A 3D Hand-drawn Gesture Input Device Using Fuzzy ARTMAP-based Recognizer[1]

Jing YANG
School of Automation Science and Electrical Engineering, Beihang University,
Beijing, 100083, CHINA

and

Won-Chul BANG, Eun-Seok CHOI, Sung-Jung CHO, Jong-Koo OH, Joon-Kee CHO, Sang-Ryong KIM
Advanced Systems Research Lab, Samsung Advanced Institute of Technology,
San 14-1, Nongseo-Dong, Giheung-Gu, Yongin-Si, Gyeonggi-Do, 449-712, SOUTH KOREA

and

Eun-Kwang KI
New Solution Team, LG Electronics,
SOUTH KOREA

and

Dong-Yoon KIM
Kionix Inc.,
36 Thornwood Drive, Ithaca, NY 14850, USA

ABSTRACT

In this paper, a novel input device based on 3D dynamic hand-drawn gestures is presented. It makes use of inertial sensor and pattern recognition technique. Fuzzy ARTMAP based recognizer is adopted to realize gesture recognition by using 3-axis acceleration signals directly instead of reproduced trajectories of gestures. The proposed method may relax motion constraints during inputting a gesture, which is more convenient for user. This prototype of input device has been implemented on a remote controller to manipulate TVs. The recognition rate of 20 gestures is higher than 97%. It clearly shows the effectiveness and feasibility of the proposed input device. As a result, it is a powerful, flexible interface for modern electronic products.

Keywords: Gesture Recognition, Neural Network, Feature extraction.

1. INTRODUCTION

A natural, efficient, powerful and flexible interface is indispensable for human interaction of electronic products. Desired interfaces facilitate a richer variety of communications capabilities between humans and electronic products. Since hand gestures are commonly used in daily life, the hand gesture-based input device is a competitive candidate due to its flexibility and naturalness.

Many different techniques are employed to implement this concept, such as vision-based gesture recognition [1]-[3]. This technique is adaptable to recognize both static and dynamic gestures. However, it is non-effective when the line of sight is obstructed. Furthermore, it is unsuitable for mobile devices because of special requirements of equipments. Instead, acceleration-based inertial sensing technique is a better choice for sensing handwritten gesture in 3D space [4-8] because of its autonomy and mobility.

While gesture-based input device is presented in [6-8], dynamic gestures in 3D space are recognized by using reproduced trajectories. The accuracy of reproduced trajectories has great influence on the recognition rate. Usually some constraint conditions are required during inputting a gesture, for error compensation to improve the accuracy of trajectory estimation. It is unnatural and inconvenient for user manipulation. In order to relax motion constraints of data capturing procedure, in this paper raw data from accelerometers are directly used to recognize the gesture.

Fuzzy ARTMAP (Fuzzy Adaptive Resonance Theory-supervised predictive MAPping), which satisfies most properties of a successful pattern classifier[9][10], is a kind of neural network architecture that automatically selects complex combinations of factors to build accurate prediction for applications such as handwritten character recognition [11] and fault detection [12]. The ARTMAP family members have

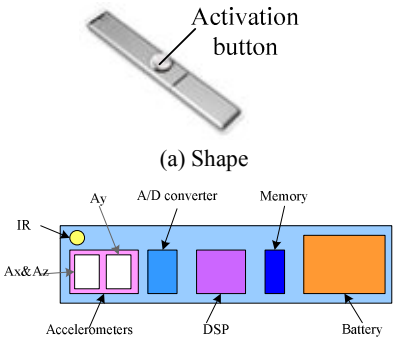
[1] This work was done while all authors were research staff members in Samsung Advanced Institute of Technology.

advantages over many other NN models and are especially suited to classification problems. One advantage is that it is faster than other neural networks due to the small number of training epochs required by the networks to "learn" the input data. Also, its classification results are easily interpretable [13].

The rest of this paper is organized as follows. Section 2 introduces the overview of the proposed input device. Section 3 introduces the algorithm of Fuzzy ARTMAP-based gesture recognition. Section 4 presents current experimental results of a remote controller for TVs which is designed according to the proposed technique. At last, conclusions of the proposed device are given.

2. DEVICE OVERVIEW

The design of the proposed input device is shown in Figure 1. To use this device, the user needs only to press the activation button, which is shown in Figure 1 (a), and draw gestures in 3D space. Different commands are represented by different gestures.



