Bulletin of Electrical Engineering and Informatics
Vol. 9, No. 3, June 2020, pp. 996~1004
ISSN: 2302-9285, DOI: 10.11591/eei.v9i3.1977

Hand gesture recognition using discrete wavelet transform and convolutional neural network

Muhammad Biyan Priatama[1], Ledya Novamizanti[2], Suci Aulia[3], Erizka Banuwati Candrasari[4]
[1,2,4]School of Electrical Engineering, Telkom University, Indonesia
[3]School of Applied Science, Telkom University, Indonesia

Article Info

Article history:

Received Sep 1, 2019
Revised Nov 13, 2019
Accepted Jan 21, 2020

Keywords:

Convolutional neural networks
Discrete wavelet transform
Hand gesture recognition

ABSTRACT

Public services are available to all communities including people with disabilities. One obstacle that impedes persons with disabilities from participating in various community activities and enjoying the various public services available to the community is information and communication barriers. One way to communicate with people with disabilities is with hand gestures. Therefore, the hand gesture technology is needed, in order to facilitate the public to interact with the disability. This study proposes a reliable hand gesture recognition system using the convolutional neural network method. The first step, carried out pre-processing, to separate the foreground and background. Then the foreground is transformed using the discrete wavelet transform (DWT) to take the most significant subband. The last step is image classification with convolutional neural network. The amount of training and test data used are 400 and 100 images repectively, containing five classes namely class A, B, C, # 5, and pointing. This study engendered a hand gesture recognition system that had an accuracy of 100% for dataset A and 90% for dataset B.

Corresponding Author:

Ledya Novamizanti,
School of Electrical Engineering,
Telkom University,
Jl. Telekomunikasi, Terusan Buah Batu, Bandung 40257, Indonesia.
Email: ledyaldn@telkomuniversity.ac.id

1. INTRODUCTION

Nowadays the technology growth leads the development of human physical recognition, one of which is the hand recognition for nonverbal communication. Signaling is a form of nonverbal communication frequently used in sign language and symbols [1]. Gestures have a major role in braiding the communication because it unconsciously-used as supplemental hint to the information that cannot be said verbally [2]. Past research was performed by taking two-dimensional hand gesture contours and converted it into one-dimensional signals using reference values. Wavelet decomposition was implemented for one-dimensional signals that were converted from two-dimensional contour images. The four statistical properties of wavelet coefficients then extracted and artificial neural networks (ANNs) were adopted to classify the gestures. It obtained the accuracy of 97% with 240 images collected from 20 people [3]. The issue with this research spotted on the system failure at recognizing the gestures with almost look-alike contours and that resulted the classifier to false-predict the output. More, the fuzzy images made it more difficult for classifier to predict the movements performed, so camera type and light intensity are very important in this research. Using both hands for experiment is also reduces accuracy.

The objective of this research is to design a hand gesture recognition system with discrete wavelet transform (DWT) method and convolutional neural network (CNN) classifier that has good performance. The benefit of DWT is that it can provide time and frequency information simultaneously and wavelets can be adjusted and adapted easily [4, 5]. CNN has the advantages that DNNs have, as it can have more than one hidden layer between the input and output [6-8]. In addition, allowing two-dimensional arrays or the dimensions in the input layer and the weight can be shared [9]. In CNN concept have feature extraction before classify, it can additionally feature to classify which has a similar gesture [10]. The problems including how to make the design and simulation of hand gesture recognition system to the datasets, how the accuracy obtained using these method and classifier, and how the input influencing the parameters on systemperformance.

Data acquisition performed by taking the sebastien marcel static hand posture database dataset as Dataset A [11] and Erizka's dataset as Dataset B [12]. The structure of this paper consists of four sections. Section 1 describes introduction, section 2 describes research method, basic formulation of DWT, CNN, and system performance parameters, section 3 describes the performance of the hand gesture system, section 4 describes conclusion.

