,
- projection onto a perceptual colorspace (Krauskopf's colorspace ACr1Cr2).

The second stage performs a perceptual subband decomposition. Firstly, a Contrast Sensitivity Function (CSF) is used to weight the image spectrum in a psychovisual way. Then a psychovisual partitioning of the resulting spectrum is applied, both in spatial frequency and orientation.

2.2. Structural Features Extraction

We construct a reduced representation of the image from its perceptual representation. This reduced representation is a set of structural features extracted around fixation points. Fixation points are selected in the perceptual subbands. Fixation points are located on a spiral covering the whole image, as detailed in [5] and shown on figure 2. Structural features are oriented segments with contrast at different resolutions, extracted by a stick growing algorithm as described on figure 2. Basically, a stick growing algorithm tries to build a centered segment in all the directions and keeps the longest one. Each segment end must be located on a value which is at least 50% of the value at the center of the segment (which is the fixation point). This processing is faster than using filer banks for example. These structural features correspond to the information extracted by the V1 area of the HVS. A set of parameters is used to describe features: orientation, length, width and contrast.
At last, features are transformed to create a structural image representation which is invariant to zoom, translation and rotation. This is done for the original image and for the image to recognize. The image to recognize is a distorted (or not) version of the original image.
Representations are stored in a database called the visual image memory. Distortions may be due to lossy image coding schemes for example but also to other image processing systems. Since no model of the degradation is used in this method, this is a "blind" image quality evaluation.

Fig. 2. Fixation points locations (top) and feature extraction in a subband (bottom) on the LENA image

3. COMPARISON BETWEEN STRUCTURAL REPRESENTATIONS

The structural representation of the original image is compared to the structural representation of the image to recognize which is a distorted (or not) version of the original image, using several similarity measures. In fact, the similarity between two images is the mean value of local similarities since we compare features at the same locations (same fixation points) in these two images. The comparison process is described in figure 3.

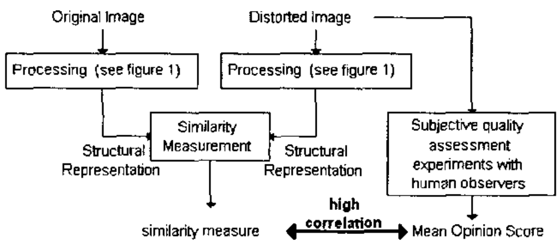

Fig. 3. Similarity Computation

For each fixation point i=1..N (N=352 for these results, which means that the spiral contains 22 circles of 16 fixation points each) and for each feature parameter, a local similarity (LS_i) is computed. Basically, local similarities are computed from normalized differences between feature parameters located at the fixation point i in the original image ($feature1_i$) and in the distorted image ($feature2_i$). Similarity measures can combine several features parameters (like the length of the feature for example) and differ by the way local similarities are combined (geometric or arithmetic average). Similarity metrics presented in this paper are functions using only one or several parameters among the following features parameters: orientation (O_i), length (L_i), width (W_i), and contrast (C_i) as described below. The contrast value is the value in the subband, at the fixation point.

$$LS(O_i) = max(0, 1 - \frac{|O_{orig,i} - O_{dist,i}| \, mod \frac{\pi}{2}}{\frac{\pi}{2}}) \quad (1)$$

$$LS(L_i) = max(0, 1 - \frac{|L_{orig,i} - L_{dist,i}|}{L_{orig,i}}) \quad (2)$$

$$LS(W_i) = max(0, 1 - \frac{|W_{orig,i} - W_{dist,i}|}{W_{orig,i}}) \quad (3)$$

$$LS(C_i) = max(0, 1 - \frac{|C_{orig,i} - C_{dist,i}|}{C_{orig,i}}) \quad (4)$$

We have tested several types of similarity measurements and compared them.

4. RESULTS

To evaluate the efficiency of a quality measurement method, we compare the quality scores it outputs (similarity) with the quality scores given by human observers during quality experiments. During these experiments, observers were asked to score the quality of the images which were presented under normalized conditions. Quality scores were taken from a scale with 5 grades ranging from 1 which is the lowest quality to 5 which is the highest. The mean value of the scores for a given image is called the Mean Opinion Score (MOS) of this image. We use the MOS data collected from previous tests made in our lab.

We used several similarity measures like the following ones:

$$S0 = \frac{\sum_{i=1}^{N} LS_i(C_i)}{N} \quad (5)$$

$$S1 = \frac{\sum_{i=1}^{N} LS_i(0_i) * LS_i(L_i) * LS_i(W_i) * LS_i(C_i)}{N} \quad (6)$$

$$S2 = \frac{\sum_{i=1}^{N} LS_i(0_i) + LS_i(L_i) + LS_i(W_i) + LS_i(C_i)}{4 \cdot N} \quad (7)$$

$$S3 = \frac{\sum_{i=1}^{N} LS_i(0_i) * LS_i(L_i) * LS_i(W_i)}{N} \quad (8)$$

$$S4 = \frac{\sum_{i=1}^{N} LS_i(0_i) + LS_i(L_i) + LS_i(W_i)}{3 \cdot N} \quad (9)$$

