A SCALABLE VIDEO COMPRESSION TECHNIQUE BASED ON WAVELET TRANSFORM AND MPEG CODING

Pao-Chi Chang and Ta-Te Lu

***Abstract*—We present a scalable coding scheme based on discrete wavelet transform (DWT) and MPEG coding for video applications. It utilizes the hierarchical pyramid structure that provides multiple resolutions. DWT decomposes the image into several bands. In each band, a fixed-size motion compensated MPEG coder with custom-designed quantization tables and scanning direction is employed. This scheme has the advantages of reusing the widely available MPEG hardware and software as well as relieving the limitation of the image size imposed by the MPEG hardware technology. The simulation results show that the DWT-MPEG coding method improves the image quality over ordinary MPEG coding by 0.3 ~ 1.5 dB.**

***Index Terms*—Video compression, Scalable coding, Discrete wavelet transform, MPEG.**

I. INTRODUCTION

With the development of video technology and high bandwidth networks such as ATM, video services are getting more popular in our daily life. A scalable coding is desirable for providing various levels of video quality in a heterogeneous environment. With scalable coding, we can reconstruct video sequences with only the partial bit stream so that the coded signal can be stored or transmitted with different resolutions or bitrates. Many multiresolution coding systems with spatial scalability have been investigated over the last decade [2]-[6]. Discrete wavelet transform (DWT) has shown great potential in video coding and its structure is particularly suitable for scalable coding [5]-[7].

In this work, we use DWT to perform band decomposition and the widely available MPEG coding to perform the compression for each band. With the finalization of MPEG standard, the implementation technology of MPEG codec has been improved quickly in both software and hardware [1] . Thus we intend to design a scalable coding scheme that only needs DWT as a pre-processor and minor modifications to existing MPEG codecs so that not only the development time can be much shortened but also the cost can be reduced since the existing MPEG hardware and software can be reused. In addition, it also provides the expansion structure for MPEG coding. Namely, the limitation of the image size imposed by MPEG hardware technology can be relieved by using multiple MPEG chips for different bands. Thus we propose a scalable DWT with MPEG coding scheme (DWT-MPEG) that employs wavelet pyramid structure, hierarchical fixed block-size motion compensation, and modified MPEG coder. In the modified MPEG coder, custom-designed quantization matrices and scanning directions are applied to different bands for achieving better performance.

The paper is organized as follows. In section II, we provide an overview of the DWT-MPEG system. Section III presents band analysis and modified MPEG coding particularly. The simulation results are discussed in section IV. Finally, conclusions are given in section V.

II. DWT-MPEG SYSTEM OVERVIEW

The block diagram of a three-layer DWT-MPEG encoding system is shown in Fig. 1 as an example. We use a 9-7 taps biorthogonal wavelet filter [7] which has the linear phase and is able to decompose the image into three layers with seven non-uniform bands, labeled as in Fig. 2. Observing that the lowest band, LL2, has similar statistic features with the original image, we apply MPEG coding with default setting to LL2 directly. However, the default parameters of MPEG are not optimal for other bands due to different statistic characteristics. Thus we need to modify MPEG scheme for higher bands to improve the coding efficiency.

Because the motion activity between different bands are highly correlated, the searching time will decrease substantially by using hierarchical fixed size multiresolution motion estimation [4], [5]. Moreover, we observe that the energy of each band is concentrated only on a small but different part of DCT coefficients. As a result, efficient coding can be achieved with custom designed quantization matrices in MPEG coder for different band. After analyzing the correlation of each band, we find spatial redundancy can be further removed by utilizing distinct scanning directions in MPEG coder. Rate control is applied to each band based on the bit allocation. Finally, the encoded bit streams are sent to multiplexer before transmission to the networks. In decoder, depending on different resolution requirements,

---

This work was jointed supported by National Science Council and Mentor Data Systems, Taiwan, under Grant NSC-87-2213-E-008-041 and M04-8603, respectively. The authors are with the Department of Electrical Engineering, National Central University, Chung-Li, Taiwan 320.

