Dynamic Gesture Recognition Based on Fuzzy Neural Network Classifier

Ching-Han Chen[1] Nai-Yuan Liu[3]

Department of Computer Science and Information Engineering, National Central University
No.300, Jhongda Rd., Jhongli City, Taiwan

[1 3]pierre@csie.ncu.edu.tw

Kirk Chang[2] Gimmy Su[4]

Delta Electronics, Inc.
No 3, Dongyuan Road, Jhongli City, Taiwan

[2 4]kirk.chang@delta.com.tw

Abstract— **This paper presents a dynamic gesture recognition method based on the combination of the fuzzy features of the dynamic gesture track changes and the fuzzy neural network inference system. This method first classified the dynamic gestures roughly into circular gestures and linear gestures. Further, gestures were classified narrowly into up, down, left, right, clockwise, and counter-clockwise gestures. These six dynamic gestures, which are commonly used in IP-TV controlling, were introduced as the recognition goal in our dynamic gesture recognition system. The results show that this method has a good recognition performance and fault tolerance, and more applicable to real gesture-controlled human-computer interactive environment.**

***Keywords*— gesture recognition, fuzzy system, neural network**

I. INTRODUCTION

Human-computer interaction is a study discussing the interaction between users and computer systems. Through human-computer interaction, computer systems and users are able to communicate with each other. Human-computer interaction makes the communication easier, fits in with users' needs, and enhances the interaction [1]. In recent years, the methods of human-computer interaction constantly improved. "The more intuitive and natural technologies will replace mouses and keyboards," Bill Gates mentioned on BBC News websites [2], which means touching, visual, and voice interfaces will be more and more important. This is so-called the natural user interface. Thus, gesture recognition has become very popular as a human-computer interaction in recent years.

Human hands, through each joint, are able to make various combinations of actions. A wide range of human-computer interactions [1] are developed by gestures. Gestural human-computer interaction interface can basically be divided into static gestures [2,3] and dynamic gestures [4]. Static gesture is a static image with static gesture information. Dynamic gesture is composed of a series of continuous static gesture images. The information includes the changing of the gestures. This paper focuses on the dynamic gesture recognition. Dynamic gesture recognition analyses the locus of hands or the changing information of the hand movements in continuous images. Presently, the main methods of dynamic gesture recognition are Hidden Markov Model (HMM) [5], Dynamic Time Warping [2], and Neural Network [6].

Track-based dynamic gesture [7] is often unable to meet the diversity of the real world if it only applies pre-established gestures database for feature extraction, training and classification. For example, it may cause identify failure by the actual gesture waving in different backgrounds from the gestures in database. Besides, the same gesture made by different people would be different, either. Even the same gesture made by one person at different times would show significant differences.

We propose a fuzzy neural network classifier taking dynamic gesture locus as feature. First, we extract the data of the gesture locus. Then we classify gestures into linear gestures and circular gestures by two fuzzy neural network classifiers. Finally, we precisely classify the six dynamic gestures, which are commonly used in IP-TV controlling, by the changing of the tracks.

II. FEATURE EXTRACTION OF GESTURE LOCUS

First, we take a gesture locus as a virtual ellipse and define three features in continuous hand locus block coordinates, which are the ratios of major and minor axes, difference between major and minor axes, and the frequency difference between clockwise and counter-clockwise.

A. Major axis and minor axis ratio

In a sequence of M continuous images, we record the center coordinates of the hand area in each image. Then, in these M data, we find out the maximum value of x (X_{max}), the maximum value of y (Y_{max}), the minimum value of x (X_{min}), and the minimum value of y (Y_{min}). Furthermore, we define the major axis as a and the minor axis as b of the hand area, as shown in Fig. 1 (a) and make $\frac{a}{b}$ as t_1.

$$\begin{cases} a = X_{max} - X_{min} \\ b = Y_{max} - Y_{min} \end{cases} \quad (1)$$

B. Major axis and minor axis difference

We define t_2 as the difference between major axis and minor axis (a-b).

