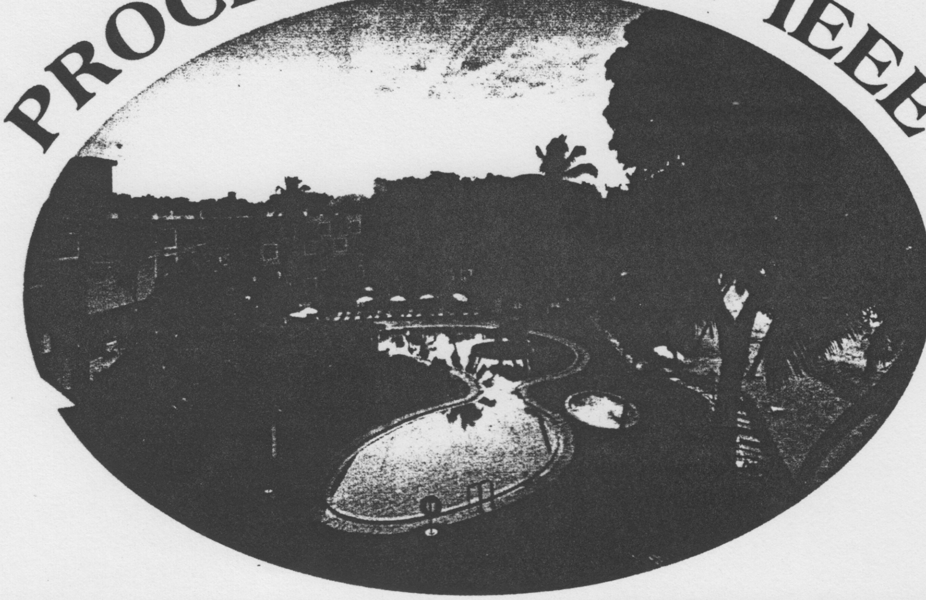
PROCEEDINGS OF IEEE



INTERNATIONAL CONFERENCE ON INDUSTRIAL TECHNOLOGY 2000

Volume 2 of 2

Goa - India

19 - 22 January 2000

B. Bandyopadhyay and N. K. Sinha
Editors

IEEE ies

Sponsored by
IEEE Industrial Electronics Society

Published by

JAICO PUBLISHING HOUSE

(4) ICIT 2000 (Goa) 87

Image Categorization and Coding Using Neural Networks and Adaptive Wavelet Filters

Subhasis Saha and Rao Vemuri
University of California Davis
Davis, CA 95616 USA

Abstract- Wavelet based compression schemes are the natural choice for the multi-resolution representation of images because of their successive approximation and better decorrelation property. Experiments conducted by compressing images through wavelet filters and integer wavelet transforms suggest that the filter performance indeed is image dependent. It is observed that no wavelet filter outperforms others uniformly while compressing sample images drawn from a large selection. In fact, a detailed analysis of the results reveals that certain wavelets perform better on certain classes of images. A Neural Network can therefore, be used to categorize the input image into one of these classes. A wavelet-based lossy or lossless coder is then used to compress the image using the most "appropriate" wavelet filter or integer-transform suitable for that class.

I. INTRODUCTION

There is a need for digital technologies that strive to deliver, automatically, just the right amount of image data for a given application while retaining image quality. Such a technology will enable to use and reuse a single image and make it available to multiple users at multiple resolutions via multiple transmission channels. Existing compression techniques require users to make a compromise between image quality and compression ratio(CR) at the time of encoding the image. This is like saving the constant 'pi' to a fixed number of decimal places, whether it is to be used to calculate the circumference of a drainpipe or the machining tolerance of a component of a high-precision watch. Dynamic resolution management at the user level allows the user to save all the image files at full precision and provide resolution on demand, enabling a layout application or web browser to "chop" or truncate the file to the precision required, when it has received sufficient data to display the image at the desired quality level. Dynamic resolution allows the user to choose between faster display speed and highest quality at the time of output, depending on how the image is to be used.

