Multilayered Image Representation: Application to Image Compression

François G. Meyer, *Member, IEEE*, Amir Z. Averbuch, *Member, IEEE*, and Ronald R. Coifman

***Abstract*—The main contribution of this work is a new paradigm for image representation and image compression. We describe a new multilayered representation technique for images. An image is parsed into a superposition of coherent layers: piecewise smooth regions layer, textures layer, etc. The multilayered decomposition algorithm consists in a cascade of compressions applied successively to the image itself and to the residuals that resulted from the previous compressions. During each iteration of the algorithm, we code the residual part in a lossy way: we only retain the most significant structures of the residual part, which results in a sparse representation. Each layer is encoded independently with a different transform, or basis, at a different bitrate, and the combination of the compressed layers can always be reconstructed in a meaningful way. The strength of the multilayer approach comes from the fact that different sets of basis functions complement each others: some of the basis functions will give reasonable account of the large trend of the data, while others will catch the local transients, or the oscillatory patterns. This multilayered representation has a lot of beautiful applications in image understanding, and image and video coding. We have implemented the algorithm and we have studied its capabilities.**

***Index Terms*—Adaptive coding, cosine transforms, image coding, multilayered coding, wavelet transforms.**

I. Introduction

THE underlying assumption behind transform coding is that the basis $\{\psi_n\}$ (e.g., a discrete cosine transform (DCT) basis, or a wavelet basis) used for compression is well adapted to most images. This assumption is clearly violated by the following observation, made by the authors in [1]: at low bit rates, the distortion depends on the ability of the basis to approximate the image f with a very small number of coefficients α_n

$$\min_{\alpha_n} \left\| f - \sum_{n=0}^{M} \alpha_n \psi_n \right\|; \qquad M \ll N. \tag{1}$$

A consequence of this observation is that one should be able to reduce the distortion by replacing an orthonormal basis with a richer library of basis functions. The size of such a library is typically much greater than the effective dimension, N, of the input space. One can then exploit this redundancy by choosing among many possible representations that expansion of the image which results in the best approximation with a very small number of coefficients. This fact has been observed for a long time, and it was one of the motivations for the construction of the wavelet packets, and cosine packets libraries. Because the size of these libraries become exponentially large as the size of the image increases, the authors in [2] have devised an astute dynamic programming strategy in order to extract the "best basis" from a large library of basis functions. As was observed in [3] the best-basis is always trying to find a compromise between two conflicting goals: 1) describe the large scale piecewise smooth regions and 2) describe the local textures. For this reason, the best basis rarely provides the optimal transform to compress large classes of images (such as the so-called "natural images"). As was observed in [4], if the signal is composed of highly nonorthogonal components, then the method may not yield a sparse representation. By dropping the "orthonormal basis" constraint, it becomes in principle possible to match the local textures with localized cosine functions (for instance), and the piecewise smooth regions with wavelets. Mallat and Zhang [5] proposed to use the projection pursuit algorithm [6] to identify such components. Because the projection pursuit algorithm is only a greedy algorithm it suffers from some serious limitations: 1) even if the initial image is a finite (possibly small) linear combination of the vectors of the library, the projection pursuit algorithm is not guaranteed to recover the components and 2) the algorithm converges very slowly as one increases the number of terms in the approximation. An interesting alternative was proposed by Chen [4]: for any given signal f, one constructs a basis from vectors ψ_n of the library in such a way that the l_1 norm of the coefficients, $\sum_{n=0}^{N-1} |\alpha_n|$, is minimized. Minimizing the l_1 norm will result in a very sparse representation of f. Furthermore, the authors proved that if the original signal f is a very sparse combination of some vectors of the library, then the coefficients α_n can be recovered with perfect accuracy. Unfortunately the overall complexity of the algorithm is very high: $\mathcal{O}(N^{3.5} \log_2^{3.5} N)$, which makes this approach of little practical use for image compression.

In this work we propose a general framework for image representation. Our hypothesis is that an image can be decomposed as the sum of two layers: a "cartoon image" and a texture map. The cartoon image provides a description of the salient parts, or edges, inside the image, as well as the piecewise smooth changes in the illumination. The texture map permits to fill in the texture in the regions enclosed by edges. We advocate that the cartoon and texture map should be represented with two different sets of basis functions. We propose to represent the cartoon image with

Manuscript received October 1, 1998; revised May 23, 2002. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Kannan Ramchandran.

F. G. Meyer is with the Department of Electrical Engineering, University of Colorado, Boulder, CO 80309-0425 USA (e-mail: francois.meyer@colorado.edu).

A. Z. Averbuch is with the Department of Computer Science, School of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel.

R. R. Coifman is with the Department of Mathematics, Yale University, New Haven, CT 06510.

Publisher Item Identifier 10.1109/TIP.2002.802527.