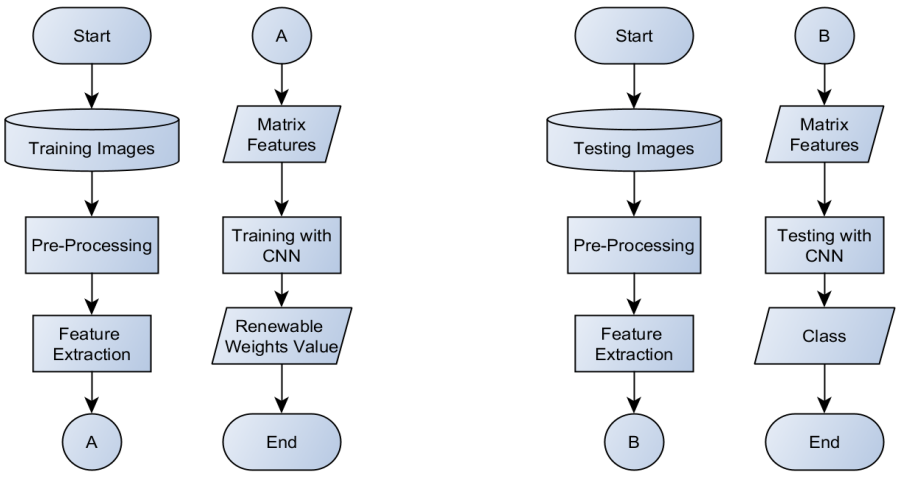
2. RESEARCH METHOD

A system has been designed to recognize hand gestures. In general, the purposed system is illustrated in Figure 1. The general scheme illustrated in Figure 1 would be the basis of this study. A description of the process in Figure 1(a) of the systemdesign as follows:

- Input was a training image database from a dataset that had red green blue (RGB) layer.
- Pre-processing, which was the process of resizing the image and segment the skin color of the input image. Image resizing with size of 76×66 pixels for dataset A and 128×128 pixels for dataset B. The segmentation skin color by determining the pixel value threshold of YCbCr (Cr), filling noise, and closing. Then result from pre-processing would generated a hand contour image.
- Feature extraction with DWT.
- Parameter training for forward and backward of training images in each class using CNN with input in the form of feature vector from training images. Result from this process is renewable weights value.

In addition to the training process, there was also a testing process. A description for the process in Figure 1(b) of the systemdesign was as follows:

- Input was a training image database from a dataset that had RGB layer.
- Pre-processing, which was the process of resizing the image and segment of skin color of the input image which would generated a contour image.
- Feature extraction with DWT.
- Class prediction with CNN. The process that happened was a calculation for forward parameters and predict the class with the highest probability.



(a) (b)

Figure 1. System flowchart, (a) Training, (b) Testing

2.1. Discrete wavelet transform

Wavelet is a wave oscillation with zero initial amplitude, which can increase or decrease, thus form a fluctuating wave [13] illustrated in Figure 2. A crucial feature in wavelets is time-frequency localization [5]. The benefit of time-frequency localization is that it has variations in wavelet analysis on the time-frequency aspect ratio [14]. Then decompose the input, interpreted as a signal, using $g(n)$ (low pass filter decomposition) and $h(n)$ (high pass filter decomposition) [15]. The next step is down-sampling of two. The output is a low and high frequency signals. These two processes are done for the row section. It then repeated for the column, which then produces four sub-bands of output containing low and high frequency information [16]. A few examples of wavelet applications is data compression, voice signal recording, and music signal. The experiment result showed that the wavelet transform-based approach was better than existing minutiae-based methods and it required less response time which was more suitable for online verification [5, 17].

