Combining Image-Processing and Image Compression Schemes

H. Greenspan[1] and M.-C. Lee[2]
Communications Systems Research Section

An investigation into the combining of image-processing schemes, specifically an image enhancement scheme, with existing compression schemes is discussed. Results are presented on the pyramid coding scheme, the subband coding scheme, and progressive transmission. Encouraging results are demonstrated for the combination of image enhancement and pyramid image coding schemes, especially at low bit rates. Adding the enhancement scheme to progressive image transmission allows enhanced visual perception at low resolutions. In addition, further processing of the transmitted images, such as edge detection schemes, can gain from the added image resolution via the enhancement.

I. Introduction

There is a new trend developing in the image-processing and image compression fields that has to do with the convergence of the two fields. This convergence has now become known as "second generation" image coding. It is the result of a growing need to handle large amounts of image data either in transmission or in automated image handling—such as image database query and retrieval—while the classical compression schemes are reaching their limits. It is now accepted that, in order to achieve more advanced compression schemes, we need to use our knowledge about images and their characteristic behavior to advantage in compression.

In this article, we present an initial attempt to combine an image-processing scheme, specifically, an image enhancement scheme, with existing image compression schemes. At the image-processing end, we use our knowledge about the behavior of edges across scale (across different resolutions) in order to extrapolate in scale and increase the resolution of a blurred input image. The ability to extrapolate in scale is very useful for compression. We can envision data rate savings by not transmitting at certain frequencies and trying to reconstruct the information back at the receiver's end; we can think about combining image processing with existing image compression schemes, such as the subband coding (SBC) and pyramid schemes, to achieve additional savings; and, finally, we can use this ability in progressive transmission applications, whereby the lower resolution images get enhanced and, thus, information can be extracted at earlier stages of the transmission.

These are some of the issues that we investigate in this article. In Section II, we describe the pyramid compression schemes, specifically a variation on the Burt and Adelson scheme [1], and we investigate its

[1] Currently at the California Institute of Technology, Pasadena, California.

[2] Currently with Microsoft Corporation, Redmond, Washington.

combination with image enhancement. Section III follows similar steps in relation to the SBC compression scheme. Finally, in Section IV, the combination with progressive transmission for savings in analysis time is described.

II. Combining Image Processing With Pyramid Compression Schemes

A. The Pyramid Representation

The pyramid scheme codes an input image in a multiresolution representation via the generation of subimages of various scales, as shown in Fig. 1. Here, $\boxed{M_{x_k} \downarrow M_{y_l}}$ denotes subsampling by M_{x_k} and M_{y_l} in the x and y directions, respectively. Low-resolution subimages G_k are created by passing G_{k-1} through a low-pass filter, H, and the decimation box. In the encoder [Fig. 1(a)], we transmit subimages $\{L_0, L_1, \cdots, L_K, G_{K+1}\}$ obtained by

$$\begin{aligned} L_0 &= G_0 - G_{1i} \\ L_1 &= G_1 - G_{2i} \\ &\cdots \\ L_K &= G_K - G_{(K+1)i} \\ G_{K+1} & \end{aligned} \tag{1}$$

where L_k is the difference subimage at the kth level, G_k is the low-resolution subimage of the kth level, and G_{ki} is the interpolated version of G_k (using filter F). In the decoding part [Fig. 1(b)], we reverse Eq. (1) to get the original signal, G_0. The pyramid representation has been introduced in the literature for coding purposes, as it was shown to be a complete representation [1]. Perfect reconstruction is guaranteed if there is no quantization of the transmitted data, regardless of the choice of filters H and F.

B. Compression Via the Pyramid Representation

The pyramid structure can be used for compression purposes. Using the pyramid coding scheme, we decompose the original image into several subimages, with different sizes, and then apply different quantization and encoding strategies in the different subimages, depending on the signal characteristics. For example, in a linear (e.g., Laplacian) pyramid, the signal variance in different subimages tends to be different. Usually, lower-frequency subimages have higher variance. Therefore, we would allocate a different number of bits to the subimages (more bits per pixel for higher variance subimages).

There are several points to be made about compression via the pyramid scheme:

(1) As stated, this is an oversampled system. This means that the number of output pixels at the transmitter is greater than that of the original image. Specifically, if the original size is $N \times N$ and the decimation box is $\boxed{2 \downarrow 2}$ in every level (K levels overall), then the total number of output pixels is

$$\mathcal{N} = \sum_{i=0}^{K} \frac{N}{2^i} \times \frac{N}{2^i} \tag{2}$$

Compared to subband and transform coding, which are critically sampled systems, it seems that the compression ratios we can obtain using this structure are lower because we have to transmit more data. However, due to the following two properties, this is not always the case.

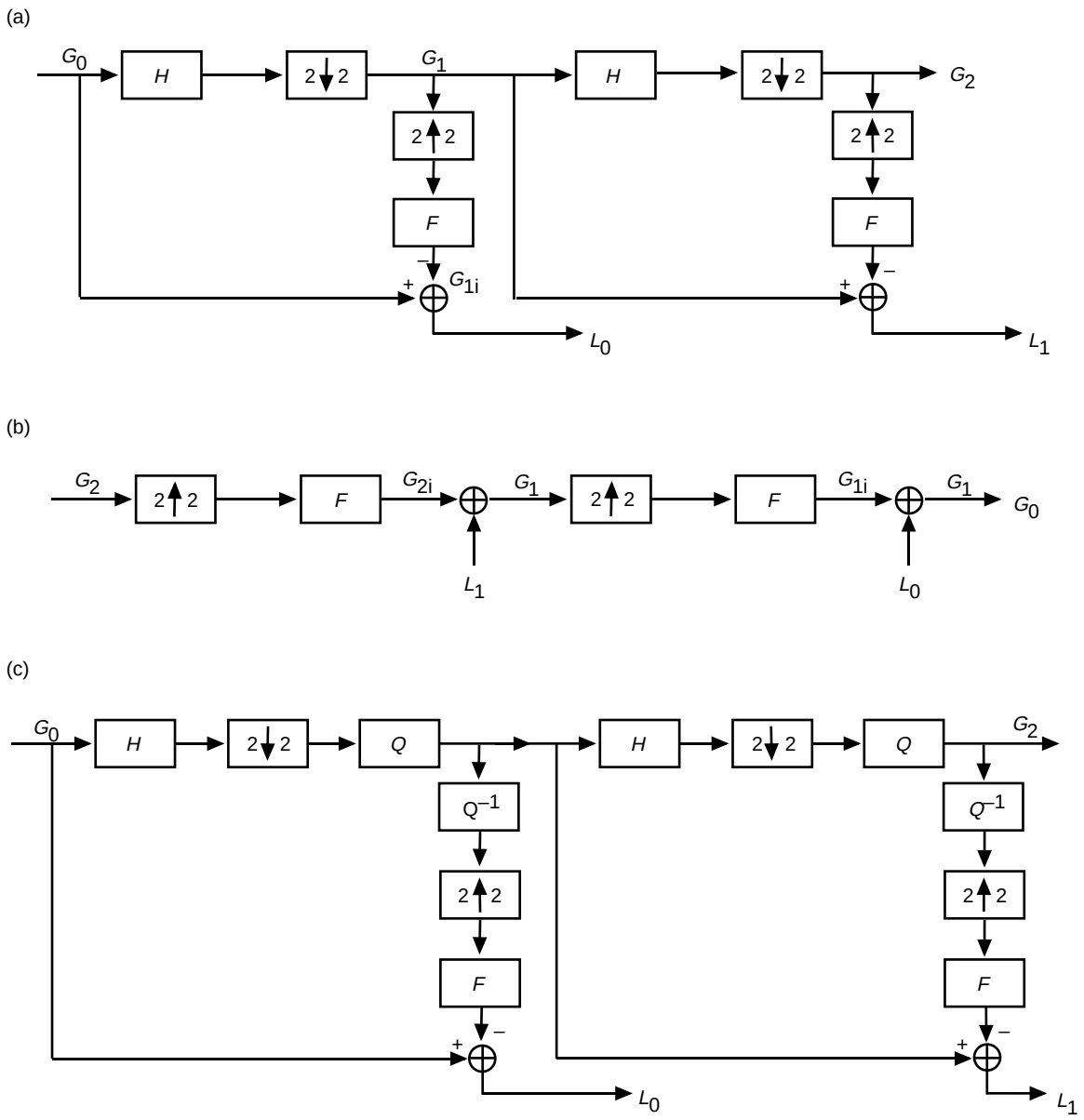


Fig. 1. Pyramid scheme: (a) encoding, (b) decoding, and (c) new concepts.

