Quality Assessment of Hdphoto - The New Proposed Compression Algorithm

Mohamed-Chaker Larabi*, Thomas Richter**, Ghislain Anciaux*; * SIC lab, University of Poitiers, France, ** University of Stuttgart, Germany

Abstract

Objective quality assessment of lossy image compression codecs have become an important part of the recent call of the JPEG committee for Advanced Image Coding. The aim of this work is twofold: First, we present state-of-the-art still image codecs, JPEG [1] in two variations, a visual and PSNR optimal JPEG2000 [2, 3] version, H.264/AVC [4, 5] and the recently proposed HDPhoto format [6, 7] introduced by Microsoft. We measure the performance of these codecs by subjective, full-reference ordering tests and two visual image quality metrics on a carefully selected test set. Second, we evaluate the prediction quality of the metrics by comparing them with the outcome of the subjective tests.

Introduction

Image compression aims at representing digital images as for example obtained from digital cameras in the shortest possible size. Unlike lossless compression, lossy coding techniques are used that remove irrelevant image information -irrelevant to an estimated standard observer-. This means that an image compression code needs to define a suitable compromise between image quality at one hand and required file size at another. While image size has an obvious and precise definition, image quality is a much more delicate item. The purpose of a (full reference) image metric is exactly that: Compare the reconstructed image with the original and assign a number to this pair that predicts how a standard observer would judge the quality of the compression result.

Traditionally, the mean square error has been often used as such a metric: It is mathematically feasible, it is easy to optimize for it, but its prediction to observed image quality leaves lots to deserve. The search for improved image metrics is thus an important scope of the Advanced Image Coding call of the JPEG committee.

The aim of this paper is to evaluate the new proposed codec (HDPhoto) and compare its rate-distortion performance to existing standards. The objective image quality metrics discussed in this paper are Multiscale M-SSIM [4] by Wang, Simoncelli and Bovik, and the high-dynamic range version of VDP [5, 8]. The results obtained from these tests are then compared to the results of subjective testing obtained from an ordering test.

The test image set is a subset of the set carefully selected by the authors for JPEG-internal testing -we had to restrict the set for the purpose of the subjective tests. The test set includes mostly natural images of varying content and structural complexity, but also one compound image and an ultrasonic sample, representing all different image types at least once in the set. The image set is currently restricted to 8bpp grey-scale or RGB images, no high-dynamic range images have been considered for this specific test. The test set originates from the ITU and the JPEG set used for JPEG and JPEG2000 evaluation, the test set kindly provided by Microsoft for HDPhoto evaluation, the ISO 400 test set, the Kodak test image set, and a collection of additional high-resolution images that complements Microsoft test by images containing high-texture, low-contrast images.

Used Codecs

Test images have been compressed with five important standards currently available on the market: Traditional baseline JPEG [1], JPEG with arithmetic coding option, JPEG 2000 [2, 3], JPEG 2000 with visual optimization, the H264 I-frame compression [4, 5] and the recently proposed HDPhoto image compression by Microsoft [6, 7].

JPEG uses the almost traditional DCT for energy compaction, followed by a uniform scalar quantization. Quantization bucket sizes can be freely selected, but what we use here for testing are the tables optimized for a human observer recommended by the JPEG in its specification. The quantization process is followed by either a Huffman-based VLC entropy coder in the baseline specifications, or an arithmetic coding option using the QM binary coder. In both cases, we run the IJG implementation for the tests.

H264 is the coding technology deployed in the MPEG-4 video coding standard. Its features are mainly tuned towards moving-image (video) compression, and it thus uses a DCT like block transformation for energy compaction which fits fine into the motion prediction scheme of video compression. Each block carries its own quantization parameters to adapt the code to spatial image variations. The quantized parameters are encoded by an entropy coder backend. For this work, we used the reference implementation FRExt provided by Fraunhofer.

JPEG2000 is the latest still image compression standard of the ISO. While it clearly aims at improving the performance, its design goals also target at uttermost flexibility. Most other codecs discussed here, its transformation is not block-based, but uses a discrete wavelet transformation, followed by scalar quantization. A flexible embedded bitplane rate allocation and compression scheme, the EBCOT, forms the backend where the arithmetic encoding is performed by the MQ coder.