Wavelet based compression schemes are the natural choice for this multi-resolution representation of images because of their successive approximation property[1]. Multi-resolution successive approximation corresponds to the human visual system, which helps the multi-resolution techniques in terms of perceptual quality. Wavelet-based image coders are typically comprised of three major components: a wavelet filter bank decomposes the image into wavelet coefficients which are then quantized in a quantizer and finally an entropy encoder encodes these quantized coefficients into an output bit stream, as shown in Fig. 1(a). Fig. 1(b) shows a typical wavelet-based lossless image coder, where the wavelet transform is replaced by integer-to-integer(I2I) transforms[2] derived from the corresponding wavelet filters and the quantization step is eliminated. Although one has the freedom to choose each of these components from a pool of candidates, it is often the choice of either the wavelet filter or the I2I transform that is crucial in determining the ultimate performance of the coder. If the performance of the wavelet filter or I2I transform is poor in the first place, the schemes for quantization (for lossy coder) and entropy encoding, however elegant they are, cannot generally provide adequate compensation to maintain significant picture quality or bit-rate (BR).

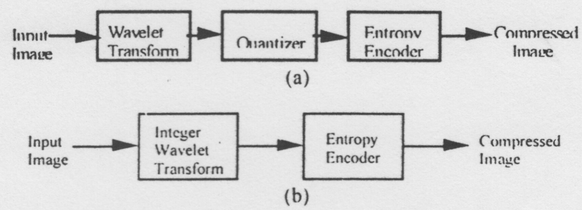


Fig. 1. A typical wavelet-based (a) lossy and (b) lossless image coder

All well known lossy and lossless image coding algorithms developed so far, use either a specific filter bank or a specific I2I transform chosen from a large number of such filters and transforms designed and developed by researchers over the years. Once chosen, the coefficients are hard coded into the algorithm. In other words, the same filter/I2I transform is used for coding and decoding all types of images whether it's a natural image, synthetic image, medical image, aerial image, scanned image, compound image or any other image for that matter. Our extensive image compression experiments using sample test images from all of the above categories show that this generic approach of filter/transform selection may not always give the best quality of service (image quality or compression) from the viewpoint of a specific application. For example, in lossy coders while Haar filters perform the best in compound images, it performs rather poorly in natural images presumably because of its low order. For natural images higher order filters like Villasenor's-18/10 filters perform better. CDF-9/7 filters are good for coding medical images. Similar observations have been made for lossless coders as well. So, for both lossy and lossless compression schemes the performance has been found to be much more image dependent.

II. FILTER AND I2I TRANSFORM FEATURES

When deciding on a filter bank for lossy image compression, there are many choices. The two well-known wavelet filter families used in wavelet-based lossy image coders are

0-7803-5812-0/00/$10.00©2000IEEE

orthogonal[3] and biorthogonal[4] wavelets. The following are the key features that distinguish one wavelet filter from the other, and need be considered while making a choice for image compression.

- Orthogonality
- Linear phase (Symmetric)
- Length of the filters
- Smoothness (Number of zero moments)
- Regularity measure (Holder regularity)
- Order of the filters
- Energy compaction (Coding gain)
- Wavelet coefficient distribution statistics

The following wavelet filters have been used to compress various images in the lossy experiment.

- Orthogonal filters- Haar, Daubechies filters of order two, four, and eight with four, eight, and sixteen coefficients respectively, and Adelson's symmetric filters with nine coefficients.
- Biorthogonal filters- Cohen, Daubechies, and Feauveau (CDF) filters e.g. CDF-9/7, CDF-9/11, CDF-13/3 filters, Villesenor filters e.g. Vill-18/10, Vill-13/11 and Vill-6/10 filters, Odegard 9/7 filter, and Brislawn 10/10 biorthogonal filter.

For lossless compression using I2I wavelet transforms there are many choices as well[2]. Orthogonality of such transforms is not as important as in lossy image coders. Since there is no quantization involved, arbitrary filters can be chosen as long as it can reduce the average entropy at the output. The I2I transforms used in the lossless coding experiments include S, S+P, TS, TT, CDF(2,2), CDF(2,4), CDF(4,2), CDF(4,4), and CDF(2+2,2) transforms. Here, the notation (M,N) represents a transform with M and N vanishing moments in the analysis and synthesis high pass filters, respectively.

