A comparative study of color image compression standards using perceptually driven quality metrics

Francesca De Simone, Daniele Ticca, Frederic Dufaux, Michael Ansorge, Touradj Ebrahimi
Multimedia Signal Processing Group
Institute of Electrical Engineering
Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland

ABSTRACT

The task of comparing the performance of different codecs is strictly related to the research in the field of objective quality metrics. Even if several objective quality metrics have been proposed in literature, the lack of standardization in the field of objective quality assessment and the lack of extensive and reliable comparisons of the performance of the different state-of-the-art metrics often make the results obtained using objective metrics not very reliable. In this paper we aim at comparing the performance of three of the existing alternatives for compression of digital pictures, i.e. JPEG, JPEG 2000, and JPEG XR compression, by using different objective Full Reference metrics and considering also perceptual quality metrics which take into account the color information of the data under analysis.

Keywords: Image compression, codec performance, Full-Reference quality assessment, perceptual quality metrics.

1. INTRODUCTION

When considering the lossy compression of digital still pictures and video sequences, the performance of different compression algorithms (i.e. codecs) can be compared in terms of compression efficiency, computational requirements, and additional functionalities like those offered by the bit-stream format. The compression efficiency of an image or video codec expresses the ability of the algorithm 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. It is usually expressed in terms of Rate-Distortion (RD) curves which plot the quality measure versus the bit-rate of the compressed data expressed in bit per pixel (bpp), i.e. bit per sample (bps), value. Since the end users of multimedia data are often human beings, subjective tests can be performed in order to evaluate the quality of compressed multimedia data and build the RD curves. These tests, where a significant sample of human subjects are asked to rate the quality of the processed material, are difficult to design, time consuming, and expensive. Thus, an important issue in codec performance evaluation methodologies is to define adequate objective quality metrics to estimate quality. This metric is called Full Reference (FR) metric, since it assumes as inputs both the original image or video sequence (i.e. reference data), and its compressed version (i.e. test data). The objective metric should provide a reliable estimate of the visual quality of compressed data as perceived by an end user.

So far, in the state of the practice, the visual quality metric used in the Rate-Distortion studies for compression of both still pictures and video sequences is the Peak Signal to Noise Ratio (PSNR). Even if widely used, the PSNR does not provide a perceptual visual quality measure as it is based on purely pixel by pixel difference measure, the Mean Square Error (MSE) [1]. In particular, this measure does not take into account the features of the area where the changes in pixel values are located and its perceptual importance in the quality evaluation of the overall scene. Furthermore, the PSNR metric is usually applied only on the luminance component, assuming that the objective assessment of the luminance channel provides a good approximation of the overall visual quality of the full color data. Any information related to the impairment of the color channels is thus usually discarded. When this is not the case, i.e. the PSNR values are computed on all the color channels, it is not easy to establish a relation between the different RD curves and the overall subjective assessment of the full color data.

In order to solve the open issues related to the performance of PSNR, many perceptual metrics have been proposed in recent years. Some of these algorithms have been proven to significantly outperform PSNR in terms of correlation with subjective quality judgment [2]. Nevertheless, mainly due to lack of standard guidelines for objective visual quality assessment, and unavailability of publicly accessible implementations of different metrics, just a very few examples of

coding performance evaluation studies considering metrics other than PSNR are available in literature [3] [4]. Furthermore, the analysis of the results expressed in terms of quality indexes other than PSNR values can be difficult to interpret when the reader is not familiar with the quality metric under consideration.

In this paper, performance of three different algorithms for the compression of still pictures, namely, JPEG [5], JPEG 2000 [6], and JPEG XR [7], are compared by means of a compression efficiency evaluation. The study focuses on the comparison of these three codecs when considering compression of high resolution low dynamic range (i.e. 24bpp) pictures with no sub-sampling of the chroma components (i.e. 4:4:4 coding). The images in our dataset are all natural scenes. A set of objective FR visual quality metrics is used to produce different RD curves. These curves are then analyzed and compared, while advantages and drawbacks of each metric in this scenario are highlighted.

This paper is structured as follow. Section 2 provides an overview of the quality metrics selected for this study. The test conditions, in terms of original pictures and coding configuration used to produce the dataset, are detailed in Section 3. Finally, the Rate-Distortion evaluation approach and the resulting curves for the codecs comparative study are discussed and analyzed in Section 4. Conclusive remarks are presented in Section 5.

2. SELECTED FULL REFERENCE METRICS FOR OBJECTIVE QUALITY ASSESSMENT OF STILL PICTURES

The objective metrics applied in this study can be classified into two sets: metrics which do not take into account the existence of different color channels of the visual data and are thus applied on one single color channel of the image; and metrics which do take into account the existence of several color channels in the captured data and the interaction between them, thus providing a quality measure which is intended to predict the perceived quality of the full color data. We will refer to them as mono-channel and multi-channels metrics, respectively.

