Very Low Bit Rate Video Image Coder Using the Fractals

Yong Hoh Kim*, Jong Whan Jang*, Jae Gil Jeong*, Dooyeong Park*

ABSTRACT

New very low bit rate segmentation video image coding technique is proposed by segmenting image into textually homogeneous regions. Regions are classified into one of three perceptually distinct texture classes(perceived constant intensity, smooth texture, and rough texture) using the Human Visual System(HVS) and the fractals. To design very low bit rate video image coder, it is very important to determine the best block size for estimating the fractal dimension and the thresholding of the fractal dimension for each texture class. Good quality reconstructed images are obtained with about 0.10 to 0.21 bit per pixel(bpp) for many different types of imagery.

I. Introduction

Toward very low bit rate environment [9, 11], new image compression techniques are strongly required for various applications such as in the areas of digital video codecs for desktop multimedia computers, electronic publishing, and video teleconferencing etc. Symbolically based image compression technique which is a promising solution [2, 5, 7-8] employs properties of the HVS and tools of image analysis to achieve good image quality at very low bit rates. One approach to symbolically based image compression techniques is segmentation based image compression. In segmentation based image compression, the image to be compressed is segmented, i.e. the pixels in the image are separated into regions having widely differing perceptual importance. The importance of a region corresponds to the amount of information it conveys to the viewer. Typically, the amount of local detail or high frequency content, is considered a reasonable measure of this importance; however, this by itself is an inadequate measure for efficient very low bit rate image compression [7]. It is also desirable to consider other significant characteristics, such as texture and the global context of the local region, in order to assess the local information content. Certain regions are critical to our subjective evaluation of quality, and relatively small errors can perceptually have a major degrading effect on the overall reproduction quality. Such regions tend to dominate the viewer's attention and are intrinsically less compressible than background segments. Consequently, when the overall bit rate is low, a uniform allocation of bits across image implies that the spatial distribution of perceptual degradation is highly nonuniform; some regions have a starvation diet of bits, causing a significant degradation, while other regions have been coded with far more bits than needed for perceptually transparent quality.

In the proposed new technique, we overcome the texture representation problem and determine the best block size for estimating the fractal dimension and the thresholding of the fractal dimension. The segmentation technique we present segments image into texturally homogeneous regions with respect to the degree of perceived roughness using the HVS and the fractals. After segmentation, the image can be viewed as being composed of region boundaries and texturally homogeneous regions. As image coding system with high compression and good image quality is achieved by developing an efficient coding technique for the region boundaries and the three textural classes. The proposed algorithm is applied to different types of imagery.

In section 2, determination of the best block size for estimating the fractal dimension and the thresholding of the fractal dimension is presented. In section 3, we describe the proposed texture image segmentation and the proposed actual coding scheme. Finally, conclusions are provided in section 4.

*Division of Computer, Informations & Communications, and Electronics Engineering

Manuscript Received June 4, 1996.

※This paper was supported by NON DIRECTED RESEARCH FUND, Korea Research Foundation

Ⅱ. Determination of the best block size for estimating the fractal dimension (D) and the thresholding of D

In the proposed compression algorithm, an image is divided into blocks for the efficient implementation for the proposed coding system and D of each block is computed. It is very important to choose the best block size in the image so that good estimates of D are obtained and good image quality can be maintained. When the block size is small, there may not be enough pixels to describe the texture within the block. For example, if the block is 1×1, there is only one pixel in the block and it is not possible to characterize its texture. When the block size is large, several different textures may be present within the block and the estimated D will not accurately represent the characteristics of the multiple textures. Another issue to be considered when choosing the block size is the computation requirements. The computation requirement for the large block is more expensive than the one for the small block. For example, consider the block sizes 8×8 and 16×16. The number of pixels in the block size 16×16 is four times larger than the number in the block size 8×8. Therefore, the computation time to compute D in the larger block is more expensive. Considering the issues discussed above, we conclude that the smallest feasible block size is the best block alternative.

2.1 Fractal dimension (D)

The definition of D is a set for which the Hausdorff-Besicovich dimension is strictly greater than the topological dimension [1]. We consider object X in an E-dimensional space. $N(\varepsilon)$ is the number of E-dimensional sphere of diameter ε needed to cover X, where E is an integer and the E-dimensional space is the minimum integer dimensional space among all possible integer dimensional spaces which can envelop X. Thus, if $N(\varepsilon)$ is given by

$$N(\varepsilon) = K \cdot (\frac{1}{\varepsilon})^{D}, \qquad as \ \varepsilon \rightarrow 0, \tag{1}$$

where K is a constant and X has Hausdorff dimension D. If D is fractional, D is also called the fractal dimension. For fractal objects, D is independent of ε [1].