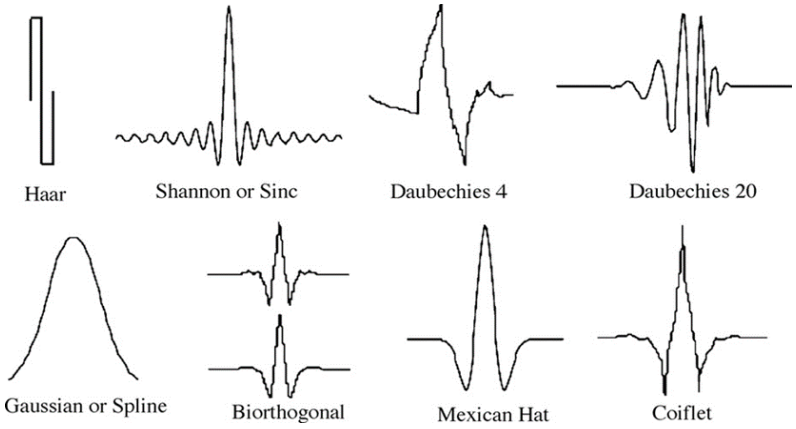


Figure 2. Wavelet for continuous wavelet transform (CWT) and DWT applications

2.2. Convolutional neural network

Deep learning algorithm has a set of algorithms that would check a high-level abstraction model on data using a processing layer with a complex structure and consists of several non-linear transformations [18]. Deep learning can be supervised, semi-supervised, or unsupervised [8, 16, 19]. One example of deep learning is a deep neural networks (DNNs) [6]. The effective use of this type of neural network is in the case of unlabeled or unstructured data [20]. There are also a much faster yet simpler learning algorithm called extreme learning machines (ELM). But the main problem of this learning algorithm is hidden-neuron-sensitive [21].

In deep learning, convolutional neural networks (CNN, or ConvNet) is DNNs classes, which most commonly applied to analyze visual images. CNN is a neural network that subsist of a combination of convolutional layers, pooling layers, and fully-connected layers [22]. One of the uniqueness of CNN is that not all neurons in the earlier layer are connected to the next layer, making computing and training time faster [23]. The convolution layer is unique to CNN, where it has a collection of filters that can be used to learn input images illustrated in Figure 3.

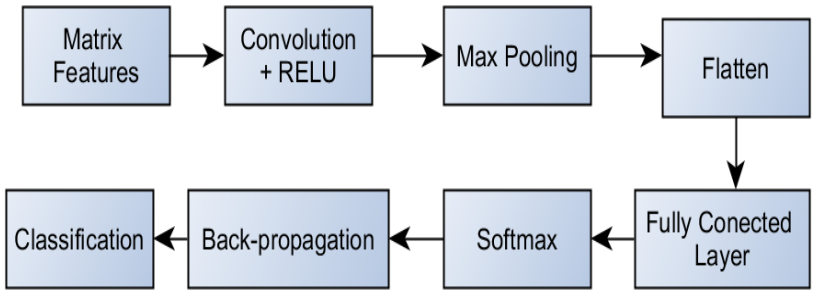


Figure 3. CNN architecture

Convolutional layer processes data with a grid topology [7]. Through the convolution layer, features will be extracted and advance to the next layer in the hope of a more complex features would be extracted [19]. The process at the CNN convolution layer is [1, 24].

$$c[i,j] = (I * K)[i,j] = \sum_m \sum_n I(m,n)K[i-m, j-n] \quad (1)$$

Pooling layer, in the basis, is a resizing process which purposed is to reduce the number of parameters and calculation time needed when the training a network [25]. The third layer in CNN is a fully-connected layer, where this layer takes all the neurons in the preceding layer (convolutional layer and pooling layer) and connects them to every single neuron that exists in this layer [26]. The activation functions used are of two kinds, namely a rectrified linear unit (ReLU) with [27]:

$$f(z) = max(0, z) \quad (2)$$

with x is the input value to the neuron and this function is used only in the convolution layer. Then the second is the softmax function [10] with:

$$f(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \; for \; i = 1, \dots, K \; and \; z = z_1, \dots, z_K \in R^K \quad (3)$$

after the forward pass, there is one process functioned to improve the weight value so that it would be the same as expected. This process is called back-propagation [28]. The basic idea of back-propagation is the utilization of chain-rule to calculate the effect of changing-weight in the network to a cost function and minimizing the error function by employing gradient descent, so that the output from the network is the same as the expected output [29]. Error function is used to estimate the difference between the output of a network with the expected output [23, 30].