Four different mono-channel metrics have been considered in our study: the PSNR, as reference state of the practice metric; the Mean Structural SIMilarity index (MSSIM) [8]; the Visual Information Fidelity measure in Pixel domain (VIF-P) [9]; and the PSNR Human Visual System Masking metric (PSNR-HVS-M) [10]. These metrics have been selected from the wide variety of mono-channels metrics proposed in the literature based on two criteria:

1) They include a representative selection of different approaches proposed so far in the state of the art for FR objective quality assessment, namely: the state of the practice non perceptual but widely used metric (PSNR), a perceptual metric based on the content features extraction and abstraction (MSSIM), a statistical natural scene analysis based metric (VIF-P), and finally a Human Visual System (HVS) low level features model based metric (PSNR-HVS-M);

2) Their implementations are publicly available to be used by the research community, [11] [12] [13].

The approach for choosing the multi-channel metrics is quite different, since in the state of the art, when considering the multi-channel case, few algorithms have been proposed, for example those proposed in [14] [15] [16]. In fact, the management of the color information when evaluating the quality of natural scenes is still a very open research challenge. Furthermore, very few multi-channel metrics have been contributed to the research community with publicly available implementations and their performance have not been tested in a reliable and extensive way. Three multi-channel metrics have been used in our study: the DCTune metric [17], and the multi-channel versions of two mono-channel metrics, i.e. PSNR and MSSIM, namely, the Weighted PSNR (WPSNR) and the Weighted MSSIM (WMSSIM). The WPSNR and WMSSIM are based on the simple weighted average of the mono-channel measures applied to the three components of the image considering a fixed color space. In this study, the considered color space is Y'CbCr [18] and the weighting factors for the luminance channel and the chrominance channels, have values 0.8, 0.1 and 0.1, respectively [19]. Three variants of the WPSNR have been considered: WPSNR based on the weighted average of the PNSR values computed on the three channels, that we refer to as WPSNR; the WPSNR based on the weighted average of the MSE values computed on the three channels, that we call WPSNR-MSE; and finally the WPSNR based on the weighted average of the pixel values of the three channels, that we denote as WPSNR-PIX.

A general overview of each selected metric is presented in the following subsections.

2.1 Mono-channel metric: Peak Signal to Noise Ratio

The Peak Signal-to-Noise Ratio metric is defined as:

$$PSNR = 10 \, log_{10} \frac{\left(2^B - 1\right)^2}{MSE}$$

where the Mean Square Error (MSE) is defined as:

$$MSE = \frac{1}{MN} \sum_{y=1}^{M} \sum_{x=1}^{N} \left[Im_a(\,x,\,y\,) - Im_b(\,x,\,y\,)\right]^2$$

and:

- B is the bit depth of the considered pictures (i.e. 8 bpp per component for the pictures used in this work);

- M and N are the dimensions of the pictures;

- Im_a, Im_b are the two pictures to be compared.

The PSNR can have values in the range [0 dB, $+\infty$ dB), with PSNR equal to $+\infty$ dB when the two compared images are identical.

2.2 Mono-channel metric: Mean Structural SIMilarity

The Structural SIMilarity (SSIM) quality metric is built on the hypothesis that the human visual system is adapted to extract structural information from the scene [8]. Structural information can be defined as "the attributes that represent the structure of objects in the scene, independent of the average luminance and contrast" [21]. Thus, the perceived image distortion can be approximated by the structural information change detected between the reference and the test image. The similarity measure compares the original and the distorted signal considering three main features of images: the luminance, the contrast and the structure. Considering the luminance channel of the reference and test images, these three features are modeled respectively as: the mean pixel value, the standard deviation of the pixel values, and the structural map computed by subtracting the mean value from the entire image channel and normalizing this difference by the standard deviation value. The SSIM map, computed on two images, is thus defined by multiplying the values of three comparison functions: the luminance comparison function, the contrast comparison function and the structural comparison function. Each comparison function is computed as the correlation between the corresponding feature in the reference and the test images. The SSIM metric is applied locally to the image, introducing sliding windows. As result, a SSIM index map is produced, and in order to obtain a representative quality value for the whole image, the Mean SSIM value is calculated by computing the mean value over the whole picture. For the mathematical details of the SSIM metric the reader is referred to [8]. The MSSIM implementation used in our simulation can be downloaded from [11]. In our study the default values of the parameters involved in the MSSIM formula have been considered.

The MSSIM index can have values in the range [0, 1], with MSSIM equal to 1 when the two compared images are identical. Values in the range [-1, 0) are possible when the contrast of the test image is reversed.