If D is to be used to characterize the texture in an image, we need a method for estimating D from the given dataset. Many different estimators have been proposed; box counting [1], yardstick [3], and blanket [10]. In our case, a blanket method is adopted since it is computationally efficient. The blanket method is described in detail in paper [10]. A brief explanation of the procedure for estimating D is given here. All points in the three-dimensional space at distance ε from the surface are considered, covering the surface with a blanket of thickness 2ε. The surface area $A(\varepsilon)$ is then the volume $V(\varepsilon)$ occupied by the blanket divided by 2ε. The area $A(\varepsilon)$ is given by

$$A(\varepsilon) = \frac{V(\varepsilon)}{2\varepsilon} = \varepsilon^{2} \cdot N(\varepsilon) = K \cdot \varepsilon^{2-D} \tag{2}$$

where K is a constant.

From a theoretical viewpoint, if a surface is a perfect fractal surface, then D will remain constant over all ranges of scale ε. In practice, there are scale range limitations of D due to limitations in textural images. For example, the resolution limit of the image system sets a lower limit on the fractal scaling behavior. An upper limit may be set by the structure being examined. Thus, a real surface will be fractal over some range of scales rather than over all scales. These limiting scales can be expressed as upper (ε_{max}) and lower cutoff (ε_{min}) scales.

To compute D, we apply the log function to both sides of Eq. (2). A least square linear regression is applied to fit a straight line to the plot of $\log A(\varepsilon)$ vs. $\log(\varepsilon)$ for ε_{min} and ε_{max}. The fractal dimension D is equal to 2 minus the slope of the straight line.

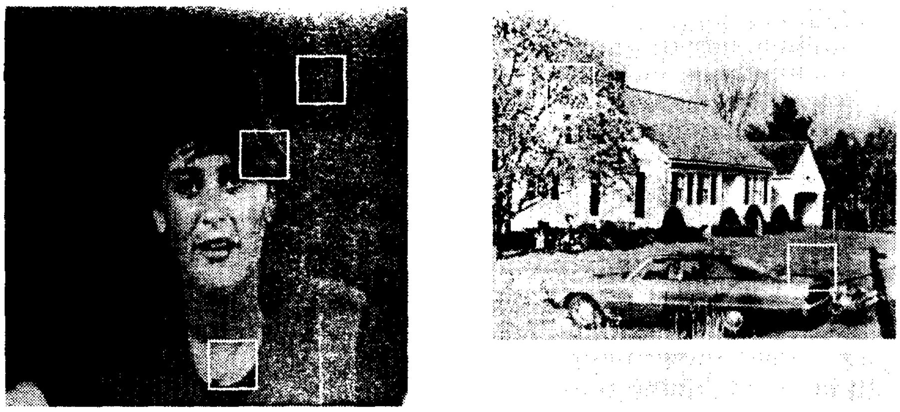
$$\log A(\varepsilon) = (2-D)\log(\varepsilon) + K \tag{3}$$

From Eq. (3), we can deduce a procedure to estimate D of an image surface using a least-square linear regression.

2.2 Experimental results for determining the best block size for estimating D

To determine the best block size in terms of the fractal dimension (D), each three 30×30 subimages of two different types of test images with 256×256 pixels with 256 gray levels given in Fig. 1 are taken. The one in Fig. 1(a) is a head and shoulder image with little texture variation. This image is typical of video teleconferencing applications. The other in Fig. 1(b) is a natural outdoor image with highly textured areas. To achieve the segmentation efficiently, three subimages of Miss USA in Fig. 1(a) on the top, middle, and bottom belong to class Ⅰ, class Ⅱ, and class Ⅲ respectively. Three subimages of House in Fig. 1(b) on the right and top, bottom, and left and top belong to class Ⅰ, class Ⅱ, and class Ⅲ respectively.

