High-quality still color image compression

F. Truchetet, MEMBER SPIE
B. Joanne
F. Pérot
O. Laligant
LE2I, Université de Bourgogne
IUT, 12 rue de la fonderie
71200 Le Creusot, France
E-mail: f.truchetet@iutlecreusot.u-bourgogne.fr

Abstract. A wavelet packet transform algorithm for color still images is presented and we show how and with what performances the transformed image tree can be pruned and the imagets altered to obtain a compression/decompression algorithm respectful of human psychovisual image perception. The basic assumptions for human vision on which the algorithm has been constructed are presented; and we deal with the color transformation used before applying wavelet packet transform. We also point out that a quasi-lossless compression/decompression scheme can be easily obtained with a compression ratio up to 1:10 (quantization step was not considered here). Finally, a new quality criterion based on the results of tests on human sensitivity to various scale and color image details is proposed. This figure of merit can be considered a multiresolution release of the most commonly used peak SNR (PSNR) criterion. [S0091-3286(00)01102-8]

Subject terms: image compression; color image; wavelet packets; psychovisual image assessment; color image metric.

Paper 980200 received May 20, 1998; revised manuscript received July 20, 1999; accepted for publication July 30, 1999.

1 Introduction

Still image compression can be considered from various points of view, but for many applications, such as image storage, image transmission, and image indexing for multimedia, the psychovisual perception of these images is exactly the point. This is especially true when one is dealing with real-world images. Therefore, the human eye assessment is used in this work as the quality control criterion for the compression/decompression scheme under consideration. We aim for a very low distortion in reconstructed image from a psychovisual point of view. An algorithm is proposed for lossy compression of color still images based on multiresolution analysis on a color space designed for its psychovisual relevancy. Multiresolution analysis enables us to take into account the variations of human eye/brain spatial resolution with color contrast and with global luminance.[1] Previous experiences[2] connected with gray-value still image compression/decompression scheme design have shown that the wavelet transform, the Mallat algorithm,[3] is a very efficient method for this purpose, particularly if real-time implementation is under consideration.[4] Hence a wavelet packet transform algorithm for color images is presented in this paper and we show how and with what performances the transformed image tree can be pruned and the imagets altered. According to Fig. 1, the study is limited to the three first stages of the classical compression scheme; we do not pretend to have anything new to add to quantization (scalar or vector) and coding algorithms. These parts are not specific to the color image compression problem and a rich literature is available on these subjects.

In the next section of the paper, the basic assumptions for human vision on which the algorithm has been constructed are presented; the following section deals with the color transformation used before applying wavelet packet transform; and in the third part, we show that a quasi-lossless compression/decompression scheme can be easily obtained with compression ratio up to 1:10 (quantization step was not considered here). Finally, we propose a new quality criterion based on the results of our tests on human sensitivity to various scale and color image details. This figure of merit can be considered as a multiresolution release of the most commonly used peak SNR (PSNR) criterion.[2]

2 Color Human Vision and Resolution

It is well known that the retinal cells responsible for color vision are the cones and that there are three sorts of such cells each one being sensitive to a particular range of wavelengths: the S cone for the short ones (blue), the M cones for the medium ones (green), and the L cones for the longest (red). For artificial vision, the same design is used, and in color cameras, there are three or sometimes four color channels. But while in these devices, the number of pixels for each color is the same and they are regularly dispatched on the sensitive layer, in the eye, the number of cells of each sort is very different and they are dispatched in a very irregular manner. Thus, spatial resolution depends on the

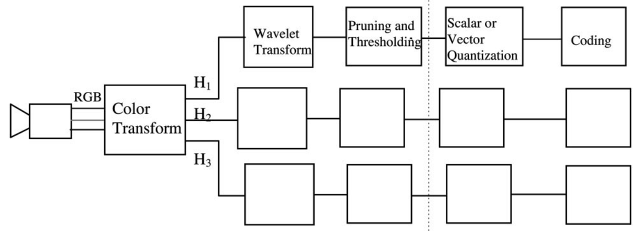


Fig. 1 General color image compression scheme.

