A MULTI-CHANNEL OBJECTIVE MODEL FOR THE FULL REFERENCE ASSESSMENT OF COLOR PICTURES

Francesca De Simone, Michael Ansorge and Touradj Ebrahimi

Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

ABSTRACT

This paper presents a new approach for the design of a full reference objective quality metric for the assessment of color pictures. Our goal is to build a multi-channel metric based on the perceptual weighting of single-channel metrics. A psycho-visual experiment is thus designed in order to determine the values of the weighting factors. This metric is expected to provide a new useful tool for the quality assessment of compressed pictures in the framework of codec performance evaluation.

1. INTRODUCTION

The compression efficiency of an image coding algorithm expresses its ability to maximize the visual quality of the compressed data while minimizing the number of bits used to store the data, for a range of compression rates. Being the human subjects the end users of the digital data, subjective tests can be performed, where a significant sample of human subjects is asked to rate the quality of the processed material. Since these tests are time consuming and expensive, usually objective metrics are used in order to assess the quality of the compressed images. These metrics are called Full-Reference (FR) quality metrics, because they assume as input both the original image (i.e. reference) and the compression version of it. A substantial effort has been recently deployed by the research community to design objective visual quality metrics which achieve a good correlation with the subjective quality evaluation. Nevertheless, most of the well-known and widely used FR quality metrics take into account only the luminance channel of the picture under analysis. Hence, in the state of practice, the quality performance evaluation and optimization of full color image algorithms are usually done by mean of methods which are applied on the luminance component only of the signals. Due to the evident influence of the color information on the human perception of the quality of visual data, this approach is of course limiting *a priori* the correlation that can be met with such a single-channel metric compared to the subjective judgment.

This paper presents a general model for designing a metric for the assessment of color pictures, by exploiting data collected from human subjects by means of a properly designed psycho-visual experiment. The idea is to perform an experiment with human subjects so as to understand "how the color information affects the overall assessment of the distorted image with respect to the perceived quality on the luminance-only version of the same distorted image". The test methodology is described in Section 2. The task is then to fit an objective model to the subjective data resulting from this test, as described in Section 3.. We refer to this objective model as "multi-channel metric", since it consists of the weighted average of single-channel quality measures on the luminance channel and the two chrominance channels of the picture. Conclusions are presented in Section 4.

2. PSYCHO-VISUAL TEST METHODOLOGY

In our experiment the test material is presented to the subjects according to a slightly modified version of the Double Stimulus Continuous Quality Scale (DSCQS) method [1]. Pairs of pictures are shown, with the reference picture on one side of the screen and a distorted version on the other side. The assessor is asked to judge the quality of the distorted picture with respect to the reference, and to rate it by choosing a rate on a continuous scale in the range going from 0 (Very bad quality) to 100 (Excellent quality). This test is performed by considering pairs where the true color images are shown (i.e. reference image and distorted version in the same screen); and the corresponding pairs where just the luminance components of each true color image previously considered are displayed (i.e. luminance component of the reference shown together with the luminance component of the distorted image). The luminance-only and true color pairs are randomly mixed while assuring that luminance-only and full color displays are always consecutive. The test room conditions, as well as the number of subjects, the test duration, the training session details and the processing of the subjective data are assumed to be standard compliant [1].

By computing for each test condition the mean score over the entire set of subjects' rates, two sets of Mean Opinion Score (MOS) results are obtained: one set of subjective scores is referred to the quality evaluation of luminance only stimulus and we will call it MOSluma; the other set is related to the quality assessment of the full color stimulus and we call it MOScolor. The first interesting analysis made at this stage is related to the influence of the color information on the overall quality assessment by the human subject: the same distorted picture can be rated in different ways if the subject sees the true color image or just the luminance component of the true color image.