Manuscript received June 28, 1999

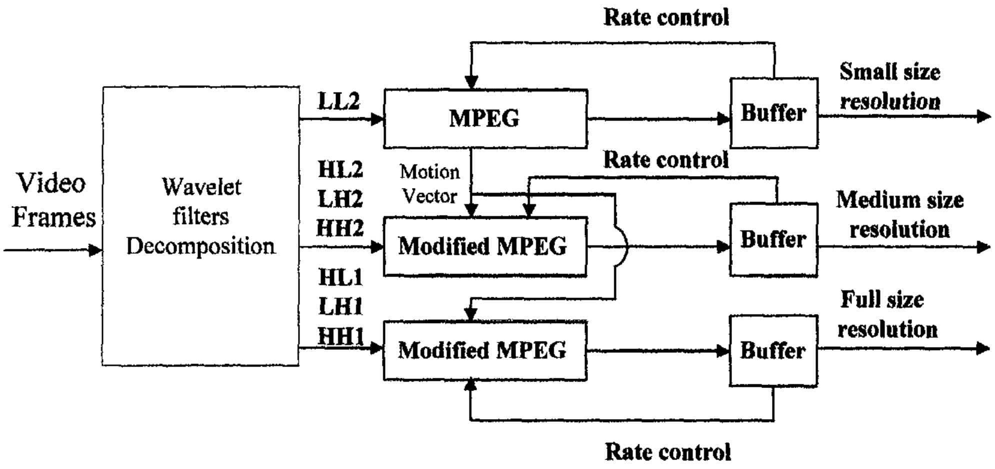


Fig. 1. Encoder of a three-layer DWT-MPEG video coding system.

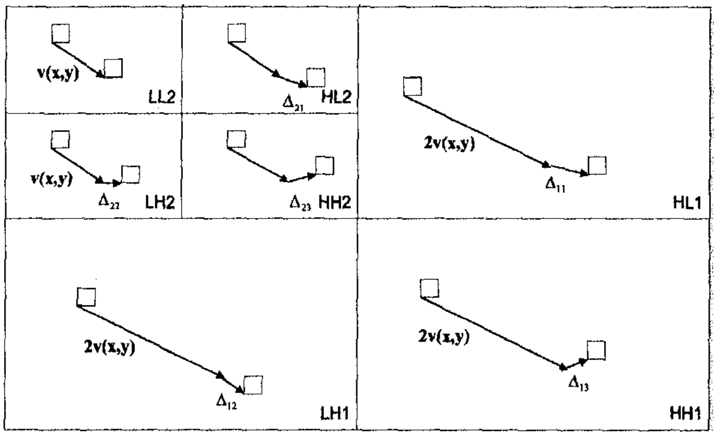


Fig. 2. Fix block-size multiresolution motion estimation.

some or all of these bands can be synthesized to reconstruct video frames.

III. MODIFIED MPEG CODING

The modified MPEG encoder is shown in Fig. 3. It has the same architecture as the original MPEG, including DCT, scalar quantization, scanning and variable length coding (VLC), with the differences in hierarchical motion estimation/compensation, custom-designed quantization matrices, and distinct scanning directions.

A. Hierarchical Motion Estimation and Compensation

The hierarchical fixed motion estimation is shown in Fig. 2. The motion estimation in the lowest band, LL2, is performed as in MPEG. Because the LL2 band is already down sampled twice in each direction, the search area for motion estimation in this band is only ±6 pixels horizontally and vertically. To be compatible with MPEG, the same block size, 16x16 pixels, is applied. Because motion activities in different bands are highly correlated, the motion vector in LL2 can be reused in high frequency bands. In LH2, HL2, and HH2, it is sufficient to use this motion vector directly as the initial value in both directions since they have the same resolution with LL2. In LH1, HL1, and HH1 bands, each LL2 motion vector is used for 4 adjacent blocks with the size of the 16x16 MPEG motion compensation block size. We take the LL2 motion vector multiplied by two as an initial guess, and refine the motion vector by searching within a smaller area, ±3 pixels, in both directions. The motion compensation in each band is equal to what MPEG performs. By reusing the motion information, this structure is easy to implement with low computational complexity.

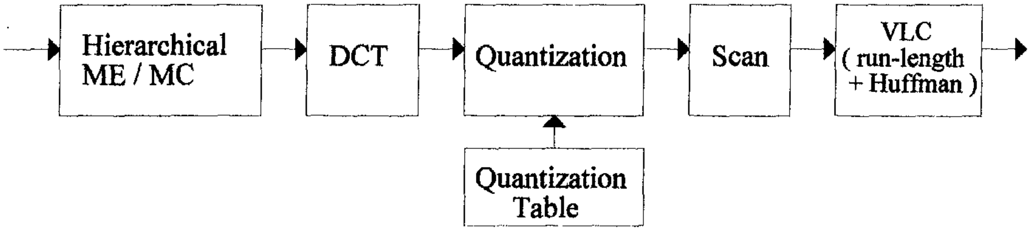


Fig. 3. Modified-MPEG encoder.