A plot of D versus block size is given in Fig. 2. The curves with a diamond symbol(◇), a cross symbol(+), and a square symbol(□) correspond to class Ⅰ, class Ⅱ,

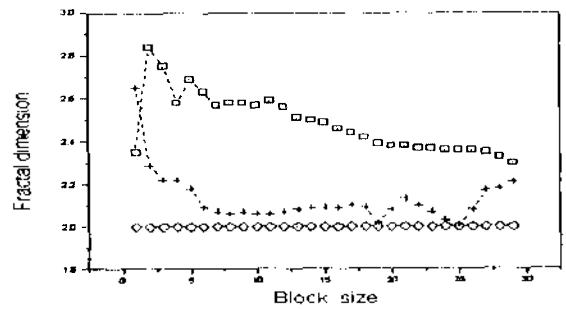


(a) Miss USA (b) House

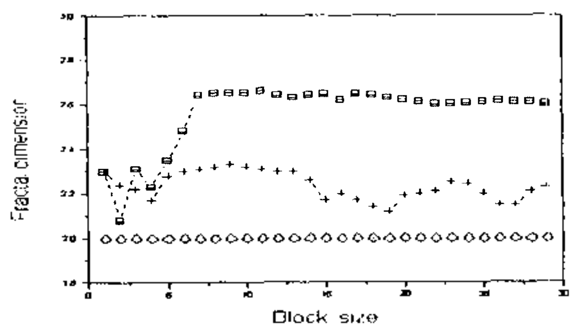
Fig 1. Two test images of Miss USA and House with each three 30×30 subimages. Three subimages of Miss USA on the top, middle, and bottom belong to class Ⅰ, class Ⅱ, class Ⅲ respectively.

and class Ⅲ respectively. We investigate the variation of D versus the block size for each class. Block size is varied from 2 to 30 and is increased from the left, top corner. Curves of D versus block size are given in Fig. 2. In each plot, the x-axis represents the block size and the y-axis D. D corresponding to ◇ has almost a constant value, 2.0, for blocksize 2, ..., 30. That is, there exists only a single texture of class Ⅰ in each block. The shape of the curve corresponding to + is quite variable for blocks between (2, ..., 7) and (19, ..., 30) but is nearly constant for the middle block size (8, ..., 18). The reason for variability in the smaller 2, ..., 7 blocks is the small number of pixels to characterize the texture. In larger 19, ..., 30 blocks more than one texture is present. The middle 8, ..., 18 blocks have only one texture and provide the least variable estimate of D. Note, 30×30 subimage at the bottom of Fig. 1(a) has three different textures; the neck and two sweaters. In general, when the block size is large, there is more likelihood that several textures will be in the block, and the value of D will not remain constant. The shape of the curve corresponding to □ looks similar to that of class Ⅲ. At the small block sizes, there are too few pixels to estimate the texture and at the large block size, multiple textures are present in the block. The curve is nearly constant for middle (8, ..., 18) block sizes. In summary, the larger block sizes may not give good estimates of D because they contain several textures and the smaller block may not contain enough pixels to characterize the texture.

Through extensive experimentation, we have found that block sizes of 8×8 up to 14×14 have almost a constant value. Thus these blocks meaningfully represent the textural characteristics of a region. The means of D's and the standard deviations of D for the blocks in each class for the test images as a result of these simulations are given in table 1 and table 2.



(a)



(b)

Fig 2. Plot of D versus block size in (a) Miss USA and (b) House. The curves with a diamond symbol (◇), a cross symbol (+), and a square symbol (□) correspond to class Ⅰ, class Ⅱ, and class Ⅲ respectively.

Table 1. Statistics of variation of D in Miss USA

class	block size	mean of D	standard deviation
class Ⅰ	2×2 to 28×28	2.000605	0.000015
class Ⅱ	8×8 to 19×19	2.271142	0.000154
class Ⅲ	8×8 to 14×14	2.665640	0.000075

Table 2. Statistics of variation of D in House

class	block size	mean of D	standard deviation
class Ⅰ	2×2 to 28×28	2.000000	0.000000
class Ⅱ	8×8 to 19×19	2.296040	0.000094
class Ⅲ	8×8 to 14×14	2.650985	0.000172

We choose an 8×8 block size for the block-by-block segmentation algorithm in the following section since the smaller block size reduces the computation and storage requirement and as will be seen later is consistent with giving the best image quality. Furthermore, by comparing curves in the plot, curves on □, +, and ◇ symbols are the top, middle, and bottom respectively for the mid sized blocks from 8×8 to 14×14 rougher texture produces higher fractal dimension (D).

