Implementing Discrete Wavelet and Discrete Cosine Transform with Radial Basis Function Neural Network in Facial Image Recognition

Samuel Lukas[1], Aditya Rama Mitra[2], Ririn Ikana Desanti[3], and Dion Krisnadi[1]
[1]Informatics Department, Universitas Pelita Harapan, Karawaci, Indonesia
[2]Computer System Department, Universitas Pelita Harapan, Karawaci, Indonesia
[3]Information System Department, Universitas Pelita Harapan, Karawaci, Indonesia
Email: {Samuel.lukas, Aditya.mitra, Ririn.desanti, Dion.krisnadi}@uph.edu

***Abstract*—Face image recognition has been widely used and implemented in many aspects of life, such as in the field of investigation or security. However, research in this area is still rarely done. Source images in this paper are taken directly from 41 students with a total of 131 faces in JPG format, each with a dimension of 256×256. By applying Discrete Wavelet Transform and Discrete Cosine Transform, an image can be represented as a number of DCT coefficients efficiently. Recognition process is done using Radial Basis Function Neural network. The experiment results show that the best configuration for RBF is 8×41×41 with recognition rate of student faces is 100% and 98% of the sample face images are identified perfectly.**

***Index Terms*—discrete wavelet transform, discrete cosine transform, radial basis function neural network**

I. Introduction

Feature-based approach and brightness-based approach can be applied in face recognition system. The feature-based approach uses key point features of the face, such as the edges, eyes, nose, mouth, or other special characteristics. Therefore, the calculation process only covers some parts of the image that have been extracted in prior. On the other hand, the brightness-based approach calculates all parts of the image. It is also known as holistic-based or image-based approach.

Since all parts of the image have to be considered, the brightness-based approach needs a longer time to process and is also complicated. To make it short and simple, the image has to be transformed to a certain model. Many models have been proposed. One of that model was introduced by Turk and Pentland, 1991, using the Principle Component Analysis [1]-[3]. Other proposed model applied such as Discrete Wavelet Transform (DWT) [4] and Discrete Cosine Transform [5]-[8]. In the present paper, combined model of DWT and DCT is applied.

After extracting features of the image by using DWT and DCT, the system goes to a recognition phase. There are also a lot of recognition models that can be used, such as back propagation neural network [4] and Hidden Markov Models [5], [6]. This paper discusses on how face recognition can be done using Radial Basis Function Network, RBFN [8], [9]

Manuscript received August 15, 2015; revised November 2, 2015.

II. Feature Extraction and Recognition System

A. Discrete Wavelet Transform

Wavelet transform decomposes a signal into a set of basis functions. These basis functions are called wavelets. Wavelets are obtained from a single prototype wavelet called mother wavelet by dilation and shifting [10]. The DWT has been introduced as a highly efficient and flexible method for sub band decomposition of a signal. The 2D-DWT is nowadays established as a key operation in image processing. It is known as a multi-resolution analysis, and it decomposes images into wavelet coefficients and scaling function.

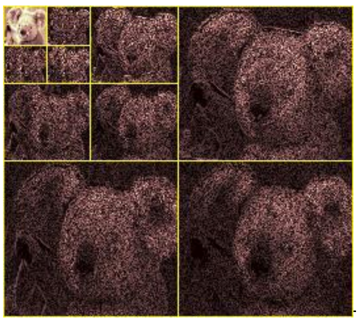

Figure 1. Three level decomposition for 2D-DWT

Wavelet transform converts the image into a series of wavelets that can be stored more efficiently than pixel blocks. Wavelets have rough edges, they are able to render pictures better by eliminating the blocks. The process is done through filters with different cut-off frequencies at different scales. It is easy to implement, and reduces the computation time and resources required [11]. A 2-D DWT can be seen as a 1-D wavelet scheme which transforms along the rows and then a 1-D wavelet transform along the columns. The 2-D DWT operates in a straight forward manner by inserting array transposition between the two 1-D DWT. The rows of the array are processed first with only one level of decomposition. This

doi: 10.18178/joig.4.1.6-10

essentially divides the array into two vertical halves, with the first half storing the average coefficients, while the second half stores the detail coefficients. This process is repeated again with the columns, resulting in four sub-bands for each decomposition level (see Fig. 1) within the array defined by filter output. Fig. 1 shows a three-level 2-D DWT decomposition of an image [12].