Accuracy calculated from the classified image using CNN [31]:

$$Accuracy\ (\%) = \frac{N_c}{N} \times 100\% \quad (4)$$

with Nc is the amount of data detected correctly by the classification system and N is the sum of all test data. While, computational time is the time used by the system to perform a series of processes.

3. RESULTS AND DISCUSSION

There are 5 classes classified in the system, namely classes A, B, C, pointing, and # 5. System testing was finalized on five class of hand gestures. Using two types of datasets namely dataset A, from Sebastien Marcel Static Hand Posture Database [11], and dataset B, the from Author dataset [12]. Table 1 shows the class specification. The differences between both dataset is that the resolution and quality. Dataset A had a diverse resolution, but mostly at 76×66 pixel and a low quality image. Whilst the Dataset B, because it was taken from the smartphone camera, had a resolution of 4128×3096 pixels and a higher quality image. Both dataset were through resizing process became 76×66 pixels.

Table 1. Data specification

Class	Amount of data used	
	Dataset A	Dataset B
A	50	50
B	50	50
C	50	50
Pointing	50	50
#5	50	50
Total	250	250

3.1. Layer testing on system performance

The test using the percentage distribution of 80-20% training and testing for both datasets. Seven layers were used to determine which one was the best for this system. Note for the YCbCr and HSV, only the Cr and V layer were used respectively. Table 2 shows the effect of layer wavelet on system performance. Based on Table 3, the difference between the brightness and contrast value in the two datasets affected

the accuracy. For dataset B the binary layer was taken, it was normal. But for the dataset A, because the accuracy was the same for Cr from YCbCr, binary, grayscale, and V from HSV layer, the binary then chosen for simplicity.

Table 2. Layer test results

Layer	Accuracy (%)		Computatation time (s)	
	Dataset A	Dataset B	Dataset A	Dataset B
Red	96	82	1,31	20
Green	98	76	1,40	21
Blue	96	72	1,40	20
YCbCr (Cr)	100	74	1,30	21
Binary	100	88	1,70	29
Grayscale	100	72	1,39	20
HSV(V)	100	74	2,02	24

3.2. Sub-band testing on system performance

In this section, testing uses the distribution of training percentages and 80-20% testing for both data sets. Input image using binary layer. Four subband at decomposition level 1, namely LL, LH, HL, and HH were tested to determine which subband wavelete is the best for this system. Table 3 shows the effect of wavelet subband on system performance. Based on Table 3, the LL subband in both datasets produces the highest systemaccuracy. Both datasets take the LL sub-band because this sub-band produces finer image contours than the other subband, as shown in Figure 4. There are negative values in the images in the LH, HL, and HH sub-bands that cause image information to be lost.

Table 3. Sub-band test results

Sub-band	Accuracy (%)		Computatation time (s)	
	Dataset A	Dataset B	Dataset A	Dataset B
LL	100	92	1,3	25
LH	94	72	1,60	36
HL	96	72	1,50	19
HH	94	60	1,6	19

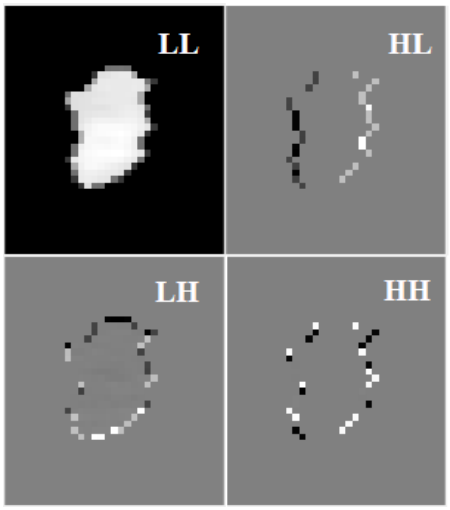


Figure 4. Display of sub-band test results using dataset A

3.3. Level decomposition testing on system performance

In this section, testing uses the distribution of training percentages and 80-20% testing for both data sets. Input images use binary layers, and subband low are used in this test. Four decomposition levels are tested to determine which wavelet decomposition level is the best for this system. Table 4 shows the effect of wavelet decomposition level on system performance. Based on Table 4, if the decomposition