Unlike S0 which only uses contrast, S1, S2, S3 and S4 use several features parameters combined with a geometrical average (S1, S3) or an arithmetic average (S2, S4). Results are shown in tables 1 and 2 for two sets of 45 JPEG and 45 JPEG2000 compressed images (at different compression rates). The VQEG has defined some metrics in order to qualify the performance of an image quality criterion in terms of accuracy (RMSE weighted by the confidence on the MOS; RMSEw), monotonicity (rank order correlation coefficient: RCC), consistency (outlier ratio: OR) and agreement (Kappa coefficient). These metrics are given in [6]. Correlation coefficients (CC) are also mentioned in the tables. A correlation coefficient of 0.966 has been found with the set of JPEG compressed images and of 0.951 with the set of JPEG2000 images. The 45 images of each set are well-known images of the image processing research community, like "lena", "boats", etc... None of these 45 images was used to adjust the parameters of this method (so the method is not dependent on these images).

Similarity results (for metric S3) for JPEG images are shown on figure 4.

	RMSEw	RCC	OR	kappa	CC
S0	0.56	0.952	0	0.70	0.956
S1	0.52	0.961	0	0.65	0.960
S2	0.52	0.950	0	0.70	0.961
S3	0.49	0.957	0	0.70	0.966
S4	0.54	0.945	0	0.65	0.956

Table 1. Results for 45 JPEG images

	RMSEw	RCC	OR	kappa	CC
S0	0.85	0.943	0.04	0.52	0.934
S1	0.81	0.951	0.04	0.55	0.938
S2	0.87	0.954	0.02	0.60	0.939
S3	0.80	0.962	0.02	0.60	0.951
S4	0.92	0.955	0.02	0.55	0.935

Table 2. Results for 45 JPEG2000 images

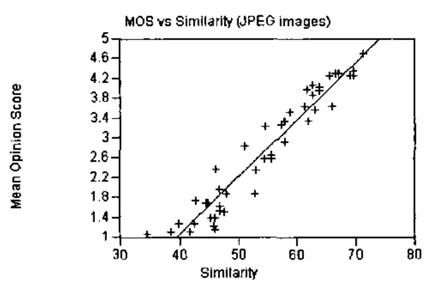

Fig. 4. Results for similarity S3 with 45 JPEG images

The different similarity metrics have quite equivalent results but S3 has proved to be the most correlated with Mean Opinion Scores. It indicates that using structural information to evaluate image quality is more relevant than using only contrast information (S0) or contrast information combined with structural information (S1 or S2). It also shows the severity of the human judgment is best reproduced using a geometrical average than a classical arithmetic average. We found equivalent results on a larger database (233 JPEG images from the LIVE database [7]): CC=0.961 for metric S3.

5. CONCLUSION

We have presented a new quality assessment method based on structural information. Results have shown to be highly correlated with Mean Opinion Scores, especially metric S3.

To compute the quality of an image, the method only needs to know the invariant representation of the original image. The number of data in the reduced reference used for these results is 1056 (real numbers) and can be lowered a lot: on the 45 JPEG images, we found equivalent performances with N=240 (CC=0.964) and slighly lower results with N=160 (CC=0.937). So we can imagine the reduced reference would be transmitted with the compressed image.

This new quality assessment method is implemented in two Windows applications. "Smart Compress" allows users to compress images into JPEG format by choosing the desired quality of the compressed images instead of choosing a meaningless compression ratio or quality factor. "Quality Assesser" evaluates the quality of JPEG or JPEG2000 images. MSE, RMSE and PSNR methods, all having lower performances, are implemented too for results comparisons purposes. These applications may be freely downloaded at *http://www.dcapplications.t2u.com.*

6. REFERENCES

[1] Daniel A. Pollen, "On the neural correlates of visual perception," *Cerebral Cortex, Vol. 9, No. 1*, pp. 4–19, January 1999.

[2] W.H. Merigan, T.A. Nealey, and H.R. Maunsell, "Lesions in macaques affect both single and multiple-viewpoint shape discriminations," *Vis Neurosci 15*, pp. 359–369, 1998.

[3] A. Cowey, "Projection of the retina on to striate and prestriate cortex in the squirrel monkey," *Saimiri Sciureus, Journal of Neurophysiology, 27*, pp. 366–393, 1964.

[4] Mathieu Carnec and Dominique Barba, "Pattern recognition based on the perception and the behavior of the human visual system," *In Proc. of Digital Signal Processing, Santorini, Greece*, 2002.

[5] Mathieu Carnec and Dominique Barba, "Simulating the human visual system: Towards objective measurement of visual annoyance," *In Proc. of Systems, Man and Cybernetics, Hammamet, Tunisia*, 2002.

[6] A. M. Rohaly, P. Corriveau, J.Libert, A. Webster, V. Baroncini, J. Beerends, J. L. Blin, L. Contin, T. Hamade, D. Harrison, A. Hekstra, J. Lubin, Y. Nishida, R. Nishihara, J. Pearson, A. F. Pessoa, N. Pickford, A. Schertz, M. Visca, A. Watson, and S. Winkler, "Video quality experts group : Current results and future directions," *vcip*, vol. 3, pp. 742–753, 2000.

[7] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database," http:// live.ece.utexas.edu/research/quality.