2.3 Mono-channel metric: Visual Information Fidelity

The Visual Information Fidelity (VIF) metric is based on the assumption that images of the human visual environment are all natural scenes and thus they have the same kind of statistical properties. These statistical properties can be represented by Natural Scene Statistics (NSS) models and in particular by using the Gaussian scale mixture (GSM) model in the wavelet or spatial domain [20]. According to this approach, perfect quality images can be modeled as the outputs of a stochastic source that passes through the HVS channel and are received by the cognitive process in the brain. The amount of statistical information that the HVS extracts from the input visual signal is computed as the mutual information between the input and the output of the HVS channel. Any kind of processing which distorts the natural scene signal makes the image appearance unnatural, thus modifying the NSS. This processing can be modeled as an additional channel which filters the signal before it passes through the HVS channel. The statistical information that the HVS extracts from a distorted image is thus computed as the mutual information between the input of the distortion channel and the output of the HVS channel. The mutual information values are quantified by building stochastic models for the source, distortion, and the HVS [9]. The visual quality fidelity measure is defined as the fraction of the reference image information that could be extracted from the test signal, expressed as the ratio between the conditional mutual information between the input and the output of the HVS channel given the particular reference signal, and the conditional mutual information between the input of the distortion channel and the output of the HVS channel given the

particular reference signal. In our simulations we used the computationally simpler multi-scale pixel domain implementation of VIF (VIF-P) detailed and available for download at [12].

The VIF measure can have values in the range [0, 1], with VIF equal to 1 when the two compared images are identical. The VIF index can also have values greater than 1, when the test image is a contrast enhanced version of the reference.

2.4 Mono-channel metric: PSNR Human Visual System Masking

The PSNR-HVS-M is a modified version of the PSNR, intended to take into account the effect of the human visual system Contrast Sensitivity Function (CSF) and contrast masking. In particular, the metric assumes the same formula as the PSNR but computing the Mean Square Error in the DCT domain, as the difference between the DCT coefficients of each 8x8 original image block and the DCT coefficients of the corresponding 8x8 distorted image block. The difference is weighted to take into account the variation of the contrast sensitivity with the frequency. Furthermore, the masking effect between the DCT coefficients of each block is modeled. The PSNR-HVS-M implementation used in our simulation can be downloaded from [13].

The PSNR-HVS-M can have values in the range [0 dB, $+ \infty$ dB), with PSNR-HVS-M equal to $+ \infty$ dB when the two images are identical. Negative values are also possible depending on the way the modified MSE value in the denominator of the logarithmic PSNR-HVS-M formula is defined.

2.5 Multi-channel metric: DCTune

The DCTune metric is based on the computation of an error map as the arithmetic difference between DCT coefficients of the test and reference images [17]. Since the error visibility at each location in the image depends upon the mean luminance of the image block, the error DCT coefficients are weighted by their absolute visibility (light adaptation) as a function of the DCT frequency, the display mean luminance and resolution. Then the errors are further adjusted by contrast masking, by attenuating each error value by a power function of the corresponding coefficient in the reference image. Finally the resulting values are pooled over frequency and space by mean of Minkowski metric. When the different color channels are considered, the pooling is also inter-channels thus the metric can be considered as a multi-channel metric. The DCTune implementation used for our simulation can be downloaded from [21].

The DCTune measure can have values in the range [0, $+ \infty$), with DCTune index equal to 0 when the two compared images are identical.

2.6 Multi-channel metric: Weighted PSNR

The well-know PSNR is usually applied on single channel basis, thus usually three values are produced when considering full color pictures, one for each color channel, assuming that a three dimensional color space is considered. The multi-channel version of the PSNR metric aims at providing a single quality index which takes into account the distortion of all the color channels. According to the definition of the PSNR, reported in Subsection 2.1, there are three different ways to define the weighted average of the contributions of the different color channels:

- the first one is based on the simple weighted average of the PSNR values computed on the three channels, and we call it Weighted PSNR (WPSNR):

$$WPSNR = w_1 PSNR_1 + w_2 PSNR_2 + w_3 PSNR_3$$

- the second one is based on the weighted average of the Mean Square Error values computed on the three channels, and we call it Weighted PSNR on MSE (WPSNR-MSE):

$$WPSNR\_MSE = 10 \log_{10} \frac{\left(2^B - 1\right)^2}{\left(w_1 MSE_1 + w_2 MSE_2 + w_3 MSE_3\right)}$$

- the third one is based on the weighted average of the pixel values of the three channel, before computing the MSE, and we call it Weighted PSNR on PIXel (WPSNR-PIX)

$$WPSNR\_PIX = 10 \log_{10} \frac{\left(2^B - 1\right)^2}{\frac{1}{MN} \sum_{y=1}^{M} \sum_{x=1}^{N} \left(PIXAV_a - PIXAV_b\right)^2}$$

where $PIXAV_n = \left(w_1\,Im_{n1}(\,x,\,y\,) + w_2\,Im_{n2}(\,x,\,y\,) + w_3\,Im_{n3}(\,x,\,y\,)\right)$ with $n \in [a,b]$ and Im_a and Im_b are the two pictures to compare.