considered retinal field and on color. Biological and behavioral measurements of the cone mosaic lead to the conclusion that there are about twice as many L cones as M cones and that the number of S cones is even smaller by several orders of magnitude.[1,5] But the brain is fed with more elaborate data than the basic three-color information and we must consider the preprocessing by the ganglion cells; studies in this field are still in progress and conclusions are not very steady, but it seems that one can admit that there are three paths: one for achromatic information and two for chromatic contrast signals[6] (red/green and blue/yellow). Another important idea to take into account is that the spatial resolution for line grids depends on the orientation of the lines and that vertically oriented details are seen with a better resolution than the horizontally oriented ones, and these latter better than the diagonally oriented details. We show in the next sections of the paper how we can benefit from these properties to compress color images via multiresolution analysis.

3 Image Color Processing

3.1 *Antagonistic Colors Model*

To use a scalar scheme with color images we must choose a projection of the 3-D vector color space on three axes.[7] According to the elements pointed out in the previous section, it is known that the usual RGB basis does not correspond to color perception-analysis done by the human visual system. A color space transformation is commonly used in color characterization, which is called the HSV (hue, saturation, value) transform. The nonlinearity of this transformation is one of its drawbacks; another one is that it is not in good relation with human color representation. Psychovisual experiments tend to favor the basis of antagonistic colors. The best known one is called the hue-cancellation experiment, more information and bibliographic references can be found in Ref. 1. In particular, the ganglionic cells that form the last step of the preprocessing before transmission to the optical nerve make such a transformation. This basis (H_1, H_2, H_3) is obtained by linear transformation of the RGB base. It is composed of an achromatic component (R+G) and of two chromatic ones (R−G and G−B). The transformation is more precisely defined as

$$\mathbf{H}=\begin{bmatrix} H_1=\dfrac{\mathrm{R}+\mathrm{G}}{2} \\ H_2=\dfrac{\mathrm{R}-\mathrm{G}}{2} \\ H_3=\dfrac{2\mathrm{B}-\mathrm{R}-\mathrm{G}}{4} \end{bmatrix}. \tag{1}$$

As previously noted this transformation is linear and easily invertible. The linearity property can be useful in implementation stage, it is also a strong support for our further asumption about the additionnal psychovisual consequences of image components alterations. The H_1 component is supposed to be close to the one used by the human brain (the parvo system and magno system) for analyzing contrast information. The parvo system processes steady

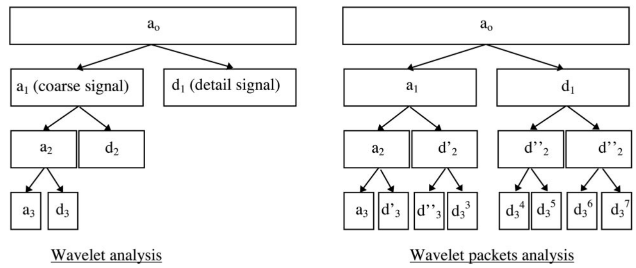


Fig. 2 Imaget analysis.

features and analyzes their shape and details. The magno system processing deals with movement and 3-D features.[1] The H_2 and H_3 components underline antagonistic colors. Here H_2 is the R−G component and H_3 is approximately the blue-yellow one. The V_4 brain area uses H_2 and H_3 components to complete scene analyzing. Its resolution is three to four times smaller than those of the parvo and magno systems.

For these reasons and after some tries with other color bases we choose to use this transformation in our compression design.

3.2 *Color Image Wavelet Transform*

The most popular algorithm for multiresolution image analysis by wavelet transform was proposed by Mallat.[3] Through orthogonal or biorthogonal projections it enables us to decompose a gray-value image into a set of detail ''imagets'' with size decreasing with resolutions in the three main directions (for separable filtering scheme). In the wavelet packet analysis, each detail imaget is analyzed, as shown in Fig. 2. The last resolution level is not decomposed and the remaining information is contained in a small imaget that corresponds to a coarse view of the original image. We have shown previously[4] that biorthogonal wavelet basis based on the B-spline function family leads to very efficient implementation design. It is, indeed, possible to perform a real-time (video rate) still-image transform with a simple hardware setup involving only some field programmable gate array (FPGA) circuits. This is due to the shortness of the filters and to the simplicity of their coefficients, which can be expressed in the form of the power of 2. In this study, we use an orthonormal wavelet basis constructed from cubic B-splines functions called the Battle-Lemarié basis. The associated digital filters are infinite impulse response (IIR) but they have been truncated to 23 taps filters. This basis is the best among the most popular ones for image compression, but our experiment with other bases (Daubechies, biorthogonal) has shown that our conclusions concerning color transform and wavelet coefficient pruning are robust against a derating in the basis choice.

