IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 22, Issue 1, Ser. III (Jan - Feb 2020), PP 36-43
www.iosrjournals.org

An Approach in Building a Vision-Based Hand Gesture Recognition System

Huy Nguyen Phuong[1], Mai Thuong Duong Thi[2]
[1]*Faculty of ElectronicsEngineering*
Thai NguyenUniversity of Technology, Vietnam
[2] *Faculty of Information Technology*
University of Information and Communication TechnologyThai Nguyen, Vietnam

Abstract: *Building a vision-based hand gesture recognition system is a very interesting problem due to its large scope of application. This paper presents an approach to this problem based on the combination of some of the effective tools in image processing. These include the use of the Particle Filter for motion monitoring and image acquisition of the hand via webcam, the extraction of the feature of a hand image by a Gabor wavelet transform, image recognition using the Multi-layer Perceptron neural network. The simulation results show that this method is highly accurate and fully applicable in practice.*

Date of Submission: 04-02-2020 Date of Acceptance: 19-02-2020

I. Introduction

Gestures are a powerful means of communication between people. Like words, gestures can be an independent means of expressing people's ideas in communication. Today, thanks to scientific advances, people can use hand gestures to create natural interactions between themselves and computer devices [4].

There are two main approaches for computers to recognize gestures. They are "Data-Glove based" and "Vision Based" approaches. The Data-Glove based methods use sensors to digitize hand and finger movements into multi-parameter data. The auxiliary sensor makes it easy to collect the configuration and movement of hand. However, these devices are quite expensive and give users uncomfortable feeling. In contrast, Vision-Based methods require only one camera to create a natural interaction between humans and computers without using any additional equipment. The advantage of this method is that it allows information collected on hand gestures to be collected in a non-invasive way and at a low cost. However, this approach also poses many challenges because these systems need to be background-invariant, lighting insensitive. Furthermore, such systems must be optimized to meet requirements, including accuracy and stability [9].

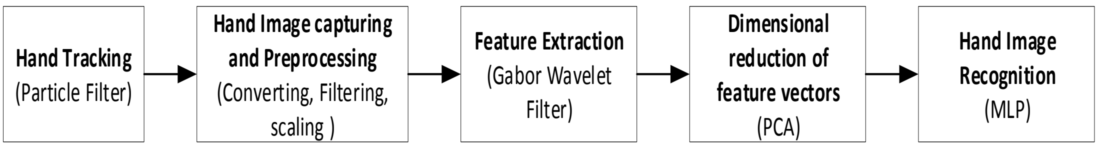

Figure 1. Steps to build the vision-based hand gesture recognition system

In this study, we propose a method to build a gesture-based identification system by incorporating some effective tools as described in Figure 1. First, we use the Particle Filter algorithm for hand tracking and capture images of hands via webcam. Next, we extracted feature vectors of the hand images by using Gabor wavelet transform algorithm. Finally, we use the Multi-Layer Perceptron (MLP) neural network for gesture recognition. The following sections of this paper will describe in detail how to combine these tools.

II. Hand tracking using Particle Filter

Filtering is the problem of estimating the state of the system as soon as a set of observations about the system is received and validated. This problem plays an extremely important role in many fields of science, technology, economics and finance. To solve this problem, we model the variability of the system and the noise in the observations. The results obtained from this process are usually nonlinear quantities and have a non-Gaussian distribution. So far, particle filters are considered superior than other filters for estimating state parameters with the above properties [6].

Particle filter is a technique for estimating state parameters based on Bayes recursive formula. The Bayes formula predicts a series of hidden parameters based on observed data only.

Assume that the state variables X_t describe the target configuration at time *t*, and Z_t is the observed quantity extracted from the image. We aim to estimate the value of based on all observations $Z_{1:t} = (Z_1, ..., Z_t)$ until time *t*. By using the Bayes formula, we can estimate the posterior probability $p(X_t | Z_{1:t})$ [5]:

$$p(X_t | Z_{1:t}) = \eta p(Z_t | X_t) \int p(X_t | X_{t-1}) p(X_{t-1} | Z_{1:t-1}) dX_{t-1} \quad (1)$$

where, η is the ratio factor, $p(Z_t | X_t)$ denotes the observation model that measures how well the observation fits the predictions, and $p(X_t | X_{t-1})$ describes the motion model that proposes the next state X_t based on the previous state X_{t-1}.

