Automatic Decision Method of Effective Transform Coefficients for Face Recognition

Jean Choi[1,2], Yun-Su Chung[1,2], Ki-Hyun Kim[2], and Jang-Hee Yoo[2]

[1] Department of Information Security Engineering
University of Science & Technology
52 Yeoeun-dong, Yuseong-gu, Daejeon 305-333, Korea

[2] Biometrics Chipset Research Team
Electronics and Telecommunications Research Institute
161 Gajeong-dong, Yuseong-gu, Daejeon 305-350, Korea
{jchoi, yoonsu, kihyun, jhy}@etri.re.kr

Abstract. In this paper, we propose a novel face recognition method using the effective transform coefficients of face images. The method is based on extraction of effective vector in Discrete Cosine Transform (DCT) domain and Linear Discriminant Analysis (LDA) of effective vector. In general, face images have characteristics that they show larger energy congestion in horizontal frequency coefficients than in vertical or diagonal frequency coefficients. However, many previous methods have shortcomings that they don't utilize the facial energy characteristics. To overcome shortcomings above, the proposed method selects the effective coefficients of the face in DCT domain and then extracts feature vector through LDA analysis on DCT coefficients. Experimental results show that our method has improvements of recognition performance over the previous methods.

1 Introduction

In face recognition, feature extraction is one of the most important steps and it performs the reduction of high dimensional image data into low dimensional feature vectors. Dimensionality reduction is essential for extracting effective features and reducing computational complexity in classification stage. In feature extraction, there are two approaches; it is a method in spatial and frequency domain. In spatial domain, the most frequently used techniques are Principal Components Analysis (PCA) and Linear Discriminant Analysis (LDA), but the way of dimensionality reduction based on PCA or LDA is also high computational cost and needs the large number of training samples [1-3]. To overcome these weaknesses, the concerns and works about feature extraction using Discrete Cosine Transform (DCT) in frequency domain has been growing for the last several years.

Over the past few years, a several number of works have been conducted on face recognition method using DCT. Recently, DCT has been employed in face recognition for dimensionality reduction by Hafed and Levine [4]. Next, Jing and Zhang introduced feature extraction method based on selection of DCT frequency bands with favorable linear separability [5]. Recently, Er, Chen and Wu also presented a way using DCT coefficients by zigzag-scanning [6-7]. These methods assumed that energy aggregated uniformly in the low-frequency region; so to speak, it is a distinction of non-face images rather than face images. Therefore these methods have shortcomings that can't utilize energy characteristics of face images. To overcome the shortcomings above, energy characteristics of face images unlike non-face images need to be checked and utilized. As a result, we propose a new feature extraction method using the energy characteristics of face images.

The proposed method consists of three steps. First, face image is transformed into the frequency domain using DCT. Second, *Energy Probability (EP)* is calculated for the dimension reduction and optimization of valid information, and then *Energy Map (EM)* is created. Finally, in order to obtain the most significant and invariant feature of face images, LDA is applied to the data extracted using the EM. Throughout the experiments, our method has shown improvements on the dimension reduction of feature space and the face recognition over the previously proposed methods.

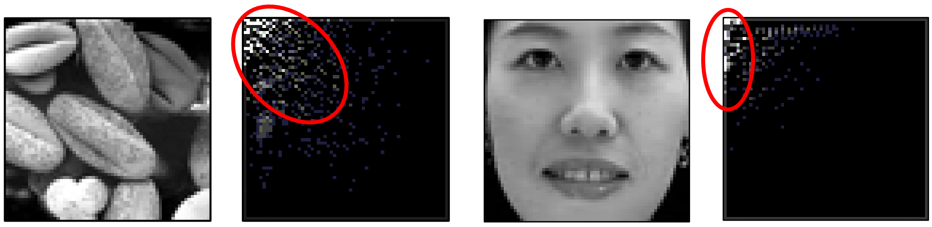
The organization of the paper is as follows. In Section 2, we define an EP as magnitude of effective coefficients and explain the creation process of the energy map using EP. and describes the face recognition procedure by our method, which is based on the EP and energy map. In addition, we interpret the background of DCT and LDA. Experimental results are presented and discussed in Section 3. Finally, Section 4 concludes this paper.

2 Face Recognition using Effective Transform Coefficients

In this section, energy probability is defined as criterion of effective information and the creation process of energy map using energy probability is explained.