(a) Shape

(b) Structure

Figure 1. Configuration of the gesture-based input device

In Figure1 (b), 3-axis acceleration of user's gesture is sensed by accelerometer module. After through A/D converter, the collected signal is sent to DSP module where gesture recognition is fulfilled. At last, this input device communicates with other controlled device by IR (Infrared LED) module to emit commands. In addition, this device includes memory module and battery module.

3. GESTURE RECOGNITION METHOD

Previous method

Figure 2 shows the main idea of our previous research [8]. Firstly, the proposed input device is to realize the customer's commands by inputting a 3D hand gestures. Accelerometers are installed on the proposed device to sense 3-axis accelerations of hand motions. By using these acceleration signals, trajectories of the inputting gestures are estimated by using inertial navigation technology. After that, *Bayesian Network* based recognizer realizes gesture recognition and determine the input intention of users by using reproduced trajectories.

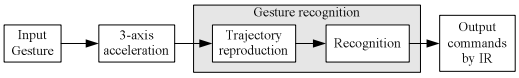


Figure 2. Previous realization of a gesture-based input device

While inertial navigation technology is adopted for trajectory estimation, accumulated error is unavoidable because of using integrations twice. In order to compensate these errors, users are required to pause a while before and after inputting a gesture to obtain reference information. It is very inconvenient and unnatural for users. Hence, in this paper raw data of accelerations are directly utilized for gesture recognition.

The difficulties to realize raw data based recognition are:

1) The gravity effect on each axis is quite different because of user's different attitude to hold the device, since the accelerometer employed on each axis senses both gravity component and user action simultaneously.
2) The gathered information is formed by the series of the sampling time and integrated accelerations along 3 axes of device only. Because the initial velocity is unknown, and there is no enough information about the attitudes to hold the device during inputting a gesture, it's difficult to estimate the instantaneous velocity and the whole trajectory.
3) Since the requirements of pauses before and after motions are not necessary in the proposed method in this paper. Such relaxation of data capturing constraints makes users more convenient. However, since it is not easy to make sure that the data gathering duration equals to the real motion period, part of date would be lost or additional data could be added. The case of losing data is more serious.

In the following, Fuzzy ARTMAP based recognizer is developed and tested. This method attempts to conquer the above-mentioned difficulties by using good robustness of Fuzzy ARTMAP.

Overview of proposed method

Figure 3 shows the flow chart of the proposed gesture recognition method based on Fuzzy ARTMAP. It includes two stages, feature extraction and recognition.

In the feature extraction stage, the raw date is processed to extract discriminative features as the first step. Then the extracted feature vectors are normalized for further processing.

In the recognition stage, the normalized feature vectors are sent to Fuzzy ARTMAP based classifier and mapped to the

corresponding gesture label. The detailed description is given in the following sections.

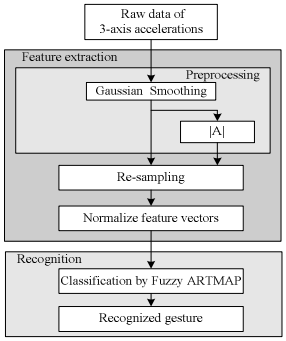


Figure 3. Flow chart of proposed gesture recognition method

Fuzzy ARTMAP

Fuzzy ARTMAP (FAM) is a kind of network architecture for supervised learning of recognition categories and multi-dimensional maps in response to arbitrary sequences of analog and binary input vector.

Figure 4 shows the architecture of the FAM network. FAM includes the map field and a pair of Fuzzy ART module, ART_a and ART_b. ART modules create stable recognition categories in response to arbitrary sequences of input patterns $\{\mathbf{a}, \mathbf{b}\}$. Then two models are interconnected by the map field. Each ART module includes three layers, input layer F_0, match layer F_1 and selection layer F_2. Take ART_a as an example, it transform the M_a vector $\mathbf{a}$ into the $2M_a$ vector $\mathbf{I}^a = (\mathbf{a}, \mathbf{a}^c)$ at F_0^a by complement coding. The constructed clustering scheme of ART_a, aggregating similar patterns in the same cluster represented by a node in F_2^a, has a prototype (a weight vector $\mathbf{w}_j^a$ connecting the node j in F_2^a to all nodes in F_1^a). The map filed, has only layer F^{ab}, links each category formed in each ART module consistently by a weight matrix $\mathbf{w}^{ab}$.