The above recursive equations and expressions look simply. However, in reality, we cannot calculate them because to calculate $\int p(X_t | X_{t-1}) p(X_{t-1} | Z_{1:t-1}) dX_{t-1}$, it is necessary to perform an integral calculation in which the data has a very large and very complex dimension.

Unlike calculus-based calculation methods in which each method always tries to find a solution to the above equations through one or more other equations, Monte Carlo methods rely on simulation and approximate the distribution functions as well as the above integrals by a set of data samples generated by the distribution functions in these integrals. This is the main reason for the Monte Carlo-based filtering algorithms such as Particle filtering to solve the nonlinear and non-Gaussian problems thoroughly and effectively.

We denote a set of particles $\left\{ \left(X_t^{(i)}, \mathrm{w}_t^{(i)} \right), i = 1, ..., N \right\}$ to represent the distribution. Where, $X_t^{(i)}$ and $\mathrm{w}_t^{(i)}$ are a state and an associated weight of the *i*th particle at time *t*, respectively. Eq. (1) can be approximated by [5]:

$$p(X_t | Z_{1:t}) = p(Z_t | X_t) \sum_i \mathrm{w}_{t-1}^{(i)} p(X_t | X_{t-1}) \quad (2)$$

Given a set of particles from the previous time $\left(X_t^{(i)}, \mathrm{w}_t^{(i)} \right)$, configurations at the current time $X_t^{(i)}$ are calculated from a proposal distribution

$$q(X_t) = p(Z_t | X_t) \sum_i \mathrm{w}_{t-1}^{(i)} p(X_t | X_{t-1}) \quad (3)$$

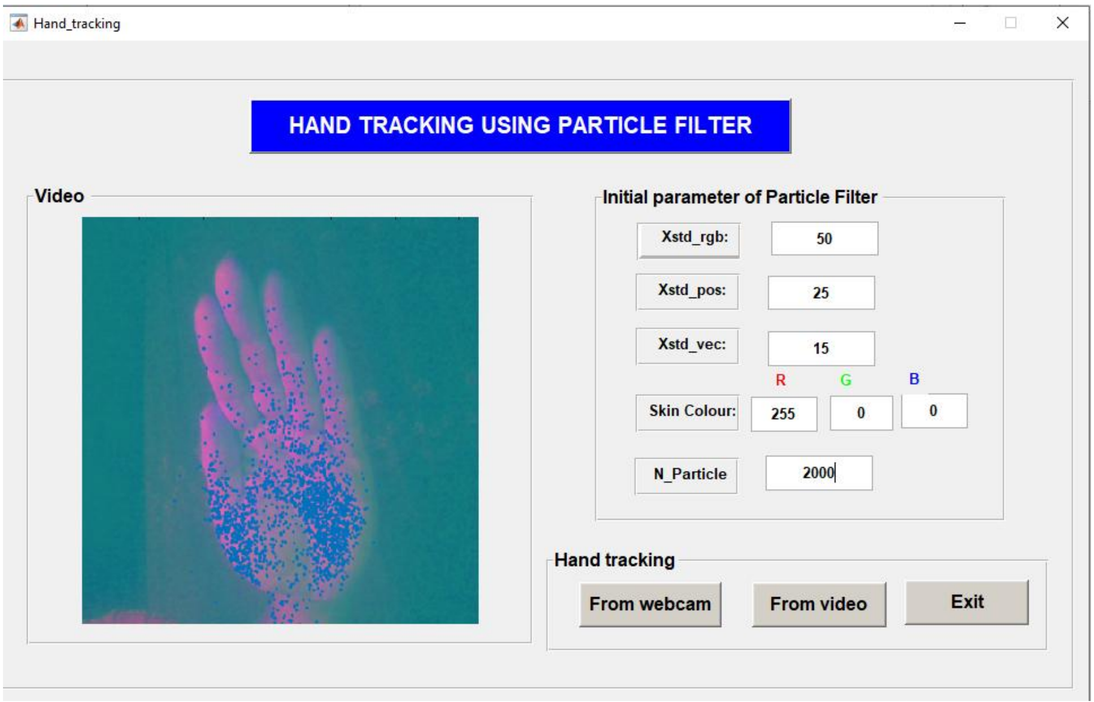

Figure 2. Simulation program interface of Particle Filter algorithm

The weights $w_t^{(i)}$ are then updated in proportion to $p(Z_t | X_t^{(i)})$. The particle position is predicted according to the motion model. The estimate of state of particles is based on a linear extrapolation of the previous state with Gaussian noise. The current state of the target is estimated as a weighted average over the states of the particles. The observation model evaluates this weight by likelihood. In hand tracking, the weight is generally established the appearance of target from the image that is utilized color, edges or texture as a feature.