We thus run JPEG2000 here in two possible variations: A PSNR optimal version that tries to minimize the mean square error, and a visually tuned version that includes fixed frequency weights and visual masking options. Both options use the Pegasus Imaging codec.

HDPhoto is a recent image compression codec proposed by Microsoft, and is currently undergoing its standardization as JPEG-XR at the ISO. HDPhoto uses an overlapped block transformation for decorrelation, scalar quantization followed by an adaptive Huffman coding entropy coding backend. The implementation we are testing with is the Device Porting Kit provided by Microsoft, which has not been tuned yet to a visual metric. We rather use the provided code as a black-box and did not intent to improve it towards human vision for the purpose of this paper. Later work will discuss the opportunities of HDPhoto for optimization towards human vision.

Quality assessment

Test preparation

The test image set was carefully collected to include images of various natures. We include images having very unstructured components like water or the sky as well as highly textured images: Grass in particular is a challenge for wavelet based codecs since its structure is easily confused with image noise and quantized to an unstructured flat, and thus unnaturally looking surface. The set includes also one compound and one grey-scale image completes it, though our focus is clearly on natural color images. Otherwise, images are only 8bpp, RGB, and neither considered high-dynamic range images.

The test images have been provided partially by the ISO and the ITU for JPEG and JPEG2000 testing, by Microsoft for HDPhoto testing and also include a collection of images collected by the authors to round the test-set up.

Preparation of images for the tests is a bit delicate. First of all, several test images have to fit on one screen for the ordering test, which means that we need to scale down most of the images in the test set. Naturally, compression has to be performed in the scaled image domain as otherwise compression artifacts can get removed or hidden in the downscaling process. Second, not all codecs provide the option to compress the images to a specific rate, but rather specify target size by an otherwise unspecific quality parameter similar to JPEG. To address this issue, we compiled a script that implements an external rate control for each codec in the set. This script first runs an initial bisection algorithm to find a rate close to the target rate, and then runs an extensive search near the found target quality to locate the best possible compression parameter generating the highest possible output rate lower or equal than the target rate. Note that this favors codecs that hit the target rate more precisely; we decided that this addresses the issue of rate allocation best as it disallows codecs to compress better at the cost of precision.

Objective measurement

In this paper, we use two visual quality metrics: The multiscale SSIM metric proposed by Bovik et al. [8] and the Visual Difference Predictor (VDP) introduced by Daly et al. [9, 10].

MS-SSIM is a low-complexity visual metric that follows an unusual bottom-up design strategy in so far as it does not attempt to model the physical properties of the human visual system. Instead, it separates the image into several scales, and then measures in each scale the local contrast, defined as the variance over 11 x 11 blocks, and the structural difference between reference and decoded image as the mean square error of the luminance and variance normalized signal. In the lowest scale, the luminance error is also measured. Errors are pooled for each scale, multiplied by experimentally obtained weighting factors and pooled again to a single number between zero and one describing the similarity of the two images.

VDP, however, follows a more traditional design pattern by dividing the signal into several frequency bands and measuring the local visibility of image errors in each band. Visual masking effects, i.e. masking of errors behind image structure, are taken into consideration. The errors are then again pooled up, using the well-known contrast sensitivity of the human eye, identifying all those pixels that are likely observed as different in the decompressed image. VDP is a very high-complexity metric, and unlike MS-SSIM not fit to become part of an image rate allocation process.

Subjective measurement

Casual or informal subjective testing by a reasonably expert viewer remains an important part of system evaluation or monitoring. Formal subjective testing has been used for many years with a relatively stable set of standard methods until the advent of digital compression subjective testing described in the ITU recommendation [11] and ISO standards [12].

The assessment procedure requires normalized test conditions such as non-reflective walls, controlled lighting, calibrated monitor,… Also, the observer must have normal visual acuity and no color blindness.

In the framework of this contribution, we have selected an ordering test allowing to assess the quality of an impaired image not only in comparison with the reference one but also the other impaired images obtained with other coders as shown by the following image. The task of the observer consists in selecting images in an increasing order of quality. A score of 1 is assigned to the image selected as the worst while a score of 6 is assigned to the best one.

