General Remote Control Based on Hand Patterns Detection Using Convolutional Neural Network

J Utama1*, H Y Biu2
1,2Departemen Teknik Elektro, Universitas Komputer Indonesia, Indonesia

Email: *Jana.utama@email.unikom.ac.id

Abstract. In this paper, the system is made to be able to recognize hand patterns to be able to control electronic devices, as a substitute for the conventional remote. This system consists of a computer containing a program with a trained Convolutional Neural Network (CNN) model and an MCU node containing a remote infrared signal database whose function we want to replace. The camera functions as a reader of hand patterns which will be classified using trained CNN models. Each hand gesture represents the infrared pulse in the database program n Node MCU. There are 4 hand patterns that represent commands in this proposed system. To be able to handle much more commands, this proposed system is designed to change mode according to the combination of hand patterns detected by the system algorithm. The changing of mode results in the changing of the infrared pulse which is represented by each hand pattern in the database. Node MCU will transmit infrared pulse corresponds to hand gesture from CNN classification after receiving a command from the computer through TCP/IP communication.

1. Introduction

At this time emerging technologies are growing and being applied in various fields is the technology that utilizes neural networks. An artificial neural network itself is inspired by the human nervous system. This makes this technology is known as part of the Artificial Intelligence (AI). Neural network algorithms have many uses in its application, one of which is for pattern recognition, either color or shape that can replace the image processing algorithms. One study that utilizes pattern recognition in a robot system is the study of ball retainers [1]. Research by the author this time using an artificial neural network algorithm as a method to classify images based on patterns of hand captured in real-time via the camera sensor.

In this study, the authors used as a substitute for hand pattern remote control conventional widely used today in electronic equipment such as televisions (TV) or air conditioning (AC). Because of its small and can be taken anywhere, the remote control can be lost due to human error, such as forgetting to put it where it is. In addition to these issues, other issues that arise in the use of portable remote control is when we have a lot of electronic devices that are controlled using the remote, it takes too many remote controls so that the possibility of eliminating the remote user will increase. Artificial neural networks have been widely applied in studies in various fields of science. In the medical field, for example, neural networks can be used to diagnose a particular disease [2]. In that study, artificial neural networks diagnose the disease based on the symptoms suffered by the patient. In addition to the medical field, the artificial neural network has also been applied in smart security systems on the basis of face recognition [3,4]. There is also a security system with fingerprint recognition that uses neural networks as an essential algorithm [5]. In addition, the neural network can
be used to identify patterns of human behavior [6]. The human behavioral pattern recognition system itself can be used to create a security system. In addition to the medical field and the security of the foregoing, the artificial neural network is also widely applied in smart vehicles. One application is the self-driving car using a deep neural network [7]. In addition, the artificial neural network has been applied to the automatic parking space detection system [8]. In a smart home automation system, the neural network is used to control electronic devices through command hand patterns [9]. There is also a remote control replacement tool is Microsoft Kinect which uses artificial intelligence to detect the movement of the human body [10]. This study uses a neural network architecture Convolutional Neural Network (CNN) to detect patterns of hand as the basis of the general system remote control commands. In his research, Joydeep ever makes a similar system that is general remote control but using the mobile as a substitute for conventional remote [11].

This research was conducted with the aim to create a general system of remote control with a command base hand pattern. This remote control system can be used for all electronic devices that use an infrared remote control as a controller, such as a TV or air conditioning. Because using the hand as a base pattern command, the system is expected to overcome the problem of losing the remote for user negligence, as often happens in the use of conventional remote control.

In order to replace the function of the remote control made an IR pulse decoder subsystem to store data to a remote command in a database that can be used to control electronic devices through the command pattern of hands. Because the hand pattern that can be used is limited while the system is created with the aim to control multiple devices, so in this study used combinations of patterns to replace hand the target device to be controlled.

2. Method
Convolutional Neural Network (CNN) architecture is basically used to process images and also sounds to classify the images or sounds. Images with a size of 100 x 100 pixels require 10000 nodes if processed using neural networks such as perceptrons [12]. But with the convolution process on CNN, the required nodes are no longer that much because the initial process that occurs at CNN is to reduce the size of the data to be processed at the fully-connected layer. There are several algorithms applied to image data in order to classify the image. The algorithms are applied to certain layers on the network that is made. These layers are the Convolution layer, the Pooling layer, and the Fully-connected layer.