2.3 Experimental results for determining the thresholding of D

Five 8×8 subimages belonging to each class in each test image with 256×256 pixels, and 256 gray levels are chosen. In this experiment, 8×8 block size is used for each subimage because the block size is the smallest one to characterize meaningfully the texture of a region as discussed in section 2.2. For Miss USA, 5 subimages in the background and sweater are chosen to represent class Ⅰ, 5 subimages in the neck, cheeks and shoulder for class Ⅱ, and 5 subimages in the hair, eyes, mouth, and nose for class Ⅲ. For House, 5 subimages in the sky and concrete wall are chosen for class Ⅰ, 5 subimages in the lawn and car for class Ⅱ, and 5 subimages in the trees for class Ⅲ.

A plot of D of the fifteen subimages for each class is given in Figure 3. In the plot, the x-axis represents D and the y-axis the number of blocks at that D. The curve with ◇ corresponds to class Ⅰ, the curve with + to class Ⅱ, and the curve with □ to class Ⅲ. D belonging to class Ⅰ are distributed around $D = 2.0$. The curves of D belonging to class Ⅱ and class Ⅲ are approximately bell-shaped around their means respectively. There are gaps between the curves for each class. Through these results, the value of D_1 should not be greater than the minimum $D_{\mathrm{II}min}$ of D belonging to class Ⅱ. The value of D_2 should be greater than the mean of D of class Ⅱ. We propose $D_1 = D_{\mathrm{I}max}/2$ and $D_2 = D_{\mathrm{II}mean}/2$, where $D_{\mathrm{I}max}$ and $D_{\mathrm{II}mean}$ are the maximum D belonging to class Ⅰ and the mean D belonging to class Ⅱ respectively. Therefore regions belonging to class Ⅰ (perceived constant intensity) have D less than D_1. The second class (smooth texture) contains regions with D between D_1 and D_2. The third class (rough texture) contains regions with D greater than D_2.

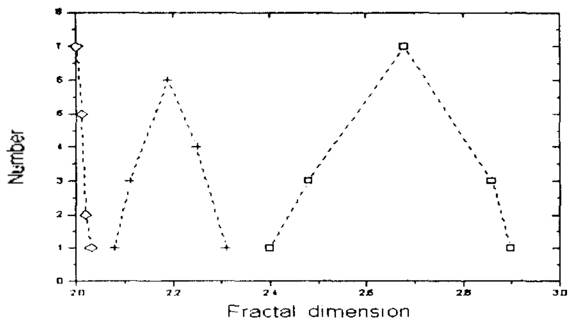


Fig 3. A plot of D of the subimages for each class. The x-axis represents D and the y-axis the number of blocks at that D. The curves with ◇, +, □ correspond to class Ⅰ, class Ⅱ, and class Ⅲ respectively.

Ⅲ. The proposed texture image segmentation

The goal of the image segmentation process is to decompose an image into texturally homogeneous regions with respect to the degree of roughness as perceived by the HVS. Textural regions are classified into three classes; class Ⅰ, class Ⅱ, and class Ⅲ. For example, the background in a head and shoulder image or the sky in a natural image is considered as class Ⅰ, the face or the shoulder is considered as class Ⅱ, and the trees and the bushes in a natural image are considered as class Ⅲ. To extract texture information for accomplishing textural-based image segmentation, the fractal dimension (D), mean, and just noticeable difference (JND) are used in the segmentation algorithm. The segmentation algorithm is based on a region growing technique [4]. A unique of feature of the region growing process used in this research is that it is directed by the texture feature distance between image blocks. The region growing is achieved through a merging test condition between texturally homogeneous neighboring blocks. If the condition for merging is satisfied, an observing block can be merged into a neighbor block. Otherwise, a new region is declared.

For our segmentation, we have used a centroid linkage region growing method because it is guaranteed to produce disjoint segments with close boundaries and provides a sequential algorithm for growing region. The centroid linkage region growing method is illustrated in paper [4]. The texture features are used the mean, JND, and the class type based on D of the image block.

Incorporating the HVS and the fractal model, the proposed texture-based image segmentation algorithm for image coder is defined as follows.

Step 1) Divide the image into $NR \times NC$ blocks (NR and NC are the numbers of row and column blocks, respectively).

Step 2) Calculate the feature set: the mean and the class type for each block and the JND lookup table.

Step 3) Calculate the distance between an observing block and its 4-connected neighboring blocks. The distance is given by

$$D(OB, NB) = \begin{cases} 0 \quad \text{if} & \begin{cases} F(OB) < D_1,\ C(OB) = C(NB) \\ |M(OB) - M(NB)| < JND(OB, NB) \\ \text{or} \\ D_1 \le F(OB) < D_2,\ C(OB) = C(NB) \\ \text{or} \\ F(OB) \ge D_2,\ C(OB) = C(NB) \end{cases} \\ 1 \quad \text{otherwise} \end{cases}$$

where $F(OB)$ is D of an observing block. $C(OB)$ and $C(NB)$ are the class types of an observing block and its neighboring block respectively. $M(OB)$ and $M(NB)$ are the means for an observing block and its neighboring block respectively. $JND(OB, NB)$ is JND between an observing block and its neighboring block.

