Color Quality Assessment of Different Implementations of JPEG2000

Adrian STOICA, Mohamed Chaker LARABI, Christine FERNANDEZ-MALOIGNE

IRCOM-SIC Laboratory, UMR-CNRS 6615, University of Poitiers
Bat. SP2MI, Téléport 2, 86962 – FUTUROSCOPE Chasseneuil Cedex, France
{stoica, larabi, maloigne}@sic.sp2mi.univ-poitiers.fr

Abstract

Colour images become abundant because of the low cost of the acquisition devices and the advent Internet. Starting from that, the definition of new algorithms for colour image processing is seen as extremely important. Among, these algorithms, the compression is of a high importance for applications such as archiving, transmission, etc. In addition, the availability of techniques allowing to compare faithfully the results quality is very necessary. The subjective evaluation of the colour image quality came into focus because of the "failure" of the objective criteria to offers good results in terms of human perception. In general, objective and subjective evaluation of compressed images can lead to significant different conclusions. This paper deals with the problem of quality assessment for the new compression standard JPEG2000. Its aim is to compare different implementation of the standard and to offer a point of view on the assessment of this kind of algorithms.

Keywords: JPEG2000, colour image quality, subjective assessments

1. Introduction

In recent years there has been a growing number of digital image and video application, which all imply displaying, storing and transmitting large volumes of image data. Obviously, image data compression is a must and various lossless or lossy compression have been developed, notably the JPEG-family standards.

The evaluation of the compression quality is a simple and straightforward task in the case of lossless compression: standard criteria (compression ratio, execution time, etc) can be used. In the case of lossy compression the main difficulty arises in describing the type and the amount of degradation induced in the reconstructed image. The need of evaluating image quality in a human-assessment-consistent way has led to several approaches in image quality evaluation. There are two major types of image quality criteria: objective criteria and subjective (human-judgement-based) criteria. Since subjective image quality measure exhibit some inherent drawbacks (the use of a normalized evaluation room, a large panel of human observers, etc), there has been a great deal of interest in developing quantitative measures, either in numerical or graphical form, that can be consistently used as a substitute [1], [2], [3].

The remainder of the paper is organized as follows: Section 2 presents an short overview of the JPEG2000 image compression standard, Section 3 presents the different JPEG2000 implementations used for this paper; Section 4 presents an overview of subjective quality assessments, Section 5 presents the experimental conditions and procedures used for this work; Section 6 presents the results of image quality evaluations with both objective and subjective methods for JPEG2000-compressed colour images based on four implementations: Jasper, JJ2000, Kakadu and our implementation (here named EIRE-codec). The paper ends with some conclusions.

2. JPEG2000 Overview

JPEG2000 is a new image-encoding standard that provides a feature set vital to many high end and emerging imaging applications [4]. JPEG2000 provides high compression with image quality superior to all existing standard encoding techniques. Following the International Organization for Standardization (ISO) meeting in New Orleans in December 2000, JPEG2000 Part 1 has been officially declared an international standard.

JPEG 2000 is based on the discrete wavelet transform (DWT), scalar quantization, context modelling, arithmetic coding and post-compression rate allocation. The DWT is dyadic and can be performed with either the reversible Le Gall (5,3) filter, which provides for lossless coding, or the non-reversible Daubechies (9,7) biorthogonal one, which provides for higher compression but does not do lossless. The quantizer follows an embedded dead-zone scalar approach and is independent for each sub-band. Each sub-band is divided into rectangular blocks (called code-blocks in JPEG 2000), typically 64x64, and entropy coded using context modelling and bit-plane arithmetic coding. The coded data is organized in so-called *layers*, which are quality levels, using the post-compression rate allocation and output to the code-stream in packets. The generated code-stream is parseable and can be resolution, layer (i.e. SNR), position or component progressive, or any combination thereof.

JPEG2000 also supports a number of functionalities [5], many of which are inherent from the algorithm itself. An example is the random access, which is possible because of the independent coding of the code-blocks and the packetized structure of the codestream. Other supported functionalities are Regions of Interest (ROI), error-resilience, random access, multi-component images, palletized colour, compressed domain lossless flipping and simple rotation, to mention a few.