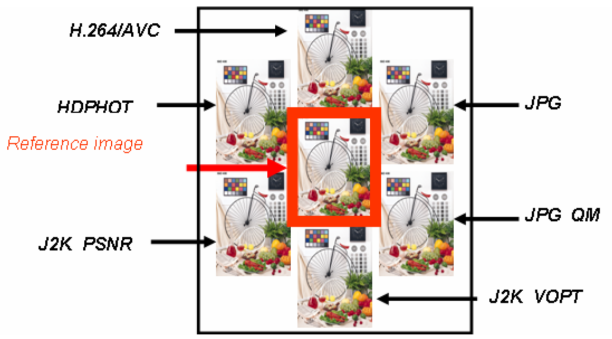

Figure 1. *Subjective assessment configuration*

The obtained results are compiled in order to have the Mean Opinion Score (MOS) that represents the average score for each coder for all bitrates and images. The MOS values are passed through a statistical analysis to first check the validity of the results and then to reject the outliers. By outlier one means an observer that has given random answers in his evaluation. This task is performed by the mean of the kurtosis test. The next step consists in recompiling the MOS and the computation of its confidence interval.

In this work, two subjective tests have been run simultaneously in order to make a contradictory study. Each of these tests is composed of :

- 480 images (6 coders x 4 bitrates x 20 original images)
- 15 observers

The order and the position of the different images were defined randomly for both tests.

Experimental results

Subjective assessment experiments have been run using the images set and the methodology described above. Figure 2 shows the average MOS results for the tests. The subjective quality affected to each coder is quite the same for the two tests but with relatively different scores. From one side, the visually optimized version of JPEG2000 is considered by the observer as giving the best qualitative results [13]. From the other side, JPEG baseline is considered as giving the worst results and this is what we were expecting. However, HDPhoto has not caught the qualitative interest of the observers since it has been classified among the lowest coder.

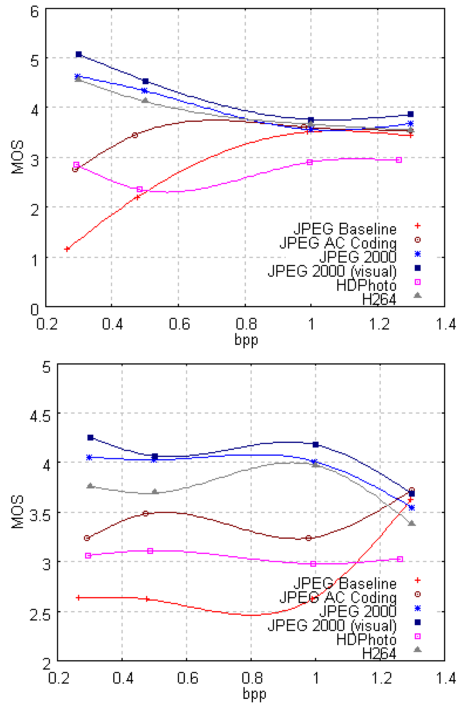

Figure 2. Subjective assessment results for test 1 (top) and test2 (bottom)

The objective measurements are given in figure 3. The interpretation of the figures is quite difficult because it represents average values for the whole image datasets and bitrates used in this study. Nevertheless, we can notice that the visually optimized JPEG 2000 gives the best results from the SSIM point of view while the VDP is in favor of JPEG-QM. The results have to be considered by image and/or by bitrate in order to be more precise. Due to pages limitation, the complete results cannot be presented.

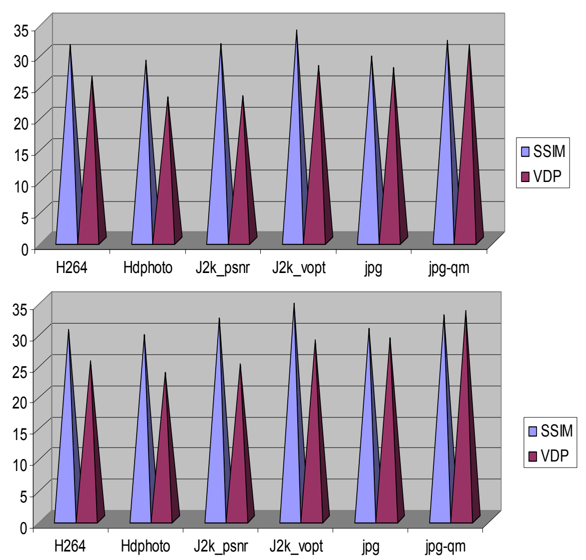

Figure 3. Objective measurements (VDP, SSIM) for test1 (top) and test2 (bottom).