Step 4) If there is a neighboring block with distance 0, then merge the observing block into it; else declare a new region. If there are more than two good neighboring blocks, merge the observing block into a neighboring block whose mean value is closest to the mean value of the observing block.

Step 5) Repeat step 3 to step 4 until all blocks are segmented and stop.

The proposed texture segmentation-based image coder system for the very low bit rate is given in detail in paper [7]. The number of segments and the number of bits representing the textures of the segments are directly proportional to the bit rate of the coded image. Thus, the main purpose of the preprocessor which is the first stage of the proposed transmitter is to alter the image in such a way that fewer segments and textures are proposed by the segmenter, but without degrading the visual quality of the segmented image. After preprocessing, the image data is segmented into texturally homogeneous regions with respect to the degree of roughness as perceived by the HVS. The segmentation is accomplished by thresholding the fractal dimension discussed in section 2.3. The last stage in the transmitter is the mixed encoding of the segments of each class and their boundaries. For boundary coding, accurate representation of the boundary is necessary to describe the location of the region boundary because of the HVS sensitivity of the edges. We choose an errorless coding scheme to represent the boundaries. A binary image representing the boundaries is created. Then, the binary data is encoded using an adaptive arithmetic code since it has been found to be superior to Huffman code, runlength code, and crack code. For regions which belong to perceived constant intensity, only the mean intensity values need be transmitted to describe the textures of the regions. In this case, lossy compression has already taken place since we are approximating each region texture with a constant value. We do not wish to introduce any further compression so a lossless adaptive arithmetic code is again employed to achieve further compression. Since a mean intensity requires 8 bits, the mean values are perceived constant regions. Regions belonging to smooth texture and rough texture are not directly encoded. To get higher compression, these regions are modeled first using the 1-D first order polynomial function. The coefficients of the polynomial functions are encoded because the variance of the coefficients is less than that of the original data. An adaptive arithmetic code is used to encode the coefficients.

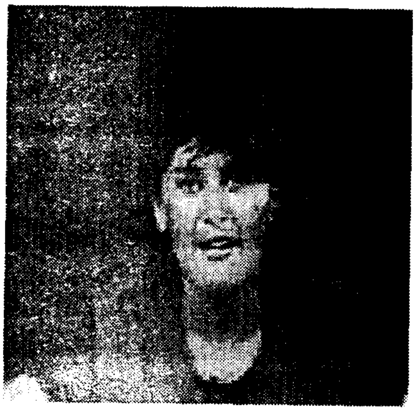
To compute the total number of bits required to transmit an image, the three numbers of bits calculated for the boundary, constant region, and smooth/rough texture are added. The bit rate is the sum of the number of bits divided by the total bits of an image. The bit rate, BR is given by

$$BR = \frac{SP + CP + BR}{256 \times 256 = 65536} \qquad (3)$$

where SP is the number of bits required for encoding of the boundaries, CP is the number of required for encoding of the constant regions, and BP is the number of required for encoding of the smooth and rough texture regions.

Ⅳ. Conclusions

Decoded images are obtained with $D_1 = 2.033$, $D_2 = 2.371$, and block size 8×8 for the proposed texture segmen-



(a) the decoded Miss USA

(b) the decoded House

Fig 4. The decoded images of the two test images using $D_1 = 2.033$, $D_2 = 2.371$, and the best block size 8×8. The compression rates of Miss USA and House are 0.11 and 0.21 respectively.

lation compression technique. Each image consists of $ 256×256 pixels with 256 gray levels. The images are viewed on a 20" SUN monitor with 256 possible gray levels. The monitor was calibrated so that there was a linear relationship between gray level numeric value and output luminance. The compression ratios for the two test images Miss USA and House are 0.11 and 0.21 bpp respectively. The decoded images for the two test images are given in Fig. 4. The CR(compression ratio)/SNR is given in the table 3. It shows that the proposed texture segmentation-based image coder performs good in terms of SNR.

Table 3. The table of CR/SNR for two test images. CR stands for compression ratio.