B. Discrete Cosine Transforms

Discrete Cosine Transform (DCT) is a transform coding mostly used in signal processing or digital image processing. It is derived from the Discrete Fourier Transform (DFT). The objects of this paper are in the form of images. Therefore 2D DCT are implemented [7], [8]. Spatial domain of an image $I(x,y)$ is transformed to frequencies domain as $C(u,v)$ which is stated as (1) & (2), and the inverse as (3).

$$C(u,v) = \alpha(u)\alpha(v)\sum_{r=0}^{row}\sum_{c=0}^{col} I(r,c)\cos\left[\frac{(2r+1)u\pi}{2N}\right]\cos\left[\frac{(2c+1)v\pi}{2N}\right] \quad (1)$$

$$\alpha(u),\alpha(v) = \begin{cases} \sqrt{\frac{1}{N}} & for\ u,v = 0, \\ \sqrt{\frac{2}{N}} & for\ u,v = 1,2,\ldots,N-1 \end{cases} \quad (2)$$

$$I(r,c) = \sum_{u=0}^{row}\sum_{v=0}^{col} \alpha(u)\alpha(v)C(u,v)\cos\left[\frac{(2r+1)u\pi}{2N}\right]\cos\left[\frac{(2c+1)v\pi}{2N}\right] \quad (3)$$

DCT has important properties. They are de-correlation, energy compaction, domain scaling, separability, and symmetry. De-correlation means that there is no correlation in calculating among all the DCT coefficients. Therefore, all DCT coefficients can be calculated independently. DCT exhibits excellent energy compaction for highly correlated images. Efficacy of a transformation scheme can be directly gauged by its ability to pack input data into as few coefficients as possible without introducing visual distortion in the reconstructed image significantly. To be noted that DCT is not scaling invariant. This implies that in an image recognition system, all of the images used either for training or identification have to be uniform in size. Separability means that the DCT coefficients can be computed in two steps by successive 1-D operation on rows and columns of an image. It is stated in (4). Another look at the row and column operations in Equation (4) reveals that these operations are functionally identical. Such a transformation is called a *symmetric transformation* [9].

$$C(u,v) = \alpha(u)\alpha(v)\sum_{r=0}^{row}\cos\left[\frac{(2r+1)u\pi}{2N}\right]\sum_{c=0}^{col} I(r,c)\cos\left[\frac{(2c+1)v\pi}{2N}\right] \quad (4)$$

C. Radial Basis Function Neural Network

The idea of Radial Basis Function (RBF) networks is derived from Multi-Layer Perceptron (MLP) networks, but RBFNs take a slightly different approach. They have five main features. They are two-layer feed-forward networks. The hidden nodes implement a set of radial basis functions (e.g. Gaussian functions). The output nodes implement linear summation functions as in an MLP. The network training is divided into two stages: first the weights from the input to hidden layer are determined, and then the weights from the hidden to output layer. The training/learning process is very fast. Configuration of RBFN for P input nodes with Q hidden nodes and R output nodes can be seen in [10]. Fig. 2 is an example scheme of RBFN.

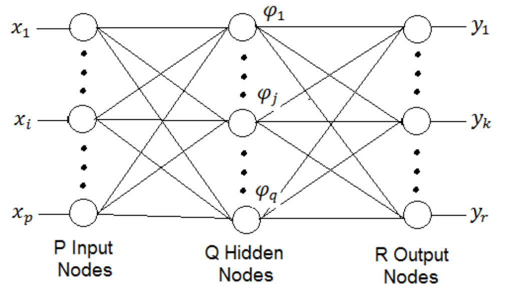

Figure 2. Structure of RBFN

The goal of RBF is to find a function $f: x^p \to y^r$ so that it can interpolate a set of N data points in a p-dimensional input space, $\boldsymbol{x} = (x_i\ x_2 \ldots x_p)$, to be mapped onto the r-dimensional output space, $\boldsymbol{y} = (y_i\ y_2 \ldots y_r)$. Every hidden node has a center vector, $\boldsymbol{x_c} = (x_{c_1}\ x_{c_2} \ldots x_{c_p})$ and a variance, σ_c^2. The output of every hidden node is stated in (5), then by doing linear combination with the weights, W_{kj}, from hidden nodes to the output nodes, the output of the RBF is in (6). Remove the first page footnote if you don't have any information there.

$$\varphi(\|\boldsymbol{x} - \boldsymbol{x_c}\|),\ \varphi(a) = exp\left(-\frac{a^2}{2\sigma^2}\right) \quad (5)$$

$$\boldsymbol{y_k} = f(\boldsymbol{x}) = \sum_{j=1}^{q} W_{kj}\varphi_j\left(\left\|\boldsymbol{x} - \boldsymbol{x_{c_j}}\right\|\right) \quad (6)$$

III. System Design

Block diagram of the system is presented in Fig. 3. They consist of training process and recognizing process. The outputs of the training process are the average and the standard deviation of each hidden node, and also the weights of RBFN from hidden nodes to the output nodes. On the other hand, for recognizing process, the output is the name of the person, while the input is a face image.