Given the knowledge of the human visual system's behaviour, the next step was to figure out how to exploit these properties effectively in a practical compression system. Some compression systems may allow a more thorough exploitation of the properties of the HVS than others may. For example, JPEG2000 supports features such as visual progressive weightings (frequency and colour), neighbourhood masking, self-masking, point-wise extended masking, local light adaptation, eccentricity and temporal frequency. The classical JPEG support only two of those features: frequency weightings and colour weightings.

One common visual optimization strategy for compression is to make use of the contrast sensitivity function (CSF) that characterizes the varying sensitivity of the visual system to 2D spatial frequencies [12]. In general, human eyes are less sensitive to high frequency errors than to low frequency errors. The CSF can be used to determine the relative accuracies needed across different spatial frequencies, where the term "weight" is used to describe the desired proportional accuracy. To use the CSF, which is usually described in visual frequencies of cycles/degree (cpd), it must be mapped to the compression domain of digital frequencies such as cycle/pixel. A proper frequency and colour weighting can usually result in significant detail and texture preservation with no introduction of colour distortions. In general, frequency weighting is more effective for large viewing distance or high dpi printing. In fact, it can also be used to reduce the flicking artefacts of Motion JPEG2000.

3. Some JPEG2000 Implementations

3.1 Kakadu

Kakadu is a complete implementation of the JPEG2000 standard, Part 1, i.e., ISO/IEC 15444-1 [9]. This new image compression standard is substantially more complex than the existing JPEG standard, both from a computational and a conceptual perspective.

The Kakadu software framework provides a solid foundation for a range of commercial and non-commercial applications. By making a consistent and efficient implementation of the standard widely available for both academic and commercial applications, the widespread adoption of JPEG2000 is considerably encouraged. The Kakadu software has been written specifically to cover a variety of different types of applications.

3.2 JJ2000

JJ2000 has been developed in a joint effort between Canon Inc., Ecole Polytechnique Fédérale de Lausanne and Ericsson Inc. [10]. The source code for the software is made freely available from the project's website. JJ2000 is written in Java. For the purposes of hardware co-processing, this is a severe limitation. Being Java software, the code runs on a Java virtual machine. This introduces two key problems. Firstly, the software is slower to execute than an implementation that is native to a PC. More importantly, there is the issue of portability.

3.3 Jasper

The JasPer Project [11] is an open-source initiative to provide a free software-based reference implementation of the codec specified in the JPEG-2000 Part-1 standard (i.e., ISO/IEC 15444-1). This project was started as a collaborative effort between Image Power, Inc. and the University of British Columbia. Presently, the ongoing maintenance and development of the JasPer software is being coordinated by its principal author, Michael Adams, who is affiliated with the Digital Signal Processing Group (DSPG) and Department of Electrical and Computer Engineering at the University of Victoria.

4. Subjective Quality Evaluation

The problem that there does not exist an objective image quality measure that perfectly reflects the subjective impression of a human observer is well-known. The various objective image quality measures indicate only roughly the image quality and are not properly defined for colour images.

Subjective tests provide the foundations for building vision models. At the same time, they are the only true benchmark for evaluating the performance of the various perception-based image processing tools. Unfortunately, perceptual responses cannot be represented by exact figures; due to their inherent subjectivity, they can only be statistically described. Even in psychophysical threshold experiments, where the task of the observer is just to give a yes/no answer, there exists a significant variation between observers. In the evaluation of supra-threshold artefacts, these differences become even more pronounced, because the objectivity of artefacts strongly relates to the observers expectations and presumptions, as to the intended application. Previous observer experiences also lead to a different weighting of the artefacts [6], [8].

Subjective assessment of visual quality has been normalized in the ITU-R Rec.500 [7], which suggests standard viewing conditions, criteria for the selection of observers and test material, assessment procedures, as well as data analysis methods. While targeted at the subjective assessment of television pictures, most of it directly applies to still images as well. In particular, it describes the Double Stimulus Continuous Quality Scale (DSCQS) and the Double Stimulus Impairment Scale (DSIS).

In a DSCQS test, viewers are shown stimulus pairs consisting of a “reference” and a “test” stimulus, which are presented twice in an alternating manner, with the order of the two chosen randomly for each trial. Subjects are not informed which the reference is and which is the test stimulus. They rate each of the two pairs separately on a continuous quality scale ranging from “bad” to “excellent”. Analysis is based on the difference in rating for each pair, which is calculated from an equivalent numerical scale from 0 to 100. The DSCQS has been shown to work reliably even when the quality of test and reference stimuli are rather similar, because it is quite sensitive to small differences in quality.