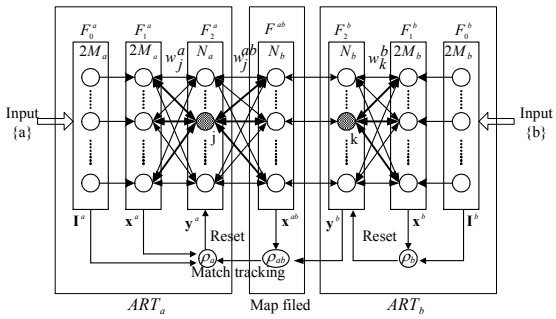


Figure 4. A schematic of the FAM network

The supervised learning algorithm is employed in this paper to train a FAM for the application of gesture recognition. During supervised learning period, FAM receives a stream $\{\mathbf{a}, \mathbf{b}\}$ of input patterns, where $\mathbf{b}$ is the desired prediction of the given $\mathbf{a}$. According to fuzzy adaptive resonance theory, ART_a and ART_b classify $\mathbf{a}$ and $\mathbf{b}$ into categories respectively. Then the map filed forms the predictive associations between categories of ART_a and ART_b by using match tracking rule. Detailed operation process of FAM was given in [9][10][11].

A FAM can be viewed as a propositional inductive learning system. For each sample, the most similar rule encode in the FAM and satisfying the vigilance criterion is chosen to generalize with this sample, if there is no rule satisfying the criterion, a new one is created from this sample. The important parameters have great influence on the FAM dynamics are: (1) the choice parameter $\alpha > 0$: the small value of α tends to minimize recoding during learning; it should be large enough to affect the values of the choice function. (2) the vigilance parameter $\rho \in [0, 1]$: The high vigilance leads to small hyperrectangle while low vigilance permits large hyperrectangle. That is, the number of categories is more while the vigilance value is higher.

As a classifier, FAM has the following features: (1) to learn and associate multiple categories with the same output, that is, multiple disconnected clusters of features with the same output classification; (2) to dynamically self-adjust its size depending on the complexity of the input. In other words, the number of recognition codes in the network will vary dynamically on the presentation of new input to the network; (3) to realize one-shot stable learning while in fast learning mode; (4) to reduce the risk of a dead cluster forming by initialization of new recognition patterns which is dependent on the input vector.

Preprocessing

Preprocessing of raw data is a preparation for achieving high quality data of gesture recognition. It includes the following operations:

Gaussian smoothing: As mentioned in section 2, collected raw data are time series signals of 3-axis accelerations. The unavoidable measure noise and unexpected user's hand trembling will influence the real acceleration signals. Hence it is necessary to get rid of such kind of disturbance which would have bad influence on feature vector generation.

One dimensional Gaussian smoothing is adopted for each axis of the sensor signals, since the noise distribution is unknown. Then assuming Gaussian kernel with zero-mean, the convolution weight for measures is given by

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-x^2}{2\sigma^2}} \quad (1)$$

where σ is the standard deviation.

Suppose that the raw data are time series based 3-axis accelerations signals given as follows:

$$\mathbf{A}^b(t_k) = [a_x^b(t_k), a_y^b(t_k), a_z^b(t_k)] \quad (\ k = 1,\ldots,\text{nsample}\ ) \qquad (2)$$

where *nsample* is the number of sampling data for a gesture. Then data after Gaussian smoothing are:

$$\mathbf{A}^f(t_k) = [a_x^f(t_k), a_y^f(t_k), a_z^f(t_k)] \quad (\ k = 1,\ldots,\text{nsample}\ ) \qquad (3)$$

After that, the corresponding norm of the 3-axis acceleration is

$$|\mathbf{A}^f(t_k)| = \sqrt{(a_x^f(t_k))^2 + (a_y^f(t_k))^2 + (a_z^f(t_k))^2} \qquad (4)$$

This reflects the amplitude of accelerations. Now available physical quantities are $|\mathbf{A}^f(t_k)|$, $a_x^f(t_k)$, $a_y^f(t_k)$ and $a_z^f(t_k)$.

Re-sampling (Segmentation): In this part, the time-series type of vectors with different length is transformed into fixed high-dimensional vectors by means of data re-sampling. It is implemented by picking out fixed number of re-sampling data from all samples by using linear interpolation with equal Euclid distance in the measurement space. Consequently, paused points which have no explicit information value for recognition are ignored and eliminated automatically. Only the core information contents reflect typical motion feature are remained.

Normalization: In order to regularize the feature scales and make the feature more distinctive, normalization is employed to standardize all data with the same scale or metric. Another reason is that FAM requires the inputting components are in the range of [0, 1].