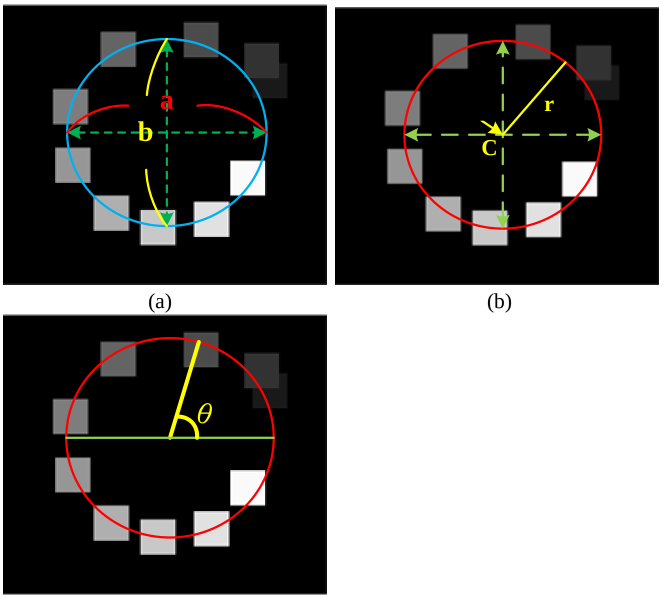
C. Clockwise and counter-clockwise frequency

After Xmax, Ymax, Xmin, Ymin are found, we calculate the intersects between a and b as the center C of the virtual circle. And take $r = \frac{a+b}{2}$ as the radius of the virtual circle, shown in Fig. 1 (b). Then we calculate the angle θ between the horizontal line across C and the line from C to each center of the hand area, as shown in Fig. 1 (c).

We can decide the gesture is clockwise or counter-clockwise by M different angels (θ). And record the numbers of times that clockwise (f_{CW}) and counter-clockwise (f_{CCW}) occur. We make the clockwise and counter-clockwise frequency t_3 as $f_{CW} - f_{CCW}$.

 ISBN: 978-1-61208-250-9

However, the three characteristics of the continuous gesture locus, t_1, t_2, and t_3, are not well-defined, susceptible to interference. Thus, we fuzzify the three characteristics and put them into the Fuzzy Neural Network to classify.



(a) (b)

(c)

Fig. 1 Definition of the three characteristic parameters of the gesture locus (a)major and minor ratio (b)major and minor difference (c)clockwise/Counter-clockwise frequency

III. Fuzzy Neural Network Classifier

Fuzzy Neural Network (FNN) combines the concept of fuzzy theory and neural network. In recent years, its applications are proposed continuously [8,9]. In general, the neural network and fuzzy theory both are able to estimate the system without mathematical models. They are used to simulate the function of the human brain. The neural network imitates the message passing mechanism in brain cells and the fuzzy system simulates the mental status and psychological reasoning of human.

Basically, in order to achieve the purpose of learning, a function estimation of the neural network is basic on the training data input, output values and the connecting bonding parameters between neurons which are adjusted through repeated error corrections. It is worth noting that there is no way to know the intranet architecture. We cannot directly encode the normal if-then rules in the network. The only method is giving a large number of training data to the system. Compare with the artificial neural networks, the fuzzy systems can directly encode the expert knowledge values in the system, with high tolerance. Thus, through the complementary nature of the neural network and fuzzy system, this combination system has the advantages of both. [10].

A. *Fuzzy System*

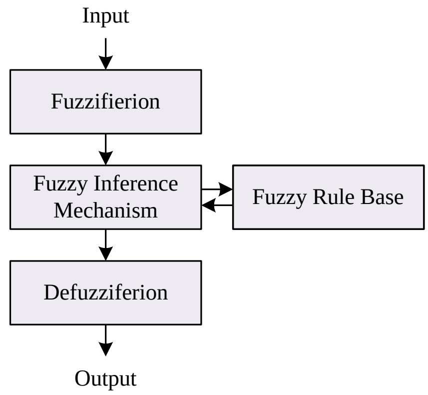


Fig. 2 The basic architecture of the Fuzzy System

Fuzzifierion is to translate the precise input data into a set of syntax fuzzy information. Fuzzy sets are composed of a number of membership functions. In this study, we use the trigonometric functions as membership functions.