Table I
Quantization matrices in different bands.

15	28	20	29	25	25	30	29
17	15	26	21	26	25	26	29
12	19	17	26	29	32	28	27
10	16	17	17	18	22	23	25
9	13	18	15	19	18	33	25
8	12	13	15	19	17	19	26
8	10	15	15	17	20	18	24
8	15	12	17	17	20	25	41

(a) LH2

13	10	18	17	17	11	10	9
20	14	14	20	23	14	16	13
23	19	17	17	16	16	15	13
21	21	24	20	21	24	26	17
28	27	24	30	23	21	22	21
36	40	27	27	26	20	32	18
39	33	38	27	27	24	23	22
40	33	32	50	26	25	50	26

(b) HL2

19	14	16	13	18	13	10	17
21	31	15	11	11	15	13	10
17	17	14	17	14	9	18	11
20	16	14	15	11	13	9	9
13	13	14	16	10	11	11	11
17	14	17	13	10	12	13	10
14	13	12	14	17	14	11	10
21	20	21	11	22	12	11	11

(c) HH2

16	13	11	11	17	18	11	16
13	15	18	18	13	16	14	15
14	13	13	16	15	13	16	18
13	15	18	16	17	17	19	19
13	22	17	19	11	14	17	19
10	21	15	19	13	13	24	19
9	14	15	13	11	13	15	22
7	13	12	11	13	15	11	18

(d) LH1

15	14	15	20	14	13	9	7
16	15	18	17	18	14	16	10
19	21	27	21	20	14	16	13
26	27	23	16	26	21	15	16
20	22	18	23	22	24	18	22
22	20	21	20	19	25	19	14
20	22	23	25	32	19	16	15
22	23	21	22	29	22	19	22

(e) HL1

25	23	21	18	22	14	11	8
30	21	14	18	17	10	13	9
24	22	18	12	16	13	12	7
33	13	17	13	13	14	6	6
23	12	17	11	16	11	6	6
23	14	12	11	12	8	6	3
25	10	16	9	6	3	4	4
11	9	9	6	5	8	3	3

(f) HH1

B. DCT transform and quantization matrix design

The statistical properties of DCT coefficients in each band are significantly different. When a block has vertical edges, its energy concentrates on the first several rows of AC coefficients. Conversely, if a block contains horizontal edges, its energy concentrates on the first several columns of AC coefficients. The default quantization matrices suggested in MPEG have large step levels at high frequency parts, i.e., fewer bits are allocated to quantize high frequency DCT coefficients, because high frequency coefficients are less important perceptually. Since the image in the lowest band has similar statistic features with the original image, we use the default quantization matrices defined in MPEG in this band. Based on the DWT structure, however, the significant data that we need to reconstruct a smooth lowpass image are concentrated on high frequency DCT coefficients in highpass bands. Therefore, we need to quantize these coefficients appropriately by using custom-designed quantization matrices for each band to improve the reconstruction performance.

The quantization tables in DWT-MPEG are designed based on energy distribution for still image [8]. It suggests to give more bits to the parts with large energy. The energy ratio $E_{i,j}$ for the (i,j)-th DCT coefficient is defined as

$$E_{i,j} = \frac{\sigma_{i,j}^{2}}{\sum_{m+n\neq 0} \sigma_{m,n}^{2}} \tag{1}$$

where $\sigma_{i,j}$ stands for the root mean square value of the energy of (i,j)-th DCT coefficient. The quantization stepsize for (i,j)-th coefficient is defined as

$$Q_{i,j} = \frac{max\left|C_{i,j}\right|}{E_{i,j}} \times \frac{1}{\Theta} \tag{2}$$

where $C_{i,j}$ represents the (i,j)-th DCT coefficient, and Θ designates the quality factor assigned, which can be adjusted dynamically. A smaller Θ results in higher

Table II
Corresponding bitrates in different bands with different scanning method under fixed PSNR

Band	LH2	HL2	HH2	LH1	HL1	HH1
PSNR(dB)	47.2	47.1	45.2	45.9	45.6	49.2
Default(bps)	273443	265795	129297	642758	608475	560822
Proposed(bps)	268115	259301	123888	622533	595127	439463

Table III
"MIT" and "Renata" the bit allocation of each band at 0.2 bpp.