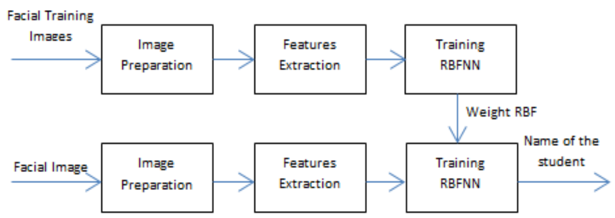

Figure 3. Block diagram of the system

Face images for training process consist of 131 images from 41 subjects. Each subject consists of at least three images. Four of them are shown in Fig. 4.

Face detection of each input image is applied and then the face image is normalized to the size of 256×256 pixels, and then it is transformed into gray scale image. Fig. 5 shows forty two of the normalized sample trained images.

Figure 4. Four sample input images

Figure 5. Forty two sample images to be trained

Feature extraction is done by implementing DWT and DCT. The output of the process is m × m DCT's coefficients. Some of these DCTs are used in training process by RBFN with p input nodes, 41 hidden nodes and 41 output nodes.

Suppose p-DCT coefficients are used as the features of an image, and there are n images in training set with k classes labeled as $c_i, i = 1,2,\dots,k$, therefore the input data set can be stated as a $n \times k$ matrix stated as $X = \{x_{ij} \,|i = 1,2,\dots,n \;;\; j = 1,2,\dots,k\}$. The training process outputs are μ_i, σ_i^2, $i = 1,2,\dots,k$ and also the weights of RBF matrix from hidden layer to output layer, labeled as W with size $k \times k$. μ_i and σ_i^2 are parameters of i^{th} node in the hidden layer, and they are vector centre in (7) and variance in (8) of training data of i^{th} class out of k classes respectively. For each data in the training set, $\boldsymbol{x}_i, i = 1,2,\dots,n$, the output of q^{th} node in the hidden layer stated in (9) and the target output of node q is in (10). It is clear that $\{h_q\}$ can be formed as matrix H with size $n\times k$, and $\{t_q\}$ in the form of T with size $n \times k$. RBF weights from hidden layer to output layer are $W = \{w_{ij}\}$ in (11).

$$\mu_{ij} = \frac{1}{|C_i|}\sum_{i\epsilon C_i} x_{ij} \qquad i = \{1,\dots,k\} \;;\; j = 1,2,\dots,p \quad (7)$$

$$\sigma_i^2 = \frac{1}{p|C_i|}\sum_{i\in C_i}\sum_{j=1}^{p}(x_{i,j} - \mu_{ij})^2 \quad i = 1,2,\dots,k \quad (8)$$

$$h_q = \varphi_q(x_{ij}) = e^{\left(-\frac{\sum_{j=1,i\epsilon C_i}^{p}\left(x_{ij}-\mu_{qj}\right)^2}{2\sigma_q^2}\right)} \quad q = 1,2,\dots,k \quad (9)$$

$$t_q(x_{ij}) = \begin{cases} 1 & if\ q = c_i \\ 0 & others \end{cases} \quad q = 1,2,\dots,k \quad (10)$$

$$W = (H^T H)^{-1} H^T T \quad (11)$$

In recognizing process, a single face image is extracted and then presented by $\boldsymbol{x} = (x_1\ x_2\ \dots\ x_p)$. Furthermore, the output of hidden layer is presented in the form $\boldsymbol{h} = (h_1\ h_2\ \dots\ h_k)$ in (12) and the RBFN output is $\boldsymbol{o} = (o_1\ o_2\ \dots\ o_k)$ in (13).

$$h_i = \varphi_i(\boldsymbol{x}) = e^{\left(\frac{\|x-\mu_i\|^2}{2\sigma_i^2}\right)} \qquad i = \{1,\dots,k\} \quad (12)$$

$$\boldsymbol{o} = \boldsymbol{h}W \quad (13)$$

It is clear that the index j with the highest value for component o_j indicates that the input face image belongs to the class with the same index number in the system.

IV. Experiment Result and Discussion

The first experiment is done to determine the minimum number of DCTs coefficient needed to achieve good recognition percentage by using three levels of DWT. The second experiment repeats the first one, but by using two levels of DWT. The results of all experiments are tabulated at Table I. No represents the subject of the sample and column total for the number of image of that subject respectively. Whereas other numbers of each column indicate number of the image is able to be recognized correctly.