III. ANALYSIS OF COMPRESSION RESULTS

A. Lossy Compression

We have experimented with a large number of wavelet filters, both orthogonal and biorthogonal with varying lengths, regularity and smoothness, and a large set of images with varying features. Test images in our experiment include natural images, synthetic images, binary/compound images, medical images, and aerial images of different sizes. An embedded quantizer, an adaptive arithmetic entropy encoder, and up to five levels of decomposition are used. Using the same filters to code an image, the performance of the image coder at four different compression ratios (8:1, 16:1, 32:1 & 64:1) is evaluated. Due to limitations of space, only a subset (for compression ratio of 16:1) of the results is presented and analyzed here. The results of the lossy image coding experiment are plotted and shown in Fig. 2. The difference between the worst and the best peak signal to noise ratio (PSNR) values for compressing the same image using different wavelets, is anywhere from 1.5 to 6 dB, which is significant. Although there are some wavelet filters, mainly biorthogonal that perform generally well for many images there is no single one that outperforms others for all images. Fig. 3(a) through 3(d), where PSNR values using different wavelets have been plotted for various image types separately, give a clearer picture. For natural images (Lena, Barbara, Baboon, and Airplane), biorthogonal filters like the CDF-9/7, and Villasenor-10/18, perform better than Haar and Daubechies' family of orthogonal filters as shown in Fig. 3(a). It is also observed that, for both Barbara & Baboon images, the PSNR values using the same wavelet filter are lower than that of the Lena and Airplane images. This is mainly due to the presence of more sharp textures and edges in those two images than Lena & Airplane. For pure binary text image (Bengali) as well as for images containing both binary data and gray scale images (Cmpnd1, Cmpnd2), Haar filters outperform the rest by more than 3 dB as shown in Fig. 3(b). Fig. 3(c) shows the PSNR values for the aerial images (Aerial, air1 and air2) where the performance of CDF-9/7, Odegard-9/7 as well as Villasenor-10/18 and Villasenor-13/11 biorthogonal filters are very close. However, they all give better PSNR (by about 2 dB) than the orthogonal filters. It can also be seen that the plots of the images in various classes follow a similar pattern. For example, in the case of aerial images in Fig. 3(c), all three images show poor PSNR values using both CDF-9/11 and Brislawn-10/10 filters. PSNR values for four medical images (mri, nervecell, us (ultrasound), and us1) are shown in Fig. 3(d). As expected, all four images don't perform the same, because of the different spatial features for those images. For both ultrasound images (us and us1), Haar filter gives the best PSNR outperforming others by more than 2 dB. However, for mri and nervecell images, the PSNR curves are similar to those of the aerial images in Fig. 3(c) with CDF-9/7 and Villasenor-10/18 giving the best results. Apart from the best filters, Villasenor-6/10 and Odegard-9/7 perform well for the medical images.

B. Lossless Compression

Similar experiments have been performed in the lossless context on the same set of test images using a number of I2I transforms[2] implemented using the lifting scheme[5]. The performance used is bit-rate, measured in bits-per-pixel(bpp). The results of the lossless experiment are plotted and shown in Fig. 4. The difference between the best and worst BR values for the same image using different I2I transforms is anywhere from 0.14 to 1.68 bpp. There is no one transform that performs the best for all images. Fig. 5(a) through 5(d) shows the plots for different image types. For most natural images the 'S+P' transform is the best, although the performances of both CDF(4,4) and CDF(2+2,2) are equally good. For compound images no other transform comes even close to 'S' transform, and so when compressing such images 'S' transform should be the transform of choice. For medical images like 'ct' and 'nervecell' CDF(2+2,2) transform outperforms others, whereas for both ultrasound images ('us' and 'us1') the 'S' transform is best.

IV. IMAGE ANALYSIS AND CATEGORIZATION USING NEURAL NETWORK (NN)

From both lossy and lossless compression results discussed in section II & III, it is observed that certain wavelet filters and I2I transforms perform well for certain category of images. To better understand this, we have summarized both lossy and lossless results in Table-I where, the performances of wavelet filters and I2I transforms for a given image are ranked first through third along with their respective PSNR and BR values. For lossy compression, results for both CR=16:1 and CR=32:1 are included. A close look at Table-I reveals that, for certain image-categories the performance of a number of wavelet-filters and I2I transforms are so close that it is rather difficult and probably not worth the cost to try to identify the best filter/transform. Instead, identifying any one of the top performing filters/I2I transforms is good enough for that category. For example, for lossy compression of natural image Lena, all three wavelet filters CDF-9/7, Odegard, and Vill-10/18 perform equally well. Similarly, for lossless compression S+P, CDF(4,4), and CDF(2+2,2) transforms are equally good. So, choosing any one of these filters or transforms is acceptable. However, for certain other image category like Compound images, no filter or I2I transform comes close to the performance obtained from the Haar filter or 'S' transform, and so there is only one choice.