Figure 2 depicts the interface of a demo program for hand tracking using the Particle Filter algorithm. The parameters that need to be observed are the position and velocity of the hand. This program allows users to set the initial parameters for the Particle Filter algorithm such as: Standard deviation of color (Xstd_rgb), position (Xstd_pos) and velocity (Xstd_vec). Here, Xstd_pos and Xstd_vec indicate how far apart the actual motion and estimation model of the object is. The skin color parameter allows the user to track to hand by color. The final parameter is the number of "particles", also known as the number of samples for estimation.

III. Pre-processing of hand image

The results of the hand motion tracking using the Particle Filter algorithm are great, even when lighting conditions and background images change. This allows us to easily capture the image of our hand for further processing. To improve the quality, this photo needs to be pre-processed and converted to the gray image with the best image condition and fixed size.

In this study, we convert the hand image into a gray image, apply basic noise reduction techniques, and convert the image to a size of 128 × 128 pixels. The simulation results are shown in Figure 3.

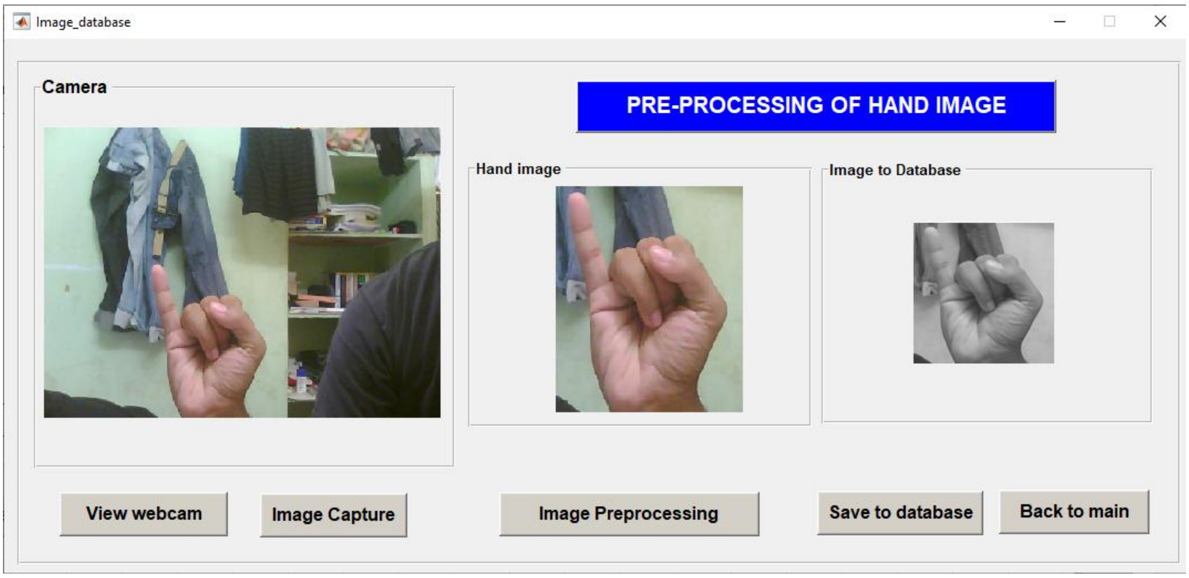

Figure 3. Extracting hand image and pre-processing

IV. Extract feature of hand image using Gabor Wavelet transform

So far, Gabor wavelet has been widely and successfully used in many fields, especially in image recognition. This is due to its ability to capture attributes of selective orientation, spatial localization, and optimization of localization in spatial and frequency domains. For image recognition, we use two-dimensional Gabor wavelet defined in Equation (4)[2],[13]:

$$G_{\mu,v}(Z)=\frac{\|k_{\mu,v}\|^2}{\sigma^2}\exp\left(\frac{-\|k_{\mu,v}\|^2\|Z\|^2}{2\sigma^2}\right)\left[e^{ik_{\mu,v}Z}-e^{-i\sigma^2/2}\right] \tag{4}$$