(2) No matter how we design the decimation filter H and the interpolation filter F (see Fig. 1), we can always obtain perfect reconstruction. Therefore, we can incorporate desired filters for H and F such that the signal characteristics in every level are better suited for compression. This provides great flexibility. For example, we can apply a very long filter or a nonlinear filter for image compression or include a motion-compensated filter for video compression. This implies that we can take advantage of some nonlinear characteristics of images to help the compression. Compared to subband and transform coding, which have a lot of constraints on designing perfect-reconstruction systems, we can get many advantages here.

(3) In tree-structured subband coding, the quantization noise of the higher level will propagate to the lower level. This is not desirable, because it is hard to describe the quantization noise behavior if we pass it through too many stages of filters. Especially when the quantization stepsize is large or the quantization noise is high, the noise behavior cannot be modeled using a simple expression. Therefore, some strange effects become apparent. However, in the Laplacian case, we can modify the structure to avoid this problem. As shown in Fig. 1(c), we can interpolate the quantized low-passed and decimated signals, and then take the difference. The advantage is that we now have exactly the same-difference subimages as in the receiving end. Therefore, if we quantize these subimages, the quantization noise will not propagate. Overall, it is much easier to control the noise behavior.

C. Gaussian and Laplacian Pyramids

In this article, we mainly concentrate on Gaussian and Laplacian pyramids, as described by Burt and Adelson, commonly referred to as the "Burt pyramid" [1], and Anderson, known as the "filter subtract decimate (FSD) pyramid" [2]. The Gaussian pyramid consists of low-pass filtered (LPF) versions of the input image, with each stage of the pyramid computed by low-pass filtering of the previous stage and corresponding subsampling of the filtered output. The Laplacian pyramid consists of bandpass filtered (BPF) versions of the input image, with each stage of the pyramid constructed by the subtraction of two corresponding adjacent levels of the Gaussian pyramid. The Burt and Anderson Laplacian pyramids differ in the details of when the subsampling step is applied and have slightly different bandpass characteristics. The Burt pyramid follows Fig. 1 exactly, whereas the Anderson pyramid, as defined in Eq. (3), is a variation leading to more computational efficiency. In the following, we refer to the input image as G_0; the LPF versions are labeled G_1 through G_{K+1} with decreasing resolutions, and the corresponding difference images are labeled L_0 through L_K, respectively. A recursive procedure allows for the creation of the Anderson pyramid, as follows:

$$\begin{aligned} G^0_{n+1} &= W * G_n \\ L_n &= G_n - G^0_{n+1} \\ G_{n+1} &= \text{Subsampled } G^0_{n+1} \end{aligned} \tag{3}$$

where G_n is termed the nth-level Gaussian image and L_n is termed the nth-level Laplacian image. Generally, the weighting function, W, is Gaussian in shape and normalized to have the sum of its coefficients equal to 1. The values used for the LPF, which is a 5-sample separable filter, are (1/16, 1/4, 3/8, 1/4, 1/16). Figure 2 presents an example of a Laplacian pyramid representation.

Fig. 2. Multiscale sequence of edge maps. Presented from left to right are the Laplacian pyramid components L_0, L_1, and L_2, respectively. The pyramid components have been appropriately expanded to match in size.

It has been shown [1] that the Laplacian pyramid forms an overcomplete representation of the image, thus enabling full reconstruction. The reconstruction process entails adding to a given LPF version of the image, G_N, the bandpass images, $L_n(n = N-1, \cdots, 0)$, thus reconstructing the Gaussian pyramid, level by level, up to the original input image, G_0. This is a recursive process, as in Eq. (4):

$$G_n = L_n + G_{(n+1)i}, \qquad n = N-1, \cdots, 0 \tag{4}$$

where $G_{(n+1)i}$ is the interpolated version of G_{n+1}.

D. How Does Image Enhancement Come In?

Several things can be noted from the above description of the pyramid representation. First, we note that the Laplacian pyramid consists of the edge maps of the input image at the different resolutions (see Fig. 2). We also note that when coding the image into its Laplacian components, most bits need to be allocated to the L_0 level. The first observation leads us to the idea of combining our knowledge about edge behavior across scale. The second observation allows the combination with compression. Our goal is to learn about the behavior of edges across scale so that we can "predict" the L_0 level of the pyramid using only the lower-resolution edge maps up to L_1.

1. The Image Enhancement Scheme. In our image enhancement work [3], we concentrate on the edge representation of an image across different image resolutions. Edges are an important characteristic of images, since they correspond to object boundaries or to changes in surface orientation or material properties. An edge can be characterized by a local peak in the first derivative of the image brightness function or by a zero in the second derivative, the so called zero crossings (ZCs). An ideal edge (a step function) is scale invariant in that no matter how much one increases the resolution, the edge appears the same (i.e., remains a step function). This property provides a means for identifying edges and a method for enhancing real edges.

We concentrate on the edge representation of an image across different image resolutions. For this we view the image in a multiresolution framework via the Gaussian and Laplacian pyramids. The Laplacian pyramid preserves the shape and phase of the edge maps across scale (see Fig. 2).

The application of the Laplacian transform to an ideal edge transition results in a series of self-similar transient structures, as illustrated in Fig. 3(a). An edge of finite resolution would produce a decrease in amplitude of these transients with increasing spatial frequency, with the magnitude of the edge going to zero at frequencies above the Nyquist limit [see Fig. 3(b)]. An edge of finite resolution can be created by starting with a low-resolution Gaussian image and then adding on all the bandpass transient structures. To create an edge with twice the resolution requires the creation of a self-similar transient at the next level, hereby referred to as L_{-1}. The most essential features of these transient structures are that they are of the same sign at the same position in space; hence, their ZCs line up, and they all have roughly the same amplitude. The precise shape of the structures need not necessarily be maintained so long as their scaled spatial frequency responses are similar. The simple procedure described next creates localized transients for L_{-1} that satisfy all these constraints except for the maintenance of constant amplitude. While more complicated procedures could handle the amplitude constraint, it was found that sharpening the stronger value edges produces in itself visually pleasing results.

The pyramid representation can be viewed as a discrete version of the scale-space description of ZC that has been introduced in the literature [4–6]. The scale-space formalism gives the position of the ZC across a continuum of scales. One of the main theorems states that ZCs of an image filtered through a Gaussian filter have nice scaling properties, one of which is that ZCs are not created as the scale increases. If an edge appears at lower resolutions of the image, it will consistently appear as we shift to higher resolutions (see Fig. 3). Although theoretically defined, not much work has yet taken advantage

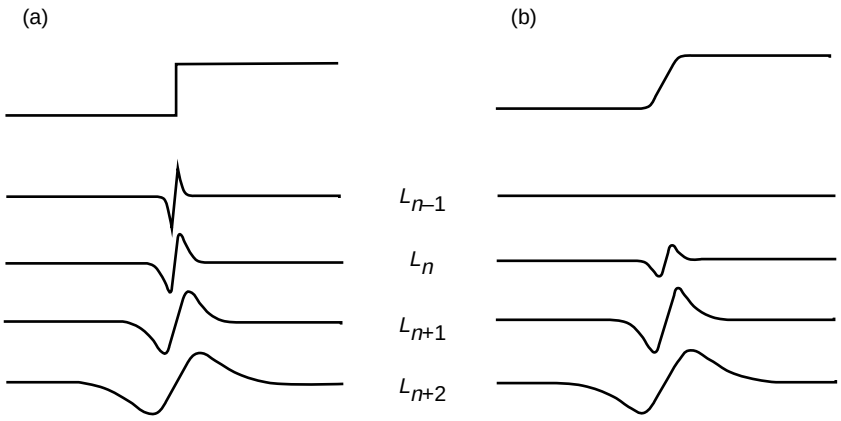


Fig. 3. Laplacian transform on (a) an ideal edge and (b) an edge of finite resolution.

of the image representation across scale. In our work, we utilize the shape invariant properties of edges across scale based on the pyramid representation and in agreement with the consistency characteristic of the scale-space formalism.

The objective is to form the next higher harmonic of the given signal while maintaining phase. Figure 4 illustrates a one-dimensional high-contrast edge scenario. The given input, G_0, is shown in (0) of the figure, together with its pyramid components, L_0 and G_1, shown in (1) and (2), respectively. From the pyramid reconstruction process, we know that adding the high-frequency component L_0 to the G_1 component can sharpen G_1 to produce the input G_0. Ideally, we would like to take this a step further. We would like to predict a higher-frequency component, L_{-1}, preserving the shape and phase of L_0, as shown in (3) of the figure, so that we can use the reconstruction process to produce an even sharper edge, which is closer to the ideal-edge objective, as shown in (4) of Fig. 4. The L_{-1} component cannot be created by a linear operation on the given L_0 component (i.e., the frequency spectrum cannot be augmented using a linear operator). We can, thus, never hope to create a higher-frequency output by a linear enhancement technique.