3. Results and Discussion
3.1. Block Diagram System
Remote control system by detecting patterns of hand has major hardware namely computers, nodeMCU ESP8266, IR receiver, and an IR transmitter. A block diagram of the designed system is shown in Figure 1.
3.2. Algorithms
Stages traversed in designing software of this system is the process of creating a model of CNN, training CNN, decoding a command from the remote control, the detection pattern of the hand that had been trained, and the wireless communication between the computer and MCU Node ESP8266.

3.2.1. CNN Model
CNN Model is an important part that is used to classify the pattern shape hand. Figure 2 shows a block diagram of the CNN-designed.

![Figure 2. CNN Model](image)

In the study, the model was designed to contain the Convolutional CNN max-pooling layer and layer, and of course a fully-connected layer by Relu Softmax activation and activation. There are three layers of Convolutional used in the model. Figure 3 shows a flowchart of the classification process carried out by the model of CNN. Here's an explanation of the flowchart.

a. The binary image size [240 x 240] is the model input CNN made.
b. In the first process, the input image is convoluted by the filter kernel size of 5 x 5 by 32 with 2 x 2 Strides strides show how far the shift when the Convolutional filter performed on the input image matrix. The output of this process is a matrix with a size obtained by using equation (1).

\[ M_w = \frac{(W_m - W_k)}{S} + 1 \]  \hspace{1cm} (1)

Where:
- \( M_w \) = The width of the output matrix
- \( W_m \) = The width of the input matrix
- \( W_k \) = The width of the kernel matrix
- \( S \) = strides

By incorporating the existing values into the equation (1) are obtained from the output of the first Convolutional process is a matrix with size [119 x 119].
c. The output of the first Convolutional process is the input of the process of pooling with matrix pooling max 2 x 2. The output of this process is a matrix of size [59 x 59].
d. Output of max-pooling is input from the Convolutional process into two. Unlike the first Convolutional process, strides are used in this process is 1 x 1. Obtained by measuring the output matrix [57 x 57].
e. The output of the Convolutional process into two and then again with max-pooling 2x2 and produces an output matrix size [28 x 28].
f. The last convolutional process is carried out to be able to produce a matrix of size [26 x 26].
g. After passing through the process of pooling max, the resulting matrix is a matrix with size [13 x 13].
h. The matrix [13 x 13] is then converted into a vector with a vector size of 169.

i. The values of this vector are fully-connected input of the layer with the number of neurons in the hidden layer is set at 128 with the activation process using the Relu activation function.

j. The output of the neurons in the hidden layer is a float value that will serve as input to the output layer. There are four neurons in the output layer defined by softmax activation. This amount represents 4 hands predetermined patterns.

k. The output of the activation softmax is a value that indicates the probability of similarity between the input images by 4 patterns that have been trained hands.
3.2.2. **CNN Training**

In this case, using 4 forms a hand pattern that will serve as a command. The forms of the hand pattern shown in Figure 4, Hands predetermined pattern, prepared by the number of 550 different images for each of the hand pattern.

![Figure 3. Flowchart Model classification by CNN](image-url)
The use of a binary image is done with the aim to accelerate the process of execution of the program. Because the trained image is a binary image, then the main program hand pattern detection in real-time, image captured by the camera should be converted into a binary image before CNN was added to the model that has been trained.

3.2.3. Decoding the remote control

The main purpose of this system design is to replace the function of remote control conventional. Command is given via keystrokes on the remote control is replaced with the hand pattern detected by the camera. Because remote control giving orders via infrared pulse, then to replace the functions we need to know the shapes of the infrared pulses transmitted by the remote control when the command button press.

To find the infrared pulse shape that represents a certain button, the process decoding required. Decoding process is what translates into the infrared pulse string or integer form as appropriate. Value string/integer obtained will be stored into the database to be paired with a specific hand pattern liking. Infrared pulse decoding process and assign process orders hand pattern as shown in Figure 5.
a. The system will store the received IR pulse as an integer value in the variable X
b. After the IR pulse is translated into an integer value, the value is paired with the handshape pattern in the database.
c. Conditions iteration stops when the IR pulses received total equals 4.