Where, $Z(x, y)$ represents the point of the fixed position (x, y); μ and v define the orientation Gabor frequency,respectively; σ stand for the standard deviation of the Gaussian; $k_{\mu,v}$ is the central frequency of the filter $k_{\mu,v}=\begin{Bmatrix} K_v\cos\phi_\mu \\ K_v\sin\phi_\mu \end{Bmatrix}$, $K_v=\frac{k_{\max}}{f^v}$, $\phi_\mu=\frac{\mu\pi}{N}$; $\|.\|$ denotes the norm operator; $\frac{\|k_{\mu,v}\|^2}{\sigma^2}$ is used to compensate the weakening of the energy spectrum; $\exp\left(\frac{-\|k_{\mu,v}\|^2\|Z\|^2}{2\sigma^2}\right)$ is the Gaussian envelop function; $e^{ik_{\mu,v}Z}$ is the vibration function, the real part of which is the cosine function and the imaginary part of which is the sine function; $e^{-i\sigma^2/2}$ stands for the DC component. About the parameters of Gabor wavelets in this study, we use five frequencies and eight orientations, $\sigma = 2\pi$, $k_{max} = \pi/2$, $f = \sqrt{2}$.

To the hand image feature extraction process, the Gabor image feature extraction is to conduct the convolution of input hand images and the Gabor wavelet described in Equation (4). We assume that the input image grey scale is $I(x, y)$ and the convolution between I and the Gabor core, $G_{\mu,v}$, is described as follows [13]:

$$O_{\mu,v}(x,y) = I(x,y) * G_{\mu,v}(x,y) \tag{5}$$

where $*$stands for the convolution operator; $O_{\mu,v}(x,y)$ is the convolution image. The hand feature information in different frequencies and the different directions can be obtained through the changes of μ and v parameter.

Figure 4. Interface of the feature vector creation program using Gabor wavelet

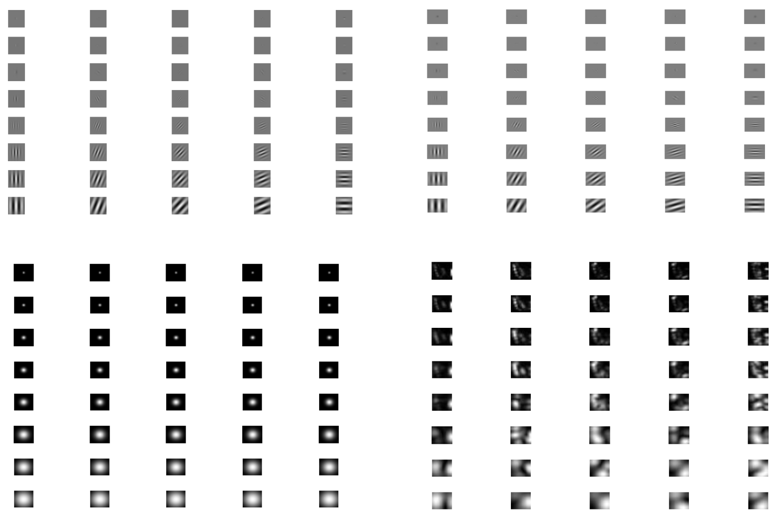

Figure 5. The real, virtual, and amplitude part of the Gabor filter and the featured image

Figure 4 illustrates the interface of a simulation program to characterize hand images using Gabor wavelets. The program allows us to examine the influence of images when changing parameters of Gabor wavelets. Figure 5 illustrates the real part, image part, the amplitude of the Gabor filter, and the output image through the filter when the number of orientations is 8 and the number of frequencies is 5.

V. Dimensional reduction of feature vectors using the PCA algorithm

When a Gabor filter is used to extract image characteristics, its output vector has a large number of dimensions. If the input image is of 128x128 pixels then the feature vector will be 40x 64x64=163840 after the input image is convoluted with the Gabor filter bank of 8 scales and 5 orientations. When this vector is directly applied to the MLP network, the training time for the network will be very long. Besides, in the set of 163840 parameters, there are some parameters whose influence is very small compared to other parameters. Therefore, in order to reduce the number of specific vector dimensions to 1000, we used the PCA method with the implementation steps described in Figure 6 [14].