Each physical quantity re-sampled is normalized by the following steps. Firstly, the smallest value is subtracted from each physical quantity. After that, the new generated physical quantities are rescaled by the biggest values. That is

$$\tilde{a}_j^f(t_m) = \frac{a_j^f(t_m) - \min(a_j^f)}{\max(a_j^f) - \min(a_j^f)} \quad (\text{j} = \text{x, y, z, } |\bullet|) \qquad (5)$$

After normalization, the feature vector is obtained.

4. EXPERIMENT RESULT

Experiment background

In this section, **MagicWand** project, which is a good application of the proposed gesture-based input device, is presented. It implements a remote controller for TV shown in Figure 5. Compared with the conventional remote controller which implements control commands by complicated layout of many buttons, the proposed system with only one button is attractively simple and slim. Obviously, it can be used as an input device for other home appliances and mobile electronic products also.

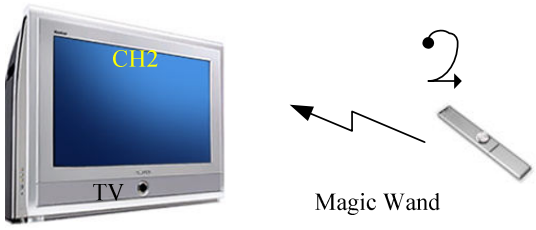


Figure 5. Remote controller for TV

Data set

Currently, 20 gestures, which are shown in Figure 6, were designed for the remote controller to manipulate TV functions.

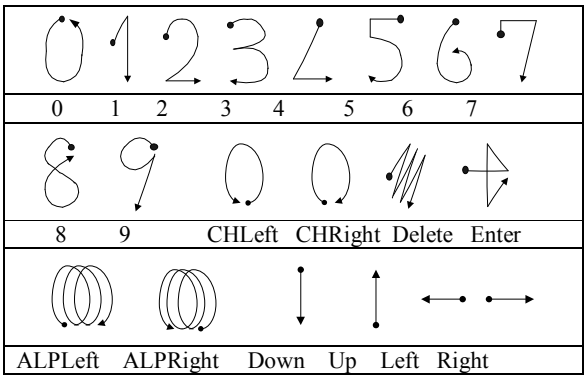


Figure 6. Gestures for TV remote controller

In the database, there are 7527 of gesture samples. All these data were collected from 45 persons by using the input device described in the section 2. The testers include both experienced users and first-time users. Each person wrote above-mentioned 20 gestures many times. Finally, there were more than 300 samples for each gesture. The sampling frequency was 28Hz.

Recognition result and discussion

Raw data collected from accelerometer were processed by the proposed gesture recognition algorithm and mapped to the corresponding input command directly. In the following, the recognition results after the generalization test will be listed to analyze the influence of different parameters of recognizer.

Figure 7 and Figure 8 show data difference of a gesture "8" before and after Gaussian smoothing. It is obviously that noise on raw signals has been removed, signal curves become very smooth. Gaussian Smoothing is suitable for deal with noise with unknown distribution.

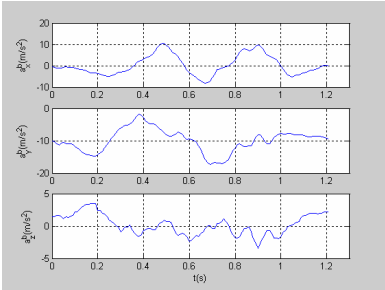


Figure 7. Raw data of accelerations

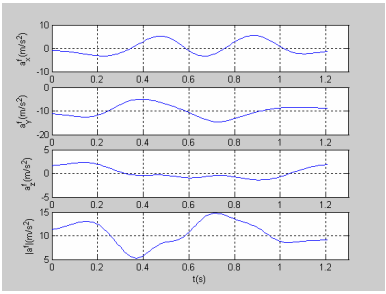


Figure 8. Accelerations and its norm after Gaussian smoothing

Figure 9 shows the process of re-sampling. The points which have explicit valuable information for recognition are kept. They reflect typical motion feature. On the other side, the points with no explicit motion information for recognition are ignored and eliminated. For example, at the beginning part of the gesture, there are many data before re-sampling, but they are very close to each other in the measurement space. After re-sampling, only the representative point "1" is pick out for building a feature vector.