Given an image, our task is to first analyze the image and then try to classify it into one of the categories mentioned earlier. A spatial domain analysis of various images show that in general, images from different categories tend to have different characteristics. For example, it is a common observation that most of the natural images are continuous in tone compared to the synthetic images most of which are of discrete tone (dynamic range of the pixel bit depth is under utilized). Such images generally have some numerical structures that are not well represented by smooth basis functions. Many medical images like MRI or CT scan contain significant low-intensity (black) regions along image boundaries. Compound images with significant amount of text are a mixture of binary and continuous tone data. Even within a particular category, images vary in many ways with widely varying first and second order Markov statistics. Whereas some are relatively flat, others are very busy having more edges and contours in them. Some are darker and others have more sharpness. So, a spatial domain analysis of these images shows different characteristics like mean, median, standard deviation etc. We use the following features to distinguish between various categories of images.

- Spatial Features e.g. Mean Median, Mode, Variance, Dispersion, Average Energy, Entropy etc.
- Edge and Boundary Information, Image Activity Code
- Higher Order Statistics e.g. Skewness and Kurtosis

We are currently implementing a 3-layer feed-forward NN [6] with supervised learning like the error back-propagation algorithm first to train and then to classify a given image into one of the several categories. The input to the NN is the image features mentioned earlier, and the output is the class the image belongs. We also plan on using an unsupervised learning technique like the Kohonen Self Organizing Feature Map (SOFM)[6] and evaluate its performance. Once the input image is identified to be from certain category, any one of the filters/I2I transforms suited for that category can be chosen to compress the image.

V. CONCLUSION AND WORK IN PROGRESS

To conclude, it is observed that for both lossy and lossless compression schemes, no specific wavelet filter or integer transform has performed uniformly better than others on the variety of test images and the performance has been found to be much more image dependent. In fact, detailed analysis of the results as shown in this paper, reveals that certain wavelets and I2I transforms perform better on certain classes of images. For example, images containing binary data (compound images) as well as certain types of medical images (ultrasound images) is compressed best using simple Haar wavelets or the 'S" transform. Natural and aerial images on the other hand are compressed best using different biorthogonal wavelet filters and corresponding I2I transforms. These observations lead us to conclude that for best results, in both lossy and lossless compression, the most "appropriate" wavelet filter should be chosen to match the image class and the characteristics of the individual image being coded. We are currently implementing a NN-based classifier to categorize the input image so that one of the best performing filter or I2I transform for that class can be chosen for subsequent compression purpose.

VI. REFERENCES

[1] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 11, No. 7, July 1989, pp. 674-693.

[2] R. Calderbank, I. Daubechies, W. Sweldens, and B. L. Yeo, "Wavelet transforms that map integers to integers", Applied and Computational Harmonic Analysis (ACHA), Vol. 5, No. 3, 1998, pp. 332-369.

[3] I. Daubechies, "Orthonormal bases of compactly supported wavelets", Comm. Pure and Applied Math., Vol. 41, Nov. 1988, pp. 909-996.

[4] A. Cohen, I. Daubechies, and J. C. Feauveau, "Biorthogonal bases of compactly supported wavelets", Comm. on Pure and Applied Mathematics, 1992, Vol. XLV, pp. 485-560.

[5] W. Sweldens, "The Lifting Scheme: A New Philosophy in Biorthogonal Wavelet Constructions", *Proc. SPIE* Vol. 2569, 1995, pp. 68-79.

[6] S. Haykin, *Neural Networks- A Comprehensive Foundation*, Macmillan College Publishing Company, New York, 1994.