The decomposition over the wavelet basis is obtained by applying the transform on each of the components H_1, H_2, and H_3. We thus obtain a multiscale decomposition corresponding to the space-frequency resolution sensitivity of the visual system. To obtain a 2-D transformation we use a separable scheme (more details can be found in Ref. 8) that leads to creating for each resolution three wavelet components (see Fig. 3). Each component is dedicated to a particular orientation detail analysis. These three decomposition orientations (v, vertical; h, horizontal; and d,

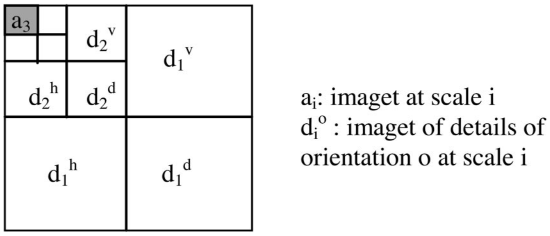


Fig. 3 The 2-D multiresolution analysis with a separable wavelet basis.

diagonal) are particularly adapted since the visual system have different detection sensitivities according to the orientations of the pattern (the lowest being along the diagonal lines). The principles of information suppression (rarely or not at all detected by the visual system) are presented in the next section.

Note especially that since the wavelet and the color transform are both linear, we can consider any permutation between these operations for the purpose of efficient implementation.

4 Compression

As previously said, in this paper we do not consider the quantization and the coding steps that are, naturally, important parts of any compression/decompression algorithm; some interesting elements for these stages can be found in Refs. 9 and 10. Therefore, the compression ratios are only those given by the entropy calculation and they must be considered as the maxima given by an ideal entropy coder. We propose a study for preprocessing under the form of a ‘‘pruning’’ of the wavelet coefficients, our aim being to maintain high visual quality for the reconstructed image. This pruning is performed in two steps, a scale pruning and then a block thresholding.

4.1 *Scale Pruning*

In the first step we set to zero the blocks of wavelet coefficients corresponding to resolution and color contrast that are not proved to be essential from a visual quality point of view. Indeed, a color image contains much more information than the human brain usually exploits.

We choose to limit our multiresolution analysis to a three-level depth. Going further in the analysis leads to underlining details that are so blurred in original image that they are of no significance for human eye (the image being seen from distance in good accordance with its size). From the wavelet coefficients obtained from the analysis we try to set to zero one set of coefficients belonging to one scale and one orientation. Then reconstruction is carried out and the reconstructed image is submitted to psychovisual appreciation. This experiment is repeated for each detail imaget of the set indexed by color axis (H_i), scale (j), and orientation (o), and each index varies from 1 to 3. It is assumed here that the global psychovisual effect of this pruning is somewhat linear and that when different acceptable alterations are added, the result is an acceptable global alteration. This assumption is certainly optimistic when pushing it to the limits, but in our work we stayed far from these limits, and we will see that the final counterexperiment confirms the results. The psychovisual appreciation gives way to a quantized assessment on a scale ranking from 1 to 6, where 1 is for a heavily degraded image, 3 corresponds to a good-looking image with a perceptible but slightly annoying impairments, and 6 is reserved for an image with no perceptible differences from the original. Table 1 presents the results obtained for three typical images. The last column shows the global resulting assessment for each scale pruning try. We consider that every result greater than 13 (enclosed cases) is acceptable for a quasi-lossless compression scheme. We notice that all details corresponding to H_2 and H_3 (chromatic plane) can be suppressed; this is in good accordance to the assumed low resolution of the human visual system for colored components (V_4 area) and to its high resolution for luminance contrasts. Only one block (a d block) of H_1, the achromatic axis, can be omitted. Note, also, that the wavelet packet analysis refinement does not lead to much better results and we can remove only four small imagets from H_1 plane, as shown in Fig. 4. These results confirm the well-known human lack of sensitivity to high-spatial-frequency components that are diagonally oriented. However, it seems more difficult to conclude about the comparative resolutions for color components H_2 and H_3.