1057-7149/02$17.00 © 2002 IEEE

wavelets. Because a cartoon is composed of edges, one should really be using brushlets [7] or ridgelets [8]. Unfortunately there exists no discrete orthonormal ridgelet transform, and brushlets are still too imprecise. Our choice of tensor products of wavelets is therefore suboptimal, but as better libraries become available they can be used within our framework. Our second hypothesis is that local cosine (or wavelet packet) bases are better suited than wavelets to represent periodic textures. Our recent experiments [9] with textured images taken from Brodatz's book, and the MIT VisTex database, indicate that an advantage can be gained by using local cosine bases over wavelets to encode images that contain periodic textures. The concept of multilayer representation was used successfully for removing noise from audio signals [10], and coding audio signals [11]. It was natural to extend the concept to the problem of sparse representation of images, or image compression. Related ideas have appeared recently in the image processing literature. In [12] the authors propose to replace the DCT used in JPEG with several different transforms that would be better adapted to the local statistics within each block. Our approach is not based on a block by block division of the image, but rather on a model of several layers. In [13], the authors describe a lossless coding algorithm that is based on the "lossy-plus-residual" concept. The image is first coded with a wavelet basis, and the residual is then encoded in a lossless manner with wavelet packets. The lossy part of their coder is in effect a simple wavelet coder. In this work we intend to address a more general and deeper problem, where one wants to approximate an image using a superposition of coherent layers: smooth-regions and edges layer, textures layer, etc. The rest of the paper is organized as follows. In the next section, we provide a detailed description of the algorithm. In Section III, we discuss a key feature of the algorithm: the ability to obtain a specialized, or tailored, basis to encode each layer R^i. In Section IV, we describe the quantization and ordering of the local cosine and wavelet packets coefficients. Results of experiments are presented in Section V.

II. Multilayered Image Compression

A. Cascade of Compressions

A block diagram of the multilayered algorithm is shown in Fig. 1. The multilayered compression algorithm consists in a cascade of compressions applied successively to the image itself and to the residuals that resulted from the previous compressions. An initial main approximation is obtained by compressing the input image with a wavelet basis. This first approximation preserves the general shape of the image, and captures the trend in the intensity function. As shown in the experiments, this first layer provides a "cartoon," or segmentation, of the original image. We then reconstruct the compressed part, and we calculate the error between the original and compressed data. This compression error defines the first remainder, or *residual*. Residuals are composed of textures, and are compressed with wavelet packets or local trigonometric bases. These bases are well adapted to texture coding [3], [9]. Once the first residual is compressed, one defines the second residual as being the compression error of the first residual. The algorithm keeps on compressing the successive residuals until we reach a residual that contains no more structure. We describe now in details the different stages of the algorithm. We consider a sequence of libraries of functions $(\mathcal{L}_i)_{i\in\mathbb{Z}}$. One can construct very large collections of orthonormal bases from $\mathcal{L}_i$. In this work $\mathcal{L}_0$ contains only one single wavelet basis, and $\mathcal{L}_1$ is a library of wavelet packets, or a library of local cosine basis functions.

1) Initialization: Let I be an image. We first compress I over the library $\mathcal{L}_0$, using the budget b_0. The approximation is performed under a budget constraint, and the result of the approximation should be described with at most b_0 bits. Let $\hat{R}^0$ be the decoded image after decompression. $\hat{R}^0$ is an approximation of the original image I, and we have

$$I = \hat{R}^0 + R^1, \qquad \text{with} \quad \hat{R}^0 = \sum_{j\in E_0} q_j^0 \psi_j^0 \tag{2}$$

where R^1 is the approximation error. At this point, we refine the approximation of I by calculating an approximation of the residual R^1. This is achieved by compressing the residual. But in order to discover different features in the image, we use a different library to compress R^1. We use a budget of b_1 bits to compress R^1. A best basis, $\{\psi_j^1,\ j\in E_1\}$, that provides the optimal compression $\hat{R}^1$ of R^1, is constructed from elements of the library $\mathcal{L}_1$

$$R^1 = \hat{R}^1 + R^2, \qquad \text{with} \quad \hat{R}^1 = \sum_{j\in E_1} q_j^1 \psi_j^1 \tag{3}$$

where $\{q_j^1,\}_{j\in E_1}$ are the quantized coefficients, and E_1 is the set of indices of the basis functions that constitute the best basis. We now reconstruct a second approximation $\hat{I}^1$ of I

$$\hat{I}^1 = \hat{R}^0 + \hat{R}^1 \tag{4}$$

where $\hat{I}^1$ is an image that can be encoded with $b_0 + b_1$ bits.

2) Main Loop of the Algorithm: Fig. 1 shows the main loop of the algorithm. Let us assume that we have carried the approximation of I up to step $n-1$. Let R^n be the residual of the approximation at step $n-1$. A best basis, $\{\psi_j^n\}_{j\in E_n}$, that provides the optimal approximation $\hat{R}^n$ of R^n with b_n bits, is constructed from the library $\mathcal{L}_n$

$$R^n = \hat{R}^n + R^{n+1} \qquad \text{with} \quad \hat{R}^n = \sum_{j\in E_n} q_j^n \psi_j^n \tag{5}$$

where E_n is the set of indices of the basis functions that constitute the best basis. Finally, we reconstruct an approximation of I using the $n+1$ compressed residual images $\hat{R}^0,\ \hat{R}^1,\ \hat{R}^2,\ \ldots$

$$\hat{I}^n = \sum_{k=0}^{n} \sum_{j\in E_k} q_j^k \psi_j^k \tag{6}$$

where $\hat{I}^n$ is an image that can be compressed with a budget of $\sum_{k=0}^{n} b_k$ bits. The coefficient q_j^k of the nonlinear approximation (6) is the quantized inner product $\langle R^k,\ \psi_j^k\rangle$.