level increases, the accuracy of the system decreases. That is because the resolution of the image will be smaller so that the features taken are less or less. The smaller image size can cause information to be lost. The illustration in Figure 5 can visually explain what happens when the decomposition level is raised. The greater the level, the smaller the resolution of the image. But for the sake of visualization, the image size is enlarged so that the image can be seen.

Table 4. Level decomposition test results

Decomposition level	Accuracy (%)		Computatation time (s)	
	Dataset A	Dataset B	Dataset A	Dataset B
1	100	92	1.4	25
2	96	74	1.5	28
3	86	78	1.2	35
4	66	74	1.2	31

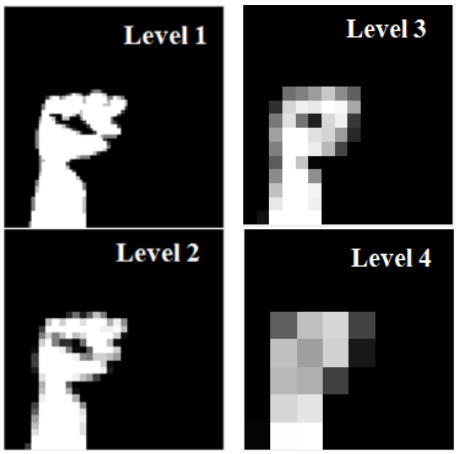


Figure 5. Displayed level decomposition for dataset B

3.4. Testing the mother wavelet types on system performance

In this section, testing uses the distribution of training percentages and 80-20% testing for both data sets. Input images is binary layers. Subband low, and 1 level decomposition wavelet are used in this test. Types of mother wavelet are tested to determine which type of wavelet is the best for this system. Table 5 shows the effect of types of mother wavelet on system performance. Based on Table 5, types of mother wavelets can make a difference in the characteristics of an object. Haar Wavelet is able to represent the characteristics of texture and shape well. Haar wavelet types produce higher system accuracy than other wavelet types.

Table 5. Mother wavelet test results

Mother wavelet	Accuracy (%)		Computatation time (s)	
	Dataset A	Dataset B	Dataset A	Dataset B
Haar	100	93	1,3	32
db3	100	82	1,6	19
db5	100	90	1,5	19
db7	98	84	1,4	19

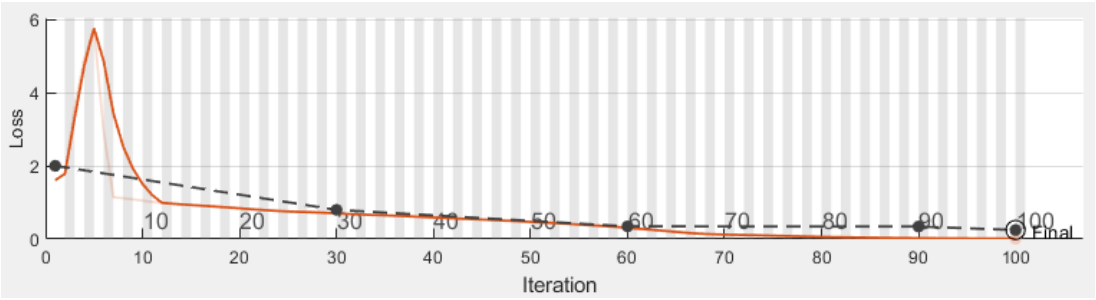
3.5. Learning rate testing on system performance