The fuzzy inference mechanism is the core of the fuzzy system. It simulates the decision model of human thinking by processing the approximate reasoning or fuzzy inference. The classifier has two inputs, and the fuzzy rules expressed as follows:

R^j:

$$\begin{aligned}&If\ x_1\ is\ A_1^j\ and\ x_2\ is\ A_2^j\\ &then\ X\ belong\ class\ \ y_j\ with\ CF = CF_j \qquad (2)\end{aligned}$$

R^j represents the j-th fuzzy rule. A_1^j and A_2^j represent the fuzzy sets where j = 1 to n with n rules. $y_j \in y\{1,2,\dots,M\}$ is the output of the j-th fuzzy rule, which is one for the M classes, CF_j represents the reliability of the fuzzy rule R^j. We apply the Generalized Modus Ponens and Max-Min composition operation, and the fuzzy inference outputs expressed as follows:

$$\mu_y(y) = \max_j \{\mu_{y,j}\} \mid y_j = y \qquad (3)$$

where,

$$\mu_{y,j} = \mu m_{A,j}(x) \cdot CF_j \qquad (4)$$

$$\mu m_{A,j}(x) = \min\left\{\mu_{A_1^j}(x_1), \mu_{A_2^j}(x_2)\right\} \qquad (5)$$

The process of converting the fuzzified value into a specific value called defuzzifierion. Formally, the outputs from the fuzzy inference may be the fuzzy sets or specific values. If the result after inference is a fuzzy set, the median method and the center area method are applied to obtain specific outputs. In this study, a single neuron is used to defuzzify, which means to connect the output value u_j of each fuzzy rule directly to the neuron and output the defuzzified result by the neuron.

B. *The Single Neuron*

In Artificial Neural Network (ANN), neurons can accept the input signal from linked unit and calculate the input value with the bond value. Then it decide whether to pass the signal down to the neuron in the next layer by checking if the calculated value is greater than a threshold. Back Propagation Network (BPN) can make the output values and the training data expectations as feedback to the network. So the network is able to adjust the bond values by the excitation from

 ISBN: 978-1-61208-250-9

external environment. This kind of learning ability can be used to generalize the fuzzy rules in the fuzzy neural network.

Considering the demand for memory and computation time in a real-time system, we decide to use single neuron. Although it is impossible to have all the ability to map nonlinear function, this can be resolved after the operation of the front-end fuzzy system. The single neuron accepts *p* external inputs, calculated with *p* bond values, and input the result to the transfer function $y = \varphi(v)$. Then it outputs the classification inference probability. The formula is as follows:

$$v = \sum_{i=0}^{p} w_i u_i \tag{6}$$

$$y = \varphi(v) = \frac{\alpha_2}{1 + e^{-\alpha_1 v}} \tag{7}$$

Where α_1 controls the shape of the function, α_2 adjusts the size of scale. Fig. 3 shows the combination of the fuzzy system and the single neuron.

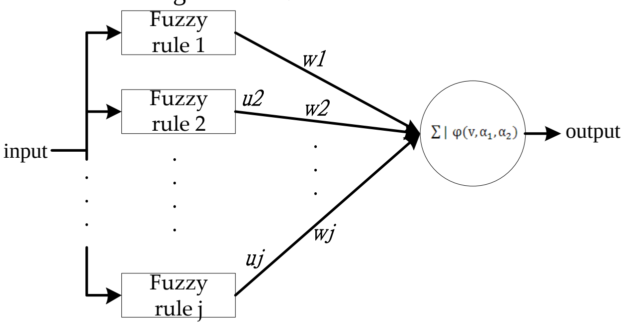


Fig. 3 The artificial neural network classifier combined with single neuron

IV. Dynamic Gesture Recognition

We use the six dynamic gestures, which are considered as the most common using gestures in IP-TV controlling interface, as the recognition goals of the classifier,