Table 1 Scale pruning.

j	**o**	H_i	‘‘Flower’’	‘‘Mandrill’’	‘‘Lenna’’	
1	*h*	H_1	3	2	3	8
1	*h*	H_2	6	6	5	$\boxed{17}$
1	*h*	H_3	6	6	6	$\boxed{18}$
1	*v*	H_1	3	3	3	9
1	*v*	H_2	4	4	6	$\boxed{14}$
1	*v*	H_3	6	6	6	$\boxed{18}$
1	*d*	H_1	5	5	6	$\boxed{16}$
1	*d*	H_2	6	6	6	$\boxed{18}$
1	*d*	H_3	6	6	6	$\boxed{18}$
2	*h*	H_1	1	1	2	4
2	*h*	H_2	6	6	5	$\boxed{17}$
2	*h*	H_3	6	6	5	$\boxed{17}$
2	*v*	H_1	2	2	1	5
2	*v*	H_2	6	6	6	$\boxed{18}$
2	*v*	H_3	6	5	6	$\boxed{17}$
2	*d*	H_1	2	4	2	6
2	*d*	H_2	6	5	5	$\boxed{16}$
2	*d*	H_3	6	6	6	$\boxed{18}$
3	*h*	H_1	1	2	1	4
3	*h*	H_2	5	5	5	$\boxed{15}$
3	*h*	H_3	5	5	4	$\boxed{14}$
3	*v*	H_1	1	3	1	5
3	*v*	H_2	4	4	4	12
3	*v*	H_3	4	4	5	13
3	*d*	H_1	1	3	1	5
3	*d*	H_2	5	6	6	$\boxed{17}$
3	*d*	H_3	5	6	6	$\boxed{17}$

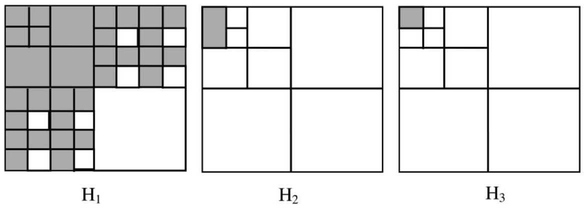


Fig. 4 Scale pruning. Imagets that can be removed are shown in white.

The images obtained after analyzing, pruning, and reconstruction were submitted to the psychovisual experiment described previously. Pruning was performed following the preceding conclusions so that all the unneeded blocks of wavelet coefficients are set to zero (see Fig. 4). Therefore, they correspond to a compression ratio of 1:4, and they were judged as five-level to six-level quality in our psychovisual test.

4.2 *Block Thresholding*

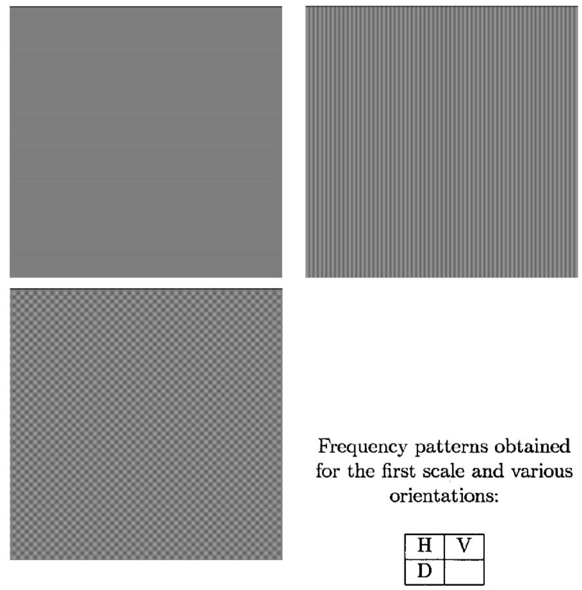
In the second step, a thresholding is applied in each block of the remaining coefficients. One threshold is chosen for each block in accordance with the dependance of visual resolution on the corresponding orientation of details.[1,5,11–13]