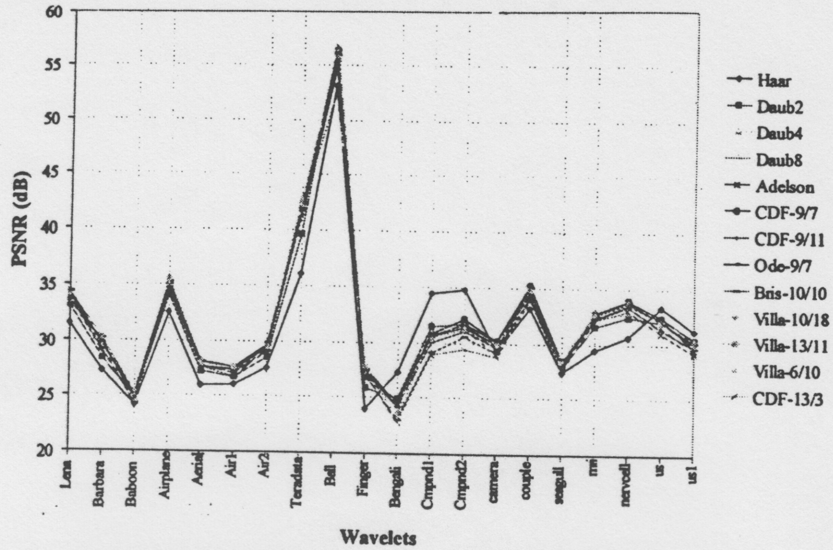


Fig. 2. Lossy compression results for various images using different wavelet filters

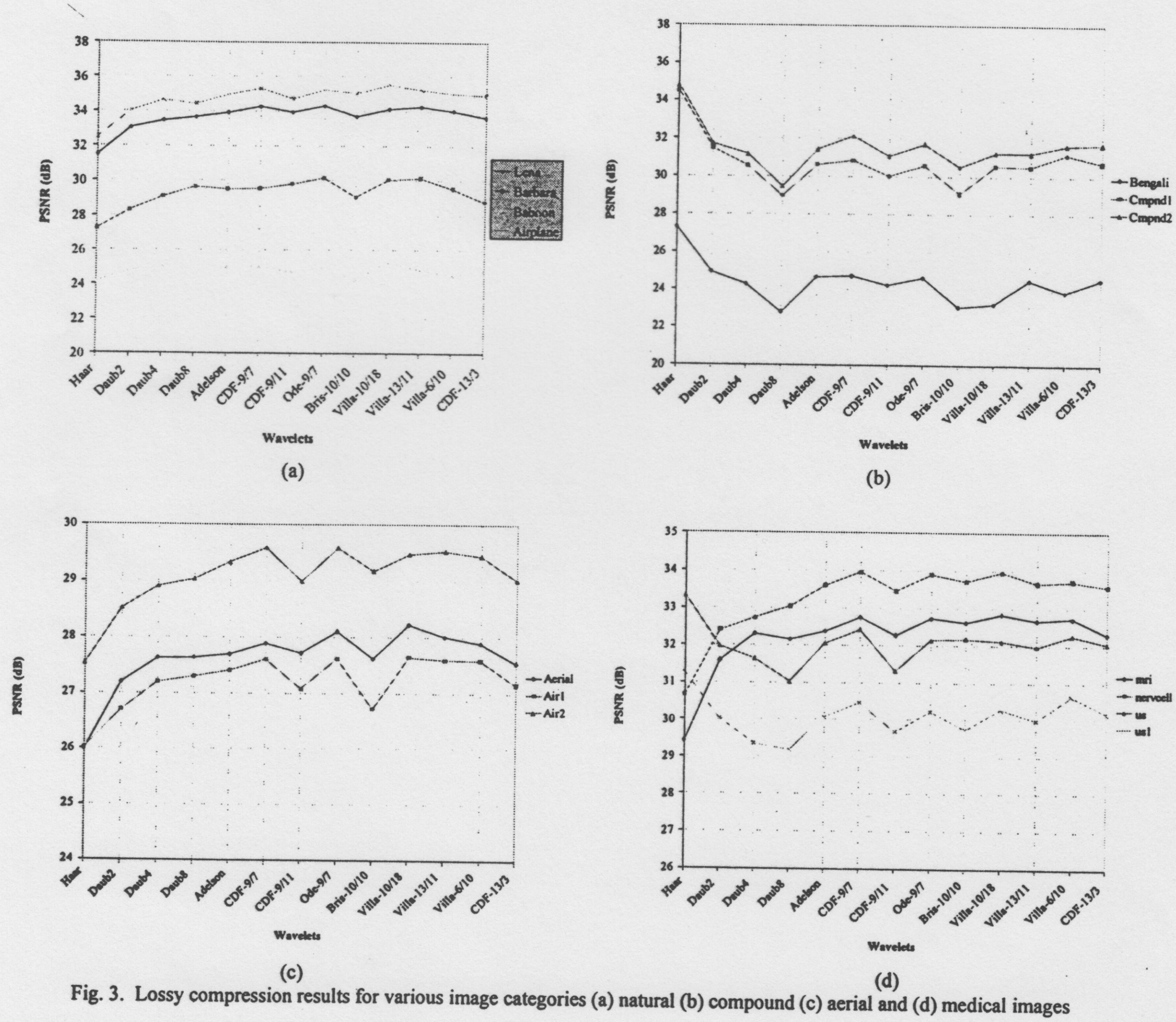


Fig. 3. Lossy compression results for various image categories (a) natural (b) compound (c) aerial and (d) medical images