Correlation study

In order to assess the performance of the metrics used for this study (SSIM and VDP), we have studied the correlation between the subjective scores (MOS) and their predicted scores. The correlation study is performed by using the Pearson Correlation Coefficient. This latter has been performed by using the obtained values but the results were surprising because the correlation was too low. By studying this problem, we have understood that this was due to the subjective scores that are obtained by ordering and not by an absolute test.

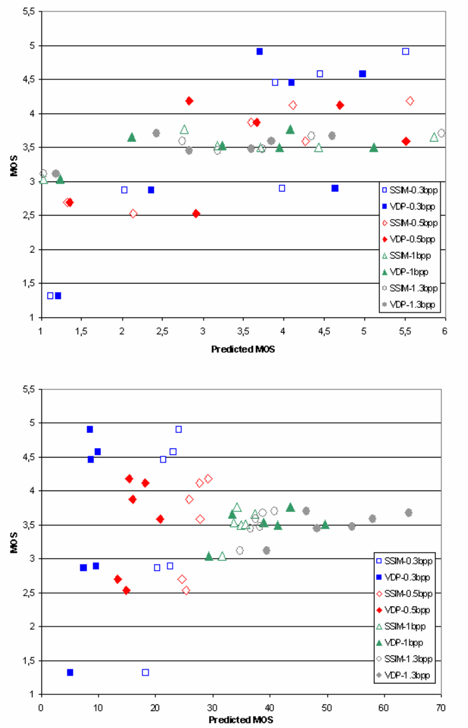

Figure 4. MOS vs. objective scores for ordered (top)and unordered (bottom) cases

So, the objective scores have been ordered and their order has been used as the new scores (see figure 4). Table 1 gives the correlation values for SSIM and VDP versus subjective MOS for ordered and unordered cases

Table 1. Pearson correlation coefficients for MOS *vs.* SSIM and VDP by the use of ordered and unordered scores.

Bitrate	SSIM		VDP	
	Ordered	Unordered	Ordered	Unordered
all bitrates	0,698	0,182	0,546	0,079
0,3 bpp	0,896	0,849	0,749	0,751
0,5 bpp	0,898	0,813	0,531	0,489
1,0 bpp	0,651	0,701	0,558	0,538
1,3 bpp	0,873	0,899	0,718	0,678

Discussion

The relative performance of HDPhoto in comparison to other codecs seems relatively consistent between objective and subjective testing, however other test results require further care. Specifically, the two objective metrics disagree on the relative performance of JPEG and the visually optimized JPEG2000, which might be caused by the different underlying technology: While the VDP metric uses a Fast Fourier Transformation and hence operates in a function test space very