3.2.4. *Hand pattern detection using the CNN model*

The results of the process CNN training has done before to produce a model of CNN with constant weights and biases. This model will be used to classify patterns of the hand is detected by the camera in real-time.

a. CNN models that have been trained to read by the program to be used to process the images captured by the camera.
b. The capture process carried out by aiming the camera to take the image data in the form of the existing conditions in front of the camera.
c. Smoothing processes are performed on the image using a bilateral filter then flip the image horizontally.
d. At the corner of the images that have been captured, made rectangular areas as regional input to read the hand pattern.
e. Pictures displayed on the screen like aid for the user to position the hand pattern on the camera.
f. The system waits for commands from the keyboard. The escape key to end the program, the b button to take the background.
g. Once the background is captured, the background is then removed with the intention to ignore the image other than the hand pattern to be detected.
h. Pictures output from the elimination of background converted to grayscale and then Gaussian blur applied to the image to make the image smoother. Here is the equation (2) that is used by OpenCV to transform RGB image into the grayscale image:

\[
\text{Grayscale} = 0.299R + 0.587G + 0.114B
\]

With R is a red pixel value of RGB image, G is a green pixel value, and B is a blue pixel value.

i. Once the image is converted again into a binary format, as input from CNN models that have in-training is a binary image. Converting a gray scale image to a binary image is done by setting a threshold value between 0 - 255. Threshold values are applied to each pixel grayscale image where pixels whose value is below the threshold will be changed to 0 and the pixel value is greater than or equal to the threshold will be changed to 255.

j. When the space key is pressed, the binary image processing results before then resize to the size of 240 x 240 in accordance with an image size that is used as input to the model of CNN. The process of resizing an image from larger to smaller sizes is done by aggregating the values of the pixels into one pixel and taking the average value. In this case, because the resized image is a binary image, the pixel is the incorporation that took the highest value.

k. The images can then be reshaping in order to have a format in accordance with the CNN model input format.

l. The image is then inserted into the model of CNN’s classified. This classification results later sent to the server (Node MCU) as a command to the server to send the IR pulses to control a specific device.

3.2.5. *TCP / IP socket programming*

Value is the output of the model CNN is a command that corresponds to a particular IR pulse. IR pulses that have been translated into a string form are sent from the computer to the Wi-Fi module ESP8266 via TCP / IP communication using socket programming. In this system, the computer acts as a client and ESP8266 acts as a server. In order for Raspberry Pi with ESP8266 can communicate, both
devices must be connected to the same network. In this case, ESP8266 connected to a Wi-Fi network Raspberry Pi.

Research through direct experiments carried out to test the remote control system that has been made. By using a PC equipped with a camera as a sensor which is used to capture an image pattern of a hand, the performance of the system was tested. System testing is done on two conditions of the room, the room with the intensity of bright light and room with low light intensity.

3.3. Testing in a bright room conditions
Each pattern is hand-tested 10 times under these conditions and test results are shown in Table 1 the lighting conditions is shown in Figure 6.

| No | Pattern shape Hand | Number of tests | Number of Correct Detection |
|----|-------------------|----------------|-----------------------------|
| 1  | C                 | 10             | 10                          |
| 2  | Okay             | 10             | 10                          |
| 3  | palm of          | 10             | 10                          |
| 4  | Peace            | 10             | 10                          |

Figure 6. Imagery that was captured in bright light conditions

From the testing that has been done, the percentage of successful detection can be calculated using equation (3).

\[
\text{Success percentage} = \frac{\text{number of correct detection}}{\text{number of tests}} \times 100 \% \\
\]

1. Hand pattern detection success percentage "C"
   10/10\times 100\% = 100\% 
2. Hand pattern detection success percentage "Okay"
   10/10\times 100\% = 100\% 
3. Hand pattern detection success percentage "palm of"
   10/10\times 100\% = 100\% 
4. Hand pattern detection success percentage "Peace"
   10/10\times 100\% = 100\%
Using equation (3) can be obtained by the percentage of total success amounted to: 40/40× 100% = 100 %

3.4. Testing in a dim room conditions

Each pattern is hand tested 10 times under these conditions and test results are shown in Table 2 the lighting conditions is shown in Figure 7.

| No | Pattern shape Hand | Number of tests | Number of Correct Detection |
|----|-------------------|-----------------|----------------------------|
| 1  | C                 | 10              | 5                          |
| 2  | Okay             | 10              | 6                          |
| 3  | palm of          | 10              | 5                          |
| 4  | Peace            | 10              | 7                          |

From Table 2, the percentage of successful detection can be calculated using equation (3).