To determine those thresholds, frequency patterns were created using the multiresolution synthesis scheme of Mallat[3] and Garcia et al.[14] In this algorithm used to reconstruct a signal after its multiresolution analysis, the approximation imaget is set to a constant midrange value $(H_1=128, H_2=0, H_3=0)$, while the detail imagets are set to zero except the one under consideration. Detail coefficients of this block are set to an arbitrary value, say, x. After reconstruction, an image that is a frequency pattern for the chosen block (fixed by scale, orientation, and color component) is obtained. Some examples of such frequency patterns obtained for the first scale are presented in Table 2. The visual sensitivity to these frequency patterns is measured by determining the minimum values of x that enable visual perception of each pattern.

The results concerning the thresholds found for frequency pattern detection are presented in Table 3. The smaller the threshold, the greater is the human sensitivity. When the threshold is equal to zero, it means that wavelet coefficients belonging to this block must be preserved in the compression process because the eye is very sensitive to this frequency pattern. Note that these results are quite consistent with the presented assumption concerning the spatial frequency sensitivity of human visual system. Indeed, on the one hand, thresholds decrease when scale increases (the greater the scale, the lower is the resolution); this reflects a greater sensitivity for low than for high spatial frequencies. On the other hand, horizontal detail blocks are better detected than vertical detail blocks, themselves being more important than the diagonal ones. Block $j=1$, $o=D$ is not tested because scale pruning has shown that it can be totally thresholded (i.e., set to zero).

Finally, the thresholdings were applied on the remaining parts of the analyzed images after scale pruning. The reconstructed images obtained after such a treatment were submitted to our psychovisual test and the results seem good because the assessments were between 4 (perceptible but not annoying impairments) and 6 (imperceptible impairments).

Table 2 Frequency patterns obtained for the first scale.



In conclusion, the results obtained after a large series of tests based on psychovisual estimations rather than on pure peak SNR (PSNR) evaluation are in good accordance with the assumed properties of the human visual perceptive system. Table 4 provides some examples of compression ratios obtained on our classical test images for quasi-lossless compression. These compression ratios were computed from the entropy of the coded images; therefore they are only the optimal values that can be obtained using entropy coding (such as the Huffman algorithm) without any scalar or vector quantization.

5 Figure of Merit for Color Image Compression

During our psychovisal experiments a very poor correlation between peak signal over noise ratio [the PSNR definition is recalled in Eq. (2)] values and psychovisual assessments for color images has been noticed. It is obvious from

Table 3 Thresholds for frequency patterns.

Scale j	Orientation o	Threshold x
1	H	2
1	V	5
2	H	0
2	V	0
2	D	1
3	H	0
3	V	0
3	D	0.2

Table 4 Compression ratios after pruning and thresholding.

Images	Size	PSNR (H_1)	PSNR (H_2)	PSNR (H_3)	PSNR ($H_1H_2H_3$)	Compression ratio
"Flower"	368×350	22.13 dB	39.96 dB	65.80 dB	26.83 dB	1:11
"Mandrill"	512×480	14.93 dB	32.81 dB	34.57 dB	19.58 dB	1:6
"Lenna"	512×480	19.75 dB	35.90 dB	44.10 dB	24.40 dB	1:10

Table 4 that the PSNR is not a reliable criterion for color image quality if we consider the human perception system as a reference.

$$\mathrm{PSNR}=10\log_{10}\left(\frac{3T\Delta^2}{\Sigma_{i\in I_\alpha}d_i^2}\right),$$

$$\Delta=\text{image data range (256 for an 8-bit coded image)}, \tag{2}$$

$$T=\text{image size}, \tag{3}$$

$$d_i^2=(r_i-\widehat{r_i})^2+(g_i-\widehat{g_i})^2+(b_i-\widehat{b_i})^2, \tag{4}$$

$$r_i,g_i,b_i=\text{pixel } i \text{ in the R,G,B (or } H_1,H_2,H_3\text{) color plane}. \tag{5}$$