B. Budget Allocation

For the class of images that we consider in this work we always choose $\mathcal{L}_0$ to be a wavelet basis. The second library $\mathcal{L}_1$ is

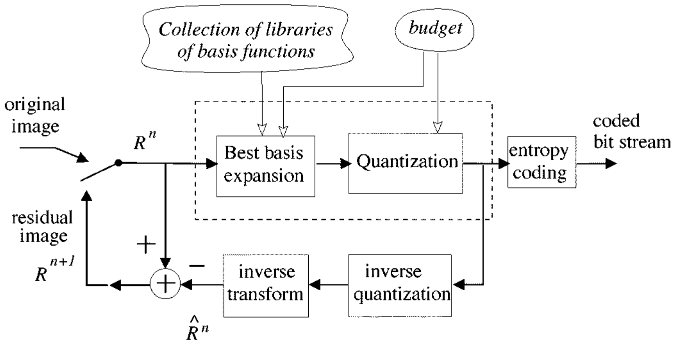


Fig. 1. Block diagram of the multilayered compression algorithm. In the first pass of the algorithm, the switch (on the left) is turned toward the original image (i.e., the image that we want to compress). In the subsequent refinement passes the switch is turned toward the residual image. The block in the dotted line compresses either the original image, or the residual. This single pass compression consists of two parts: 1) best basis selection, and calculation of the coefficients of the image using the best basis and 2) ordering of the coefficients, and quantization of the stream of coefficients. Finally, the quantized coefficients are entropy coded. The residual error is calculated, and is fed back to the compression algorithm.

either a library of local cosines, or wavelet packets. As shown in [3] and [9], these two libraries are well adapted to the representation of periodic textures. Once the library of the second layer is chosen, the only parameters that remain to be determined are the numbers of bits allocated to each layer. After a large number of experiments we made the following observation: the optimal bit allocation strategy (that we computed using an exhaustive search) is always of the following two types:

1) the main part of the bit budget is allocated to the wavelet layer, and very few bits are kept for the textural layer;
2) the main part of the budget is allocated to the textural layer, and very few bits are allocated to the wavelet layer.

One can interpret these two bit allocation policies as a partition of the set of images into two different classes:

1) images that can be easily compressed in a wavelet basis, but that also contain small areas of periodic textures;
2) images that contain mostly patches of periodic textures. The wavelet layer for such images provide a very coarse segmentation of the image.

In order to assign a new image to one of the two classes, we compute the wavelet expansion and the local cosine expansion of the image, and we compare the decay of the coefficients in each basis. Fig. 2 shows the decay of the wavelet coefficients and local cosine coefficients for two different images: Lena and roofs. The Lena image is a wavelet-friendly image that can be very easily coded with a wavelet transform. The roofs image (see Figs. 6 and 7) contains many patches of periodic texture. As was shown in [9], this image can be well compressed in a local cosine basis. For both images the decay of the coefficients is of the form $1/j^\gamma$, where j is the index or the ranked (cosine or wavelet) coefficient. The wavelet coefficients of the Lena image have a faster decay than the local cosine coefficients. The reverse phenomenon occurs for the image roofs.

Once we know what is the type of the image (wavelet versus textural), we can rapidly search for the optimal bit allocation policy. Because the budget of the coarse layer (local cosine or wavelet) varies only over a small range, one can rapidly search for the optimal number of bits allocated to the coarse layer.

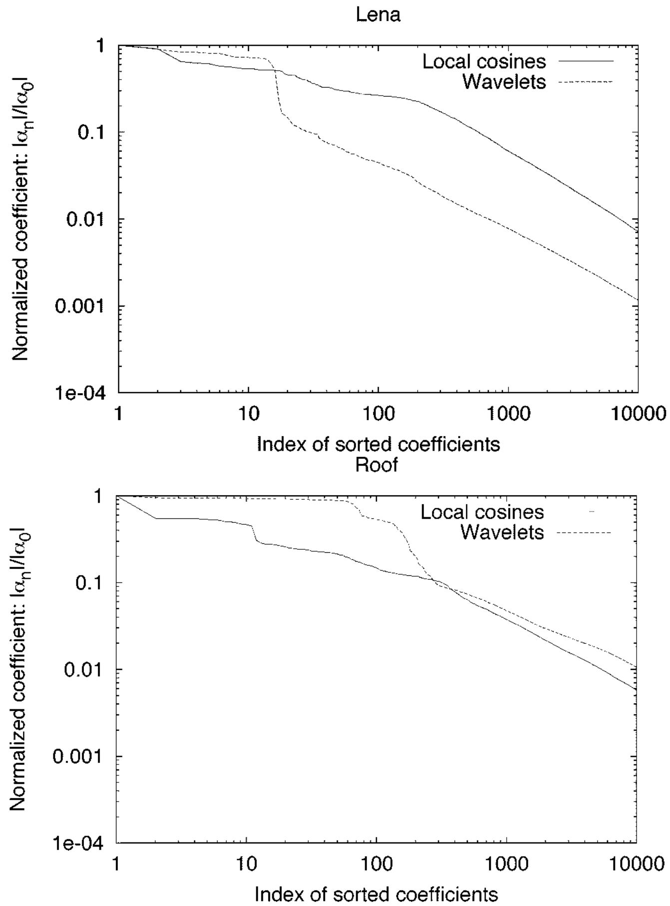


Fig. 2. Decay of the normalized coefficients in the local cosines and wavelet bases for the images (top) Lena and (bottom) roofs. The coefficients are ranked according to their magnitude, and normalized by the largest coefficient to compensate for different normalization factors in the transforms. The decay is of the form $1/j^\gamma$, where j is the index or the ranked (cosine or wavelet) coefficient.