1. Hand pattern detection success percentage "C"
   5/10× 100% = 50 %
2. Hand pattern detection success percentage "Okay"
   6/10× 100% = 60 %
3. Hand pattern detection success percentage "palm of"
   5/10× 100% = 50 %
4. Hand pattern detection success percentage "Peace"
   7/10× 100% = 70 %

By using equation (3), it also can be calculated the percentage of total success in detecting the low light conditions as follows: 23/40× 100% = 57.5 %. Tests performed in conjunction with a remote control hand pattern detection testing with the test results shown in Table 3. This testing is done at a distance of 50 cm hand pattern of the camera with bright light conditions.
Table 3. The test results in dim conditions

| No | Pattern shape Hand | Number of tests | The number of signals was successfully sent |
|----|-------------------|-----------------|------------------------------------------|
| 1  | C                 | 10              | 10                                       |
| 2  | Okay              | 10              | 10                                       |
| 3  | palm of           | 10              | 10                                       |
| 4  | Peace             | 10              | 10                                       |

By using equation (3), can be obtained a success rate of data transmission from the client to the server at the same time an IR pulse delivery success rate is based on orders given as follows: 40/40 × 100% = 100%

3.5. Room lighting conditions

With accuracy differ greatly between detection in bright conditions with poor lighting conditions indicate that the light intensity greatly affects the accuracy of the detection system. This is because the model of CNN which made receiving input of binary image. By setting the threshold constant in the process of converting grayscale image into a binary image of the change in room lighting conditions will greatly affect the shape of the input image to the model of CNN. Figure 8 shows a comparison of the input image to CNN on the condition of a bright room with dim room conditions.

![Figure 8. Comparison of a binary image in bright conditions (a) and dim conditions (b)](image)

The picture above shows a significant difference. In bright light conditions, images incorporated into the model can be said CNN perfect picture of the shape of palms, while in low light conditions that damage their image amid the shape hole of the palm. This happens because of the lack of light so that the center of the palm of the hand has a darker color than other parts that cause the pixels of the section will be translated into the black value at the time of conversion of grayscale images to binary. Because the "Ok" pattern image has a black hole in the middle, it is most likely that the CNN model will classify the imperfect palm image into the "Ok" hand pattern as happened in the tests that have been done.

From the testing that has been done, other things that affect the success of the system detects the input hand pattern is the position of the light source to the user and the camera. Although the room is arranged with high light intensity conditions, the position of the light source is behind the then most likely the system will fail to detect the pattern of a given hand. This happens because when the light source is behind the hand pattern, the palm of the hand will be darker because it is not directly exposed to light. This causes when the system converts a gray image into a black-and-white image, so there will be a part of the palm of the hand that translates to black pixels instead of white pixels. This is a binary image defect caused by the position of the hands of the source of light patterns and will result in a system difficult to classify it into the desired pattern. This problem can be solved by using the
image pattern RGB or grayscale hand as input for the model CNN. Of course, the image is used as input into the process of training CNN should fit the image that will be input into the model of CNN who have been trained during the process of detection in real-time progress. However, the use of RGB or grayscale images will make the process more slowly because of the processed data will be bigger and CNN made models require more layers in order to have a high accuracy so that the classification process will take longer. In addition, if you want to keep using the binary image as inputs to the model of CNN, the problem of image defects due to the light source can be overcome by placing the light source in front of the pattern of hands who want to be detected thus the process of converting from grayscale image to a binary image will produce a binary image that has a shape similar to RGB image captured by the camera.

Room lighting conditions affect the image to be processed in the main program and the effect can be clearly observed during the process of converting the image to grayscale image with black-and-white format. Low light conditions will result in the flawed binary image from grayscale image conversion results. This is due to the lack of light; so much a part of the palm has a darker color than other parts that cause the pixels of the section will be translated into the black value at the time of conversion of grayscale images to binary. Another thing that affects the success of the detection pattern is the position of the light source hand to hand pattern. When the light source is placed behind the hand pattern, then condition hand pattern image captured by the camera will be the same as the conditions hand pattern image when the room has dim lighting. This is because when the light source is behind the hand pattern part of the palm do not get enough light so that the palms will be darker so that when the system is converting the image of the gray image in black and white, will be part of the palm which translates pixel to black, instead of white pixels.

Binary image data used to train the model created CNN also has an influence on the level of success of the model of CNN classify patterns of hand real-time, to get the most current detection process takes place in real-time, image data is trained need to be made with conditions as close as the process of real-time detection takes place. The conditions such as the position the hand pattern in the input image, the background, the camera distance arm patterns, and hand pattern image defects that may occur. Because the model of CNN trained to classify images based on certain patterns contained in an image, the hand shape similar pattern will result in a probability that values are not far apart.