2.1. Dataset selection

For choosing the input data to be used in the psycho-visual experiment, as a starting step we restrict our field of investigation by considering a dataset of pictures including only natural images and the distortions introduced by three different JPEG compression algorithms specified hereafter. Five different contents, having different spatial and color features, have been selected for building our test set. The selected pictures are 8 bits per channel high resolution pictures chosen from the dataset established by Microsoft and T. Richter [2]. The test pictures are produced by 4:4:4 coding using i) JPEG with conventional visually optimized quantization matrix, ii) JPEG 2000 with frequency weighting of quantization steps, and iii) the new coding algorithm recently proposed by Microsoft and currently under evaluation by the JPEG standardization body under the name of JPEG XR [3]. Five levels of quality of each compressed data have been considered. Our dataset is thus composed of 5X3X5=75 test pictures, corresponding to 5 different original contents, 3 different coding techniques and 5 different samples in the perceivable quality difference range of interest. Finally, since we are using very high resolution input data, the subjective data are collected by presenting just a selected representative crop of each picture in order to fit the data in the native resolution of the screen (e.g. 1920x1440 pixels screen resolution).

3. SUBJECTIVE TO OBJECTIVE DATA FITTING METHODOLOGY: A BASIC APPROACH

The results of the subjective experiment described in the previous section can be used in order to extend any generic mono-channel metrics to the multi-channel case. In particular, to illustrate the proposed approach, we will consider, as an example, the application of our method to the widely used Peak Signal to Noise Ratio (PSNR) metric. Considering the representation of an image in the Y'CbCr color space [4], the PSNR metric can be applied on each color channel by computing the PSNR index on the luminance (PSNRy) and to the two chrominance components (PSNRcb and PSNRcr). A multi-channel metric can thus be built by simply considering the weighted average of the PSNR indexes computed on the single channels (i.e. multi-channel PSNR= w1PSNRy + w2PSNRcb + w3PSNRcr). The values of the weighting factors w1, w2 and w3 are defined by maximizing the correlation between the multi-channel model and the subjective results collected on the full color inputs (MOScolor). In particular, we assume as "reference values" the values of the Pearson correlation coefficient, the Spearman rank order correlation coefficient, and the outliers ratio evaluated between the MOSluma and the PSNRy. These three measures are the standard tools used to compare the performance of an objective metric to the subjective scores. They provide an estimate of the accuracy, monotonicity and consistency of the objective metric under analysis [1]. A system of equations is thus defined, imposing that the values of these three indexes computed between the MOScolor and the multi-channel metric are the known values computed for the luma only objective and subjective data. By solving the system of equations, the values of the weighting factors are defined.
Further conditions could be that the sum of the values of the three weighting factors is equal to 1 and that the weights of the two chrominance channels are the same. The investigation of more sophisticated conditions is also currently under progress.

4. CONCLUSIONS

The final goal of the proposed approach is to design a general method for the extension of a generic mono-channel metric to a multi-channel case, in order to achieve a better correlation with the end user quality judgment compared to the objective metric applied on the luminance channel only. One of the main advantage of our approach is that it allows the usage of well-know single-channel metrics (i.e. PSNR, SSIM) in the multi-channel scenario, thus providing a new tool to researchers who are already applying these metrics for the luminance-only objective quality assessment, by simply suggesting the right weights to be used for the weighting of the three channels.
The psycho-visual experiment described in this paper is currently running, so that the analysis of the results and the proposition of more sophisticated fitting methodologies will be reported in future publications.

5. REFERENCES

[1] Recommendation ITU-R BT.500-11, *Methodology for the subjective assessment of the quality of television pictures*, Geneva, June 2002.

[2] Test dataset available on the JPEG committee sftp server, http://www.jpeg.org.

[3] S. Srinivasan, C. Tu, S.L. Regunathan, R.A. Rossi, G.J. Sullivan, *HD Photo: a new image coding technology for digital photography,* Proceedings of SPIE, Applications of Digital Image Processing XXX, vol. 6696, San Diego, CA USA, August 2007.

[4] Recommendation ITU-R BT.601-6, *Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratio*, Geneva, January 2007.