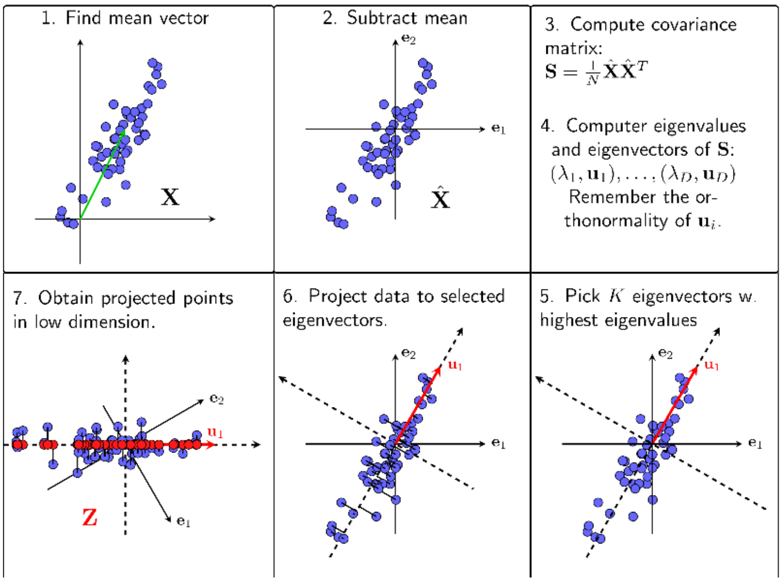

Figure 6. Dimensional reduction of feature vectors using PCA [14]

VI. Hand image recognition using MLP neural network

The classification is done using MLP neural network. The basic architecture of this network is illustrated as in Figure 7. The principles of operation and training rule of MPL neural network are referenced in [13]. In this study, we chose to use the MLP network with a hidden layer. The number of neurons in hidden layers is selected based on trial and error methods to find the optimal network structure. This process can easily be done with the help of commands in the MATLAB toolbox about neural networks [3], [8].

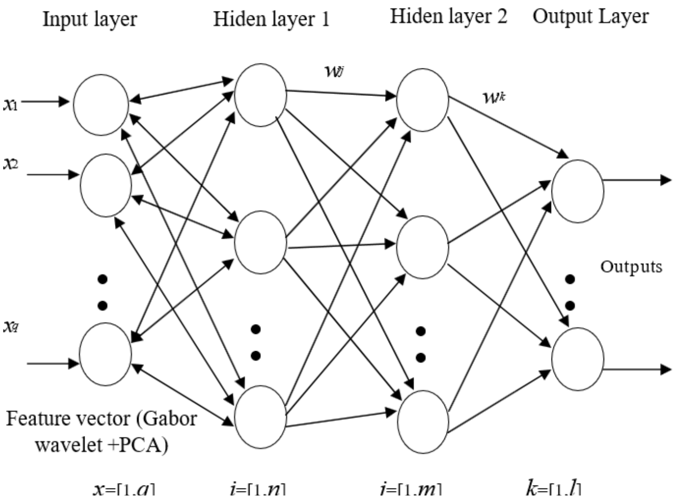

Figure 7. MLP neural network architecture

When developing the test program, we expected to identify 7 hand gestures corresponding to 7 commands (Figure 8). Therefore, to classify seven control commands, we use MLP neural network with the architecture as in Figure 9. The interface for MLP training and operation is illustrated in Figure 10.

After conducting some trial and error tests, we determine the best configuration for the neural network in terms of: the number of neurons in the hidden layer and the maximum number of epochs as follows:

- The number of neurons in the hidden layer = 150.
- Maximum number of iterations (epoch) in the learning process = 10000.
- Activation function is sigmoid function, learning rate is 0.1

The network training will stop when the maximum epoch number reaches 1000 or the average square error reaches a value as small as 0.001.