2.1 Background

If we transform face images and non-face images into the frequency domain, then we can observe differences between both in the frequency distribution pattern. In order to represent data in the frequency domain, some transform such as Discrete Sine Transform, Discrete Sine Transform, and Discrete Wavelet Transform is needed[8-11]. Here, the chosen one is the DCT as an instance. The Fig.1 (a) and (c) show the general image and the face image. In Fig. 1(b) and (d), we see diagrams about coefficients used the DCT. As shown in the Fig. 1(b), the DCT coefficients with large magnitude are mainly uniformly situated in the upper-left corner of the DCT matrix. Compared with Fig. 1(b), the DCT coefficients in Fig. 1(d) are more left located in the upper-left corner of DCT matrix. Generally, horizontal frequencies increase from left to right, and vertical frequencies increase from top to bottom. To select optimal region including this distinction, we need to define a criterion of effective information. In this paper, energy probability is used as the criterion of effective information.



(a) (b) (c) (d)

Fig. 1. Example of a non-face image (a) and its DCT transformed image (b), and example of face image (c) and its DCT transformed image (d).

The energy is frequently utilized in signal processing domain such as audio and video coding and it means characteristics of images[8]. Given a result of DCT transformed from the image, *F(u, v)*, size $N \times N$, the energy using DCT coefficients is defined as follows:

$$Energy_F = \sum_{u=1}^{N} \sum_{v=1}^{N} |F(u,v)|^2 \quad (1)$$

An important point to emphasize is the fact that $Energy_f$ is only one value, concerned with one image. It means it can be each property of each image. However, it cannot be suited our purpose, dimension reduction and optimization of valid information. For the extracted features of face, we should define Energy Probability, *EP(u, v)* and it is defined by:

$$EP(u,v) = \frac{|F(u,v)|^2}{Energy_F} \quad (2)$$

The magnitude of *EP(u, v)* is used as criterion of valid coefficients. The total number of *EP(u, v)*, N^2, indicates the ratio of an energy value holding a whole face image to an energy value held a pixel of position *(u, v)*. Hence, we can say that the large value *EP(u, v)* means more valid information than face features have.

Energy Map(EM) represents the distribution location of the effective coefficients measured by the EP. Assuming that one EM is generated from one face image, it is only optimized valid location related to the one face image rather than the general face image. Thus, if we can not only create the EM from one face image, but also create the EM from many face images, then it would be created from the general face image. The choice in the former or the latter depends on a design of systems. In this article, we limit the discussion to the EM formation about one face image.

Figure 2 summarize the generation procedure of EM. We assume that *N* is the width and height of images. The face image is transformed through DCT. The *EP*, 2-D vector of size $N \times N$, is converted into the column vector *S,* length of N^2. *S* is arranged by the value of *EP* in descending order. We assume that the value of *S(k)* is *K*. If we select the *k*-th EP, then we will be remained 1 from the first to *k*-th value. Otherwise set the value of $EP(u,v)$ to change to zero by the following equation:

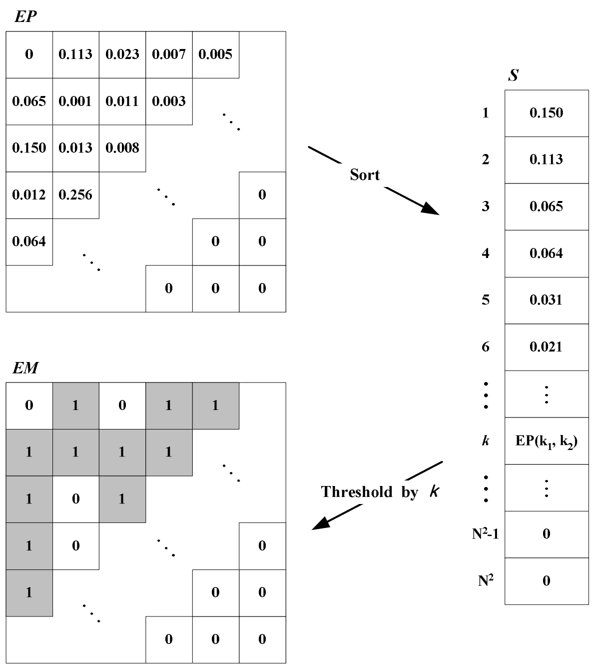
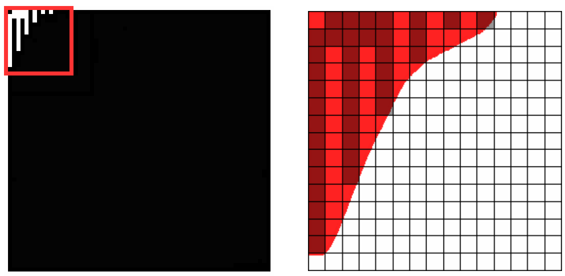