It can be concluded that the person number 13 cannot be recognized with two or three levels of DWT with 2 DCTs coefficients. Therefore, 2 DCT coefficients as a feature of a face image cannot be applied.

Processing with three level of DCT gives the same recognition result with that of two level of DCT with number of DCT coeff is 64. Using 32 DCTs shows a bit better result with three level of DWT than that of two. That is person number 32. However it is not for 16 DCTs. That is person number 22. Recognition process is the best for using 8 DCTs coefficient with three level DWT compared with two level of DWT.

Figure 6. Person number from left to right: 8, 3, 11, 29, 20, 22, 23, 30, 19, 13

Further analysis shows that, the higher the level of DWT and the smaller the number of DCTs, the faster the processes will be. By using two levels of DWT and 16 DCT coefficients, out of three samples from person #8, only one sample fails to be recognized and is recognized as person #3 as shown in Fig. 6. By using 8 DCTs, one sample from person #11 is recognized as #29; one sample from person #20, #22, #23 and #30 as #19; and one sample person from person #29 as #13. However, by using three levels of DWT and 8 DCTs, only two images, #22 and #23 are recognized as #19.

Therefore, for the best construction, we chose three levels of DWT with 8 DCT coefficients.

TABLE I. NUMBER OF IMAGE IS RECOGNIZED CORRECTLY

		DWT 3 level with # of DCTs						DWT 2 level with # of DCTs					
No	Total	64	32	16	8	4	2	64	32	16	8	4	2
1	3	3	3	3	3	3	2	3	3	3	3	3	2
2	4	4	4	4	4	4	3	4	4	4	4	4	4
3	3	3	3	3	3	2	3	3	3	3	3	2	3
4	3	3	3	3	3	3	3	3	3	3	3	3	3
5	3	3	3	3	3	3	3	3	3	3	3	3	3
6	3	3	3	3	3	3	2	3	3	3	3	3	1
7	3	3	3	3	3	3	3	3	3	3	3	3	2
8	3	3	3	2	3	2	2	3	3	2	2	1	1
9	3	3	3	3	3	3	3	3	3	3	3	3	3
10	3	3	3	3	3	3	3	3	3	3	3	3	3
11	3	3	3	3	3	3	1	3	3	3	2	3	1
12	3	3	3	3	3	3	1	3	3	3	3	1	3
13	3	3	3	3	3	3	0	3	3	3	3	3	0
14	4	4	4	4	4	4	3	4	4	4	4	4	4
15	3	3	3	3	3	2	1	3	3	3	3	2	2
16	3	3	3	3	3	3	2	3	3	3	3	3	3
17	3	3	3	3	3	3	3	3	3	3	3	3	3
18	3	3	3	3	3	3	3	3	3	3	3	3	3
19	4	4	4	4	4	3	2	4	4	4	4	3	2
20	3	3	3	3	3	3	3	3	3	3	2	3	3
21	3	3	3	3	3	3	2	3	3	3	3	3	3
22	3	3	3	2	2	3	2	3	3	3	2	3	0
23	3	3	3	3	2	3	1	3	3	3	2	2	1
24	4	4	4	4	4	4	2	4	4	4	4	4	4
25	3	3	3	3	3	3	2	3	3	3	3	3	2
26	3	3	3	3	3	3	3	3	3	3	3	3	3
27	4	4	4	4	4	2	2	4	4	4	4	4	3
28	2	2	2	2	2	2	2	2	2	2	2	2	2
29	3	3	3	3	3	3	3	3	3	3	2	2	3
30	3	3	3	3	3	2	2	3	3	3	2	3	3
31	3	3	3	3	3	2	1	3	3	3	3	2	0
32	4	4	4	4	4	3	4	4	3	4	4	4	3
33	3	3	3	3	3	3	3	3	3	3	3	3	2
34	3	3	3	3	3	3	3	3	3	3	3	3	3
35	4	4	4	4	4	4	4	4	4	4	4	4	2
36	3	3	3	3	3	3	3	3	3	3	3	3	3
37	3	3	3	3	3	3	1	3	3	3	3	3	2
38	3	3	3	3	3	3	2	3	3	3	3	3	3
39	4	4	4	4	4	3	3	4	4	4	4	3	3
40	3	3	3	3	3	2	1	3	3	3	3	3	3
41	4	4	4	4	4	4	4	4	4	4	4	3	4

ACKNOWLEDGMENT

This research is funded by Indonesian Ministry of Education and Culture, Directorate General of Higher Education, Kopertis Wilayah III, No. 108/K3/KM/2015 and Faculty of Computer Science, Informatics Department, Universitas Pelita Harapan, No. 026/LPPM-UPH/III/2015.