In this section, testing parameters CNN learning levels based on the level of accuracy and computation time system. Input images is binary layers. Subband low, 1 level decomposition, and haar wavelet are used in this test. Table 6 shows the effect of learning rate on system performance. Based on Table 6, learning rate affects how much the weight value can be adjusted by taking into account the cost function value. The lower the learning rate, the more precise the movement of the gradient in the direction to the valley (downward slope). However, by reducing the value of learning rate, the computational time needed is longer. In Figure 6, it can be seen that when the learning rate is 0.1 the training system has a very high

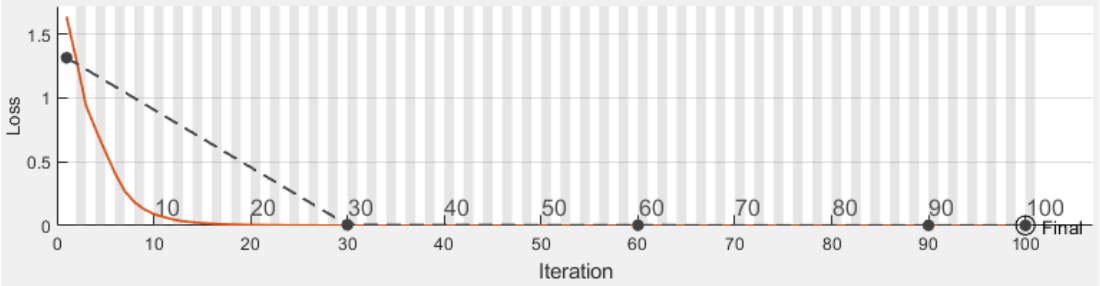
loss at the beginning of iteration. Resulting in a decrease in the accuracy of the recognition system. Whereas when the learning rate is 0.01, the training system has a lower loss at the beginning of iteration so that the systemworks more optimally.

Table 6. Learning rate test results

Learning Rate	Accuracy (%)		Computatation Time (s)	
	Dataset A	Dataset B	Dataset A	Dataset B
0,1	88	78	1.3	19
0,01	100	80	1.4	19
0,001	98	90	1.1	28
0,0001	98	70	1.1	18



(a)



(b)

Figure 6. Loss graph, (a) Learning rate at 0.1, (b) Learning rate at 0.01

3.6. Epoch testing on system performance

In this section, testing the epoch parameters used in the CNN classifier on the accuracy and computational time of the system. The epoch values to be analyzed are 10, 50, 100, 150, 200 for dataset A and dataset B. Table 7 shows the effect of epoch on system performance. Based on Table 7, the result was influenced by quality of the image parameters, such as brightness and contrast. More epoch needed to learn a higher quality images. However, if the epoch was wrong-selected, it would just elongated the time but the accuracy would not necessarily increased. The epoch taken for dataset A was 100, for simplicity. For this work, two datasets was used because of two things: first to investigate the CNN capabilities to classify the image at smaller (76×66) or bigger (128×128) pixel, secondly to check how CNN classifies image with low brightness and contrast.

Table 7. Epoch test results

Epoch	Accuracy (%)		Computatation Time (s)	
	Dataset A	Dataset B	Dataset A	Dataset B
10	98	68	1,2	19
50	96	78	1,3	19
100	100	90	1,3	19
150	100	76	1,3	19
200	100	74	1,3	19

3.7. Testing data on system performance

The goal for this test was to determine the best partition between training and testing data for DWT. The partition percentage of training data and test data to be examined were 80% -20% (40 training and 10 test data), 70% -30% (35 training and 15 test data), and 50% -50% (25 training and 25 test data) of the dataset for each type of gesture as displayed in Table 8. DWT requires a lot of training data so that the system could work optimal. It can be seen that images with good quality (dataset B) are more difficult to train and test because of more varied parameter values compared to lower one (dataset A).