- Gesture waving to the left
- Gesture waving to the right
- Gesture waving up
- Gesture waving down
- Gesture waving clockwise
- Gesture waving counter-clockwise

Based on the fuzzy neural network classifier (Fig. 3), we establish two fuzzy neural network classifiers, the circular gesture classifier and the linear gestures classifier. The input features of the linear gesture classifier are the ration and the difference of the major axis and the minor axis. The input features of the circular gesture classifier are the clockwise/counter-clockwise frequency and the ratio of the major and minor axes. In fuzzy neural network, every inference output from the fuzzy rules is linked to a neuron.

It fuzzifies the input feature values by the fuzzification model. Then process the fuzzy inference by the fuzzy rules. Finally, output the classified results computed by the neuron. Fig. 4 shows the two fuzzy neural classifier architectures. O_1 is the inference probability of the linear gestures; O_2 is the inference probability of the circular gesture.

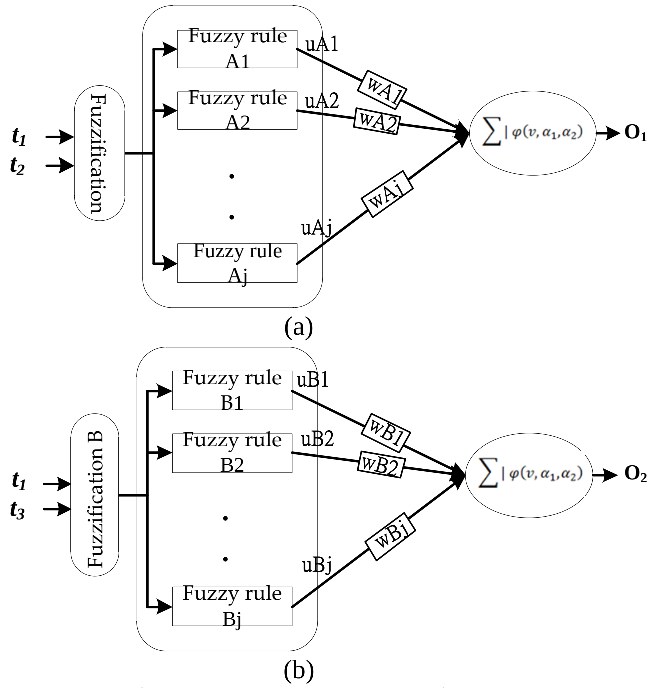


Fig. 4 The two fuzzy neural network gesture classifiers (a)linear gesture classifier (b)circular gesture classifier

The fuzzy neural network classifier can roughly determine the dynamic gestures to be either linear gestures or circular gesture. Further, in accordance with the positional relationship of the gesture trajectory coordinates, it determines the dynamic gesture to be up, down, left, right, clockwise, or counter-clockwise.

It determines the relative position in x-direction and y-direction of two adjacent hand areas of the *M* hand area coordinates, which means to make the determination by the *j*-th point coordinates (x_i, y_i) and the (*j*+1)-th point coordinates (x_{i+1}, y_{i+1}), where $y = 0,1,\dots,M-1$. The thing need to be noticed is that the waving gestures taken by the camera is left-right reversal to the direction of the user's real waving direction. We take the user's direction as determination and the formula is as following:

$$\begin{cases} C_{left} = C_{left} + 1 & ,if\ x_{i+1} > x_i \\ C_{right} = C_{right} + 1 & ,if\ x_i > x_{i+1} \\ C_{up} = C_{up} + 1 & ,if\ y_i > y_{i+1} \\ C_{down} = C_{down} + 1 & ,if\ y_{i+1} > y_i \end{cases} \tag{8}$$

$$C_{sum} = C_{left} + C_{right} + C_{up} + C_{down} \tag{9}$$

Where, C_{left}, C_{right}, C_{up}, C_{down}, are the times of waving in different directions.