In a DSIS test, the reference is always displayed before the test stimulus, and both are shown only once. Subjects rate the amount of impairment in the test stimulus according to a discrete five-levels scale ranging from “very annoying” to “imperceptible”. DSIS is the referred method when evaluating clearly visible impairments.

5. Experiments

Both objective and subjective tests were performed on 12 images from the Kodak database (768x512 pixels, 24 bpp RGB colour images). Each image was compressed at various bit-rates, ranging from 4 bpp to 0.0062 bpp (4 bpp, 3 bpp, 2 bpp, 1 bpp, 0.5 bpp, 0.375 bpp, 0.25 bpp, 0.187 bpp, 0.125 bpp, 0.062 bpp) and decompressed according to the JPEG 2000 standard. Four JPEG2000 compression algorithms were used, namely the Java implementation JJ2000 (version 5.1), the C++ implementation Kakadu (version 3.4), the C Jasper implementation and our proper one, here named EIRE-codec.

The subjective tests were performed with a panel of 30 observers, with different image processing backgrounds, which were evaluated for the visual acuity and a normal vision of colour using the Ishihara test. The observers were asked to evaluate the photographic samples of the degraded images, as displayed on a 24" Sony high-resolution computer monitor, with a Trinitron tube. The screen was calibrated using a Gretag Macbeth EYE-ONE monitor Mach 1.1 calibrator. The computer characteristics were: an AMD ATHLON XP 2100+ (1.73 GHz) processor, 1 Gb RAM, and a NVIDIA GeForce4 TI 4600 128Mo video card. This configuration allows a pixel to pixel image presentation on the screen, without image resizing or other internal interpolations. The experiments were made up in a special room, which was built according to the international standards ITU-R 500-10 and ISO 3664. The room walls were of neutral grey, in order to avoid the effect of flare. The room illumination was totally controlled variable artificial light, measured by a Minolta T-10M luminance-meter, and fixed at 62lux. The viewing distance was set at 90 cm, in order to allow a very good perception of image details.

5.1 Subjective assessment

As shown by figure 1, the observers were seated 90cm from the calibrated CRT in a darkened room in order to enhance the visualization quality.

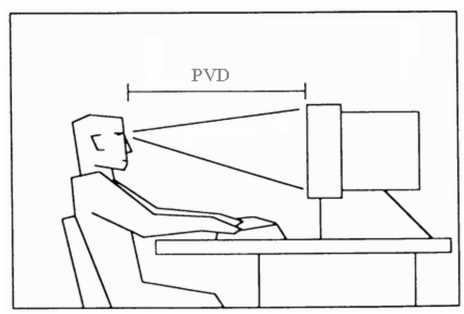

***Figure 1**. Psychophysical quality assessment protocol.*

Ordering tests

The aim of this test is to make a classification (from the best to the worst) of an image series with regard to the original image. For the needs of this psychophysical experimentation, we retained as device of study, a process that shows 9 images on the screen. In this device, the original image is placed in the centre and eight images to be studied are situated around it as showed in Figure 2.

***Figure 2**. Snapshot of an ordering test.*

Then, the human observer is asked to indicate by a mouse click, the image which he sees most qualitatively distant from the original image. This image is then masked and the same question is then asked and this, until the eight images are masked in the presentation device. This technique allows to have a classification of the images with regard to a reference one (often considered as the original image). Contrary to the classical techniques where the observer has to classify the images in increasing or lessening quality order, this technique presents the advantage to never put the observer in a complex choice position. Indeed, he just has to answer every time: "*here is the image that I don't appreciate*".

Forced choice tests

This test proposes to the human observer three images with the original in the centre and two images on both sides (Figure 3). Then, he has to choose the best of the two images situated on both sides of the original.