Table 8. Data test results

% Data testing	Accuracy (%)		Computatation time (s)	
	Dataset A	Dataset B	Dataset A	Dataset B
80%-20%	100	90	1.25	19
70%-30%	98.67	78.67	1.25	19
50%-50%	96.80	83	1.25	19

4. CONCLUSION

In this paper, a hand gesture recognition implementation system using discrete wavelet transform and convolutional neural networks was proposed. The problem at classifying similar yet different gestures, completed by adopting the DWT and CNN classification system. Dataset A's best results using the binary layer, type LL sub-band, level one decomposition, Haar wavelet in the DWT parameter, the learning rate was 0.01 and the number of epochs was 100 the CNN parameter. In dataset B, it had the best parameters using binary layer, LL sub-band type, level one decomposition, Haar wavelet on DWT parameters, learning rate value of 0.001 and the number of epochs was 100 on the CNN parameter. Hand gesture recognition system with DWT and CNN using dataset A that was accomplished from the best parameters had an accuracy of 100%. In dataset B, it had an accuracy at 90% from the best parameter.

REFERENCES

[1] A. Kendon, "Gesture," *Annu. Rev. Anthropol.,* vol. 26, pp. 109-128, 1997.
[2] J. P. Sahoo, S. Ari, and D. K. Ghosh, "Hand gesture recognition using DWT and F-ratio based feature descriptor," *IET Image Processing*, vol. 12, no. 10, pp. 1780–1787, 2018.
[3] X. Fu, T. Zhang, C. Bonair, M. L. Coats, and J. Lu, "Wavelet enhanced image preprocessing and neural networks for hand gesture recognition," *IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, pp. 838-843, 2015.
[4] H. M. Parmar, "Comparison of DCT and wavelet based image compression techniques," *Int. J. Eng. Dev. Res.*, vol. 2, no. 1, pp. 664-669, 2014.
[5] M. Sifuzzaman, M. R. Islam, and M. Z. Ali, "Application of wavelet transform and its advantages compared to fourier transform," *J. Phys. Sci.*, vol. 13, pp. 121–34, 2009.
[6] H. Karisma and D. H. Widyantoro, "Comparison study of neural network and deep neural network on repricing GAP prediction in Indonesian conventional public bank," *6th Int. Conf. Syst. Eng. Technol. ICSET*, pp. 116–22, 2016.
[7] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, 2015.
[8] F. Hussain and J. Jeong, "Exploiting deep neural networks for digital image compression," *2nd World Symp. Web. Appl. Networking, WSWAN*, pp. 1-6, 2015.
[9] D. N. Fernández and B. Kwolek, "Progress in pattern recognition, image analysis, computer vision, and applications," *Springer Berlin Heidelberg*; pp. 441–449, 2013.
[10] S. Kalita and M. Biswas, "Recent developments in machine learning and data analytics," *Springer*, pp. 397–410, 2018.
[11] S. Marcel, "Hand posture recognition in a body-face centered space," *Proceedings of the Conference on Human Factors in Computer Systems (CHI)*, pp. 302-303, 1999.
[12] E. B. Candrasari, L. Novamizanti, and S. Aulia, "Discrete wavelet transform on static hand gesture recognition," *Journal of Physics: Conference Series*, vol. 1367, pp. 1-13, 2019.
[13] A. Godfrey, R. Conway, D. Meagher, and G. ÓLaighin, "Direct measurement of human movement by accelerometry," *Med. Eng. Phys.*, vol. 30, no. 10, pp. 1364–1386, 2008.
[14] R. Ramos, B. Valdez-Salas, R. Zlatev, M. S. Wiener, and J. M. B. Rull, "The discrete wavelet transform and its application for noise removal in localized corrosion measurements," *Int. J. Corros.*, vol. 2017, pp. 1-7, 2017.
[15] S. Thakral and P. Manhas, "Image processing by using different types of discrete wavelet transform," *International Conference on Advanced Informatics for Computing Research*, pp. 499-507, 2019.
[16] M. Monteleone, "NooJ local grammars and formal semantics: Past participles vs. adjectives in Italian," *Commun. Comput. Inf. Sci.*, vol. 607, no. 8, pp. 83–95, 2016.
[17] R. C. Gonzalez and R. E. Woods, "Digital image processing (3rd Edition)," *Prentice-Hall, Inc. Upper Saddle River, USA*, 2006.
[18] S. Khan, H. Rahmani, S. A. A. Shah, and M. Bennamoun, "A guide to convolutional neural networks for computer vision," *Synthesis Lectures on Computer Vision*, 2018.
[19] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*. vol. 61, pp. 85–117, 2015.