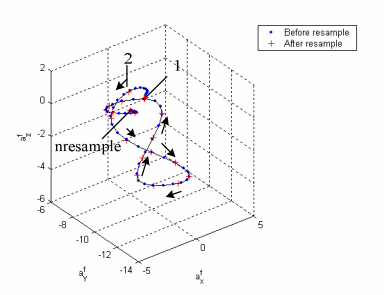


Figure 9. Re-sampling process

Table 1 lists recognition rates obtained under the condition of different standard derivation and different number of re-sampling points. It was shown that the best recognition rate was achieved while the standard deviation of Gaussian smoothing is $\sigma = 2.5$ and the re-sampling point number is 20.

Table 1. The recognition rate under different parameters for Guassian smoothing and re-sampling ($\rho_a = 0.60$)

Unit: (%)

σ / *nresample*	1.0	1.5	2.0	2.5	3.0	Average
15	96.00	96.68	96.48	96.86	95.89	96.38
20	96.28	96.43	96.90	97.00	96.32	96.59
25	96.48	96.57	96.76	96.51	96.47	96.55
Average	96.25	96.56	96.71	96.79	96.23	

Figure 10 shows feature vectors of several gestures "8" obtained after preprocessing. For each point, there are four physical quantities obtained from the norm and three components of accelerations ($|\mathbf{a}|, a_x^b, a_y^b, a_z^b$) after rescaled into the range of $[0,1]$. Consequently, 80-dimension feature vectors are generated. It is obvious that extracted features of gesture "8" are very similar. Then, gestures with similar features can be recognized.

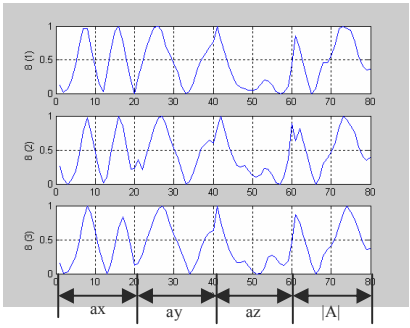


Figure 10. Feature vectors after re-sampling and rescaling

Table 2. The recognition rate under different vigilance paramenters of FAM ($\sigma = 2.5$).

ρ_a	0.30	0.40	0.50	0.60	0.70	0.80
Category Number	91	146	246	434	825	1615
Recognition Rate(%)	94.59	95.93	96.24	97.00	97.18	97.44

Table 2 shows the change of recognition rates and categories number obtained under different vigilance parameter of FAM.

1) While ρ_a increases, recognition rate increases till a maximal value, then it begins to go down. It is shown in the Fig. 11.
2) While ρ_a increases, category number increases. Consequently, computational burden increases dramatically.
   - The reason is that the high vigilance parameter ρ causes small hyperrectangle.
   - While the similarity degree of each samples in the same category increases, category number increases in order to distinguish these patterns according to subtle differences.

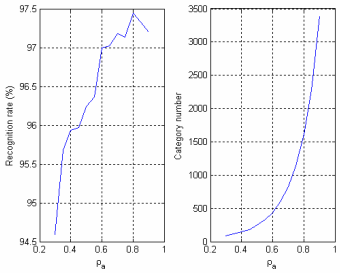


Figure 11. Recognition rate and category number of ARTa under deferent vigilance parameter

Therefore, a balance between the recognition rate and CN should be considered to select a proper ρ_a. Optimal ρ_a means this recognizer may extract the most common features from inputting pattern with low number of categories, and achieve high recognition rate at the same time**.** For real application, we selected $\rho_a = 0.60$ for our application by overall analysis according to Table 2.

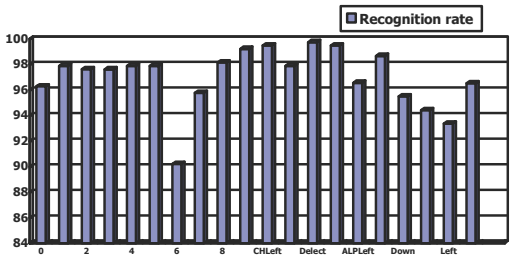


Figure 12. Recognition rates of different gestures

Figure 12 shows recognition rates of each gesture by using the proposed recognition algorithm for the generalization test. The total recognition rate is 97.00%. It is shown that the recognition rates of 17 gestures out of 20 gestures are higher than 95%. Even the lowest recognition rate of "6" is higher than 90%. Among wrong recognized gesture "6", most of them are recognized as "0", since the way to write these two gestures is very similar. If we can design discriminative gesture "6" and "0", the total recognition rate might be improved greatly.

For the recall test, the recognition rate is higher than 99%, it means that this recognizer has strong ability for learning. If training samples are enough to cover all typical features as more as possible, FAM based recognizer may provide high quality for gesture recognition.