3.6. The image data training

Another thing that affects the detection accuracy of the system is trained image data. If all the images that are trained are almost perfect hand pattern, then when the direct detection pattern image is not perfect hand or far from the perfect form will be hard classified by CNN models correctly. In addition, the pattern spacing trained hands of the camera affect detection real-time, if the images are trained drawn from a close distance from the camera, then when the detection process in real-time, distance from the camera will affect the ultimate success of the system detects patterns of hand. It actually can be overcome by adding some image processing algorithms to segment the hand pattern of the image captured by the camera. Hand pattern image segmentation results

The success system sends an IR pulse signal in accordance with the orders given, relates directly to the success of the system detected a pattern of hands properly. This is because the IR pulse transmission system works based on command obtained after the computer detects a pattern of a given hand. Therefore, the error detection result pattern will hand IR pulse transmission errors. In addition, the success of the system sends IR pulses to the device, influenced by the distance between the transmitter and the device. Because the transmitter is not equipped with an amplifier, the distance between the transmitter and the device can only transmit data with a distance of less than 1 meter [13].

4. Conclusion

In this paper, the authors have presented a hand pattern detection method as a command to replace remote control Conventional using artificial neural networks. The architecture used in this study was the Convolutional Neural Network (CNN). The percentage of successful detection in a room with
bright lighting conditions by 100%, while in a room with low light conditions is only about 57.5%. This indicates that the remote control system built suitable only implemented in a room with bright lighting. Through this study, the authors find information on some of the factors that affect the success of the model in classifying the binary image CNN hand pattern is entered into it. The factors that influence it are the lighting conditions of the room, the position of the light source to the hand patterns and camera sensors. For further development of this research, the use of RGB or grayscale images as input from CNN models can solve problems that arise in the use of the binary image. This is because the RGB and grayscale images have more patterns and characteristics compared to the binary image. In addition, the use of RGB or grayscale images will not result in problems such as image defects due to inaccuracy threshold is used as the binary image. The use of RGB or grayscale image also allows pattern trained hands more.

References
[1] Utama, J., & Riki, D. 2017. Implementation of Target Detection System Based on Color and Pattern Recognition for Ball Follower Robot, "Telekontran, 5, pp.107-117.
[2] Al-Shayea, Q. K. 2011. Artificial neural networks in medical diagnosis. International Journal of Computer Science Issues, 8(2), pp. 150-154.
[3] Mall, A., & Ghosh, S. 2010. A Neural Network Based Face Detection Approach. Int. J. Computer Technology & Applications, 3(2), pp. 823-829.
[4] Owayjan, M., Achkar, R., & Iskandar, M. 2016, October. Face detection with expression recognition using artificial neural networks. In 2016 3rd Middle East Conference on Biomedical Engineering (MECBME), pp. 115-119
[5] Soni, S. V. 2016. Fingerprint Recognition Using Artificial Neural Network: Review. Journal of Electronics and Communication Engineering (IOSR-JECE), 11(6), pp. 79-82.
[6] Ji, S., Xu, W., Yang, M., & Yu, K. 2012. 3D convolutional neural networks for human action recognition. IEEE transactions on pattern analysis and machine intelligence, 35(1), pp. 221-231.
[7] Chandra, R., & Agani, N. 2012. Self Driving Car: Artificial Intelligence Approach. Jurnal TICom, 1(1).
[8] Martinez, J., Zoeke, D., & Vossiek, M. 2018. Convolutional neural networks for parking space detection in downfire urban radar. International Journal of Microwave and Wireless Technologies, 10(5-6), pp. 643-650.
[9] Dinh, D. L., Kim, J. T., & Kim, T. S. 2014. Hand gesture recognition and interface via a depth imaging sensor for smart home appliances. Energy Procedia, 62(62), pp. 576-582.
[10] Paul, S., Basu, S., & Nasipuri, M. 2015. Microsoft kinect in gesture recognition: a short review. Int. J. Control Theory Appl, 8(5), pp. 2071-2076.
[11] Roy, J., & Roy, J. K. 2014. Design of smart universal remote using mobile for home automation. IOSR Journal of Computer Engineering, 16(5), pp. 73-80.
[12] Mureşan, H., & Oltean, M. 2018. Fruit recognition from images using deep learning. Acta Universitatis Sapientiae, Informatica, 10(1), pp. 26-42.
[13] Benet, G., Blanes, F., Simó, J. E., & Pérez, P. 2002. Using infrared sensors for distance measurement in mobile robots. Robotics and autonomous systems, 40(4), pp. 255-266.