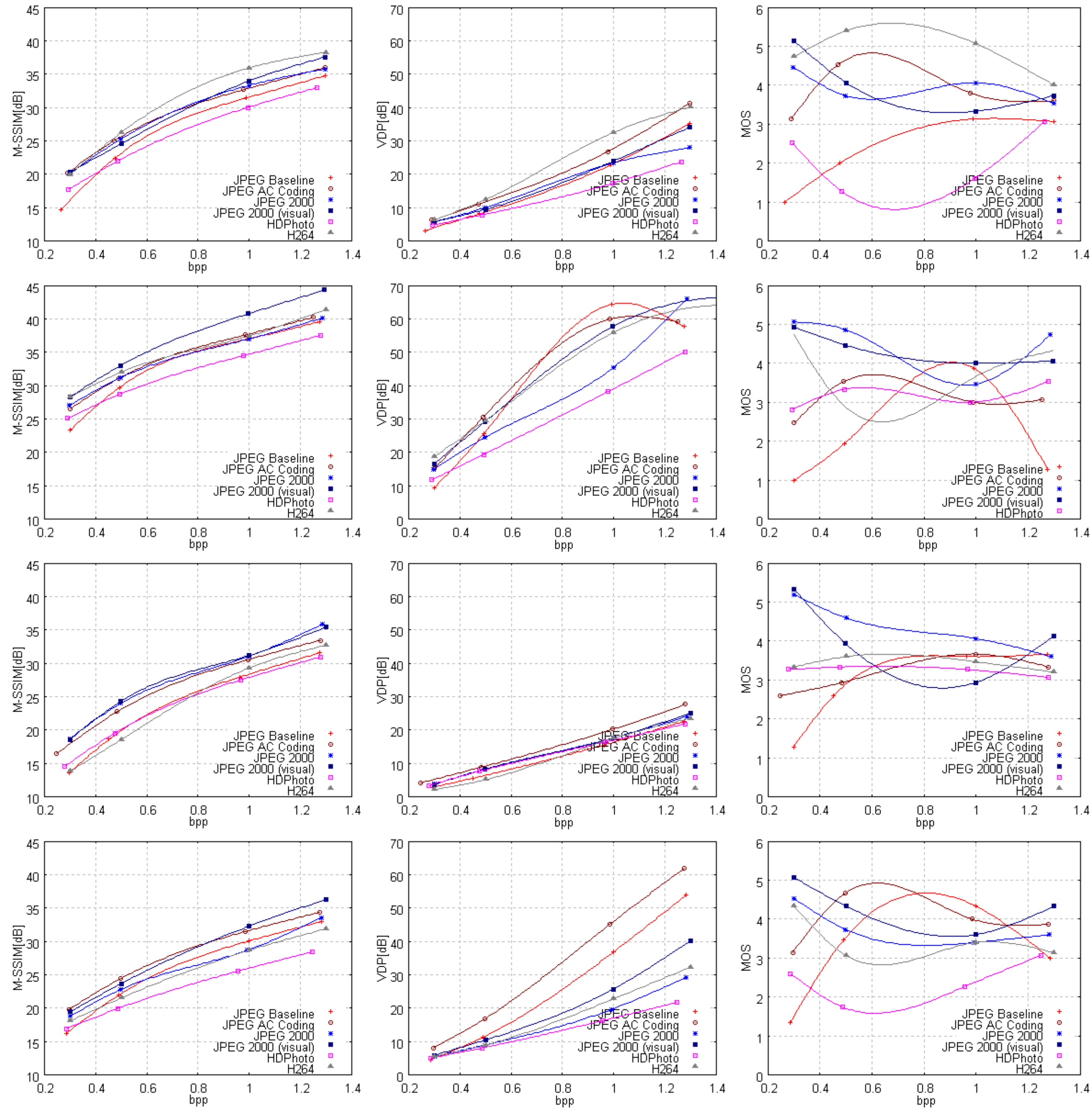

Figure 5. *Logarithmic SSIM plots (left column), VDP plots (middle column) and subjective MOS scores (right column) for the images "woman", "bike", "p01" and "Honolulu" (top to bottom).*

similar to JPEG, the M-SSIM metric separates scales by a wavelet transformation - in our case the 13/7 wavelet not used in JPEG2000 - but still by a similar transformation. This might explain the different performance of the two codecs amongst each other. However, what can be said is that visually optimizing JPEG2000 seems to have a positive effect in general, though not consistent over the full image set, as seen for subjective testing on the "woman" image.

Two sources of systematic errors should also be mentioned: First of all, we first had to scale all images before compressing and evaluating them, simply because otherwise images wouldn't have fit on the screen for the ordering test. Hence, images have been relatively small compared to what is expected to be handled in the target market of HDPhoto and JPEG2000. Note, however, that due to the larger support of the wavelet filter and hence its ability to catch longer range correlations than block-based transformations, this might have been a disadvantage for wavelet based compression. The second problem is that we had to limit ourselves to lower bitrates and thus often over-compress images beyond the limits of codecs.

Conclusion

In this paper, we have evaluated the new proposed codec (HDPhoto) in its first provided version and compare its rate-distortion performance to existing standards JPEG, JPEG 2000 and H264. This evaluation has been performed by mean of two well-known metrics (MSSIM and VDP) in addition to subjective assessment. The results show a lack of performance of the provided version of the HDPhoto codec. The addition of visual weightings and masking can improve the performance.