As in the PSNR definition, B represents the bit depth of the considered pictures (i.e. 8 for the pictures used in this work), and M and N are the dimensions of the picture. The values of the weighting factors used for the three color channels are those defined in [19] for the multi-channel MSSIM described in the following subsection: $w_1=0.8$, $w_2=0.1$, $w_3=0.1$ respectively for Y', Cb and Cr channel. Further studies will be carried out in order to investigate the proper tuning of these values.

Like PSNR, the WPSNR, WPSNR-MSE and WPSNR-PIX can have values in the range [0 dB, $+\infty$ dB), with values equal to $+\infty$ dB when the two compared images are identical.

2.7 Multi-channel metric: Weighted MSSIM

The Mean Structural Similarity index described in Subsection 2.2 can be applied in multi-channel scenario as follow:

$$WMSSIM = w_1 MSSIM_1 + w_2 MSSIM_2 + w_3 MSSIM_3$$

with: $w_1=0.8$, $w_2=0.1$, $w_3=0.1$ respectively for Y', Cb and Cr channel [19].

Like the MSSIM index, the WMSSIM can have values in the range [0, 1], with WMSSIM equal to 1 when the two compared images are identical. Negative values are also possible due to the possibility of having negative values for the mono-channel MSSIM value, which is possible in particular conditions as explained in Subsection 2.2.

3. TEST DATASET

3.1 Pictures

Fifteen different high resolution pictures have been considered for testing the performance of the coding algorithms. These pictures are typical of digital imaging application natural scenes and belong to the Microsoft dataset (image P01-P30) and to Thomas Richter's dataset (image P33 and P34) [22]. The data are all in bitmap 24bit per pixel gamma corrected standard RGB (sRGB) format. Description of the picture contents with related information on their dimensions are shown in Table 1. As measure of the amount of spatial details of each picture, the Spatial Information (SI) index, computed as the standard deviation of the output of the Sobel filtering of the luminance channel of the image [23], is also provided for each image.

Image name	Resolution (pixels)	Description and SI index
P01	4064x2704	flowers: high textured content (SI=8.0227)
P02	2268x1512	boats: both details and uniform texture (SI=8.4874)
P03	4064x2704	kid with balloon: human details and uniform texture (SI=2.5298)
P04	2268x1512	street and wooden house: details and texture (SI=9.4572)
P05	4064x2704	car1: details (SI=6.0961)
P06	4064x2704	car2: details (SI=9.7348)
P07	2592x1944	woman with complex background (SI=10.4235)
P08	2592x1944	woman with textured background (SI=9.3471)
P09	4064x2704	dog with uniform background (SI=5.723)
P10	2128x2832	Skyscraper (SI=11.4941)
P14	2704x3499	Waterfalls (SI=7.8217)
P26	4288x2848	view of Kobe city (SI=7.6728)
P30	4064x2704	flowers: textured content (SI=5.2053)

P33	3888x2592	green garden (SI=9.6837)
P34	2592x3888	man and natural background (SI=5.8758)

Table. 1. High resolution images used for the evaluation of the different codecs.

3.2 Encoders parameters

The 15 original contents described in the previous subsection have been compressed using three coding algorithms, e.g. JPEG, JPEG 2000, and JPEG XR, using configurations for coding of 4:4:4 input data.

For the JPEG coding, the IJG software version 6b has been used [24]. The 24bpp RGB input data have been compressed using the default settings of baseline profile coding for 4:4:4 data (i.e. Huffman coding, and visually optimized quantization tables). The IJG software automatically applies the pre-processing step of RGB to Y'CbCr color conversion [18] before passing in the transform domain. The compressed pictures have been produced by varying the quality parameter of the JPEG in the range 1 (minimum quality) to 100 (maximum quality).

The software KAKADU version 6.0 [25] has been used for the JPEG 2000 compression of each original 24bpp RGB image considering two coding configurations:

- coding of 24bpp RGB input data, with no chroma sub-sampling (4:4:4 coding), default settings (64x64 codeblock size, 1 tile per frame, 9x7 wavelet transform, 5 decomposition levels, 1 quality layer, no precincts), rate control and visual frequency weighting switched-off.

- coding of 24bpp RGB input data, with no chroma sub-sampling (4:4:4 coding), default settings (64x64 codeblock size, 1 tile per frame, 9x7 wavelet transform, 5 decomposition levels, 1 quality layer, no precincts), rate control and visual frequency weighting switched-on.

The KAKADU software automatically applies the pre-processing step of RGB to YCbCr color conversion [18] before passing in the transform domain. The compressed pictures have been produced by varying the output bit per pixel parameter of the rate control option between 0.01 and 8 bits per pixel.