Typically, some image judged as excellent has a global PSNR less than 20 dB! Furthermore, the PSNR measured on good-quality images depends more on image content than on psychovisual image quality. On the other hand, we noticed that the sensitivity of human vision to certain chromatic detail components is very low. Whereas some other figures of merit for color images are well known, the relevancy of the multiresolution analysis for human vision point of view tends to suggest that a quality criterion based on the experiments presented in this paper could be of interest. Therefore an estimation of the PSNR computed on multiresolution decompositions of the H_1, H_2, and H_3 images is proposed here, with the reference image being analyzed in the same manner. The global PSNR is computed from mean square error between reconstructed and reference images. This quadratic error is weighted by coefficients α_n^o ($\in[0,1]$), which depend on scale (n), detail orientation (o), and color plane. These coefficients are deduced from the assessments A_n^o given in Table 1:

$$\alpha_n^o=1-\frac{A_n^o}{18}.$$

Therefore the imagets are taken into account with respect to their relevancy from human eye point of view. The resulting multiresolution figure of merit is defined by

$$\mathrm{FM}=10\log_{10}\left\{\frac{\Delta^2\Sigma_{k=1}^3\lambda_k}{\Sigma_{k=1}^3[\Sigma_{i\in I_\alpha}(x_{i,k}-\widehat{x_{i,k}})^2+\Sigma_{n=1}^3 4^{n-1}\Sigma_{o\in O}\alpha_{n,k}^o\Sigma_{i\in I_n^o}(x_{i,k}-\widehat{x_{i,k}})^2]}\right\} \tag{6}$$

where

- $x_{i,k}$ = pixel i of the H_k plane
- Δ = transform dynamic range
- λ_k = $T+\Sigma_{n=1}^3\Sigma_{o\in O}\alpha_{n,k}^o T$
- I_α = approximation subimage
- I_n^o = subimage of level n and orientation o
- o = {horizontal, vertical, diagonal}
- $\alpha_{n,k}^o$ = imaget weight (H_k plane)
- T = image size

The term $\lambda\Delta^2$ is a normalization factor and the 4^{n-1} coefficient enables us to give the same importance to all the imagets whatever their surface is. Table 5 gives the resulting values for FM and PSNR (RGB and $H_1H_2H_3$) for our compressed images. It is clear that FM values are more consistent with the psychovisual quality of the images than PSNR ones.

The assessments given by FM and PSNR to psychovisual perception in a large series of tests are compared. It is clear that this point has to be completed by comparison

Table 5 Quality criteria comparison.

Images	PSNR (RGB)	PSNR (H_1,H_2,H_3)	FM
"Flower"	19.75 dB	26.83 dB	60.58 dB
"Mandrill"	12.49 dB	19.58 dB	40.29 dB
"Lenna"	15.86 dB	24.40 dB	47.17 dB

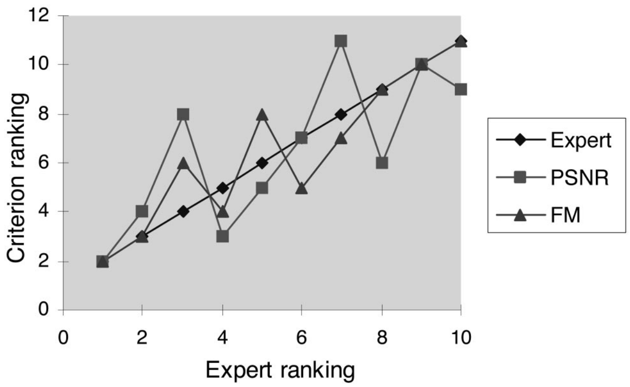


Fig. 5 Comparison of PSNR and FM.

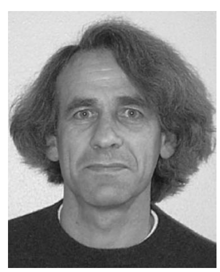
with the other classical figures of merit (CIELAB, RLAB, etc.). The short presentation given here is just to be considered as introductive to a larger study dedicated to this very point. These tests were conducted using a set of 10 variously corrupted (with noise and/or blur) images. The results are summarized in Fig. 5, where the images are ordered with respect to their quality from human expert point of view. We can see that FM gives a better estimation than PSNR. An ideal result should be a complete accordance between human expert and criterion ranking (the points should be only on the first diagonal). This promising result should be confirmed by other tests led on larger series of images affected by corruption better in accordance to our aim: Joint Photographic Experts Group-Descrete Cosine Transform (JPEG-DCT) reconstructed images, halftoned images, etc.