Combining with the times of waving clockwise $f_{CW}(C_{CW})$ and counter-clockwise $f_{CCW}(C_{CCW})$, we define the probabilities of the six dynamic gestures and the formula is as following:

 ISBN: 978-1-61208-250-9

$$\begin{cases} p_1 = \dfrac{C_{left}}{C_{sum}} \times O1 \\ p_2 = \dfrac{C_{right}}{C_{sum}} \times O1 \\ p_3 = \dfrac{C_{up}}{C_{sum}} \times O1 \\ p_4 = \dfrac{C_{down}}{C_{sum}} \times O1 \\ p_5 = \dfrac{C_{CW}}{C_{CW} + C_{CCW}} \times O2 \\ p_6 = \dfrac{C_{CCW}}{C_{CW} + C_{CCW}} \times O2 \end{cases} \quad (10)$$

Where, $p_{1,\ldots,}p_6$ are the probabilities of different dynamic gestures. O_1 and O_2 are the outputs of the linear classifier and the circular classifier. The flow chart of the process recognizing is as following:

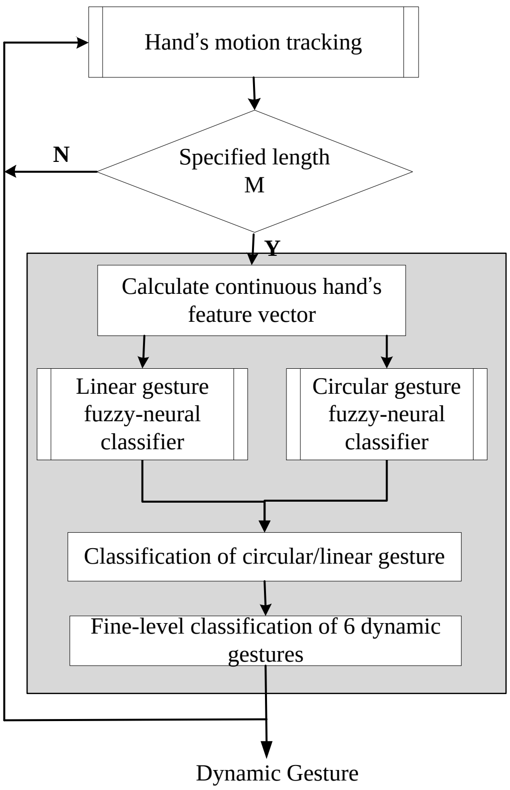


Fig. 5 The flow chart shows the process of recognizing the dynamic gestures by the fuzzy neural network

V. Experiment

We create a set of continuous images of dynamic gestures in a database. 10 different people were recorded under the six dynamic gestures, each gesture 10 times, a total of 600 gestures and each gesture has 100 samples. The image sequence length to recognize the dynamic gestures is 10 frames.

TABLE I
THE RESULTS OF DYNAMIC GESTURE RECOGNITION

Gesture	Recognizable Probability
Left	92%
Right	95%
Up	89%
Down	79%
Clockwise	96%
C-Clockwise	95%
Average	**91%**

The experimental results show that the up and down gesture recognizable probability is lower. Thus, we analyze the gestures to be mistaken as what kinds of gesture. The experimental results are shown in Table 2.

TABLE III
ERRONEOUS GESTURE RECOGNITION ANALYSIS

Gesture	Recognized as		
Up	Clockwise	Counter-Clockwise	Down
	3%	2%	6%
Down	Clockwise	Counter-Clockwise	Up
	6%	6%	9%

We observed that the experimenter's waving habits are related to the erroneous recognition. For example, when some people wave upward, their hands fall naturally after they finish the waving action. However, the up-down waving is recorded and recognized. So does the down gesture. The down-up gesture is recorded, too. And the erroneous recognition occurs.