It remains to show how the L_{-1} component of the pyramid can be generated. We extrapolate to the new resolution by preserving the Laplacian-filtering waveform shape, together with sharpening via a nonlinear operator. The waveform as in (5) of Fig. 4 is the result of clipping the L_0 component, multiplying the resultant waveform by a constant, α, and then removing the low frequencies present (via bandpass filtering) in order to extract a high-frequency response.

Equation (5) formalizes the generation of L_{-1}:

$$L_{-1} = \alpha(C(L_0)) \tag{5}$$

where $C(S)$ is defined as

$$C(S) = \begin{cases} T & \text{if } S > T \\ S & \text{if } -T \leq S \leq T \\ -T & \text{if } S < -T \end{cases}$$

Here, $T = 0.04(G_0)_{max}$.

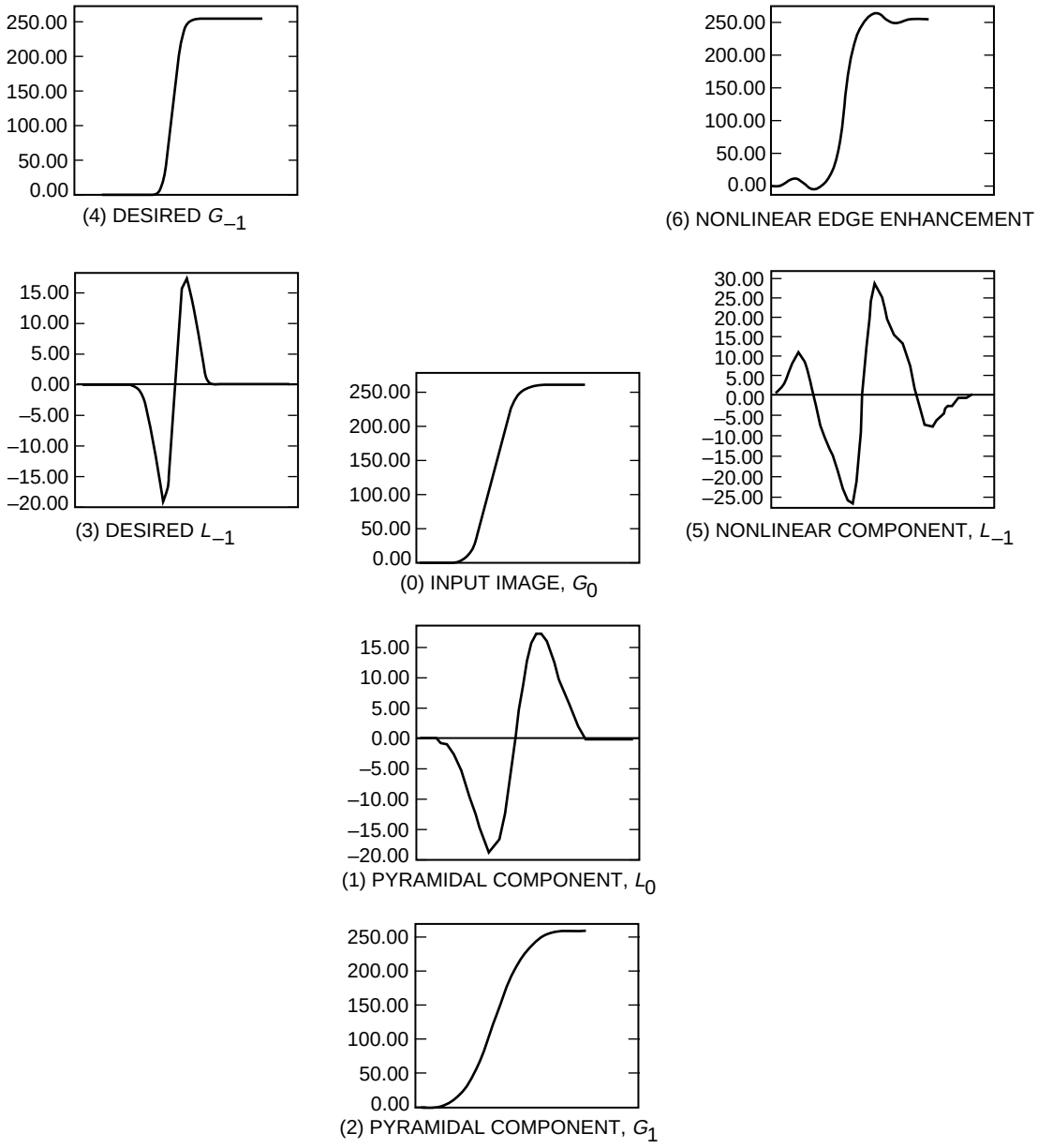


Fig. 4. The one-dimensional ideal-edge scenario.

Generating the new output image entails taking the L_{-1} image as the high-frequency component of the pyramid representation. Based on the reconstruction capability of the pyramid representation [Eq. (4)], the new output is generated next as the sum of the given input, G_0 and L_{-1}, as in Eq. (6):

$$Enhanced\ Image\ = G_{-1} = L_{-1} + G_0 \tag{6}$$

2. Enhancement Results. We next show experimental results that indicate that the enhancement scheme augments the frequency content of an input image, achieving a visually enhanced output.

A rock scene example is displayed in Fig. 5. Figure 5(a) presents the enhancement results; Fig. 5(b) displays the corresponding power spectral characteristics. The blurred input, which can be the result of

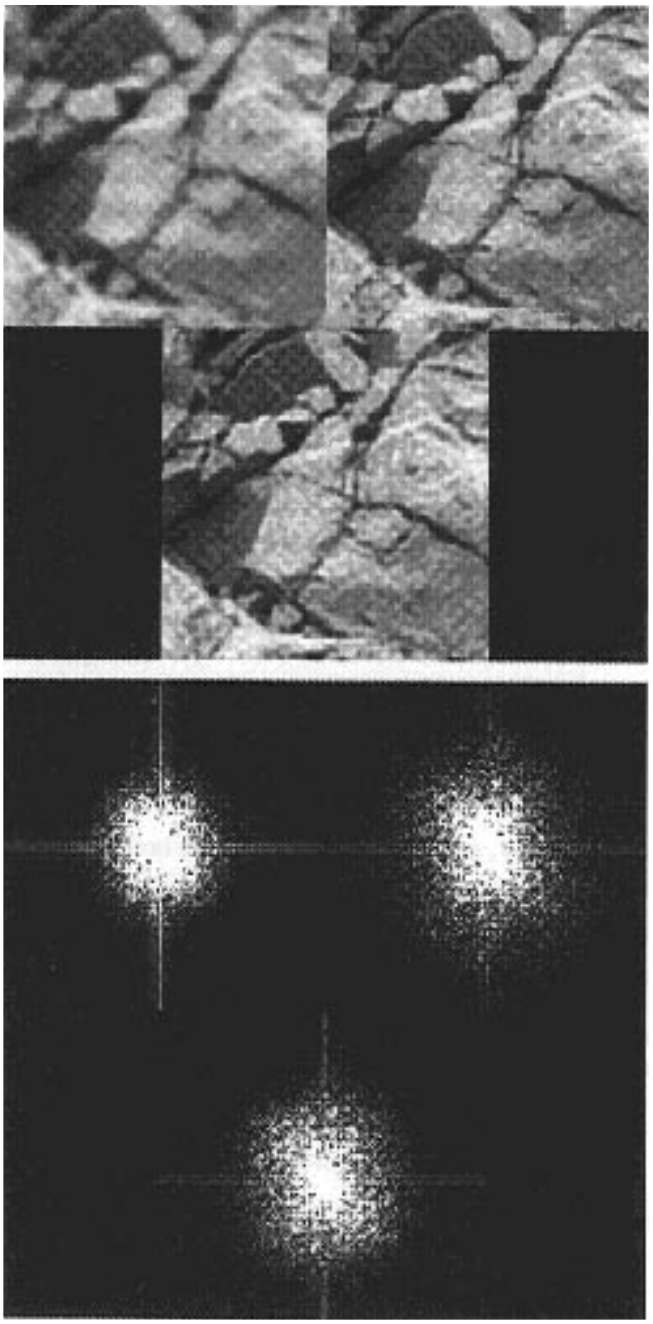


Fig. 5. A rock scene example: (a) enhancement results and (b) corresponding power-spectrum characteristics. In each of the above figures, the blurred input and original image are presented (top left and top right, respectively), followed by the enhanced output (bottom). Both visual perception enhancement and power-spectrum augmentation are evident.

cutting off high frequencies due to bandwidth considerations or a "zoom-in" application, is presented at the top-left corner. The original image, which we are assuming is not available to the system and which we wish to reproduce, is presented at the top right. The result of applying the previously presented algorithm to the blurred input is depicted in the bottom of each figure. We get an overall enhancement

perception. The enhanced image very closely matches the original one, and the power spectrum of the enhanced image is very close to the original power spectrum. For additional enhancement results, the reader is referred to [3,10].