The current results are very much taken from a work in progress that has to be extended and refined in future and ongoing research. Specifically, we need to address the limitations of the methods above, namely have to extend the tests to higher bitrates and larger images. This implies that the testing methodology has to be modified for subjective tests: We will complement our ordering tests by flicker tests: Here reference and distorted image are observed on the screen one at a time, and the observer can toggle between them shortly, making image artefacts visible by flickering back and forth between the image. This test has two advantages, namely one only has to fit one image on the screen at once, meaning that larger images can be tested, and second image artefacts become more obvious, thus allowing us to go beyond the 1.3 bpp bitrate limitation in the test described above. It should also be noted that the relation between bitrate and quality depends very much on the image: While for some images a bitrate of 1.7 bpp provides visually lossless quality, on others one can go up to 2.3 bpp and still observe visible artifacts. Subjective tests should thus be performed on harder images.

A second extension of this test to be considered is to include images beyond an 8bpp precision, i.e. true HDR images. At the time of writing it is unclear how much the above measurements carry over to higher dynamic ranges and higher bit precision. Furthermore, the HDPhoto codec we had for comparison contained no optimizations for the human visual system whatsoever, so future optimizations might offer more quality.

References

[1] W. B. Pennebaker, J. L. Mitchell: "JPEG Still Image Data Compression Standard" Van Nostrand Reinhold, New York (1992).

[2] M. Boliek (Ed.): "Information Technology — The JPEG2000 Image Coding System: Part 1", ISO/IEC IS 15444-1, (2000)

[3] A. Skodras, C. Christopoulos, T. Ebrahimi, "The JPEG 2000 still image compression standard", Signal Processing Magazine, IEEE vol. 18, Issue 5, 36-58 (September. 2001).

[4] S. Srinivasan, C. Tu, S. L. Regunathan, R. A. Rossi, Jr., 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).

[5] S. Srinivasan, C. Tu, Zhi Zhou, D. Ray, S. Regunathan, G. Sullivan: "An Introduction to the HDPhoto Technical Design", JPEG document WG 1 N4183.

[6] G. J. Sullivan, H. Yu, S. Sekiguchi, H. Sun, T. Wedi, S. Wittmann, Y. Lee, A. Segall, and T. Suzuki, "New standardized extension of MPEG-4AVC/H.264 for professional-quality video applications,", ICIP 2007, San Antonio, USA, (September 2007).

[7] T. Wiegand, G.J. Sullivan, G. Bjøntegaard, A. Luthra: "Overview of the H.264/AVC Video Coding Standard", IEEE Magazine vol. 18, No. 5, pp. 36-58 (Sept. 2001)

[8] Z.Wang, E. P. Simoncelli, and A. C. Bovick, "Multi-scale structural similarity for image quality assessment," 37th IEEE Asilomar Conf. on Signals, Systems and Computers 37, 2003.

[9] R. Mantiuk, S. Daly, K. Myszkowski, H.-P. Seidel. Predicting Visible Differences in High Dynamic Range Images - Model and its Calibration. In: Proc. of Human Vision and Electronic Imaging X, IS&T/SPIE's, pp. 204-214, 2005.

[10] R. Mantiuk, K. Myszkowski, H.-P. Seidel: "Visible Difference Predicator for High Dynamic Range Images" Proc. of IEEE International Conference on Systems, Man and Cybernetics, 2004 pp. 2763-2769.

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

[12] ISO 3664:2000, "Viewing conditions-graphic technology and photography," ISO, Geneva, Switzerland, Tech. Rep., 2000.

[13] C. Larabi, subjective assessment of HDPhoto vs. other coders. ISO/ WG1 register of the JPEG committee. Nov 2007

Author Biography

Chaker Larabi received his master in automatics from the University of Lyon 1(1998) and his PhD in image processing from University of Poitiers (2002). Since then he is teaching and research assistant at the same university and manage the work on quality assessment methodologies, color appearance modeling, integrating the HVS models in compression applications. He is currently managing a work on digital cinema in the framework of a European project named EDCine. He is senior member IEEE, SMPTE member and member of IS&T. He is co-chair of AIC adhoc group of the JPEG committee and participates to TC1-60.