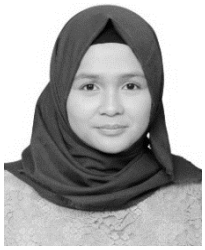
[20] H. Yi, S. Shiyu, X. Duan, and Z. Chen, "A study on deep neural networks framework," *IEEE Adv. Inf. Manag. Commun. Electron. Autom. Control. Conf. IMCEC*, pp. 1519-1522, 2016.
[21] F. Albu, D. Hagiescu, L. Vladutu, and M-A. Puica, "Neural network approaches for children's emotion recognition in intelligent learning applications," *EDULEARN15 7th Annu Int Conf Educ New Learn Technol Barcelona, Spain, 6th-8th*, 2015.
[22] I. B. K. Sudiatmika, F. Rahman, T. Trisno, and S. Suyoto, "Image forgery detection using error level analysis and deep learning," *TELKOMNIKA Telecommunication Comput Electron Control*, vol. 17, no. 2, pp. 653-659, 2018.
[23] Nielsen M. A., "Neural Networks and Deep Learning. Artificial Intelligence," *Determination Press*, 2015.
[24] J. Heaton, "Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning," *Genet Program Evolvable Mach*, vol. 19, no. 1–2, pp. 305-307, 2018.
[25] V. Bui and L. Chang, "Deep learning architectures for hard character classification," *Proceedings on the International Conference on Artificial Intelligence (ICAI)*, pp. 108-114, 2016.
[26] N. Jmour, S. Zayen, and A. Abdelkrim, "Convolutional neural networks for image classification," *Int. Conf. Adv. Syst. Electr. Technol. IC_ASET*, pp. 397-402, 2018.
[27] R. S. Sutton and A. G. Barto, "Reinforcement learning-second edition," *The Lancet. MIT Press, Cambridge, MA*, 2018.
[28] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986.
[29] T. Rashid, "Make your own network," *J. Cell. Biol.*, vol. 169, no. 4, 2005.
[30] T. S. Gunawan et al., "Development of english handwritten recognition using deep neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 10, no. 2, pp. 562-568, 2018.
[31] S. Fairuz, M. H. Habaebi, and E. M. A. Elsheikh, "Pre-trained based CNN model to identify finger vein," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 3, pp. 855-862, 2019.

BIOGRAPHIES OF AUTHORS

Muhammad Biyan Priatama obtained a bachelor degree in telecommunication engineering at Telkom University, Indonesia in 2019. The topic for his final essay was about image recognition. His research interest is image processing.

Ledya Novamizanti received the Bachelor of science degree (S.Si.) in mathematics from Andalas University, Indonesia in 2005. She received the Master of engineering degree (M.T.) in electrical engineering from Telkom University, Indonesia, in 2018. She has been working as a lecturer in Telkom University since 2010. Her current research interests include signal processing, computer vision, and pattern recognition, and artificial intelligence.

Suci Aulia is a lecturer at Telkom University, Departement of Applied Science. Her concern is in signal processing since 10 years ago. She has published several papers in her research especially the image signal processing field. She graduated in 2010 as a bachelor of engineering and graduated in 2012 as a master of engineering at Telkom University. She has been currently taking doctoral degree in School of Electrical Engineering and Informatics, Bandung Technology Institute (ITB).



Erizka Banuwati Candrasari obtained a bachelor degree in telecommunication engineering at Telkom University, Indonesia in 2019. The topic for finish her study in Telkom University was Information signal processing. Her research interest includes machine learning, analysis data, data visualization and image recognition.