In conclusion, the enhancement scheme addresses the most important features (edges) required in producing enhanced-resolution versions of existing images. The simplicity of the computations involved and ease of implementation enable it to be incorporated into real-time applications.

E. Combining Image Enhancement With Pyramid Coding

In this section, we combine the image enhancement scheme, described above, with the pyramid coding scheme. We have shown the possibility of predicting the L_0 level of the Laplacian pyramid using lower-resolution edge maps. The next step is to code an image with and without the L_0 component and evaluate the corresponding rate-distortion performance, i.e., investigate the compression savings versus the output image results that we can achieve.

We decompose the original image, G_0, into $\{L_0, L_1, L_2, G_3\}$. We scalar quantize L_2, L_1, and L_0 and then compute the entropy of the quantized signals. For G_3, we first apply differential pulse-code modulation (DPCM), then compute the entropy. The average entropy of these subimages represents the rate (bits per pixel). Here we use peak signal-to-noise ratio (PSNR) as the distortion criterion, defined as

$$PSNR = 10 \log_{10} \frac{255^2}{\frac{1}{XY} \sum_{i=1}^{X} \sum_{j=1}^{Y} (I_{ij} - \widehat{I}_{ij})^2} \tag{7}$$

where I_{ij} is the ijth pixel of image I, and X and Y represent the horizontal and vertical size of the input image, respectively.

The Lenna image is used for this coding task. Figure 6 presents the rate-distortion curves for Lenna. We concentrate first on the two pyramid-coding curves. We note that in using the L_0 component we have all pyramid levels and, thus, the reconstruction would be exact apart from the quantization errors induced. Using a *predicted* L_0 (i.e., the actual L_0 component is not being used), we introduce additional noise in the reconstruction process. In general, we note the slow degradation of the rate-distortion curve using the enhancement scheme, as opposed to the almost linear drop of the original (nonenhanced) curve. Of even more interest is that, at very low bit rates, the ability to estimate the L_0 component from the given L_1 component, or the ability to extrapolate in frequency space, allows for *better* PSNR.

An example of two images, with and without the L_0 component (top left and top right, respectively), is shown in Fig. 7, as compared with the original Lenna image (bottom). Both images are coded with approximately 1 bit/pixel. We have 0.99 bit/pixel with $PSNR = 34.77$ dB for the enhanced image and 1.053 bits/pixel with $PSNR = 33.54$ dB for the image decompressed with all its L_i components. In this case, we get a better PSNR and better perceived similarity to the original for the enhanced image with the predicted L_0 component than for the image with *all* components present. This is a very interesting and encouraging result.

Next, we compare the pyramid compression with the discrete cosine transform (DCT) (refer again to Fig. 6). The DCT clearly "wins" the PSNR comparison. We note that, at the very low bit rates, the differences are quite minimal. In addition, we need to compare the actual images, as opposed to the PSNR ratios, as is shown in Fig. 8. We note that the blockiness with the DCT is very evident and possibly more distracting to the eye than the artifacts introduced by the pyramid-plus-enhancement scheme. A zoom-in image taken from Fig. 8 is presented in Fig. 9. In the DCT coding scheme, we can see strong blocking

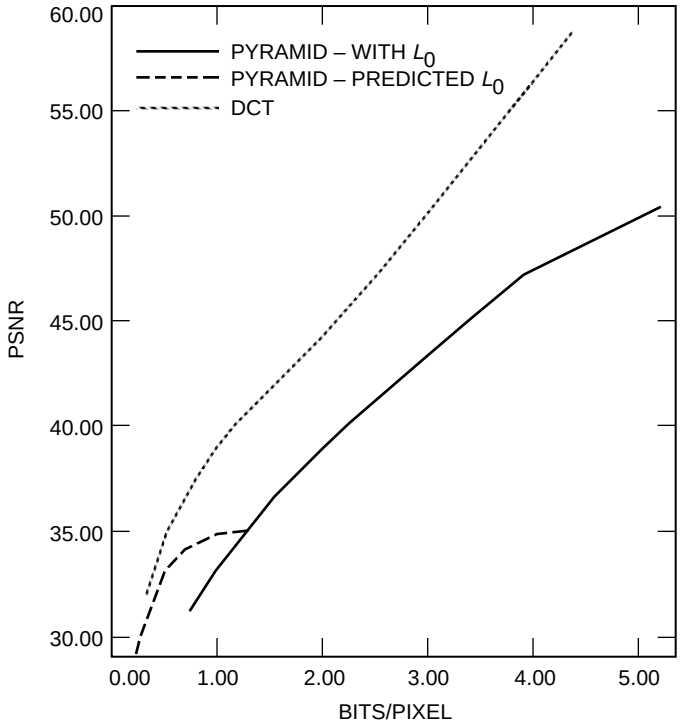


Fig. 6. Rate distortion curves for the Lenna image.

effects in the quantized image (this is the case especially when the bit rate is low or when we zoom the image up). This phenomenon results from the independent quantization of blocks. This reinforces the claim that the PSNR does not in all instances match our visual perception.

A similar investigation is done on a moon image, whose rate distortion curves are shown in Fig. 10. As before, we note the slow degradation of the rate-distortion curve with the enhancement. We see again that, at low bit rates, better PSNR is achieved by *predicting* the L_0 component via the enhancement processing stage. When comparing it with the DCT rate-distortion curve [Fig. 10(b)], we notice that at very low bit rates we actually achieve *better* performance than the DCT.

We conclude this section with a few of the moon images. Figure 11 displays the slow degradation phenomenon. Two images are displayed: The left one has 1.27 bits/pixel with $PSNR = 31.69$ dB, and the right image has almost *half* the bit rate, at 0.65 bit/pixel, with a very similar PSNR value of 31.1 dB. The two images look identical. In Fig. 12, we compare the pyramid scheme (left) to DCT (right) at the low bit rate of 0.47 bit/pixel. The PSNR ratio is larger for the pyramid coding in this case, $PSNR = 30.53$ dB, whereas the DCT case has $PSNR = 30.49$ dB. The blockiness of the DCT is certainly visible here. Note the blocks on the main rocks, which actually degrade the possibility of identifying rock boundaries, etc.

In this section, we have shown encouraging results in combining the image enhancement scheme with pyramid coding. The results are interesting especially at low bit rates. Possible modifications to the pyramid structure can help us in achieving higher compression ratios. Different filters can be applied in the pyramid structure to get better performance. We know that no matter what we put in the decimation and interpolation filters (see Fig. 1), perfect reconstruction is always guaranteed if there is no quantization and transmission loss. This provides great flexibility, since we can design a better filter to achieve better rate-distortion performance however the distortion is defined. For example, we can apply a nonlinear filter to take advantage of the nonlinear features of the human visual system. We can then obtain better

Fig. 7. Comparison of the Lenna image with and without L_0. The top left image includes L_0 in the compression, the top right uses a predicted L_0, and the bottom is the original Lenna image.

image quality with a greater perceptual effect. The use of a median filter has been shown to achieve such an improvement [7]. The combination of the enhancement scheme with modified pyramid coding schemes remains to be investigated.

III. SBC Schemes

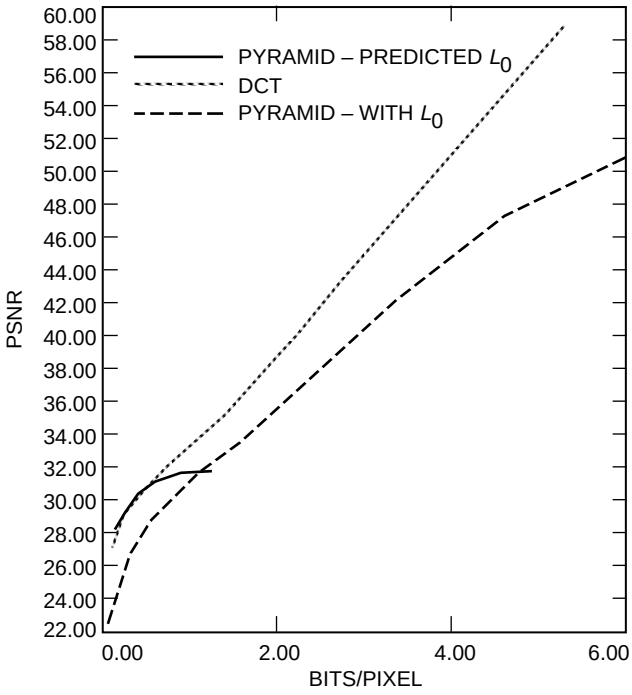
A. Introduction

SBC schemes have recently aroused much attention in the areas of image and video compression. There are several advantages to these coding schemes that make the technique attractive. Recently, the Motion

Fig. 8. The Lenna image compressed with pyramid scheme plus enhancement (top left) and with DCT (top right). The original image is on the bottom.