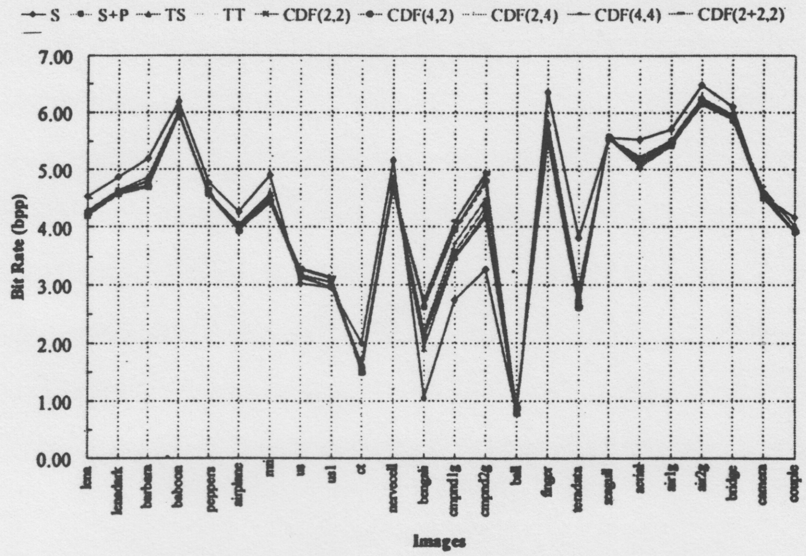


Fig. 4. Lossless compression results for various images with different integer wavelet transforms