REFERENCES

[1] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.

[2] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks: Theory and Applications*, John Wiley & Sons, Inc., 1996.

[3] A. Pentland, B. Moghaddam, and T. Starner, "View-Based and modular eigenspaces for face recognition," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1994.

[4] M. Alwakeel and Z. Shaaban, "Face recognition based on Haar wavelet transform and principal component analysis via Levenberg-Marquardt back propagation neural network," *European Journal of Scientific Research*, vol. 42, no. 1, pp. 25-31, 2010.

[5] V. V. Kohir and U. B. Desai, *Face Recognition Using a DCT-HMM Approach*, Mumbai, India: Indian Institute of Technology, 1998.

[6] S. E. Handy, S. Lukas, and H. Margaretha, "Further tests for face recognition using discrete cosine transform and hidden Markov model," in *Proc. International Conference on Electrical Engineering and Informatics*, Makasar, 2012.

[7] S. A. Khayam, *The Discrete Cosine Transform (DCT): Theory and Application*, Michigan State University, 2003.

[8] M. H. Muchri, S. Lukas, and D. H. Hareva, "Implementation discrete cosine transform and radial basis function neural network in facial image recognition," in *Proc. International Conference on Soft Computing Intelligent Systems, and Information Technology*, 2015.

[9] B. J. Oh, "Face recognition using radial basis function network based on LDA," *International Journal of Computer, Information Science and Engineering*, vol. 1, pp. 401-405, 2007.

[10] A. A. Blessie, J. Nalini, and S. C. Ramesh, "Image compression using wavelet transform based on the lifting scheme and its implementation," *International Journal of Computer Science Issues*, vol. 8, no. 1, 2011.

[11] Y. S. Bute and R. W. Jasutkar, "Implementation of discrete wavelet transform processor for image compression," *International Journal of Computer Science and Network*, vol. 1, no. 3, 2012.

[12] D. Gupta and S. Choubey, "Discrete wavelet transform for image processing," *International Journal of Emerging Technology and Advanced Engineering*, vol. 4, no. 3, March 2015.

Samuel Lukas received his Bachelor degree in Electrical Engineering from Trisakti University, Indonesia, Jakarta, in 1985. He got his master degree in Computer Science from University of Tasmania, Tasmania, Australia, in 1992. He completed his doctorate degree in Science Education in 2005 from Curtin University, Curtin, Australia.
He is a senior lecturer in Computer Science faculty for Universitas Pelita Harapan, Karawaci, Indonesia. He wrote many papers and two books chapter. The latest book chapter was titled "*Real-World Application of Genetic Algorithms, Chapter 15*". He is interested in artificial intelligence with research area in computer vision. Dr. Lukas is a member of Indonesian Higher Education Association for Information and Computer Studies.

Aditya Rama Mitra received his undergraduate degree in Mathematics from Institut Tekologi Bandung (ITB), Bandung, Indonesia, in 1994. He completed his master in Informatics at the same university in 1999.
He works as a full-time lecturer at the Computer System department of the Faculty of Computer Science, Universitas Pelita Harapan (UPH), Tangerang, Indonesia. His interests are particularly shown in the area of multimedia device configuration, digital signage audience measurement, as well as music and data visualization. Aditya Rama Mitra is also a member of Indonesian Higher Education Association for Information and Computer Studies.

Ririn Ikana Desanty received her Bachelor degree in Information System in 2002 from Universitas Pelita Harapan, Indonesia. She got her master degree in Information Technology in 2006 from University of Indonesia, Indonesia.
Since 2007, she worked for Universitas Pelita Harapan as a researcher in information system studies. Her research interests are Information System, Decision Support System, and Database System. Ririn Ikana Desanty is a member of APTIKOM (Indonesian Higher Education Association for Information and Computer Studies).

Dion Krisnadi received his Computer Science degree and Mathematics degree in 2012 from Universitas Pelita Harapan, Indonesia. He got his master degree in Information Technology in 2014 from University of Indonesia, Indonesia.
Since 2014, he worked for Universitas Pelita Harapan as a researcher in computer science studies. His research interest is artificial intelligence. Mr. Krisnadi is a member of APTIKOM (Indonesian Higher Education Association for Information and Computer Studies.