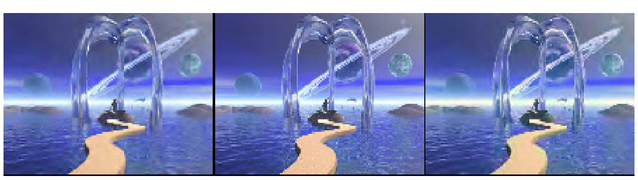

***Figure 3**. Forced-choice test.*

6. Results

In terms of subjective quality, the JPEG2000 implementations using the visual frequency weightings offer generally better results. Figure 4 presents the results of an image quality psycho-visual evaluation campaign, based on an ordering procedure, for 10 images from the Kodak database. There were evaluated four JPEG2000 implementations. Two codec were using visual weightings, the Kakadu and the EIRE-codec. The EIRE-codec was also used without weightings. Figure 4 present the Mean Option Score (MOS) and his Standard Deviation of the different JPEG2000 implementations.

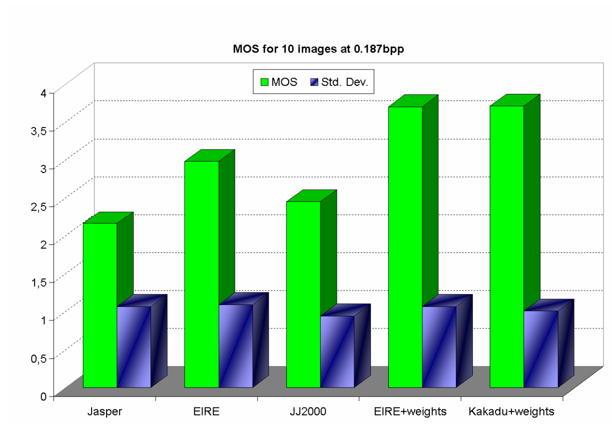

***Figure 4**. Mean Option Score and his Standard Deviation, of the different JPEG2000 implementations, for 10 images compressed at 0.187bpp.*

As the one can see from the results presented in Figure 4, according to the subjective quality evaluation, the implementations that offer the best results are the ones using the visual weightings. As one ca see, the Standard Deviation values are acceptably small, which means that the MOS values are correct. In spite the fact, that generally the implementations using weightings performs better, it is important to mention that they are some cases when no-weightings implementations were preferred. In Figure 5, we present the MOS for each of the 10 images and the 5 codec implementations.

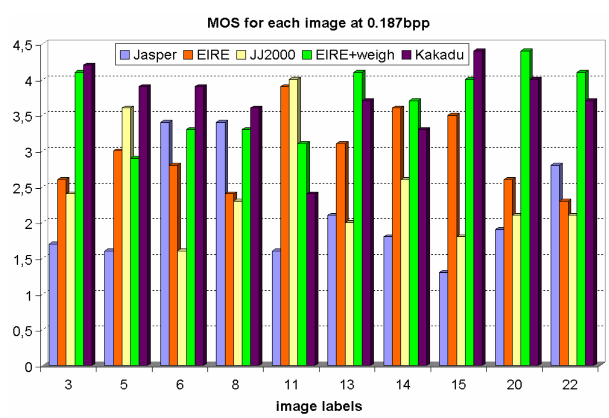

***Figure 5**. MOS of the different JPEG2000 implementations, for each of the 10 compressed images at 0.187bpp.*

From the Figure 5, one can easily identify the different cases where the no-weightings implementations were evaluated as offering better results. This fact proves that the compression quality improving by the use of the visual weightings, depends in certain cases, on the image content. For example in the case of image 11, the result resented in Figure 5 is justify by the fact that for the large quasiuniform regions of the image, the implementations using weightings produced a little blur effect, which is obviously detected by the human observes. It is well known that the human eye is very sensible of image contrast, but also different degradations in uniform regions. Therefore, in spite the fact that for the image 11, the implementations using visual weightings performs better from an image object details point of view, the human observers preferred the other implementations.