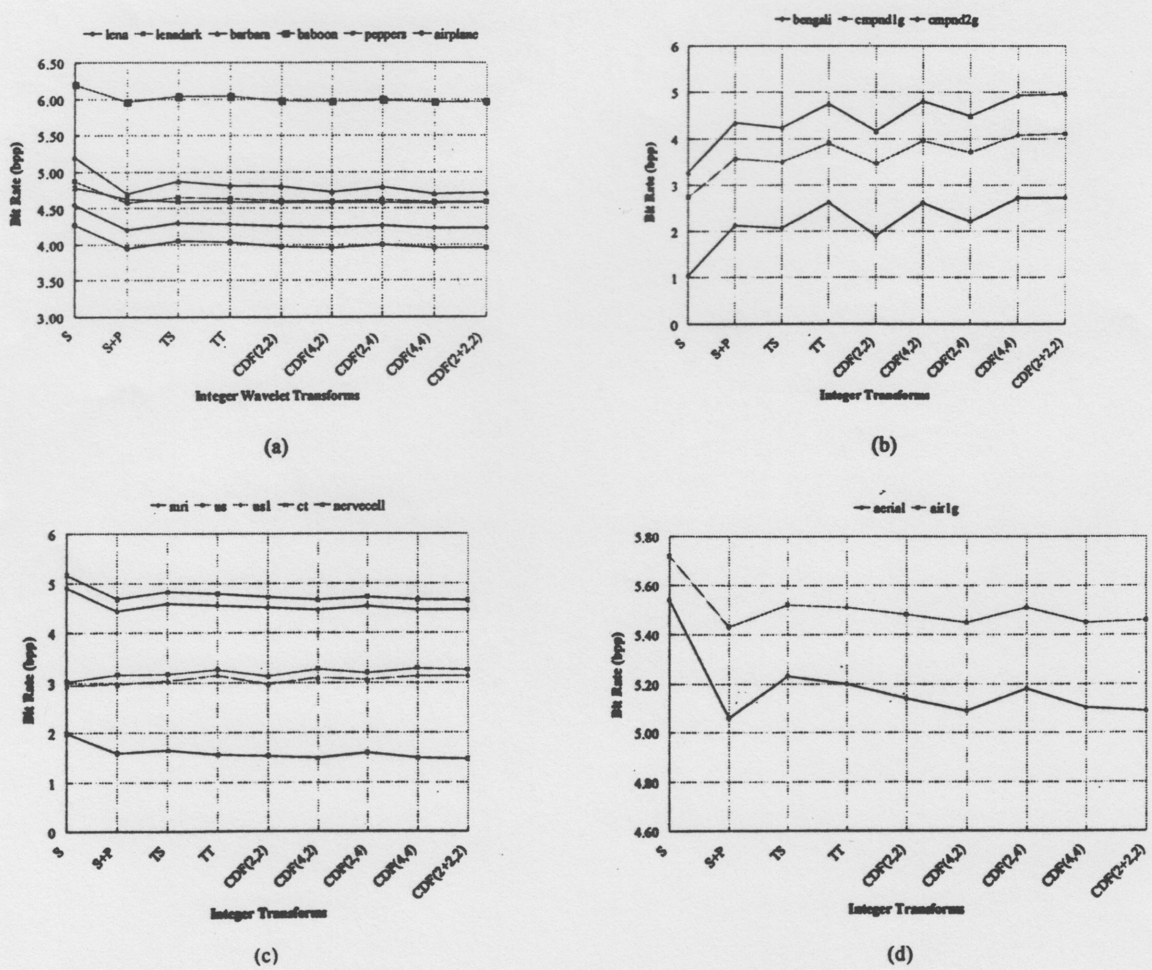


Fig. 5. Bit rates for lossless compression (a) Natural (b) Compound (c) Medical, and (d) Aerial Images