C. Geometric Interpretation

We consider the case where we only have two layers: a wavelet layer, and (for instance) a local cosine layer. For the sake of simplicity we replace the quantization by a linear projection on the subset of basis vectors for which the quantized coefficients are nonzero. The multilayer algorithm can then be interpreted as a sequence of projections on successive subspaces (see Fig. 3). The first subspace, W_0, is spanned by the subset of wavelet basis functions ψ_j^0 for which the quantized coefficients q_j^0 of the original image are nonzero. The second subset, W_1, is formed by the subset of local cosine basis functions ψ_j^1 for which the quantized coefficients q_j^1 of the residual image R^1 are nonzero, etc.

Unfortunately, the two subspaces W_0 and W_1 are not orthogonal to each others, and therefore one should replace the orthogonal projection on each subspace by an oblique projection parallel to the other subspace. As shown in Fig. 3, if one uses orthogonal projections then $\hat{R}^0$ is overestimated. The final

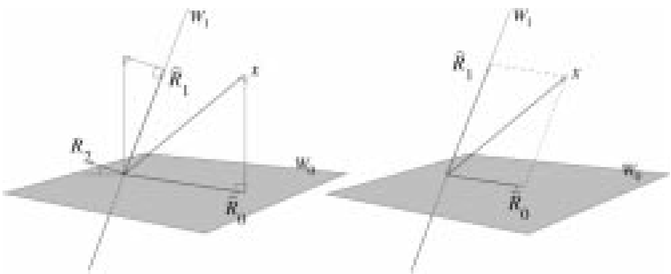


Fig. 3. Left: the vector x is decomposed into $\hat{R}^0$, $\hat{R}^1$ and R^2. At each time the residual is projected on a new subset of vectors. Right: we replace orthogonal projections by oblique projections. The vector x is now decomposed into exactly two vectors $\hat{R}^0$ and $\hat{R}^1$.

residual R^2 is highly correlated with $\hat{R}^0$, and should be added back to $\hat{R}^0$ to correct the initial overestimation.

D. Toy Example

Fig. 4 illustrates the principle of the algorithm. The first iteration of the algorithm is shown in the top of Fig. 4 where the piecewise smooth variations of the intensity is described by the layer $\hat{R}^0$ with very few wavelet coefficients (compression ratio $= 750$). During the second pass of the algorithm, we compress the residual $R^1 = I - \hat{R}^0$ with an adapted local cosine basis (compression factor $= 70$). The result of the compression, $\hat{R}^1$, is shown in the center of Fig. 4, and the second residual $R^2 = R^1 - \hat{R}^1$, is shown on the bottom. $\hat{R}^1$ constitutes the second layer. Nearly all the texture has been removed from the image $\hat{R}^0$, and is coded in the second layer $\hat{R}^1$. As a result, most of the features present in the image have been coded in either one of the first two layers, and the final residual R^2 (see Fig. 4—bottom) appears as random noise. When $\hat{R}^0$ and $\hat{R}^1$ are added together, we obtain an image which is compressed by a factor of 64 (see Fig. 7).

Visual inspection of the second layer $\hat{R}^1$ in Fig. 4—center suggests that this textural layer should be coded with two-dimensional (2-D) oscillatory patches, such as 2-D local trigonometric bases. In order to corroborate this visual and geometric intuition, we show in Fig. 5 the decay of the wavelet and local cosine coefficients of the residual image R^1. The coefficients were sorted by decreasing order of magnitude, and only the first 10 000 coefficients are displayed. It is clear from Fig. 5 that the wavelet coefficients have a slower decay than the local cosine coefficients, indicating that the local cosines are better suited than the wavelets for coding R^1.

III. Best Basis Selection

A. Libraries of Basis Functions

One of the key tenet of the multilayered representation is a mechanism to obtain a specialized, or tailored basis to encode each layer R^i. Examples of such libraries include the following.

• **Local Trigonometric Functions** [14]. Local trigonometric transforms provide an adaptive segmentation of the spatial domain in terms of oscillating patterns. An image is decomposed into overlapping blocks of different sizes within which a local Fourier expansion, or a DCT, is performed. Instead of abruptly cutting blocks in the image, we use *smooth orthogonal projectors* [14], [15]. We conducted an extensive study [9] using images with periodic textures, and we were able to demonstrate that an advantage can be gained by using local cosine bases over wavelets to encode periodic texture.

• **Wavelet Packets** [16]. Loosely speaking, wavelet packets make it possible to adaptively tile the frequency domain into different bands of arbitrary size. Wavelet packets have been used to characterize textures, and code textural images [3], [17], [18]. However, an elementary 2-D wavelet packet always displays a "criss-cross" pattern, which comes from its two symmetric peaks in the Fourier domain. As a result one needs a combination of several wavelet packets to characterize a single 2-D pattern oscillating along one direction [7].

Another example of library is the collections of

• **Brushlets** [7]. Brushlets are new families of steerable wavelet packets that adaptively segment the Fourier plane to obtain the most concise and precise representation of the image in terms of oriented textures with all possible directions, frequencies, and locations.