Fig. 2. The creation procedure of energy map



(a) (b)

Fig. 3. (a) is the *Energy Map* and (b) is magnification of 15×15 block.

$$EM(u,v) = \begin{cases} 1, & if\ EP(u,v) \le K \\ 0, & otherwise. \end{cases} \quad (3)$$

Figure 3 shows the *EM* and its magnification of 15×15 block. These diagrams represent that the lengthwise extended shape has a characteristic of large horizontal variation in face images. Universal face images have a distinguishing mark of these shapes.

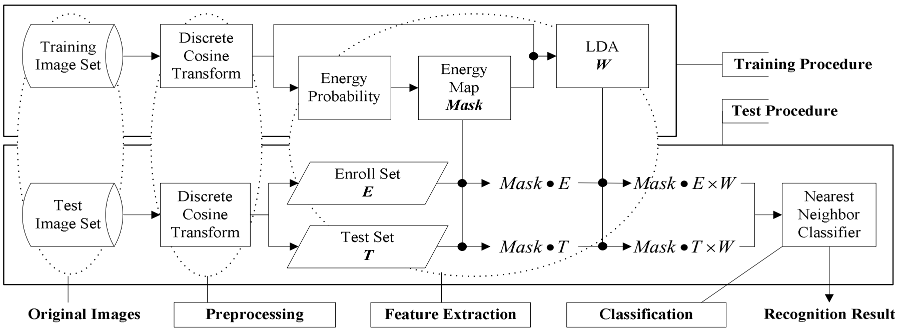


Fig. 4. Face Recognition Procedure

2.2 Face Recognition

In this paper, face recognition procedure consists of three steps; 1) preprocessing, 2) feature extraction, and 3) classification. First, original images are divided into two categories, the first is the training image set and the second is test image set. Then, each image set is transformed into the frequency domain using DCT. Second, EP is calculated for the dimension reduction and optimization of valid coefficients, and then EM is generated. Next, to obtain the most significant feature of face images, LDA is applied to the data extracted from EM. Finally, to determine the recognition rate, we perform classification by the Nearest Neighbor classifier. The Euclidean Distance is employed as a similarity criterion.

In order to explain the preprocessing, we must refer to DCT. It is frequently used solution of various problems in image and signal processing. It expresses data as the sum of cosine function for reduced size of data. Under the assumption that gray scale matrix of face image as $f(x, y)$ of size $N \times N$, its DCT as $F(u, v)$ of size $N \times N$, is calculate from:

$$F(u,v) = \alpha(u)\alpha(v)\sum_{x=1}^{N}\sum_{y=1}^{N} f(x,y) \times \cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right] \tag{4}$$

where,

$$\alpha(u),\alpha(v) = \begin{cases} \sqrt{\frac{1}{N}}, & u,v=1 \\ \sqrt{\frac{2}{N}}, & otherwise \end{cases} \tag{5}$$

with $u, v, x, y=1,2,3,\ldots,N$, where x and y are coordinates in the spatial domain, and u and v are coordinates in the frequency domain[1].

The first element, $F(1, 1)$, is the average of all the samples in the input image and is called Direct Current (DC) component. The remaining elements in $F(u, v)$ indicate the amplitude corresponding to the frequency component of $f(x, y)$, and are define as Alternate Current (AC) coefficients. It is well-known that the DC coefficient only depends on the brightness of the image. Consequently, it becomes DC-*free* (i.e., zero mean) and invariant against uniform brightness change by simply removing the DC coefficient [3]. The proposed method sets zero-valued DC coefficient and the remaining AC coefficients for eliminating illumination characteristic of the image.

Figure 4 appears the entire procedure of face recognition. First, face images are transformed into DCT domain through DCT. Second, DCT domain acquired from face image is applied on EP for the purpose of dimension reduction of data and optimization of valid coefficients. The EP is defined as criterion of effective facial features in section 2.2. Third, in order to obtain the most invariant feature of face images, the LDA is applied in the data extracted from the EM. It can facilitate the selection of effective coefficients for image recognition, because all the pixels are not useful in classification. At last, linear discriminative features are extracted by LDA and classification is performed by the nearest neighbor classifier. We divide original images into two parts of the training and test image set. The training image set is applied to DCT and it is computed EP. Then, we generate the EM. When the number of valid coefficients of EM is n and the number of training images is M, it creates *Mask* vector of size $n \times M$.