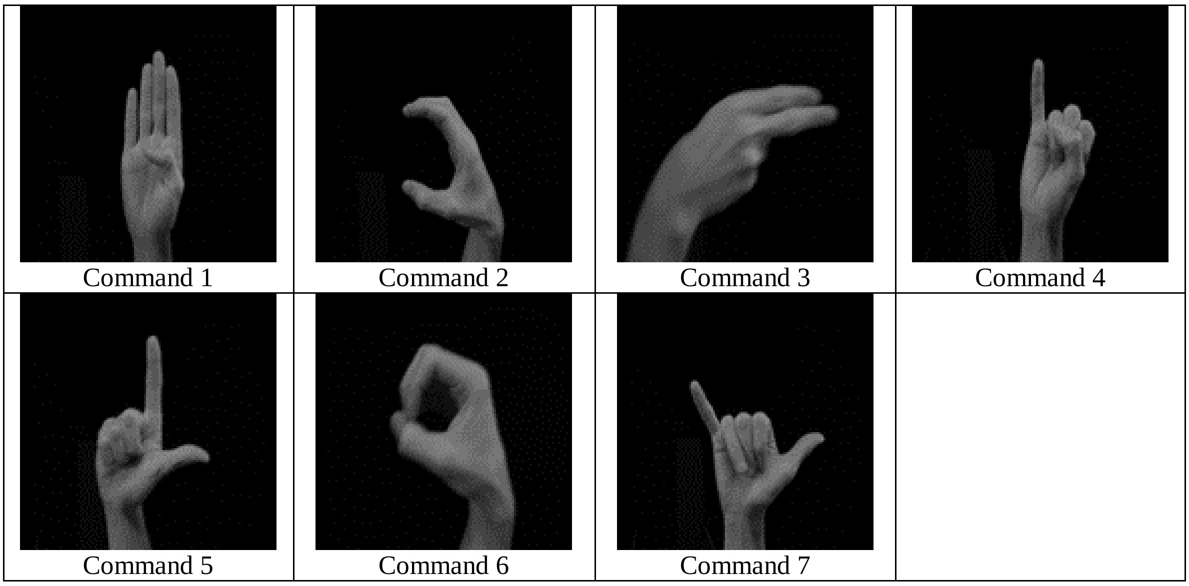

Figure 8. The samples represent computer control commands

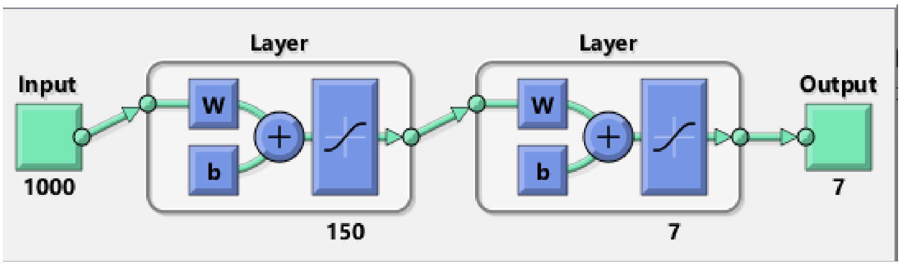

Figure 9. MLP neural network structure for hand image recognition

Each control command (from 1to 7) is practiced 40 times. The recorded data is divided into two groups, 50% for network training and 50% for testing. Thus, each gesture command will have 20 samples for network training and 20 samples are used to check the accuracy of identification.