For the JPEG XR coding, the HD Photo Device Porting Kit software version 1.0 has been used to encode each original 24 bits per pixels RGB image, considering two coding configurations:

- coding of 24 bpp BGR input data (bitmap file), with no chroma sub-sampling (4:4:4 coding), default settings (i.e. all subbands included), and Photo Overlap Transform (POT) enabled for one stage (one level overlapping)

- coding of 24 bpp BGR input data (bitmap file), with no chroma sub-sampling (4:4:4 coding), default settings (i.e. all subbands included), and Photo Overlap Transform (POT) enabled for two stages (two levels overlapping)

The DPK software automatically applies the pre-processing step of RGB to YCoCg color conversion [28] before passing in the transform domain. The compressed pictures have been produced by varying the quality parameter between 2 (maximum quality) and 255 (minimum quality).

4. RESULTS

In this section, the Rate-Distortion curves, built on our set of test images for comparing the performance of JPEG, JPEG 2000 and JPEG XR are analyzed by using the metrics described in the Section 2. The curves are shown in Figure 1-3, and are cumulative plots produced by averaging the results obtained on all the fifteen images of the dataset. In order to average the produced curves over the entire dataset of pictures, the curves obtained in terms of each quality metric index vs bpp value are interpolated using the Piecewise Cubic Hermite Interpolating Polynomial (PCHIP), by fixing an interpolation step of 0.0499 bpp. Then, for each coding configuration and each metric, the average Rate-Distortion curve is computed by simply averaging over the complete set of test pictures the results obtained after the interpolation. The curves are shown in the range [0, 2] bpp since for bit per pixel values larger than 2 bpp the perceptual visual quality of the coded pictures saturates, meaning that there is no perceivable improvement in terms of visual quality for bpp values

larger than 2 bpp. The curves in Figure 1 refer to the mono-channel metric described in Section 2, applied only on the luminance component of the images.

4.1 Quality metrics comparison on single distortion basis

At first, the analysis of the RD curves focuses on the comparison of the different objective metrics behavior when considering each single distortion, i.e. each codec, individually. The performance of a generic state of the art objective visual quality metric is usually acceptable when the analysis is restricted to one single codec, i.e. the metric predicts the quality variation when considering different version of the image produced by varying the intensity of one single kind of artifact. Considering different kinds of distortion, i.e. codec, separately, different behaviors can be detected when comparing the curves produced by using different metrics in our study case: while the curves expressed in terms of PSNR and its multi-channel variants, as well as those in terms of PNSR-HVS-M and VIF-P are growing smoothly and with almost constant step of quality index variation with respect to a constant variation of the bpp value, the DCTune and both the mono-channel and multi-channel MSSIM produced curves which rapidly saturate for bpp values greater than 1 bpp. From a first experimental visual inspection of the coded pictures, independently from the codec, i.e. from the kind of artifact, and also quite independently from the image content, we can say that this quality saturation appears in the test pictures for bpp values greater than 1.

4.2 Quality metrics comparison and codecs performance comparison

The objective metrics comparison can be extended when the RD curves produced for different codecs are analyzed. The first evidence, when considering the results in terms of performance comparison of different codecs, is the separation between the overall behavior of the non-perceptual metrics (i.e PSNR and its variants) versus the perceptual metrics (i.e. MSSIM, WMSSIM, PSNR-HVS-M, and VIF-P). While the non-perceptual measures highlight the JPEG 2000 without frequency weighting of the quantization steps as the best performing codec, the RD curves in terms of perceptual metrics show an overlapping of the curves related to the two JPEG 2000 coding configurations.

Considering just the perceptual measures, it can be noticed that the frequency weighting option slightly outperforms the other JPEG 2000 configuration only in terms of DCTune and PSNR-HVS-M quality measures. On the other hand, these two metrics both show an outlying behavior with respect to the slope of the JPEG RD curve: according to these measures, the quality of the pictures produced by JPEG compression rapidly increases from bpp values around 0.6 bpp, becoming competitive with that obtained by using JPEG 2000 with frequency weighting and better than the quality obtained by using the others codecs. This evident instability of the quality measures is very probably related to the fact that both these metrics are based on the data analysis in the Discrete Cosine Transform domain used by JPEG compression.

Finally, an interesting comparison is that regarding the three multi-channel variants of the PSNR. The PSNR and WPSNR curves are very similar, meaning that the distortion measure on the chrominance channels does not modify the quality evaluation. This could be due to the fact that the distortions on the chrominance channels caused by the coding are not relevant, or that the weighting approach is not correctly tuned to take into account the contribution of the chrominance channels. Further accurate analysis of this multi-channel scenario will be carried out in future studies. While all the PSNR multi-channels models, as well as the mono-channel PSNR, agree in placing the JPEG curve significantly below all the others and the JPEG 2000 without visual frequency weighting above all the others, the overlapping curves of the two JPEG XR configurations and the curve of the other JPEG 2000 configuration are ordered in a variable way. In particular, the three curves are overlapped in terms of PSNR and WPSNR, while they are significantly spaced when expressed in terms of WPSNR-MSE and WPSNR-PIX, with advantage of the JPEG XR codec over the JPEG 2000 in the first case, and vice versa in the second. These two alternative models will also be investigated in future work. As concluding remark, it can also be noticed that, quite differently from the other metrics, the distances among the different curves related to the different codecs remain constant for bpp values greater than approximately 0.5 bpp. A similar behavior is present also when considering the other metrics, but usually shifted for larger values of bpp, thus in a range of values related to an overall higher quality of the test pictures.