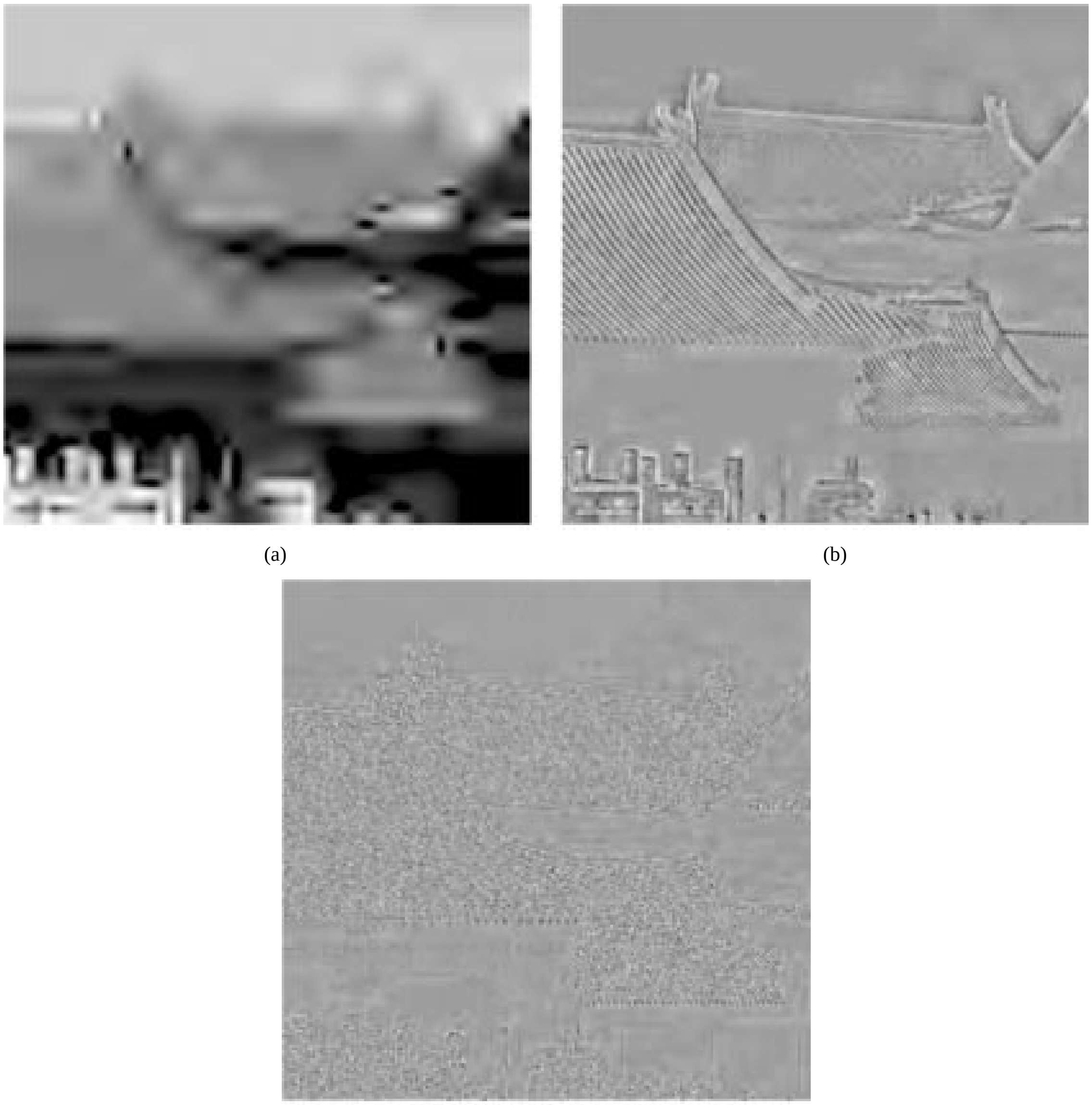
Because each library is overcomplete, it is possible to obtain a very sparse representation of each layer R^i by "tailoring" its representation. Even though one could work with other overcomplete representations that do not necessarily contain orthogonal bases [19], [5], the libraries of orthonormal bases offers many advantages: i) they provide very large subcollection of orthogonal bases, ii) in an orthogonal basis the decomposition, and the reconstruction can be performed using very fast algorithms, which are numerically exact and stable, and iii) there exist some fast algorithms that can be applied in real time, to select the optimal decomposition over the library [2].

B. Choice of a Cost Function $\mathcal{M}$

Coifman and Wickerhauser [2] suggested to use a fast dynamic programming algorithm to search for that *best basis* which is optimal according to a given cost function $\mathcal{M}$. In this work, one basis is better than another if it provides a better reconstruction quality for the same number of bits spent in coding the coefficients. Ramchandran and Vetterli [18] wedded the bit allocation algorithm of Shoham and Gersho [20] to the best basis algorithm [2]. Unfortunately, their approach is extremely computationally intensive: the problem in [18] involves three layers of nonlinear approximations, only one of which lends itself to a fast algorithm. Instead of using the rate distortion framework, we designed a cost function that returns an estimate of the actual rate achieved by each node. The cost function mimics the actual scalar quantization, and entropy coding, which are presented in Section IV. However, the cost function is much faster to compute. It is composed of two complementary terms:

- $c_1(\mathbf{x})$, the cost of coding the sign and the magnitude of the nonzero output levels of the scalar quantizer;
- $c_2(\mathbf{x})$, the cost of coding the locations of the nonzero output levels (significance map).