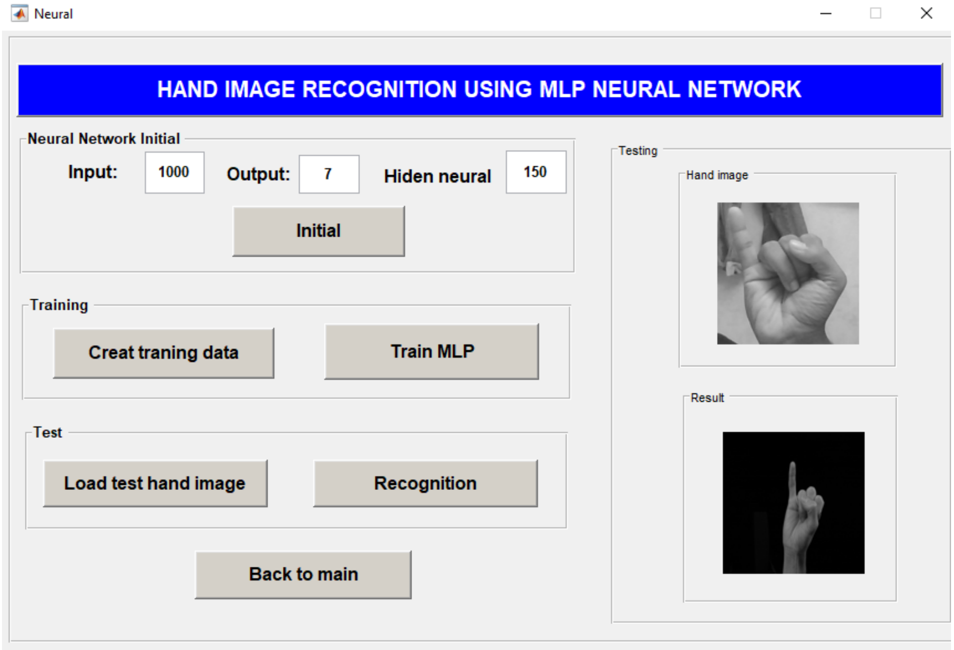

Figure 10. Interface for MLP training and operation

VII. Simulation results

To verify the operation of the vision-based hand gesture recognition method proposed above, we conducted a test simulation program to identify 7 different statements.

Figure 11 depicts the interface of a program that illustrates computer control using gestures. The "View webcam" function collects videos about the operator's operation. The "Hand tracking" function allows tracking of hand movement according to the Particle Filter algorithm. The "Hand image capture" function will perform typical image preprocessing operations and creates feature vectors by combining Gabor wavelets with PCA. The "Recognition" function will take the characteristic vector of the hand image via the MLP network and give the most appropriate conclusion. From this labeled image, the computer will know what command to execute. By training the system with 20 samples and testing with the remaining 20 samples, the results are shown in Table 1. From these results, the system shows a relatively high accuracy.

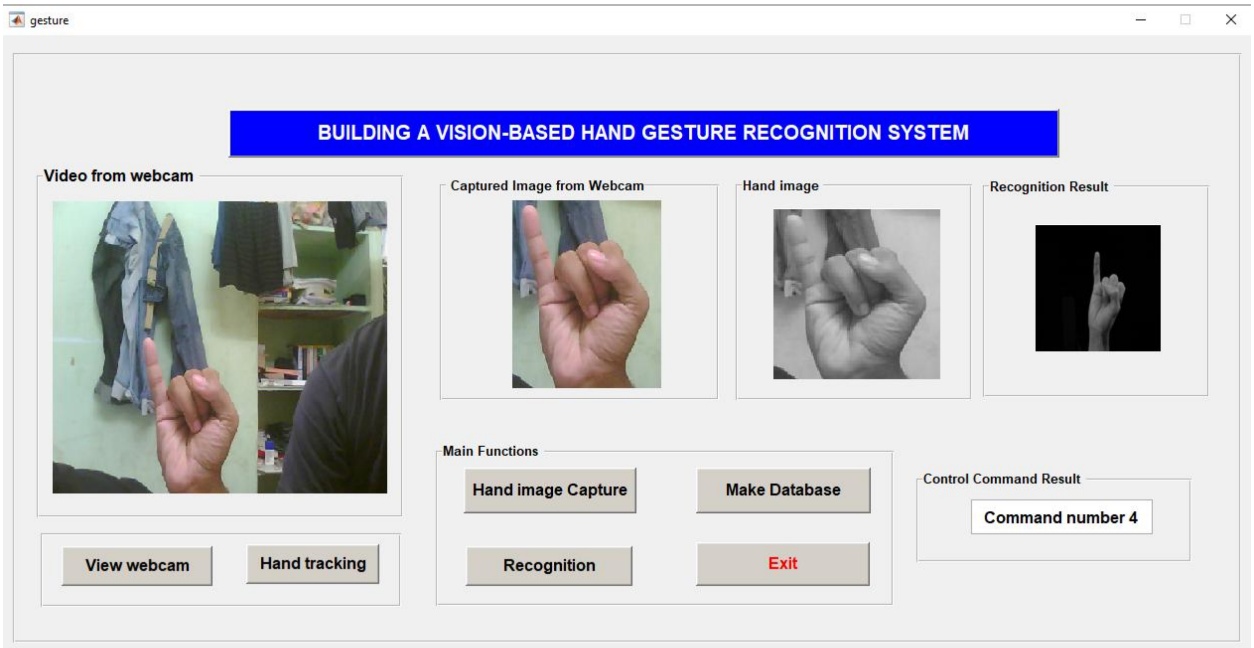

Figure 11. Program interface illustrating computer control with gestures

Table 1. Some test results

Command	**Number of test samples**	**Number of correct identifications**	**Accuracy rate (%)**
Command 1	20	18	90
Command 2	20	16	80
Command 3	20	19	90
Command 4	20	17	85
Command 5	20	17	85
Command 6	20	16	80
Command 7	20	17	85

VIII. Conclusions

It can be seen that the vision-based hand gesture recognition is a very complex problem but it will be solved if we know how to apply appropriate tools such as hand motion monitoring, image processing, feature extraction and classification.