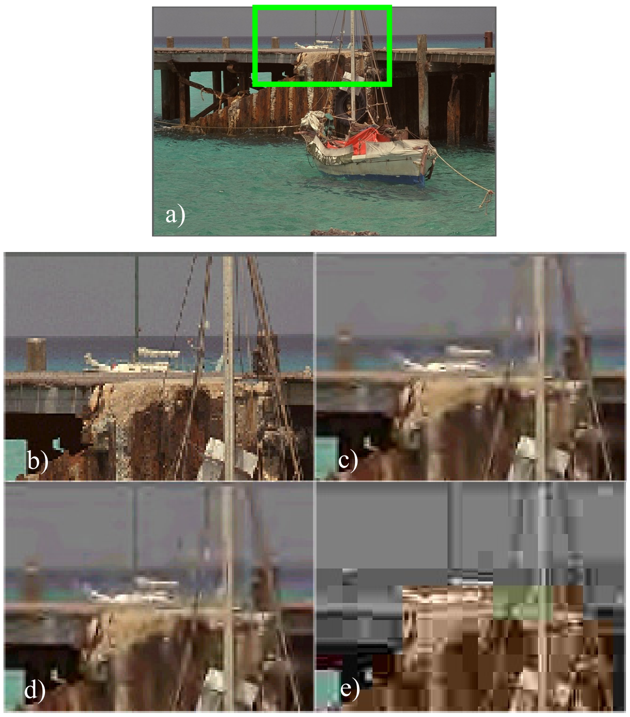

***Figure 6**. Sub-images (a) and (b) show the original image and a magnified sub-region. The sub-images (c), (d) and (e) presents the magnification of the compression result for JJ2000 (c), Kakadu (d) and JPEG-LS (e) for the subjective evaluation.*

Figure 6 presents a particular subjective evaluation situation, from we can easily understand some of the choices made by the human observers. From a visual point of view the image compressed with the Kakadu algorithm (Figure 6 *d*) was preferred to the one compressed with the JJ2000, especially because of the missing structural elements in image *c)* (the JJ2000 compression). This choice is in total contradiction with the classical PSNR measure, for which the JJ2000 perform better.

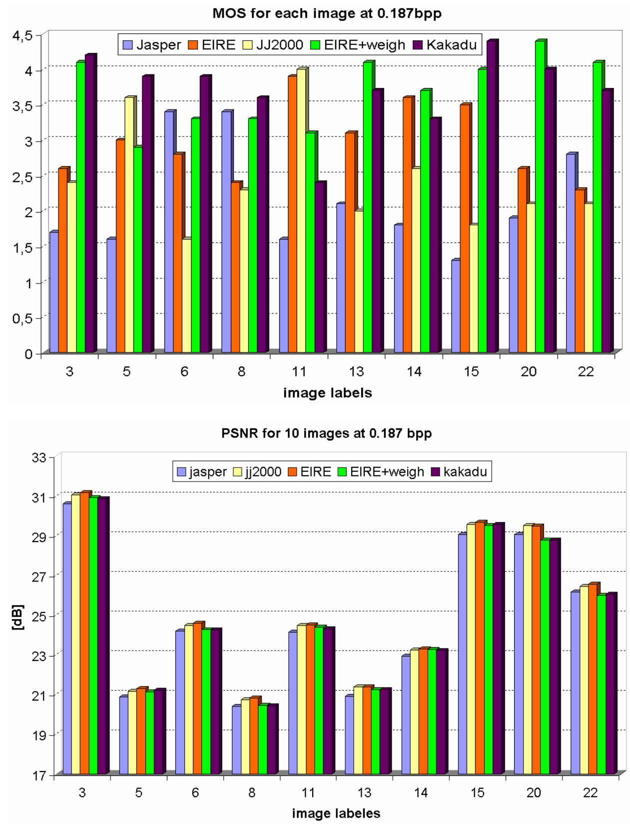

***Figure 7**. **MOS and** PSNR of the different JPEG2000 implementations, for each of the 10 compressed images at 0.187bpp.*

Figure 7 presents the MOS and the classical PSNR measure for each of the 10 images and the 5 codec implementations. As one can see, the PSNR is generally in contradiction with the subjective evaluations (MOS). That can also be seen from the mean general results presented in the Figure 8. The MOS and the PSNR are the mean values computed from all the results of the four codec implementations, for 12 images and all the compression bit rates.