Band	LL2	LH2	HL2	HH2	LH1	HL1	HH1
Variance "MIT"	2122.8	50.82	52.02	16.98	19.79	22.38	2.55
Bit rate (bpp)	0.888	0.58	0.592	0.213	0.097	0.119	0.013
Variance "Renata"	4097	96.5	75.5	9.87	126.06	51.32	34.15
Bit rate (bpp)	0.888	0.731	0.581	0.074	0.137	0.056	0.037

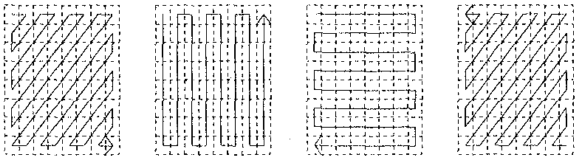

(a)Zig-zag (b)Vertical (c)Horizontal (d)Inverse zig-zag

Fig. 4. Custom-designed scanning directions for different bands.

compression ratio while a larger Θ results in better quality of image. Table I shows the custom-designed quantization matrices which are designed based on this method.

C. Scanning directions and variable length coding

In order to increase coding efficiency of variable length coding (VLC), an appropriate scanning method is selected to convert the 2-D signals into 1-D sequences in the order of descent significance. The zig-zag scan is commonly used in JPEG and MPEG standards. However, this is not the optimal choice for high frequency bands because of different energy distributions. Therefore, we apply four scanning methods to different bands, according to the energy distribution shown as in Fig. 4. The zig-zag scan is used in LL2 band. The vertical and horizontal scans are applied to LH1, LH2 bands, and HL1, HL2 bands, respectively. Finally, the inverse zig-zag scan is chosen for HH1 and HH2 bands. Table II shows the corresponding bitrates with zig-zag scan and proposed scanning methods under fixed PSNR. Obviously, the suggested scanning directions can make variable length coding (VLC) more efficient and reduce bitrates.

D. Optimal bit allocation

Before encoding, we need to determine the bit rate for each band. The objective of bit allocation is to adjust the bit rate in each band to minimize the reconstruction error while the total rate is fixed. We adopt the perceptual weighting bit allocation in [9]. It gives different weightings to each band based on the sensitivity of human visual system [10].

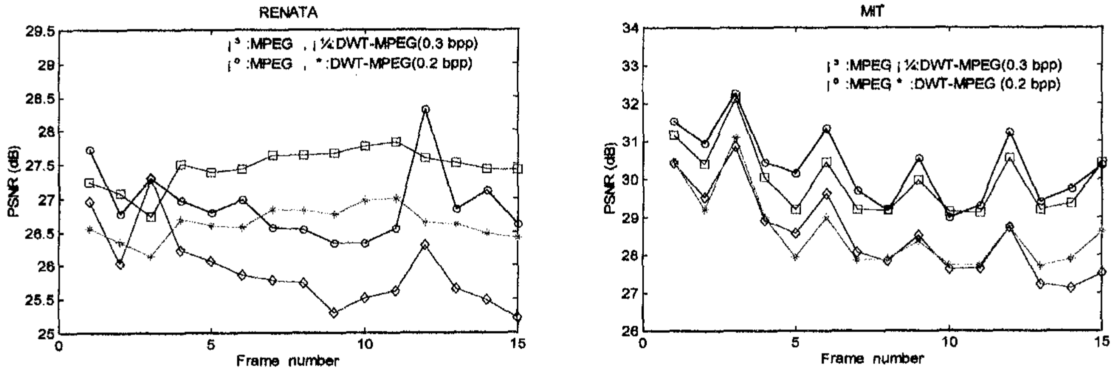
The weighting factor λ is calculated as

$$\lambda = 2^{\sum_{K=1}^{M} \alpha_K \log_2((2\log_e 2)W_K\sigma_K^2)-2R_C)} \tag{3}$$