5. CONCLUSIONS

In this paper a comparative Rate-Distortion study of three compression algorithms for the coding of 4:4:4 high resolution pictures is presented. The codecs are compared by mean of nine different objective Full-Reference visual quality metrics
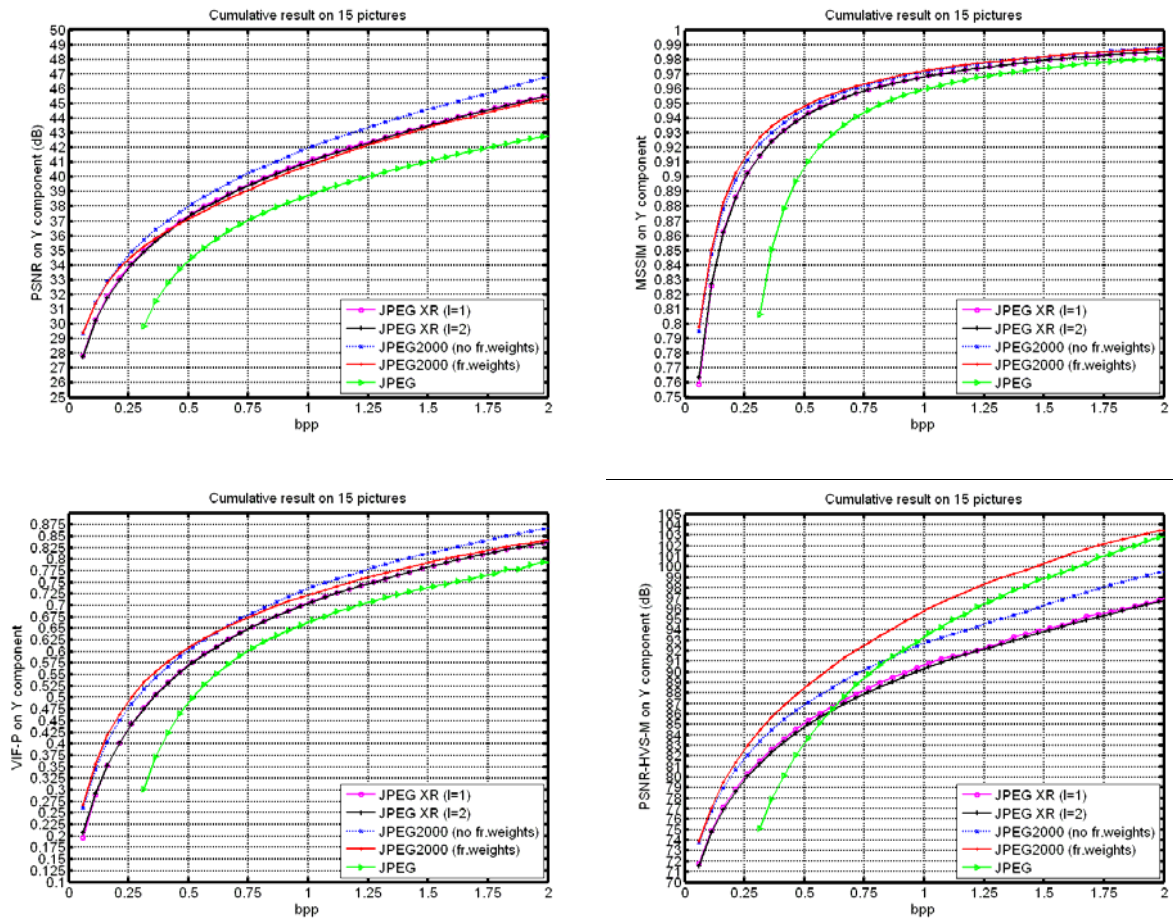
Fig.1. Rate-Distortion cumulative curves produced by using the mono-channel metrics described in Section 2. The mono-channel metrics have been applied on the luminance channel of the images.
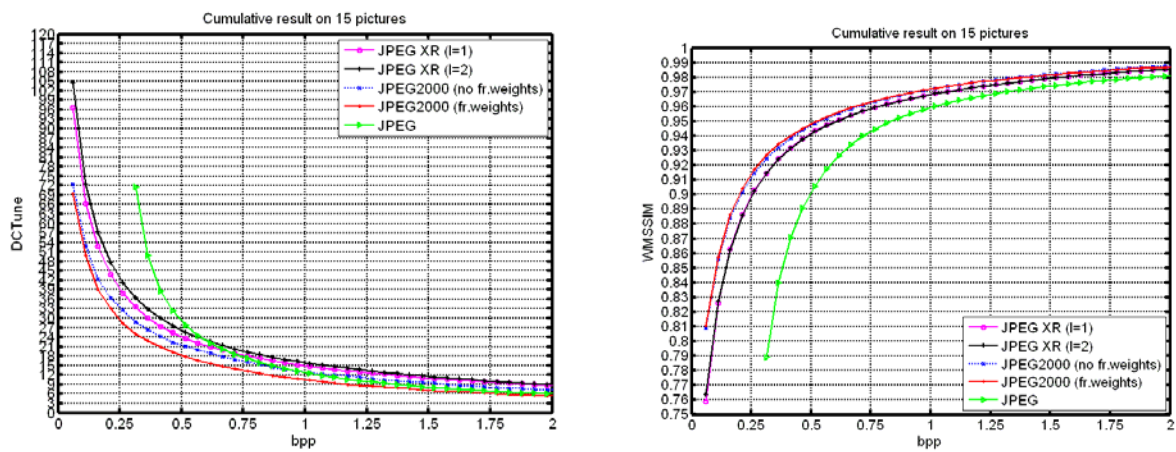