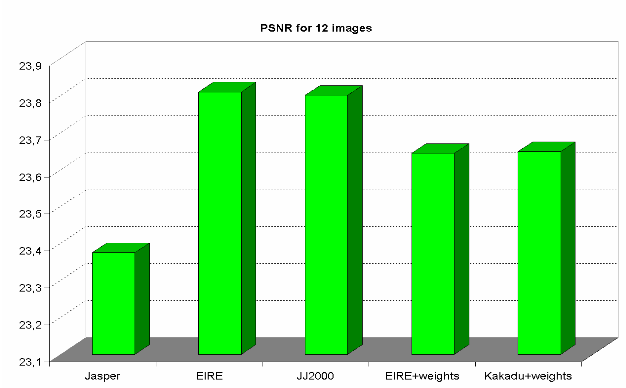

***Figure** 8. **MOS and** PSNR of the different JPEG2000 implementations, for all images and all compression bit rates.*

Thus, the implementations without weightings offer better results in terms of the quantity of original signal present in the final image. Nevertheless, as we have seen, the human observers do not judge the quality with respect to the quantity of signal, but with his quality. Thus, if we want to produce compressed images addressed to human users, the use of the classical mathematical measures in judging the image quality is not a good option, especially for colour images.

7. Conclusions

In this paper, we showed that the use of frequency visual weightings in JPEG2000 compression scheme, offers generally a better perceptual image quality. The image quality was evaluated by performing subjective assessments, with a representative panel of human observers. We have also showed, that objective image quality measures do not offer the same results as the subjective quality evaluation tests, being quite in opposition. The human subjective "criteria" are far too complex for being mathematically characterized at this time. We could only hope that, future researches may offer the solution to the actual problem. Meanwhile, judging the improvements of image quality performed by different algorithmic implementations, or choosing the most adapted compression algorithm to use in a specific application, remain some rather difficult tasks without subjective evaluations. Therefore, subjective evaluations of still colour image quality have to be seriously taken into account in future researches.

8. Acknowledgements

This work takes part in a French National Research project (RNRT), named EIRE (**E**tudes d'opt**I**misation algo**R**ithmiques de JP**E**G2000), funded by the French Research and Industry Ministry, involving different universities and industrial partners and managed by Thales Communications. The involved partners are active members of JPEG2000 French National Body.

9. References

[1] S. Daly, "The Visible Differences Predictor : an Algorithm for the Assessment of Image Fidelity", in *Digital Images and Human Vision*, edited by A. B. Watson, MIT press, pp. 197-206, 1993.

[2] M. Esckicioglu and P. Fisher, "Image Quality Measures and their performances", *IEEE Transactions on Communications*. Vol. COM-43, No. 12, pp. 2959-2965, Dec., 1995.

[3] P. Carrai, I. Heynderickx, P. Gastaldo, R. Zunino, "Image quality assessment by using neural networks", in IEEE International Symposium on Circuits and Systems, ISCAS 2002, Volume 5, pp. 253-256, 2002.

[4] A. Skodras, C. Christopoulos, and T. Ebrahimi, The JPEG 2000 still image compression standard, *IEEE Signal. Processing Mag.*, vol. 18, pp 36-58, Sept. 2001.

[5] "Information Technology – JPEG 2000 Image Coding System," ISO/IEC FDIS15444-1: 2000 (Aug. 2000).

[6] G. Deffner et al.: "Evaluation of display-image quality: Experts vs. non-experts". in *SID Symposium Digest*, vol.25, pp.475-478, 1994.

[7] ITU-R Recommendation BT.500-10: "Methodology for the subjective assessment of the quality of television pictures. ITU, Geneva, Switzerland, 2000

[8] VQEG, "Final report from the video quality experts group on the validation of subjective models of video quality assessment", http://www.vqeg.org/, March 2000.

[9] "Kakadu: A comprehensive, heavily optimized, fully compliant software toolkit for JPEG2000 developers", Kakadu Web Site, http:// kakadusoftware.com/.

[10] "JJ2000 – An Implementation of the JPEG2000 standard in Java", JJ2000 Web Site, http://jj2000.epfl.ch/.

[11] JasPer Project Home Page, http://www.ece.ubc.ca/ ~mdadams/jasper/

[12] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor. Visual thresholds for wavelet quantization error. *SPIE Proceedings - Human Vision and Elect. Imaging*, 2657, 1997.

10. Biography

Adrian Stoica received his B. S. degree in Electronics and Telecommunications from the Polytechnic University of Bucharest in 2000 and a Master in Transmission of Information: Informatics, Image and Automatic from University of Poitiers in 2001. Actually he is a Ph.D. student at the University of Poitiers and Polytechnic University of Bucharest. His work focused on Human Visual System models, colour image compression and quality evaluation, with application to JPEG2000 image compression standard.