In feature extraction, The LDA is one of the most popular linear projection methods for feature extraction. It is used to find a linear projection of the original vectors from a high dimensional space to an optimal low dimensional subspace in which the ratio of the between-class scatter and the within-class scatter is maximized. The W is the LDA optimal projection matrix and calculated by *Mask* vector. In the same way, the test image set is applied to DCT and divide test image set into two parts of the enroll set E and test set T. $Mask \bullet E$ denotes E applied on the frequency mask. The extracted final feature vector is $Mask \bullet E \times W$ and $Mask \bullet T \times W$ and is calculated the Euclidian distance between them.

3 Experimental Results

In order to evaluate performance of proposed face recognition algorithm empirically, two kinds of face database, our lab database and ORL database were used. 1) The first database, our lab database contains 1100 images of 55 individuals. Fig. 5 shows sample faces from this database. Twenty images were taken for each individual. This database includes both males and females and the size of each image is 64×64 with 256 gray levels. Few constraints on facial expression and pose were imposed. Furthermore, some of the captured images were subject to illumination variations. 2) The second database, the ORL database (http:// www.cam-orl.co.uk), contains images from 40 individuals, each providing 10 different images. For some subjects, the images

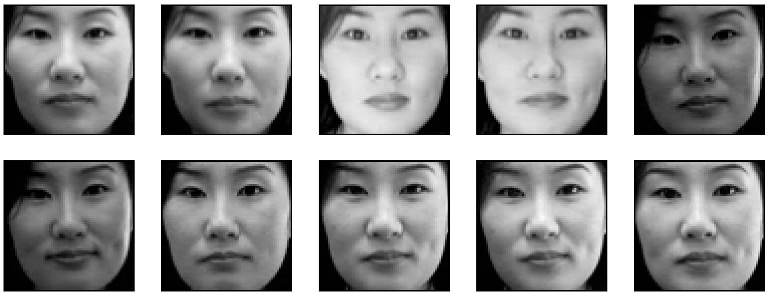


Fig. 5. Example faces in the first database

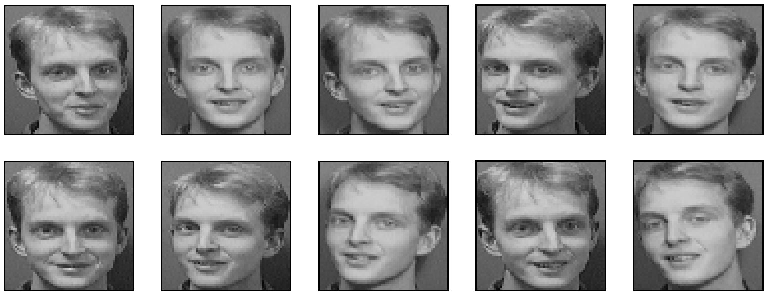


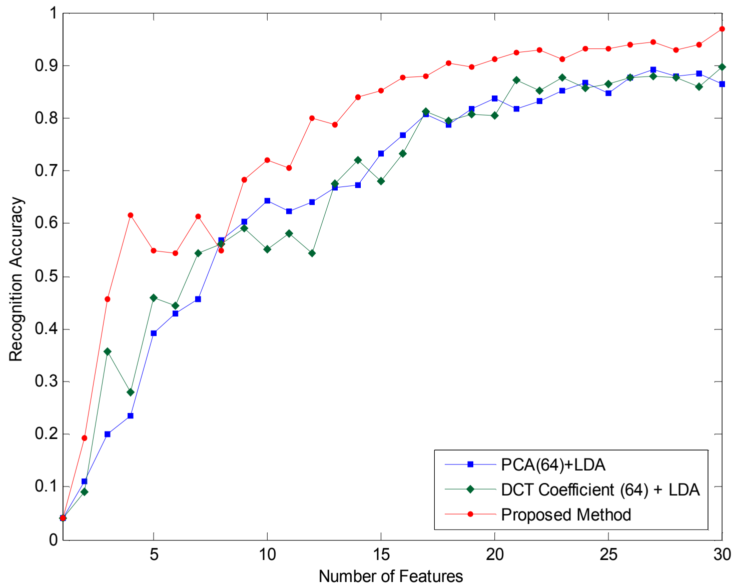
Fig. 6. Example faces in the second database (ORL database)

were taken at different times. The facial expressions (open or closed eyes, smiling or nonsmiling) and facial details (glasses or no glasses) also vary. The images were taken with a tolerance for some tilting and rotation of the face of up to 20 degrees. Moreover, there is also some variation in the scale of up to about 10 percent. All images are grayscale and normalized to a resolution of 92×112 pixels and each image is scaled to 64×64. Ten sample images of one person from the ORL database are shown in Fig. 6.