Let $\mathbf{x} = \{x_k\}$. A first order approximation to the cost of coding the magnitude of the output levels $\{|Q(x_k)|\}$



(a) (b)

(c)

Fig. 4. (a) The first layer, $\hat{R}^0$, is the decoded image after a wavelet compression by a factor 750 of the original. The difference between the original image, and this layer constitutes the first residual R^1. (b) The second layer, $\hat{R}^1$, is reconstructed after a compression by a factor 70 of the first residual R^1 using a local cosine basis. (c) The second residual R^2 is the difference between the first residual R^1 and the second layer $\hat{R}^1$.

is given by the number of bits needed to represent the set $\{|Q(x_k)|,\, k/Q(x_k) \neq 0\}$

$$c_1(\mathbf{x}) = \sum_{k/Q(x_k)\neq 0} \max\left(\log_2|Q(x_k)|,\, 0\right). \tag{7}$$

The second term, $c_2(\mathbf{x})$, is calculated using the first order entropy of a Bernoulli process: each coefficient x_k is significant with a probability p, and we assume that the significance of the coefficients are independent events. We get

$$c_2(\mathbf{x}) = -N\left(p\,\log_2(p) + (1-p)\,\log_2(1-p)\right). \tag{8}$$

The computation of the cost function requires to quantize the coefficients. A first estimate of the quantization step is required to compute the cost function. A second pass uses the actual quantization step obtained after quantization.

IV. Quantization

A. Laplacian Based Salar Quantization

The distributions of the cosine and wavelet packet coefficients are approximated with a Laplacian distribution. The Laplacian distribution yields tractable computations of the optimal entropy constrained scalar quantizers [21]. We use a particularly efficient near optimal scalar quantizer, with a symmetric dead-zone, and a reconstruction offset [21].

B. Ordering of the Coefficients and Entropy Coding

After quantization, the positions of the nonzero output levels are recorded in a significance map. The lossless compression of the significance map takes advantage of the fact that large output levels often appear in clusters. If one uses wavelet packets, subbands are scanned by increasing frequency. We then scan all the pixels inside any given subband using a Hilbert space filling

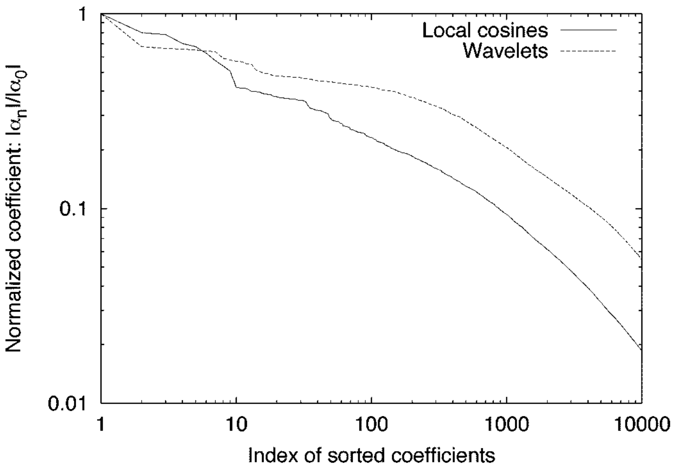


Fig. 5. Decay of the normalized coefficients in the local cosines and wavelet bases of the residual R^1. The wavelet coefficients have a slower decay than the local cosine coefficients.

Fig. 6. Roofs at 0.125 bit per pixel, SPIHT, and PSNR = 23.77 dB.

curve [3]. If one uses local cosines, one gathers together coefficients with similar two dimensional frequencies [9]. One first divides each block into a fixed number of frequency subsets: in each subset the coefficients have similar two-dimensional frequencies. We then gather from all the blocks all the coefficients that are in the same subset. The signs of the output levels are not entropy coded, but are simply packed. The magnitude of the output levels are variable length encoded, using an arithmetic coder to encode the length. The best basis geometry is described by a quadtree. We code the quadtree, with an adaptive arithmetic coder.

V. Experiments

We have implemented the coder and decoder, and an actual bit stream is generated by the coder. For each experiment we

Fig. 7. Roofs at 0.125 bit per pixel, multilayer, and PSNR = 25.23 dB.

generated a compressed file with a size equal to the targeted budget. We present the results of the multilayer compression algorithm, using the following test images.

1) Roofs, 8 bpp, 512 × 512. This image is part of the MIT VisTex database. It is composed of a mixture of periodic texture (roofs), as well as smooth regions (façades and sky).

2) Barbara, 8 bpp, 512 × 512. It is the standard image of the lady with the stripes and checker tablecloth.

3) Clown, 8 bpp, 512 × 512. It is the standard image of the clown.

These images are difficult to compress because they contain a mixture of large smooth regions, and long oscillatory patterns. In order to evaluate the performance of our algorithm, we compared it to one of the best wavelet coder that was available to us: the SPIHT wavelet coder of Said and Pearlman [22].

Roofs: Because this image contains large regions with periodic texture such as the tiles on the roofs of the buildings, we expect the combination of wavelets and local cosines to perform well. The optimal choice for the second layer was indeed the local cosine library. As shown in Fig. 6, the wavelet coder could not preserve the tiles on the roof of the buildings. In fact, our coder outperformed SPIHT by 1.1 to 1.63 dB (see Table I). While some ringing artifacts are visible Fig. 7 in the multilayer image on the top of the roof (where the intensity abruptly changes), similar artifacts are also visible in the wavelet coded image. Fig. 8 illustrates the ability of the local cosine to preserve the periodic texture on the roof, even at small bit-rates.