5. CONCLUSION

In this paper, a hand drawn gesture-based input device based on FAM recognizer is introduced. Hand gesture is recognized by using 3-axis acceleration from accelerometers directly. There is no special constraint condition for inputting gesture as previous work.

With the proposed FAM-based recognizer, the total recognition rate of 97.00% in the generalization test has been achieved. At the same time, the proposed recognizer, with perfect learning ability, can achieve higher than 99% recognition rate for the recall test. That is, FAM based recognizer may learn all patterns and provide high quality for gesture recognition. The above results validated that FAM based recognizer can achieve high recognition rate with only raw data of 3 axis acceleration directly.

The effectiveness and feasibility of the proposed input device has been verified by the prototype of a remote controller to manipulate functions of TVs with gestures.

Further work is to improve the performance of current system to support more gestures.

6. REFERENCES

[1] Pavlovic, V. I., Sharma, R., Huang, T. S., Visual Interpretation of Hand Gestures for Human-computer Interaction: a Review. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol. 19, 1997, pp. 677 – 695.

[2] Ueda, E., Matsumoto, Y., Imai, M., Ogasawara, T., Hand-pose Estimation for Vision-based Human Interfaces. **IEEE Transactions on Industrial Electronics**, Vol. 50, 2003, pp. 676 – 684.

[3] Licsar, A., Sziranyi, T., Supervised Training Based Hand Gesture Recognition System. **Proceedings of 16th International Conference on Pattern Recognition**, Vol. 3, 2002, pp. 999 – 1002.

[4] Miyagawa, T., Yonezawa, Y., Itoh, K., and Hashimoto, M., Handwritten Pattern Reproduction Using 3D Inertial Measurement of Handwriting Movement. **Transactions of the Society of Instrument and Control Engineers**, vol. 38, 2002.

[5] Mantyla, V. M., Mantyjarvi, J., Seppanen, T., Tuulari, E., Hand Gesture Recognition of a Mobile Device User, **Proc. IEEE Int. Conf. on Multimedia and Expo 2000**, NY, Vol. 1, 2000, pp. 281-284.

[6] Bang, W. C., Chang, W., Kang, K. H., Choi, E. S., Potanin, A. and Kim, D. Y., Self-contained Spatial Input Device for Wearable Computers. **Proceeding of 7th IEEE International Symposium on Wearable Computers 2003**, pp. 26-34.

[7] Chang, W., Yang, J., Choi, E.-S., Bang W. C., Kang, K. H., Cho S. J., and Kim, D. Y., A Miniaturized Attitude Estimation System for a Gesture-based Input Device with Fuzzy Logic Approach. **Proceeding of 4th Int. Symp. On Advanced Intelligent Systems 2003**, pp. 616-619.

[8] Cho, S.-J. , Oh, J. K., Bang, W.-C., Chang, et al., Magic Wand: A Hand-Drawn Gesture Input Device in 3-D space with Inertial Sensors, **Int. Workshop on Frontiers in**

Handwriting Recognition 2004 (IWFHR2004), pp. 106-111.

[9] Carpenter, G.A., Grossberg, S., Markuzon, N. , and Reynolds, J.H., Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps, **IEEE Trans. on Neural Networks**, Vol.3, No. 5, 1992, pp. 698-713.

[10] Carpenter, G.A., Grossberg, S., and Reynolds, J.H., ARTMAP: A Self-organizing Neural Network Architecture for Fast Supervised Learning and Pattern Recognition, **International Joint Conference on Neural Networks' 1991**, IJCNN-91-Seattle, Vol. i, 1991, pp. 863-868.

[11] Carpenter, G.A., Grossberg, S., Iizuka K., Comparative Performance Measures of Fuzzy ARTMAP, Learned Vector Quantization, and Back Propagation for Handwritten Character Recognition, **International Joint Conference on Neural Networks' 1992**, Vol. 1, pp. 794 - 799.

[12] Yang, J., Liu, X. Q., et al., Fault Diagnosis of DC Motor Based on Parameter Estimation and Fuzzy ARTMAP Neural Networks, **Journal of Beijing University of Aeronautics and Astronautics**, Vol. 26, No. 3, 2001, pp. 284-288.

[13] Chralampidis, D., Kasparis, T., Georgiopoulos, M., Classification of noisy signals using fuzzy ARTMAP neural networks, **IEEE Transactions on Neural Networks**, Vol. 12, No. 5, 2001, pp. 1023 -1036.