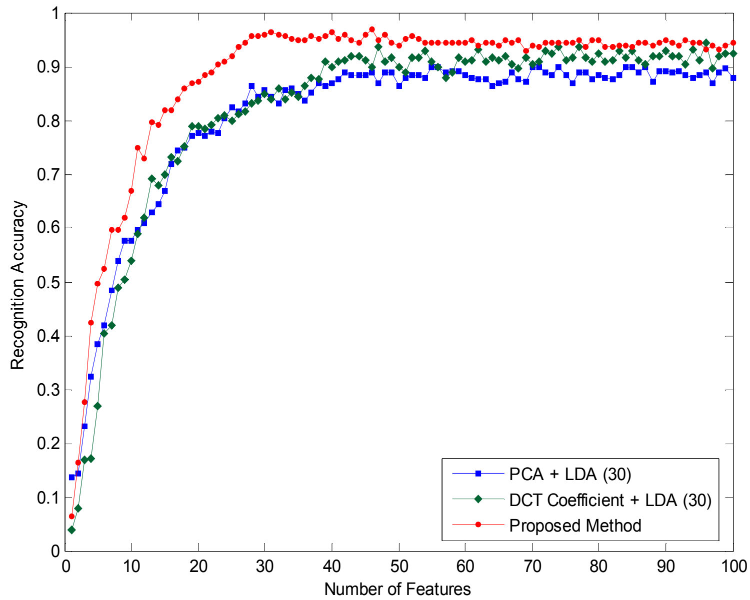
In the experiments, ten training images and ten test images per person are selected from the first database, and a training set of 550 images and a test set of 550 images are decided for the following experiments. In the second database, it is choose as five training images and five test images per person, and each set has 200 images. That is to say, the training images and test images is not overlapped between the two sets in both cases. Finally, the nearest neighbor classifier is used for classification.

The proposed method is compared with the results from the following conventional feature extraction methods: PCA + LDA and DCT coefficients + LDA. Fig. 7 and Fig. 8 show that is altered the number of reduced effective coefficient and fixed the number of extracted valid vector by LDA. The number of extracted feature used in LDA is 30 and 20 in the first and second database.

Figure 7 and 8 appear recognition rate with the increments of the number of the dimension reduction data and the final features applied to the LDA with fixed number. For the rank 1, the best recognition rates of PCA + LDA, DCT coefficients + LDA, and proposed method are 90.0%, 94.4%, and 96.8% in the first database (Table1) and



(a)



(b)

Fig. 7. (a) is recognition accuracy with the number of reduced data, 64 and the increments of the number of extracted features (b) is recognition accuracy with the increments of the number of reduce data and number of extracted features, 30.

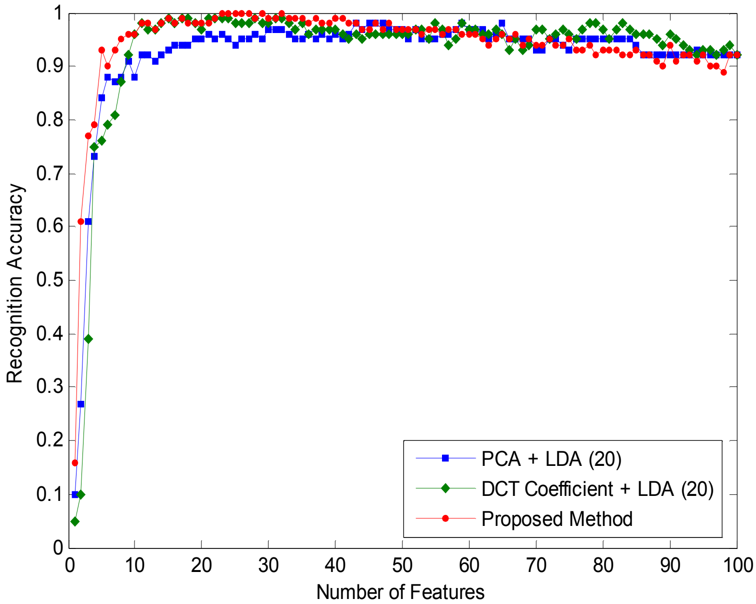


Fig. 8. The recognition accuracy with the increments of the number of reduce data and number of extracted features, 20