Fig. 9. Pyramid (left) versus DCT (right).

(a)



(b)

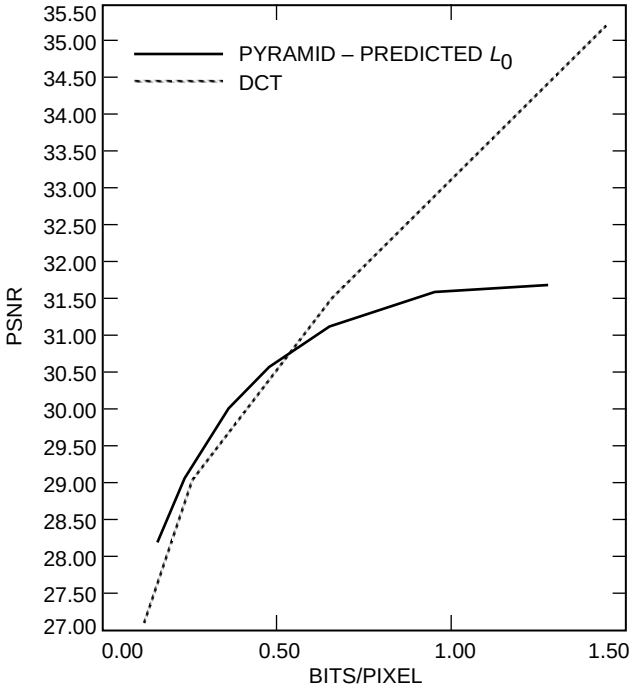


Fig. 10. The Moon image: (a) rate distortion curves and (b) comparison with DCT—zoom in.

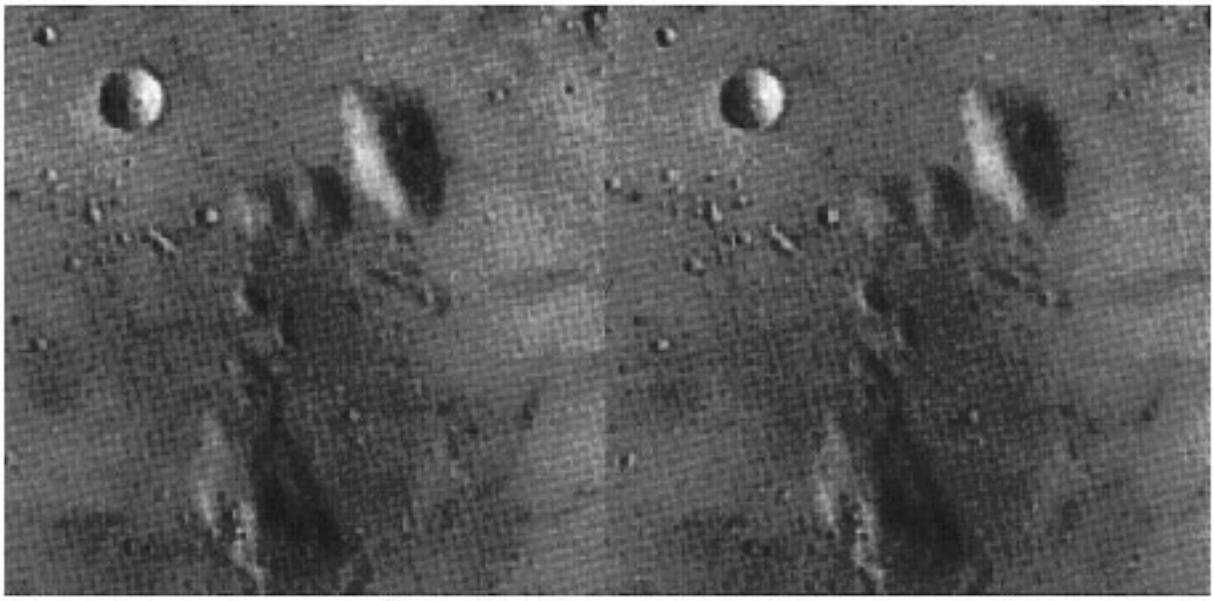


Fig. 11. Slow degradation phenomenon at low bit rates.

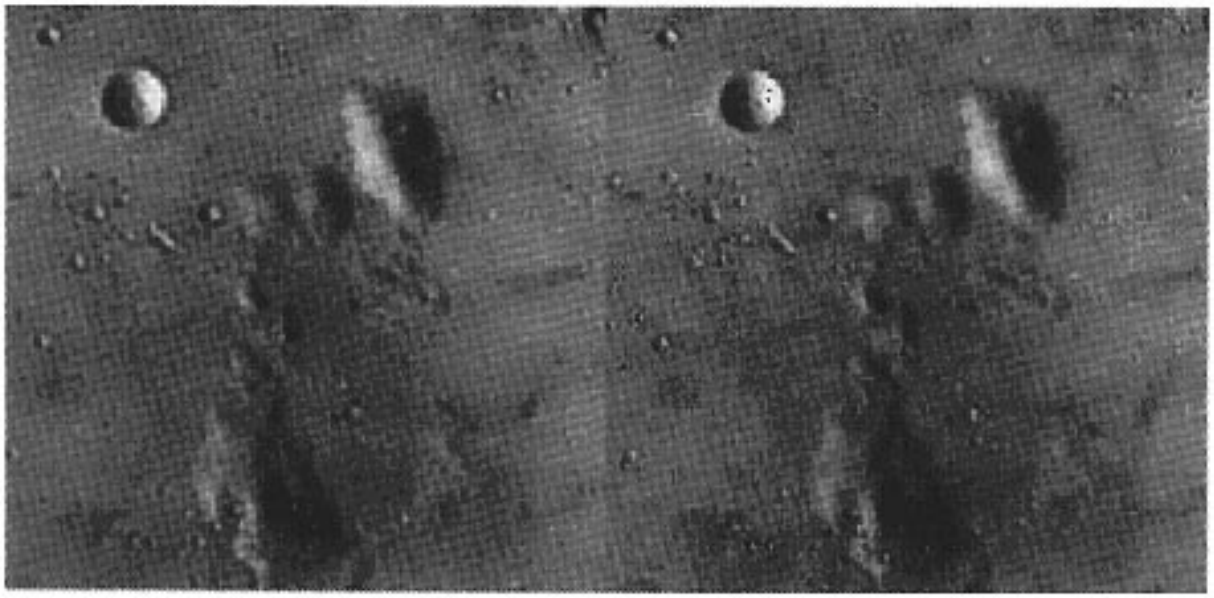


Fig. 12. Comparison of pyramid (left) versus DCT (right) at low bit rates.

Pictures Experts Group (MPEG) has proposed using this technique for its audio compression part. It is the belief of many people that SBC may take over DCT as a new image and video compression standard. Generally speaking, the compression capability of SBC is fairly good. Among all the linear multiscale techniques—such as DCT, the Laplacian pyramid, and SBC—it has been shown that SBC can provide the

best compression ratio in the rate-distortion mean-squared-error (MSE) sense, although the computational complexity is typically higher than that of DCT in achieving this. The picture quality generated by this technique is good compared to the annoying blocky effect generated by DCT. One whole frame is processed and quantized at a time as opposed to the block-by-block DCT process. However, there is another kind of distortion based on *aliasing effects* that degrades the picture quality substantially when the compression ratio is high or the bit rate is low. This is due to the signal loss in the subbands, which results in the aliasing cancellation effect provided by the filter bank being lost. This phenomenon gets more visible when the quantization noise is higher. It is generally agreed that the picture quality provided by SBC is better than that of block coding techniques. An additional advantage of SBC is that it is intrinsically progressive. This is achieved by dividing the original images into subimages in different frequency bands. One then sequentially transmits subimages with, usually, increasing frequencies. Progressive transmission is a desired property for many applications, such as data browsing and image frame conversion between different signal formats like high-definition television (HDTV) and standard TV. Although DCT can be implemented progressively, it is not as immediate a process.