Fig.2. Rate-Distortion cumulative curves produced by using the multi-channel metrics DCTune and WMSSIM described in Section 2.
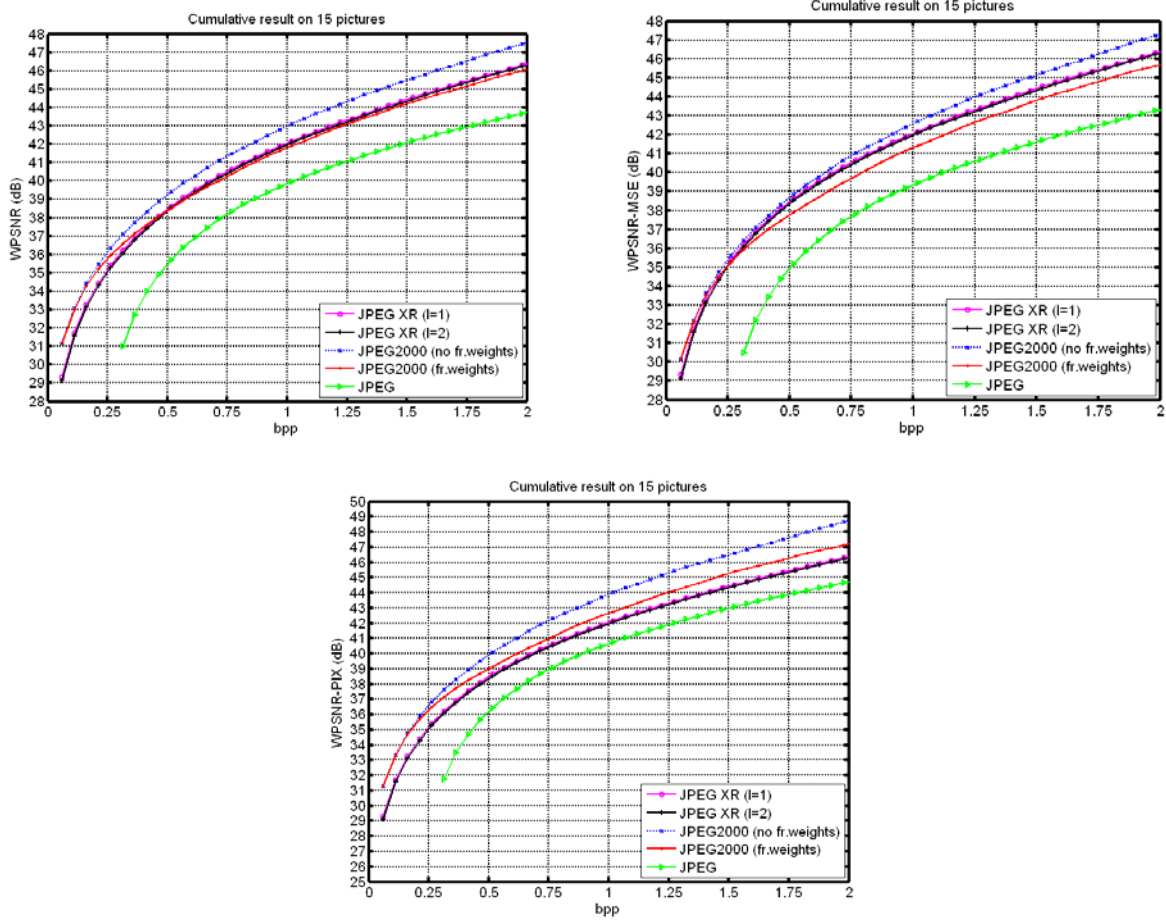
Fig.3. Rate-Distortion cumulative curves produced by using the different multi-channel metrics WPSNR, WPSNR-MSE and WPSNR-PIX described in Section 2.