Table 1. Comparison of classification performance usign the ETRI database

Methods	PCA + LDA	DCT Coefficient + LDA	Proposed method
Number of Reduced Data	74	96	46
Number of Extracted Feature	29	30	30
Recognition Accuracy (%)	90.0	94.4	96.8

Table 2. Comparison of classification performance usign the ORL database

Methods	PCA + LDA	DCT Coefficient + LDA	Proposed method
Number of Reduced Data	46	23	23
Number of Extracted Feature	20	20	19
Recognition Accuracy (%)	98.5	99.0	100.0

98.5%, 99.0%, and 100% in the second database (Table 2), respectively. Accuracy is computed by the number of reduced data in the second row and the number of extracted features in the third row in Table 1 and 2. The recognition accuracies in the forth row are the minimum number of reduced data with the best performance. By fixing the shortcomings of conventional methods, proposed methods showed the best performance among the tested methods. It can be also seen that the size of data dimension was reduced effectively.

4 Conclusions

In paper, we presented a new energy probability based feature extraction method. Our method consists of three steps. First, face images are transformed into DCT domain. Second, DCT coefficients acquired from face image is employed to energy probability for the purpose of dimension reduction of data and optimization of valid information. Third, in order to obtain the most silent and invariant feature of face images, the LDA is applied to the data set extracted from the frequency mask that can facilitate the selection of useful DCT frequency bands for image recognition, because all the bands are not useful in classification. Finally, it will extract the linear discriminative features by LDA and perform the classification by the nearest neighbor classifier. For the purpose of dimension reduction of data and optimization of valid information, the proposed method has shown better recognition performance than PCA plus LDA and existing DCT method.

References

1. A. Martinez and A. Kak: PCA versus LDA. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 23, no. 2. (2001) 228–233
2. H. Yu and J. Yang: A direct LDA algorithm for high-dimensional data with application to face recognition. Pattern Recognit., Vol. 34, no. 12. (2001) 2067–2070
3. J. Lee: Automatic Video Management System Using Face Recognition and MPEG-7 Visual Descriptors. ETRI Journal, Vol.27, no.6, (2005) 806-809
4. Z.M. Hafed and M.D. Levine: Face Recognition Using the Discrete Cosine Transform. International Journal of Computer Vision, Vol. 43, no.3, (2001) 167–188
5. X. Jing and D. Zhang: A Face and Palmprint Recognition Approach Based on Discriminant DCT Feature Extraction. IEEE Transactions on Systems, Man, and Cybernetics-part b: Cybernetics, Vol. 34, no. 6 (2004)
6. M. J. Er, W. Chen, and S. Wu: High-Speed Face Recognition Based on Discrete Cosine Transform and RBF Neural Networks. IEEE Transactions on Neural Networks, vol. 16, no. 3 (2005)
7. M. J. Er, and S. Wu: PCA and LDA in DCT domain. Pattern Recognition Letters, Vol. 26. (2005) 2474-2482
8. G.H. Park and K.H. Kim: Adaptive Scanning Methods for Fine Granularity Scalable Video Coding. ETRI Journal, Vol. 26, no.4, (2004) 332-343
9 S. Koç, E. Erçelebi : Image Restoration by Lifting-Based Wavelet Domain E-Median Filter. ETRI Journal, Vol.28, no.1, (2006) 51-58
10. W.H. Kim, S.Y. Jeong and K.H. Kim: Scalable Interframe Wavelet Coding with Low Complex Spatial Wavelet Transform. ETRI Journal, Vol.28, no.2, (2006) 145-154
11. W. K. Pratt: Digital Image Processing. John Wiley & Sons, Inc., New York (2001) 173-181
12. Gordon E. Carlson: Signal and Linear System Analysis. John Wiley & Sons, Inc., New York (1998) 152-163
13. R. C. Gonzalez and R. E. Woods: Digital Image Processing. Addison Wesley, Massachusetts (1992) 472-47713.