TABLE I
PERFORMANCE OF VARIOUS WAVELET FILTERS AND INTEGER TRANSFORMS IN LOSSY AND LOSSLESS COMPRESSION

		Lossy Coding						Lossless Coding					
		1		2		3		1		2		3	
Images	CR	Filter	PSNR	Filter	PSNR	Filter	PSNR	Transform	BR	Transform	BR	Transform	BR
Lena	16:1	CDF-9/7	36.18	Odegard	36.17	Vill-13/11	36.16	S+P	4.20	CDF(4,4)	4.22	CDF(4,2)	4.23
	32:1	"	33.17	"	33.17	Vill-10/18	33.16			CDF(2+2,2)	4.22		
Lenadark	16:1	Odegard	34.3	CDF-9/7	34.28	Vill-13/11	34.28	S+P	4.57	CDF(4,4)	4.59	CDF(4,2)	4.60
	32:1	Vill-10/18	31.57	Odegard	31.56	"	31.54			CDF(2+2,2)	4.59		
Barbara	16:1	Vill-13/11	30.18	Odegard	30.16	Vill-10/18	30.09	S+P	4.70	CDF(2+2,2)	4.72	CDF(4,2)	4.73
	32:1	"	26.76	"	26.71	CDF-9/7	26.65	CDF(4,4)	4.70				
Baboon	16:1	Vill-10/18	25.33	CDF-9/7	25.05	Odegard	24.98	S+P	5.96	CDF(2+2,2)	5.97	CDF(2,2)	5.98
	32:1	Odegard	23	Vill-13/11	22.98	CDF-9/7	22.87	CDF(4,4)	5.96	CDF(4,2)	5.97	CDF(2,4)	5.99
Peppers	16:1	Vill-10/18	35.19	CDF-9/7	35.19	Vill-6/10	35.16	CDF(4,4)	4.57	CDF(2+2,2)	4.58	CDF(2,2)	4.58
	32:1	Vill-6/10	32.75	Vill-10/18	32.67	CDF-9/10	32.60			CDF(4,2)	4.58	CDF(2,4)	4.58
Airplane	16:1	Vill-10/18	35.55	CDF-9/7	35.34	Vill-13/11	35.27	S+P	3.94	CDF(4,4)	3.95	CDF(4,2)	3.95
	32:1	Vill-10/18	31.98	Vill-6/10	31.86	CDF-9/7	31.71			CDF(2+2,2)	3.95		
Ball	16:1	CDF-13/3	56.88	Vill-6/10	56.49	Daub2	56.38	S	0.78	CDF(2,2)	0.79	S+P	0.82
	32:1	"	52.41	"	52.09	CDF-9/7	51.81					CDF(2,4)	0.82
Finger	16:1	Vill-13/11	27.74	Vill-10/18	27.66	CDF-9/11	27.57	CDF(2+2,2)	5.50	CDF(4,4)	5.51	S+P	5.60
	32:1	Odegard	24.13	CDF-9/7	24.07	Vill-13/11	24.07	CDF(4,2)	5.50				
Teradata	16:1	Vill-10/18	42.83	Vill-13/11	42.42	Odegard	42.41	CDF(4,2)	2.60	CDF(4,4)	2.61	CDF(2+2,2)	2.62
	32:1	"	38.05	Odegard	37.84	CDF-9/7	37.83						
Bengali	16:1	Haar	27.3	Daub2	24.92	CDF-9/7	24.71	S	1.04	CDF(2,2)	1.90	TS	2.07
	32:1	Vill-6/10	19.41	Daub2	19.21	Daub4	19.20						
Cmpnd1	16:1	Haar	34.51	Daub2	31.47	Vill-6/10	31.16	S	2.74	CDF(2,2)	3.46	TS	3.50
	32:1	"	25.92	"	25.92	CDF-9/7	25.53						
Cmpnd2	16:1	Haar	34.81	CDF-9/7	32.14	Odegard	31.72	S	3.27	CDF(2,2)	4.17	TS	4.24
	32:1	Odegard	25.53	"	25.45	Daub2	25.39						
US	16:1	Haar	33.32	CDF-9/7	32.43	Vill-6/10	32.27	S	3.02	CDF(2,2)	3.12	S+P	3.16
	32:1	"	28.57	Vill-6/10	28.26	CDF-9/7	27.94					TS	3.17
CT	16:1	Vill-10/18	51.37	Vill-13/11	51.19	Odegard	51.12	CDF(2+2,2)	1.46	CDF(4,4)	1.49	CDF(2,2)	1.53
	32:1	"	45.77	CDF-9/7	45.52	Vill-6/10	45.47			CDF(4,2)	1.49		
Aerial	16:1	Vill-10/18	28.23	Odegard	28.11	Vill-13/11	28.01	S+P	5.06	CDF(2+2,2)	5.09	CDF(4,4)	5.10
	32:1	CDF-9/7	25.24	Vill-10/18	25.23	Odegard	25.17			CDF(4,2)	5.09		
Air1	16:1	Vill-10/18	27.65	Odegard	27.62	CDF-9/7	27.61	S+P	5.43	CDF(4,2)	5.45	CDF(2+2,2)	5.46
	32:1	CDF-9/7	25.05	"	25.01	Vill-13/11	25.01			CDF(4,4)	5.45		
Air2	16:1	CDF-9/7	29.59	Odegard	29.59	Vill-13/11	29.54	CDF(2+2,2)	6.13	CDF(4,2)	6.14	S+P	6.19
	32:1	"	26.75	"	26.74	Vill-10/18	26.70			CDF(4,4)	6.14		
Bridge	16:1	CDF-9/7	26.23	Vill-10/18	26.16	Odegard	26.11	S+P	5.87	CDF(2+2,2)	5.87	CDF(4,2)	5.88
	32:1	Vill-10/18	24.17	Vill-6/10	24.17	CDF-9/7	24.13	CDF(4,4)	5.87	CDF(2,2)	5.87		
Camera	16:1	CDF-9/7	30.19	Odegard	30.09	Vill-6/10	29.89	CDF(2,2)	4.51	S+P	4.52	CDF(2+2,2)	4.54
	32:1	"	26.94	Odegard	26.78	Vill-10/18	26.68					CDF(2,4)	4.54
Couple	16:1	CDF-9/7	35.26	CDF-13/3	35.13	Odegard	35.04	CDF(2,2)	3.91	CDF(4,2)	3.92	CDF(4,4)	3.93
	32:1	"	31.76	Odegard	31.64	Vill-10/18	31.63	CDF(2+2,2)	3.91			S+P	3.93
Seagull	16:1	CDF-9/7	28.5	Vill-10/18	28.48	Vill-6/10	28.46	S+P	5.55	CDF(4,4)	5.55	CDF(4,2)	5.56
	32:1	"	25.89	"	25.78	"	25.78	CDF(2,4)	5.55	CDF(2,2)	5.55	CDF(2+2,2)	5.56
Mri	16:1	Vill-10/18	32.85	CDF-9/7	32.77	Vill-6/10	32.73	S+P	4.43	CDF(4,2)	4.45	CDF(4,4)	4.45
	32:1	"	29.46	"	29.15	"	29.12					CDF(2+2,2)	4.45
Nervecell	16:1	CDF-9/7	33.96	Vill-10/18	33.96	Odegard	33.90	CDF(2+2,2)	4.66	CDF(4,2)	4.67	CDF(4,4)	4.67
	32:1	"	31.08	Odegard	31.02	Vill-13/11	31.01					S+P	4.68