and tested on a dataset of 15 color images with natural scene contents. As highlighted by the analysis of the cumulative RD curves presented in Section 4, the results of the comparison vary significantly according to the metric used in the quality evaluation. Even if the correlation of each objective metric with the subjective assessment needs to be evaluated by producing subjective results with properly designed subjective tests, we believe that the objective evaluation of the coding algorithm already provides a powerful tool for testing and improving their performance, assuming that a study of the objective metrics used to carry the analysis is carried out at the same time.

Nevertheless, an extensive and reliable study of the different approaches for objective quality assessment is needed. In particular, future study will focus on the study of objective metrics for the analysis of color data. In the framework of codec comparison, further analysis will be carried out including the subjective analysis of the artifacts introduced by each codec and its dependency on the content color and spatial features and the testing of the metrics robustness when considering mixture of artifacts.

ACKNOWLEDGEMENTS

REFERENCES

1. B. Girod, "What's wrong with mean-squared error", Digital Images and Human Vision, MIT Press, Cambridge, MA, 1993.

2. H. R. Sheikh, M.F. Sabir, and A.C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms", *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440-3451, November 2006.

3. F. De Simone, M. Ouaret, F. Dufaux, A.G. Tescher, and T. Ebrahimi, "A comparative study of JPEG 2000, AVC/H.264, and HD Photo", *Applications of Digital Image Processing XXX*, *Proceedings of SPIE*, vol. 6696, San Diego, CA USA, 28 – 30, August 2007.

4. T. Richter and C. Larabi, "Subjective and Objective Assessment of Visual Image Quality Metrics and Still Image Codecs," dcc, p. 541, Data Compression Conference (dcc 2008), 2008.

5. G.K. Wallace, "The JPEG Still Picture Compression Standard", Communications of the ACM, vol. 34, no. 4, pp. 31-44, 1991.

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

7. A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 still image compression standard", *Signal Processing Magazine IEEE*, vol. 18, Issue 5, 36-58, September. 2001.

8. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity", *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, April 2004.

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

10. N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions", *Proceedings of the Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics* (VPQM-07), Scottsdale, Arizona, USA, 25-26 January, 2007.

11. http://www.ece.uwaterloo.ca/~z70wang/research/ssim/ssim_index.m.

12. http://live.ece.utexas.edu/research/quality/.

13. http://www.ponomarenko.info/psnrhvsm.htm.

14. X. Zhang, D.A. Silverstein, J.E. Farrell, B.A. Wandell, "Color image quality metric S-CIELAB and its application on halftone texture visibility," *compcon*, p. 44, Proceedings of COMPCON 97, 1997

15. C. Van den Branden Lambrecht and J.E. Farrell , "Perceptual Quality Metric for Digitally Coded Color Images", Proceedings of the European Signal Processing Conference (EUSIPCO), 1175-1178, 1996.

16. C. Chou and K. Liu, "A Fidelity Metric for Assessing Visual Quality of Color Images", Proceedings of 16th IEEE International Conference on Computer Communications and Networks, (ICCCN 2007), 1154-1159, August 2007.

17. A.B. Watson, "DCTune: A Technique for visual optimization of DCT quantization matrices for individual images", Society for Information Display Digest of Technical Papers XXIV, 946-949, 1993.

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

19. Z. Wang, L. Lu and A.C. Bovik, "Video quality assessment based on structural distortion measurement", *Signal Processing: Image Communication, Special issue on Objective video quality metrics*, vol. 19, no. 2, February 2004.

20. Martin J. Wainwright and Eero P. Simoncelli, "Scale mixtures of gaussians and the statistics of natural images," *Advances in Neural Information Processing Systems*, vol. 12, pp. 855–861, 2000.

21. http://vision.arc.nasa.gov/dctune/.

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

23. Recommendation ITU-T P.910, "Subjective Video Quality Assessment Methods for Multimedia Applications - Series P: Telephone Transmission Quality, Telephone Installations, Local Line Networks - Audiovisual Quality in Multimedia Services", September 1999.

24. http://www.ijg.org/.

25. http://www.kakadusoftware.com.

26. http://www.microsoft.com/whdc/xps/wmphoto.mspx.

27. Microsoft Corporation, "HD Photo: Device Porting Kit Specification" version 1.0 (June 2006).

28. H. S. Malvar and G. J. Sullivan, "Transform, Scaling & Color Space Impact of Professional Extensions", ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6 Document JVT-H031, Geneva, May 2003.