CR \ Image	Miss USA	House
8	27.3	15.5
10	26.4	14.8
20	25.9	12.0
40	21.2	9.7
80	18.1	8.5
100	15.2	6.9

These results indicate that, using the new texture-based segmentation image compression system, compression ratios in the neighborhood of 0.11 to 0.21 bpp are attainable with good image quality for the various imagery in very low bit rate.

One advantage of the proposed block by block method is that it allows more readily for compression ratio and image quality trade-offs. By varying parameter, the compression ratios can be easily controlled.

References

1. M. Barnsley, *Fractal everywhere*, Academic Press, Inc. 1988.
2. M. J. Biggar, O. J. Morris, and A. G. Constantinides, "Segmented-image coding: performance comparison with the discrete cosine transform," Proceedings of IEE, Part F, 135 (2): 121-132, April 1988.
3. T. J. Dennis and N. G. Dessipris, "Fractal modeling in image texture analysis," Proceedings of IEE, Part F, 136(5): 227-235, Oct. 1989.
4. R. M. Haralick and L. G. Shaprio, "Image segmentation techniques," Comput. Graphics Image Processing, 29: 100-132, 1985.
5. A. E. Jacquin, "Image coding based on a fractal theory of iterated contractive image transformations," IEEE Trans. on Image Processing, Vol. 1, No. 1, Jan. 1992.
6. A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall. 1989.
7. J. W. Jang and S. A. Rajala, "Performance comparison of texture segmentation-based image coding technique with DCT image coding technique," IEEE Workshop on Visual Signal Processing and Communications, pp. 200-205, Sep. 1992.
8. N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," Proceedings of the IEEE, Vol. 81, No. 10, Oct. 1993.
9. K. N. Ngan and W. L. Chooi, "Very low bit rate video coding using 3D subband approach," IEEE Trans. on Circuits and Systems for Video Technology, Vol. 4, No. 3, June 1994.
10. S. Peleg, J. Naor, R. Hartley, and D. Avnir, "Multiple resolution texture analysis and classification," IEEE Trans. on Pattern Anal. Machine Intell., PAMI-6(4): 518-523, July 1984.
11. P. Gerken, "Object-based analysis-synthesis coding of im-

age sequences at very low bit rates," IEEE Trans. on Circuits and Systems for Video Technology, Vol. 4, No. 3, June 1994.

▲Yong Hoh Kim

Yong Hoh Kim he received the B. S. degree in Physics from Yonsei University in 1975 and the M.S. degree in Electrical and Computer Engineering from University of Wyoming in 1985. From 1985 to 1990, he worked for the Electronics and Telecommunications Research Institute, Since 1990, he has been with PaiChai University, where he currently is an Assistant Professor of the Division of Computer, Electronics and Information Communications Engineering. His research interests involve digital signal processing and digital communications.

▲Jong Whan Jang

Jong Whan Jang he received the B.S. degree in Electronic Communication Engineering from Hanyang University in 1979. He also received the M.S.E.E. and Ph.D. degrees in Electrical and Computer Engineering from North Carolina State University in 1986 and 1990 respectively. He served the director of Computer Center in PaiChai University from 1992 to 1996. Since 1990, he has been with PaiChai University, where he is currently an Assistant Professor of the Division of Computer, Electronics, and Information Communications Engineering. His research interests include image processing, multimedia communications, and multimedia database.

▲Jae Gil Jeong

Jae Gil Jeong he received the B.S. degree in Electronics Engineering from Hanyang University in 1980, and the M.S. and Ph.D. degrees in Computer Engineering from North Carolina State University in 1987 and 1991 respectively. From December 1979 to June 1985, he was with the Agency for Defence Development, Korea. From August 1991 to August 1992, he was with the Video Communication Section at Electronics and Telecommunications Research Institute, Korea. Since September 1992, he has been with PaiChai University, where he is currently an Associate Professor of the Division of Computer, Electronics, and Information Communications Engineering. His current research interests include VLSI architectures and parallel algorithms for digital signal processing, real time video signal processing.

▲Dooyeong Park

Dooyeong Park he received the B. S. degree in Electronics Engineering from Hanyang University in 1981, and He also received the M.S. and Ph.D degrees in Electrical and Computer Engineering from North Carolina State University in 1987 and 1993 respectively. Since 1994, he has been with PaiChai University, where he is currently an Assistant Professor of the Division of Computer, Electronics, and Information Communications Engineering. His main research interests include performance modelling and evaluation of computer communication networks, broadband ISDN, and queueing systems.