The basic principle of SBC is, like other linear techniques, to take advantage of the nonuniform distribution of the signals' power spectrum. It is well known that the power spectrum of image signals tends to be nonuniform. We first use a filter bank (and decimators) to decompose the original image into several subimages in different frequency bands. One then allocates different numbers of bits in different bands depending on the signal variance in that band, thus achieving compression. It can be shown that in the uniform filter bank case, the quantization step size in every band has to be equal in order to obtain minimum MSE. However, taking account of the features of the human visual system (HVS), the quantization step sizes in different frequency bands should be different. It has been shown in testing that HVS is more sensitive to the lower frequency components. Therefore, we usually set the step sizes in lower bands to be smaller. Although not always true, typically, longer tapped filters can provide better rate-distortion performance due to better energy compaction.

There is much literature available on the principles, implementation techniques, and different types of multirate filter banks and SBC [8]. We use octave wavelet (tree-structured wavelet)-based decomposition. As shown in Fig. 13, we decompose the low-low subimage into four subimages of the next level, i.e.,

$$LL_i \longrightarrow LL_{i+1}, HL_{i+1}, LH_{i+1}, HH_{i+1}$$

by a specific filter (e.g., for the HL subimage, it is a high-pass filter in the x direction and a low-pass filter in the y direction) and a decimator that decimates the signals by 2 in both x and y directions. Therefore, if we decompose the image up to N levels (level $0, \cdots, N-1$), then there will be $3N+1$ subimages. At the receiving end, we recursively reconstruct the signals by

$$LL_i = \mathcal{E}(LL_{i+1}) + \mathcal{E}(LH_{i+1}) + \mathcal{E}(HL_{i+1}) + \mathcal{E}(HH_{i+1})$$

where $\mathcal{E}()$ represents expanding the signals first (inserting a zero between two pixels), then passing them through the corresponding filter that was originally used in the decomposition.

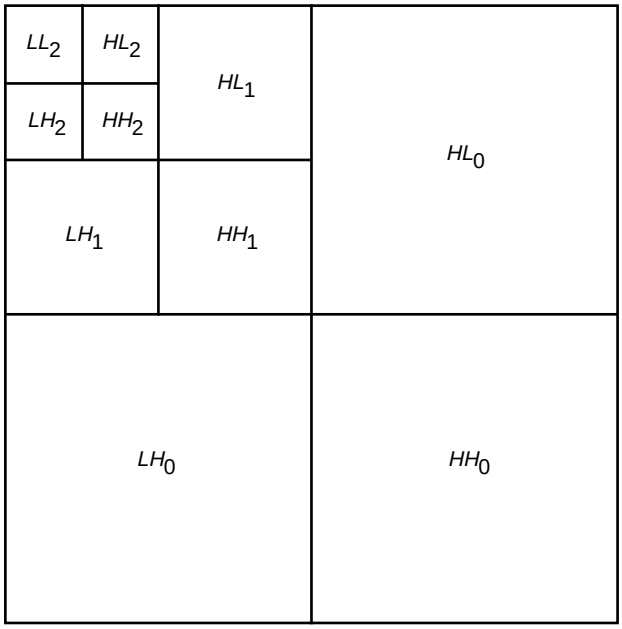


Fig. 13. SBC block diagram.

B. The Combination With Image Processing

As for the pyramid concepts of the previous section, we wish to explore the possibility of using signal correlations across frequency bands to be able to extrapolate from a lower frequency band to a higher frequency one, thus eliminating the need of transmitting all the bands.

The main difference between this case and the Laplacian pyramid case is that here we have the three difference subimages (LH_i, HL_i, HH_i) per scale instead of the single bandpass (BP) L_i component. We propose to estimate a higher-frequency level from a lower-frequency one by expanding the low-frequency level (using $\mathcal{E}$) and adding the different subimage components, as in the SBC reconstruction scheme, with an additional intermediate enhancement step.

Several possibilities come to mind in this new scenario:

(1) We can estimate each higher-level, i, subimage component from its corresponding component in the previous level, $i+1$, as follows:

$$LH_{i+1} \longrightarrow LH_i, \qquad HL_{i+1} \longrightarrow HL_i, \qquad HH_{i+1} \longrightarrow HH_i$$

where LH_i represents the low-high (x–y directions) difference subimage in level i.

(2) We can combine all BP components in level $(i+1)$ and then estimate the corresponding subimage summation in level (i):

$$LH_{i+1} + HL_{i+1} + HH_{i+1} \longrightarrow LH_i + HL_i + HH_i$$

(3) We can use only the low-low components and try to predict their behavior across the levels, i.e., $LL_{i+1} \longrightarrow LL_i$.

Continuing work on the Lenna image, Fig. 14 (top row) displays level 2 of Lenna in the SBC decomposition. Figure 14 (center row) shows the estimation of the level-1 subimages, following expansion and enhancement of the corresponding level-2 images, while the bottom row presents the original level-1 subimages. There is both a resemblance and a difference between the estimated bandpass images and the original ones. We wish to see how well we can estimate level-0 images based on the *estimated* level-1 components, i.e., using information from components in level 2 alone. Adding level-2 components together, we generate an *exact* LL_1 image (see Section III.A). We next expand that image and each of the estimated level-1 BP subimages of Fig. 14 (center row) and sum these together to generate the estimated LL_0 image [option (1) above] as displayed in Fig. 15 (left). Aliasing effects are evident. Another experiment is to take the expanded LL_1 and enhance it [option (3)]. This produces the LL_0 image at the right of Fig. 15. Again, very strong aliasing effects are evident. In order to first add all the bandpass components together and then expand [as in option (2)], we need to first define an appropriate filter for the combined subbands. The $\mathcal{E}$ function of the original SBC decomposition uses specific filters for the specific subbands and, thus, is not suitable for the task.

In the remaining experiments, we examine the possibility of expanding a given LL_1 image via *Gaussian* interpolation followed by the enhancement procedure. Figure 16 presents the results of such a procedure on Lenna. The expanded image is presented in the top left, and the enhanced image is shown in the top right. This result very closely matches the original LL_0 image, which is displayed at the bottom of the figure. We have thus achieved a good reconstruction of the 0-level image based on the information in level 2 alone, eliminating the need to transmit the level-1 subimages. Quite surprisingly, the similarity we see between the generated and original images is not reflected in the PSNR ratios. Considering the Lenna image, the PSNR value for the blurred input (top left) as compared to the bottom image is 25.1 dB. The PSNR for the enhanced image (top right) is 24.58 dB. This is quite unexpected, especially as we look at the global statistics of the three images as presented in Table 1. We note that the enhanced image has statistical characteristics that closely match the original input image.

We can conclude the following:

(1) The PSNR estimate relates to local pixel-value discrepancies between two images. Since the enhancement scheme does not attempt to reconstruct back exact pixel values of the original image, we are well aware that it is not ideal for the PSNR measure. Still, it is very apparent that the PSNR also does not represent the human perception (as in the example of Fig. 16). It is quite clear with this example that the estimated (enhanced) image is very close to the original. Because of this mismatch, we choose not to produce rate-distortion curves with the PSNR measure.

(2) Overall, we can conclude that there is no immediate step to be taken from the pyramid representation case (Section II) to the SBC case. This is probably most evident due to the aliasing effects. Ignoring certain bandpass components seems to be more crucial since these components are needed for dealiasing. All SBC subimages are needed in order to eliminate the aliasing (a theoretical analysis of this characteristic can be found in [9]). We can compensate for the missing frequencies, but unless more work is done on how to resolve the aliasing problem, this issue remains the main obstacle.

The overall conclusion at this time is that the particular image-processing scheme suggested in this article is not applicable, without major modifications, to the SBC scheme.

IV. Image Enhancement and Progressive Transmission

We conclude this article with an additional aspect of possible interest, which is the combination of image-processing with progressive image transmission schemes. Here, instead of looking for additional