and the bit rate b_k of k-th band is dertermined by

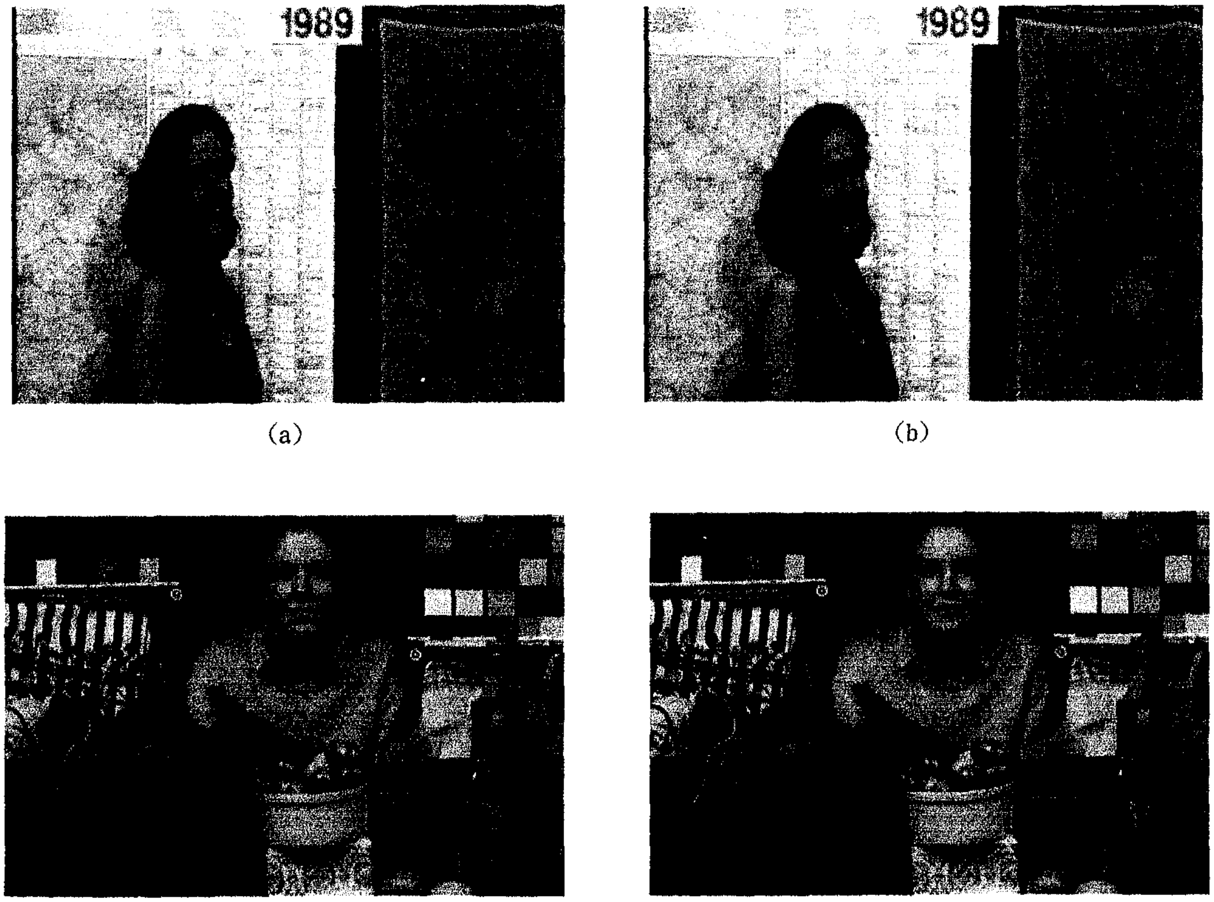
$$b_K = \frac{1}{2}\log_2(\frac{(2\log_e 2)w_K\sigma_K^2}{\lambda}) \tag{4}$$

where R_C is the total bitrates, w_K is the weighting factor for perceptual coding. In level 1 we choose $w_K = 1$ and in level 2 we choose $w_K = 2$ expect LL2. σ_K^2 is the energy of k-th band, M is the number of bands, and α_k is the relative band size. The bit allocation of each band at 0.2 bpp calculated based on "MIT" and "Renata" are shown in Table III.



(a) (b)

Fig. 5. Comparisons of PSNR between MPEG and DWT-MPEG at 0.3 and 0.2 bits/pixel. (a) Renata. (b) MIT.



(a) (b)

(c) (d)

Fig. 6. Reconstructed image quality of Renata and MIT at the rate of 0.2 bpp. (a)DWT-MPEG (26.41dB). (b) MPEG (25.22dB). (c) DWT-MPEG (28.64 dB). (d) MPEG (27.53 dB).