In this paper, through the sections detailed above on theory as well as illustrations, we have proposed a method of constructing a vision-based hand gesture recognition system.The system works with high precision and can be deployed in other practical applications such as controlling devices for smart homes, controlling robots based on gestures.

Recently, there have been many published works around improving the above systems such as improving Particle Filter algorithm to allow hand tracking in large noise, using 2DPCA algorithm instead of PCA, using Deep learning networks instead of traditional MLP networks. This is also the next development direction of this research.

Acknowledgment

This research was supported by the TNUT research grant (under project # T2018-B36). We thank our colleagues from the TNUT who provided insight and expertise that greatly assisted the research.

References

[1]. Badi, H. Recent methods in vision-based hand gesture recognition. Int J Data Sci Anal 1, 77–87 (2016). https://doi.org/10.1007/s41060-016-0008-z.

[2]. Chelali, Fatma. (2014). Face Recognition Using MLP and RBF Neural Network with Gabor and Discrete Wavelet Transform Characterization: A Comparative Study. Mathematical problems in engineering. 2015. 10.1155/2015/523603.

[3]. Ha Nguyen Thanh. "Design of Brainwave Controlled Electrical Wheelchair Using Wavelet Transform and Neural Network". IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE) 14.6 (2019): PP 54-63.

[4]. Jay Prakash, Uma Kant Gautam, Hand Gesture Recognition, International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-7 Issue-6C, April 2019, pp 54-59.

[5]. Lee, Sang-Eun & Horio, Keiichi. (2013). Human Tracking using Particle Filter with Reliable Appearance Model. Proceedings of the SICE Annual Conference, pp. 1418-1424.

[6]. Liu, Bin & Cheng, Shi & Shi, Yuhui. (2016). Particle Filter Optimization: A Brief Introduction. 95-104. 10.1007/978-3-319-41000-5_10.

[7]. Murthy, G.R.S. & Jadon, R.s. (2010). Hand Gesture Recognition using Neural Networks. Advance Computing Conference (IACC), 2010 IEEE 2nd International. 134 - 138. 10.1109/IADCC.2010.5423024.

[8]. Phuong Huy Nguyen, Thu May Duong, Thu Huong Nguyen, Thi Mai Thuong Duong, Combination of Wavelet and MLP Neural Network for Emotion Recognition System, International Journal on Future Revolution in Computer Science & Communication Engineering, Volume: 4 Issue: 11, pp. 105-109, 2018.

[9]. Samantaray, Ashutosh & Nayak, Sanjaya & Mishra, Ashis. (2013). Hand Gesture Recognition using Computer Vision. 4. 1602-1609.

[10]. S. M. Pati, S. B. Kasturiwala, S. O. Dahad, and C. D. Jadhav, (2011)." DaubechiesWavelet Tool: Application for Human Face Recognition", International Journal ofEngineering Science and Technology (IJEST), Vol. 3 No. 3, pp. 2392-2398.

[11]. Yasen, Mais & Jusoh, Shaidah. (2019). A systematic review on hand gesture recognition techniques, challenges and applications. PeerJ Computer Science. 10.7717/peerj-cs.218.

[12]. Z. Xia et al., "Vision-Based Hand Gesture Recognition for Human-Robot Collaboration: A Survey," 2019 5th International Conference on Control, Automation and Robotics (ICCAR), Beijing, China, 2019, pp. 198-205.doi: 10.1109/ICCAR.2019.8813509

[13]. Xu, Yajun & Liang, Fengmei & Zhang, Gang & Xu, Huifang. (2016). Image Intelligent Detection Based on the Gabor Wavelet and the Neural Network. Symmetry. 8. 130. 10.3390/sym8110130.

[14]. https://machinelearningcoban.com/2017/06/15/pca/

Huy Nguyen Phuong, etal. “An Approach in Building a Vision-Based Hand Gesture Recognition System." *IOSR Journal of Computer Engineering (IOSR-JCE)*, 22.1 (2020), pp. 36-43.