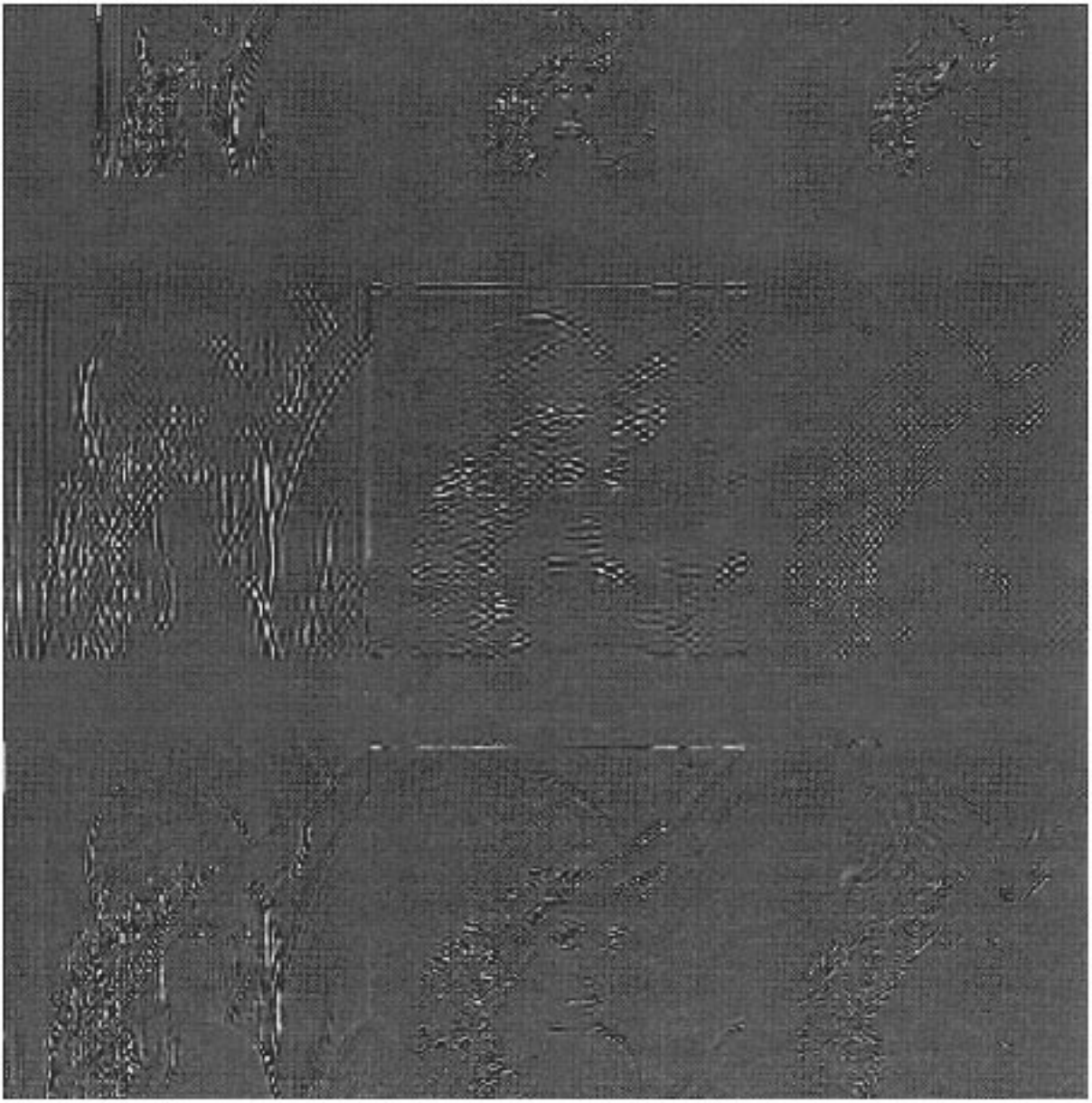


Fig. 14. SBC decomposition of Lenna. Difference subimages of levels 2, estimated level 1, and original level 1, top to bottom, respectively.

Table 1. Global statistics of the images in Fig. 16.

Characteristic	Original	Expanded image	Expanded plus enhanced image
Mean	1.344×10^2	1.344×10^2	1.344×10^2
Standard deviation	41.3	39	40
Power	140.6	140	140.3

Fig. 15. Prediction level 0 of Lenna—aliasing effects.

bit compression capabilities, we are interested in achieving successive approximations in *time*. By this we are referring to progressively transmitting information, from low resolution to high resolution, with the desire to extract information during the transmission without waiting to receive the high-resolution image. Moreover, we would like to determine at an early stage of the transmission process if the image is at all of interest, so as to determine if the high-resolution image is to be transmitted.

In Fig. 17, we demonstrate the combination of the integer subband coding (ISBC) scheme[3] of the Gaspra image with the enhancement scheme. We note the possibility of detecting the craters and other points of interest much more clearly in the enhanced images, even at extreme compression ratios. For the scientist, this can be a tool for determining interest in the region. If it does look interesting, the full-resolution image can be transmitted, without any loss.

Another domain of interest is detection of objects in a given scenery. For this task, an initial phase of edge detection is usually performed. We have looked into the combination of an edge detection scheme with the enhancement scheme to allow for better and quicker object detection.

A. Combining Edge Detection and Image Enhancement

The purpose of combining edge detection and image enhancement is two-fold: First, this scheme can be applied to progressive multiresolution image-compression systems (and progressive transmission), such as the (integer) subband coding schemes and the Laplacian-based pyramid coding scheme, to detect the edges of low-resolution images received at the early stages of transmission. We need an edge detector to catch the locations of the desired objects as soon as the low-resolution images are available. Scientists can then select and send back only the images (and acquired resolutions) of interest, thus saving in the required transmission power.

The second purpose is that, after combining the edge detection and image enhancement schemes, we get more enhanced and clear images as compared with the original images. By doing so, we can capture many more details in the received images and get a highly detailed edge map.

[3] K.-M. Cheung, "Low-Complexity Progressive Image Transmission Schemes for Space Applications," JPL Interoffice Memorandum 331-93.2-064 (internal document), Jet Propulsion Laboratory, Pasadena, California.

Fig. 16. Predicting level 0 of Lenna image from level 2 information: expanding LL_1 level (top left) with enhancement (top right) as compared to the original image (bottom).

An edge is usually defined as the point where a significant change (normally, intensity) occurs. In the results presented, we use a gradient-based method. Here, several difference (high-passed) filters are designed at 45-deg orientation interval preferences. After passing the original images through these filters, we add all the filtered (absolute) values in every direction and take a threshold. If the added sum is larger than the threshold, then the corresponding pixel is claimed as an edge point. This scheme achieves results comparable to common edge-detection schemes found in the literature. Additional details can be found in [7].

We apply the edge detection scheme on a low-resolution image: Figure 18(a) shows a low-resolution image of an oilfield image, and Fig. 18(b) shows the detected edge map. It is expected that fewer details (edges) can be captured, due to lost high-frequency components. However, we can improve this by first passing the low-resolution image through the image enhancer and then performing the edge detection. Figure 18(c) shows the enhanced image, and Fig. 18(d) shows the detected edge map. We see that more

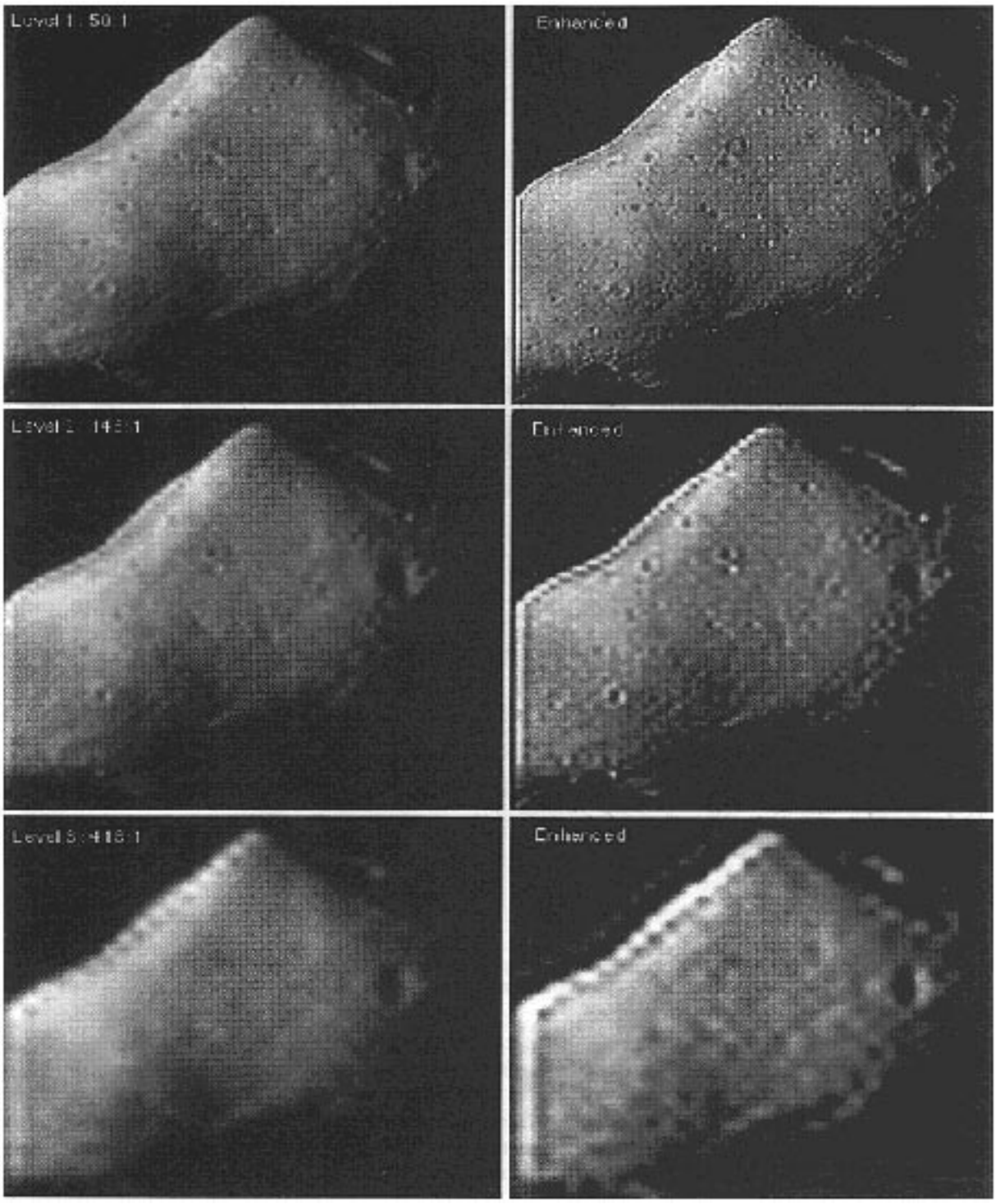


Fig. 17. Combining image enhancement with progressive transmission.