IV. SIMULATION RESULTS

We perform simulations of the proposed system on a Sparc-20 workstation. The video sequences “MIT” and ”Renata” with the size of 1408x960 pixels and 4:2:0 format are used as the test sequences. Other parameters include that the number of frames in a GOP is 12 (M=12), the frame distance between I and P frames is 3 (N=3), and the frame rate is 24 frames/sec. Fig. 5. shows the PSNR of Y component of MPEG and DWT-MPEG at the bit-rate of 0.3 bpp and 0.2 bpp. At a relatively high bit rate 0.3bpp, there is no clear winner. However our proposed DWT-MPEG yields 1.5~0.3 dB improvement over MPEG for “Renata” and 1.1~0.1 dB improvement for “MIT” at the rate of 0.2 bpp. Thus it is particularly suitable for relatively low bit rates. Fig. 6. shows the quality of reconstructed video sequences from MPEG and DWT-MPEG, respectively, at 0.2 bpp for ”Renata” and ”MIT”.

V. CONCLUSION

We have proposed the DWT-MPEG video coding scheme that supports resolution scalability. It has the advantages of reusing the existing technology of MPEG codec in software and hardware as well as short development time for implementation. The constraint of the maximum image size imposed by the hardware capability is also relieved. At a relatively low bitrate, our proposed DWT-MPEG coding provides significantly better reconstructed video quality than the original MPEG.

REFERENCES

[1] ISO-IEC/JTC1/SC29/WG11, MPEG93/457, Coded Representation of Picture and Audio Information, Test Model 5,April 1993

[2] T. Naveen and J. W. Woods, ”Motion Compensated Multiresolution Transmission of High Definition Video”, *IEEE Trans. Circ. And Syst. Video Tech.*, vol. 4, pp. 29-41, Feb. 1994.

[3] S. G. Mallat, “A theory for multiresolution signal decomposition: The wavelet representation,” *IEEE Trans. Pattern. Analysis and Machine Intell.*, vol. 11, pp. 674-693,July 1989

[4] S. Zafar, Y. Zhang, and B. Jabbari, “Multiscale Video Representation Using Multiresolution Motion Compensation and avelet Decomposition,” *IEEE Journal on Selected Areas in Communications*, vol. 11, pp. 24-35, Jan. 1993.

[5] Y. Q. Zhang and S. Zafar, “Motion-compensated wavelet transform coding for color video compression,” *IEEE Trans. Circ. And Syst.Video Tech.*, vol. 2, pp. 285-296, Sept. 1992.

[6] S. Kim and S. Rhee, “Interframe Coding Using Two-stage Variable Block-Size Multiresolution Motion Estimation and Wavelet Decomposition,” *IEEE Trans. Circ. And Syst. Video Tech.*, vol. 8, pp. 399-409, August. 1998.

[7] M. Antonini, M. Barlaud, “Image Coding Using Wavelet Transform”, *IEEE Trans. on Image Processing,* vol. 1, no 2, April 1992

[8] H. W. Paik and A. Khubchandani, “Quantization Scheme for JPEG Baseline Sequential Encoding of Still Images”, *Proceeding of 35th Midwest Symposium on Circuit and System*, vol. 2, pp. 976-979, 1992.

[9] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley Cambridge, 1996.

[10] K. R. Rao and J. J. Hwang, Techniques, standards for Image,video, and Audio Coding, Prentice Hall PTR,1996.

Pao-Chi Chang received the B.S. and M.S. degrees from National Chiao Tung University, Taiwan, in 1977 and 1979, respectively, and the Ph. D. degree from Stanford University, California, 1986, all in electrical engineering.

From 1986-1993, he was a research staff member of the department of communications at IBM T. J. Watson Research Center, Hawthorne, New York. At Watson, his work centered on high speed switching system, efficient network design algorithms, network management, and multimedia teleconferencing. In 1993, he joined the faculty of National Central University, Taiwan. His current interests are in the area of speech/image coding, video conferencing, and video coding over high speed networks and wireless communications.

Ta-Te Lu was born in Taiwan on March, 19, 1966. He received the B.S degree in Electrical Engineering from Fu Jen University in 1989 and the M.S. degree in Electrical Engineering from Chung Cheng Institute of Technology in 1991. He is currently a Ph.D candidate in National Central University.

His current interests are wavelet image/video coding, error concealment techniques.