Barbara: The local cosines provided again the optimal library to encode the second layer. Our coder outperformed SPIHT by 0.7 to 1.60 dB (see Table II). The periodic texture on the pants of the lady is very well preserved by the multilayer. The texture on the tablecloth is also well rendered (compare Figs. 9 and 10). Unfortunately, some criss-cross patterns appear on the face of Barbara (see Fig. 10). These patterns come from

TABLE I
ROOFS: PSNR (IN DECIBELS) FOR VARIOUS BIT RATES

Rate (bpp)	0.0625	0.125	0.25	0.50	0.75	1.00
Multi-layer	22.61	25.23	28.51	32.79	35.69	38.05
SPIHT	21.51	23.77	27.02	31.18	34.18	36.68

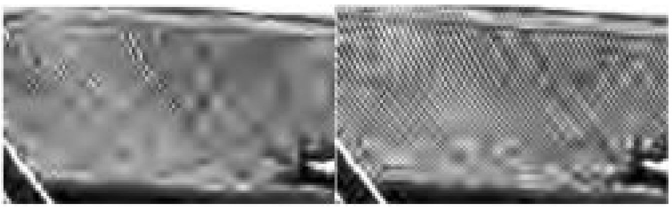


Fig. 8. Detail of the roofs. Left: wavelet coder and right: multilayer.

TABLE II
BARBARA: PSNR (IN DECIBELS) FOR VARIOUS BIT RATES

Rate (bpp)	0.0625	0.125	0.25	0.50	0.75	1.00
Multi-layer	24.04	26.21	29.18	32.92	35.30	37.27
SPIHT	23.35	24.86	27.58	31.39	34.25	36.41

Fig. 9. Barbara at 0.125 bit per pixel, SPIHT, and PSNR = 24.86 dB.

blocks that are at the boundary of the face and the scarf (see Fig. 11). The distribution of coefficients in such blocks is dominated by the periodic texture of the scarf. Because the intensity in the area of the face is smooth, one needs many other coefficients to cancel the periodic pattern of the neighboring scarf. Many of these coefficients are set to zero by the quantization.

Fig. 10. Barbara at 0.125 bit per pixel, multilayer, and PSNR = 26.21 dB.

Fig. 11. Detail of Barbara. Left: wavelet coder and right: multilayer.

A "wavelet friendly" image: Clown. While the goal of this work is the coding of images that contain a mixture of piecewise smooth regions as well as periodic textures, we wanted to benchmark our coder against a "wavelet-friendly" image. We report in Table III the results obtained with the image "Clown." For this image, the wavelet packets provided the optimal library for encoding the second layer. Our coder only marginally outperformed SPIHT in terms of PSNR. Figs. 12 and 13 show the decoded Clown at 0.125 bbp using SPIHT, and the multilayer algorithm. The decoded SPIHT image is more blurred than the multilayered image. In particular, the reflection of the clown in the mirror, as well as the right hand of the clown have been smeared by the wavelet coder. SPIHT was also unable to preserve the texture of the wallpaper or the texture of the clown's shirt. Overall, the multilayer algorithm reconstructs an image that is better in terms of visual quality than the image reconstructed by the wavelet coder.

TABLE III
CLOWN: PSNR (IN DECIBELS) FOR VARIOUS BIT RATES

Rate (bpp)	0.0625	0.125	0.25	0.50	0.75	1.00
Multi-layer	25.75	28.64	32.34	36.37	38.67	40.18
SPIHT	25.57	28.23	31.95	35.93	38.27	40.05

Fig. 12. Clown at 0.125 bit per pixel, SPIHT, and PSNR = 28.23 dB.

Fig. 13. Clown at 0.125 bit per pixel, multilayer, and PSNR = 28.64 dB.

VI. CONCLUSION

We have addressed the problem of efficiently coding images that contain a mixture of smooth and textured features. We have shown that a new solution to the image coding problem is provided by "multilayered" representations. An image is parsed into a superposition of coherent layers: smooth-regions layer, textures layer, etc. A coder based on this new paradigm was studied: it offers the advantage of being scalable, both in term of spatial resolution, and in terms of quality of reconstruction. The evaluation of the algorithm indicates that this new coder outperforms one of the best wavelet coding algorithms [22], both visually and in term of the quadratic error. Furthermore in error-prone environment at low-bitrate (such as wireless networks), this decomposition permits to efficiently protect the first layer (which corresponds to a very small number of bits), and could provide robust transmission over mobile channels.

REFERENCES

[1] S. Mallat and F. Falzon, "Analysis of low bit rate image transform coding," *IEEE Trans. Signal Processing*, vol. 46, pp. 1027–1042, Apr. 1998.

[2] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inform. Theory*, vol. 38, pp. 713–718, Mar. 1992.

[3] F. G. Meyer, A. Z. Averbuch, and J.-O. Strömberg, "Fast adaptive wavelet packet image compression," *IEEE Trans. Image Processing*, vol. 9, pp. 792–800, May 2000.

[4] S. S. Chen, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, 1998.

[5] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397–3415, Dec. 1993.

[6] P. J. Huber, "Projection pursuit," *Ann. Statist.*, vol. 13, no. 2, pp. 435–475, 1985.