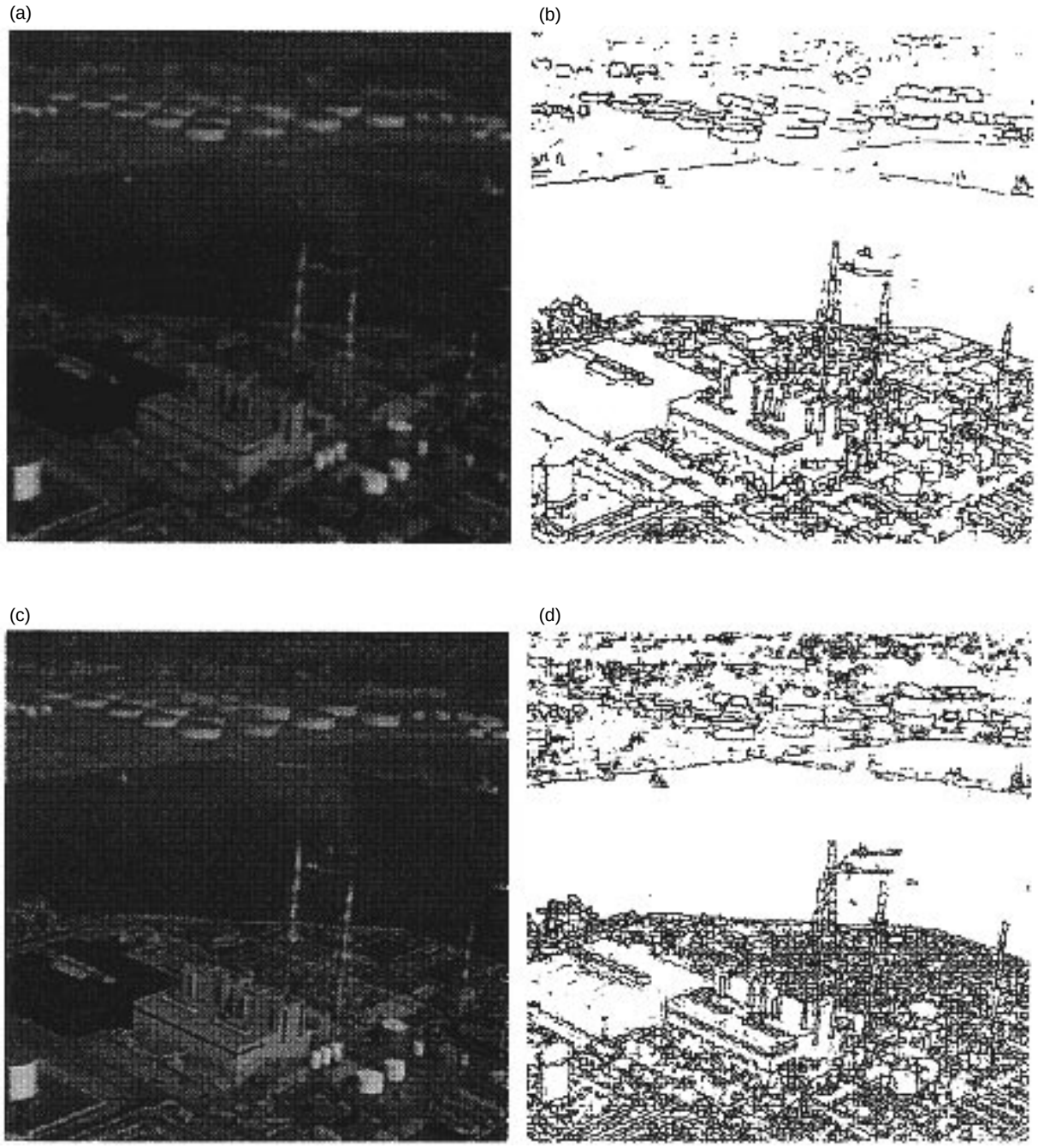


Fig. 18. Edge detection of the oilfield image: (a) a low-resolution image; (b) the detected edge map for (a); (c) an enhanced image; and (d) the detected edge map for (c).

details can indeed be captured. For example, we can count more oil tanks (top), detect the ships more clearly (center), and perceive more of the building structures (bottom).

V. Summary and Conclusions

In this article, we have done a preliminary analysis on the combination of image compression with image-processing schemes, specifically image enhancement. Encouraging results have been achieved with

the pyramid compression scheme, especially at low bit rates (which is the new frontier in the compression field). The combination of the image enhancement scheme and SBC was not as successful. Overall, we conclude that a scheme that needs all frequency bands for dealiasing cannot be easily altered and, specifically, cannot have bands removed and processed without introducing aliasing effects. Finally, the case for including enhancement in progressive image transmission was made, with results indicating the enhanced visual perception at low resolutions. In addition, further processing of the transmitted images, such as basic edge detection schemes, can gain from the added image resolution via the enhancement.

Several issues for further exploration stem from this work. In the pyramid scheme, we are interested in the possibility of pursuing steps similar to the ones described, at lower resolutions, i.e., predicting one level from a lower-resolution one, at the low resolutions of the pyramid, thus extending the compression from the L_0 level to L_1, etc. Initial investigation indicates that this is not a simple extension to the existing algorithm. As the resolution is decreased substantially, it is more difficult to locate the edges. In addition, the sharpening process will require more investigation as to how to "fill-in" the regions in the image that have been blurred and now have been sharpened. Overall, this idea requires further research.

From the SBC investigation, we learn that it could be the case that the compression and image-processing schemes cannot be combined as is. It is not always possible to take the existing algorithms and combine; rather, we might need to rethink the compression scheme together with the image-processing algorithms to generate new compression schemes.

The work presented is very preliminary work. Still, we believe that the results are interesting enough to support future work in this direction. We have only touched upon one category of image-processing schemes, image enhancement. Other processing, such as actual segmentation of images based on content, "model-based coding," and more, is attracting much interest in the research community as the new frontier for image compression.

References

[1] P. J. Burt and E. A. Adelson, "The Laplacian Pyramid as a Compact Image Code," *IEEE Transactions on Commununications*, vol. COM-31, pp. 532–540, 1983.

[2] C. H. Anderson, *A Filter-Subtract-Decimate Hierarchical Pyramid Signal Analyzing and Synthesizing Technique*, United States Patent 4,718,104, Washington, D.C., 1987.

[3] H. Greenspan and C. H. Anderson, "Image Enhancement by Non-Linear Extrapolation in Frequency Space," *Proceedings of SPIE on Image and Video Processing II*, vol. 2182, pp. 2–13, 1994.

[4] A. Witkin, "Scale-Space Filtering," *Proceedings of IJCAI*, Karlsruhe, West Germany, pp. 1019–1021, 1983.

[5] A. L. Yuille and T. Poggio, "Scaling Theorems for Zero-Crossings," A. I. Memorandum 722, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1983.

[6] A. L. Yuille and T. Poggio, "Fingerprints Theorems for Zero Crossings," A. I. Memorandum 730, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1983.

[7] M.-C. Lee, *Still and Moving Image Compression Systems Using Multiscale Techniques*, Ph.D. Dissertation, California Institute of Technology, Pasadena, California, 1994.

[8] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Englewood Cliffs, New Jersey: Prentice-Hall, 1993.

[9] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable Multi-Scale Transforms," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 587–607, 1992.

[10] H. Greenspan, *Multi-Resolution Image Processing and Learning for Texture Recognition and Image Enhancement*, Ph.D. Thesis, California Institute of Technology, Pasadena, California, 1994.