Shan et al. [11] applied a MHI conversion to the tracking results by Mean Shift Embedded Particle Filter (MSEPF), which has the advantages of both Particle Filtering and Mean Shift. Then, it extracts features by using the seven constant torque variables that Hu proposed. Further, the template matching method proposed by Bobick & Davis [12] was applied and Mahalanbis Distance was taken as matching basis. We experiment with the same database and compare the identification performance between this study and Shan algorithm. The results are shown in Table 3. The recognition rate is higher than the method Shan et al. [11] proposed, although the average recognition time of this algorithm is longer.

TABLE IIIII
THE COMPARISON WITH OTHER METHOD

Algorithm	Recognition Probability	Average Recognition Time (msec)
This Study	94%	4.54
Shan[11]	56%	2.57

VI. Conclusions

In this paper, we proposed a dynamic gesture recognition method based on fuzzy features, fuzzy neural network, and fuzzy inference system. First, it roughly classifies the dynamic gestures to circular gestures and linear gestures. Further, it sub-classifies the gestures to up, down, left, right, clockwise, and counter-clockwise. The experimental results show that this method has good recognition performance, makes the system to achieve better fault tolerance, and is more applicable to real gesture-controlled human-computer interactive environment.

ACKNOWLEDGMENT

The authors are grateful for the supporting by the National Science Council of the Republic of China [grant numbers NSC 101-2220-E-008 -002] and Funding from the Joint Research Center of National Central University and Delta Electronics Inc., under the project NCU-DEL-101-A-08.

REFERENCES

[1] C. Manresa, J. Varona, R. Mas and F. Perales, 2005, "Hand tracking and gesture recognition for human-computer interaction", ELCVIA, Volume 5, No. 3, p.96-104

[2] C. Manresa, J. Varona, R. Mas and F. Perales, 2005, "Hand tracking and gesture recognition for human-computer interaction", ELCVIA, Volume 5, No. 3, p.96-104

[3] W. T. Freeman and M. Roth, 1995, " Orientation Histograms for Hand Gesture Recognition", IEEE Intl. Wkshp. on Automatic Face and Gesture Recognition, Volume 50, Issue 2, pp.174.

[4] K. J. Chang, 2005, "Computer Vision Based Hand Gesture Recognition System",Master Thesis, National Tsing Hua University, Department of Electrical Engineering, page 21-32 .

[5] L.K. Chang, 2002, "Hand Gesture Recognition Based on Hausdorff Distance", Journal of Image and Graphics, Volume 7, No.11.

[6] M. Elmezain, A. Al-Hamadi, B. Michaelis, 2009, "Hand Locus-based Gesture Spotting and Recognition Using HMM", ICIP 16th IEEE International Conference , pp. 3577-3580.

[7] K. Murakami and H. Taguchi, 1991, "Gesture Recognition using Recurrent Neural Networks", Proc. of the SIGCHI conference on Human factors in computing systems, pp.237-242.

[8] Kosko, Bart (1992). Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence. Englewood Cliffs, NJ: Prentice Hall. ISBN 0-13-611435-0.

[9] F. J. Lin, W. J. Hwang, and R. J. Wai, 1999, "A supervisory fuzzy neural network control system for tracking periodic inputs," IEEE Trans. Fuzzy Systems, Volume 7, No.1, pp. 41-52.

[10] Y. C. Chen and C. C. Teng, 1995, "A model reference control structure using a fuzzy neural network," Fuzzy Sets and Systems, Volume 73, pp.291-312.

[11] C. Shan, T. Tan,Y. Wei, 2007, "Real-time hand tracking using a mean shift embedded particle filter", Pattern Recognition, Volume 40, Issue 7, pp.1958-1970.

[12] A. Bobick, J. Davis, 2001, "The recognition of human movement using temporal templates", IEEE Trans. Pattern Anal. Mach. Intell., Volume 23, No.3, pp.257-267.