[7] F. G. Meyer and R. R. Coifman, "Brushlets: A tool for directional image analysis and image compression," *Appl. Comput. Harmon. Anal.*, pp. 147–187, 1997.

[8] D. L. Donoho, "Orthonormal ridgelets and linear singularities," *SIAM J. Math. Anal.*, vol. 31, no. 5, pp. 1062–1099, 2000.

[9] F. G. Meyer, "Image compression with adaptive local cosines: A comparative study," *IEEE Trans. Image Processing*, vol. 41, pp. 616–629, June 2002.

[10] J. Berger, R. R. Coifman, and M. J. Goldberg, "Removing noise from music using local trigonometric bases and wavelet packets," *J. Audio Eng. Soc.*, vol. 42, no. 10, pp. 808–818, 1994.

[11] M. Goodwin, "Residual modeling in music analysis–synthesis," in *Int. Conf. Acoustics, Speech, Signal Processing*, vol. 2, 1996, pp. 1005–1008.

[12] M. Helsingius, P. Kuosmanen, and J. Astola, "Image compression using multiple transforms," *Signal Process.*, vol. 15, pp. 513–529, 2000.

[13] D. Marpe, G. Blättermann, and P. Maass, "A two-layered wavelet-based algorithm for efficient lossless and lossy image compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 1094– 1102, Oct. 2000.

[14] P. Auscher, G. Weiss, and M. V. Wickerhauser, "Local sine and cosine bases of Coifman and Meyer," in *Wavelets—A Tutorial*. New York: Academic, 1992, pp. 237–256.

[15] M. V. Wickerhauser, *Adapted Wavelet Analysis From Theory to Software*. New York: Peters, 1995.

[16] R. R. Coifman and Y. Meyer, "Size properties of wavelet packets," in *Wavelets and Their Applications*, Ruskai *et al.*, Eds. Boston, MA: Jones and Bartlett, 1992, pp. 125–150.

[17] T. Chang and C. C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans. Image Processing*, vol. 2, pp. 429–441, Oct. 1993.

[18] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. Image Processing*, vol. 2, pp. 160–175, Apr. 1993.

[19] S. S. Chen, "Basis pursuit," Ph.D. dissertation, Dept. Statist., Stanford Univ., Stanford, CA, Nov. 1995.

[20] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1445–1453, Sept. 1988.

[21] G. J. Sullivan, "Efficient scalar quantization of exponential and Laplacian random variables," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1365–1374, Sept. 1996.

[22] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partioning in hierarchical trees," *IEEE Trans. Circuits Syst., Video, Techol.*, vol. 6, pp. 243–250, June 1996.

François G. Meyer (M'94) recived the M.S.(Hon.) degree in computer science and applied mathematics from Ecole Nationale Superieure d'Informatique et de Mathématiques Appliquées, Grenoble, France, in 1987. He received the Ph.D. degree in electrical engineering from INRIA, France, in 1993.

From 1993 to 1995, he was a Postdoctoral Associate in the Departments of Diagnostic Radiology and Mathematics, Yale University, New Haven, CT. In 1996, he was an Associate Research Scientist with the Departments of Diagnostic Radiology and Mathematics, Yale University. From 1997 to 1999, he was an Assistant Professor with the Departments of Radiology, and Computer Science, Yale University. He is currently an Assistant Professor with the Department of Electrical Engineering, University of Colorado, Boulder. His research interests include image processing, biomedical signal and image analysis, image and video compression, and diverse applications of wavelets to signal and image processing. He is the editor (with A. Petrosian) of *Wavelets in Signal and Image Analysis, From Theory to Practice* (Norwell, MA: Kluwer, 2001).

Dr. Meyer is a member of Sigma Xi.

Amir Z. Averbuch (M'01) was born in Tel Aviv, Israel. He received the B.Sc. and M.Sc. degrees in mathematics from the Hebrew University, Jerusalem, Israel, in 1971 and 1975, respectively. He received the Ph.D. degree in computer science from Columbia University, New York, in 1983.

From 1966 to 1970 and 1973 to 1976, he served in the Israeli Defense Forces. From 1976 to 1986, he was a Research Staff Member in the Department of Computer Science, IBM T. J. Watson Research Center, Yorktown Heights, NY. In 1987, he joined the Department of Computer Science, School of Mathematical Sciences, Tel Aviv University, where he is now Associate Professor of Computer Science. His research interests include wavelet applications for signal/image processing and numerical computation, multiresolution analysis, scientific computing (fast algorithms, parallel and supercomputing software and algorithms).

Ronald R. Coifman received the Ph.D. degree from the University of Geneva, Switzerland, in 1965.

He is Phillips Professor of Mathematics at Yale University, New Haven, CT. Prior to coming to Yale in 1980, he was a Professor at Washington University, St. Louis, MO. His recent publications have been in the areas of nonlinear Fourier analysis, wavelet theory, numerical analysis, and scattering theory. He is currently leading a research program to develop new mathematical tools for efficient transcription of physical data, with applications to feature extraction recognition and denoising. He was chairman of the Mathematics Department, Yale University, from 1986 to 1989.

Dr. Coifman is a member of the National Academy of Sciences, American Academy of Arts and Sciences, and the Connecticut Academy of Sciences and Engineering. He received the DARPA Sustained Excellence Award in 1996, the Connecticut Science Medal in 1996, the Pioneer award from the International Society for Industrial and Applied Mathematics in 1999, and the National Medal of Science in 1999.