6 Conclusion

This study is the first stage of a complete approach of the compression problem of color still image with good visual restitution. The results already obtained are consistent with the known features of human vision characteristics and they show that the multiresolution analysis provided by the wavelet packet transform on a well-chosen color space is a good way to take into account the dependency of the resolution on the color contrast and on the detail orientation. It is also a good way for quantized quality evaluation of color images. This work must be continued to define, with the same psychovisual requirements, the best quantization stage for a complete compression/decompression algorithm. Our multiresolution figure of merit should be a good criterion to be used in this task.

References

1. B. A. Wandell, *Foundations of Vision*, Sinauer Associates, Sunderland (1995).
2. M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, ''Image coding using wavelet transform,'' *IEEE Trans. Image Process.* **1**(2), 205–220 (1992).
3. S. Mallat, ''A theory for multiresolution signal decomposition; the wavelet representation,'' *IEEE Trans. Pattern. Anal. Mach. Intell.* **11**(7), 674–693 (1989).
4. F. Truchetet and A. Forys, ''Implementation of still-image compression-decompression scheme on FPGA circuits,'' in *Still Image Compression*, *Proc. SPIE* **2669**, 66–75 (Jan. 1996).
5. F. Truchetet, B. Joanne, O. Laligant, and L. L. Yan Voon, ''A color image compression scheme based on psychovisual criteria,'' in *Proc. IS&T/SIDs 5th Color Imaging Conf.*, pp. 136–140, Scottsdale (Nov. 1997).
6. L. Brun and J. P. Braquelaire, ''Une amélioration des méthodes de quantification de couleurs par partition dynamique,'' *2ème Journées AFIG*, Toulouse, France (1994).
7. A. Gottshalk and G. Buchbaum, ''Information theoretic aspects of color signal processing in the visual system,'' *IEEE SMC*, **13**(5), 864–873 (1983).
8. I. K. Eom, H. S. Kim, and K. S. Son, ''Image coding using wavelet transform and human visual system,'' in *Still-Image Compression*, *Proc. SPIE* **2418**, 176–183 (1995).
9. S. A. Rajala, H. J. Trussell, and B. Krishnakumar, ''Visual sensitivity to color-varying stimuli,'' *Proc. SPIE* **1666**, 375–386 (1992).
10. S. A. Rajala, ''Impact of human visual perception of color on very low bit-rate image coding,'' *Proc. SPIE* **2308**, 39–46 (1994).
11. H. R. Wilson and D. Regan, ''Spatial-frequency adaptation and grating discrimination: predictions of a line-element model,'' *J. Opt. Soc. Am.* **1**, 1091–1096 (1984).
12. M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Prentice-Hall, Englewood Cliffs, NJ (1995).
13. J. M. Foley and G. E. Legge, ''Contrast detection and near-threshold discrimination in human vision,'' *Vision Res.* **21**, 1041–1053 (1981).
14. N. Graham and J. Nachmias, ''Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple-channel model,'' *Vision Res.* **11**, 251–259 (1971).
15. A. Garcia, F. Truchetet, O. Laligant, C. Dumont, E. Verrecchia, and M. A. Abidi, ''Multiscale analysis of 3D surface image: application to clam shell characterization,''in *Three-Dimensional Image Capture and Applications*, *Proc. SPIE* **3313**, 126–133 (1998).



Fred Truchetet received the masters degree in physics at Dijon University, France, in 1973 and a PhD in electronics at the same university in 1977. He was with Thomson-CSF for two years as a research engineer and he is currently full professor and the head of Le2i, Laboratory of Electronic, Computer and Imaging Sciences, at Université de Bourgogne. His research interests are focused on image processing for artificial vision inspection and particularly on wavelet transforms, multiresolution edge detection, and image compression. He is member SPIE, IEEE, and ISIS (a research group of the CNRS).

Olivier Laligant received his PhD degree in 1995 from the Université de Bourgogne, France. He is assistant professor of the Computing, Electronic, Imaging Department (Le2i) in Université de Bourgogne. His research interests are focused on multiscale edge detection and wavelet transforms.