21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Gesture Recognition System Based on Adaptive Resonance Theory

Paul K. J. Park[1], Jun Haeng Lee[1], Chang Woo Shin[1], Hyun-Surk Ryu[1], Byung-Chang Kang[1], Gail A. Carpenter[2], Stephen Grossberg[2]
[1]Frontier IT Lab, Samsung Advanced Institute of Technology, Yongin, Korea
{k.j.park, junhaeng2.lee, changwoo.shin, eric_ryu, bckang}@samsung.com
[2]Department of Cognitive and Neural Systems, Boston University, Boston, USA
{gail, steve}@cns.bu.edu

Abstract

We report on the moving hand gesture recognition technique using Adaptive Resonance Theory (ART). To detect the start and end points of a continuous moving gesture (known as "gesture spotting" problem), we propose the adaptive distributed prediction technique. Our results show that, unlike conventional non-recurrent neural networks, the proposed technique can be utilized usefully in reliable real-time learning (2000 times faster than with alternative methods) and recognition of continuously moving patterns.

1. Introduction

Recent progress in understanding the dynamic behavior of the human brain facilitates the development of smart IT devices by incorporating novel design ideas from brain-inspired cognitive processing. Previously, artificial neural networks such as the Multi Layer Perceptron (MLP) with the Back Propagation (BP) algorithm have been proposed as a pattern classifier in smart IT devices. However, this network has always suffered from the stability-plasticity dilemma (for example, one cannot have both the stability (retainability) of old, previously learned patterns and the plasticity to rapidly learn new patterns). Adaptive Resonance Theory (ART) has been applied to various systems because it can overcome this problem by mimicking human cognitive processing [1]. This algorithm uses an interaction between complementary processes of resonance and reset, which are predicted to control properties of attention and memory search just like a human brain (i.e., in multiple cortical areas) [2]. ART is often used as a static pattern classifier in applications where its architecture was not designed for analyzing continuously moving patterns in the manner of recurrent neural networks. For this reason, the Hidden Markov Model (HMM) has been utilized as the classification algorithm in moving gesture recognition system [3]-[5]. In this paper, we demonstrate moving hand gesture recognition based on ART. Since ART carries out match-based learning and prediction, continuous moving hand gesture recognition can be performed by adjusting the decision threshold of distributed prediction in the classification process. This decision threshold can be determined automatically before testing. The results show that 95 % recognition accuracy can be obtained by optimizing the decision threshold and the measurement period. In addition, our results confirm that the learning time of ART was about 2000 times faster than HMM while maintaining comparable accuracy.

2. Method and results

Figure 1 shows the structure of moving hand gesture (time-varying pattern) recognition systems. The moving objects were detected by using a dynamic vision sensor [6]. The tracking of moving objects can be done simply by spatiotemporally correlating the output of the vision sensor [7]. Figure 2 shows the hand gestures captured by our vision system when a subject is doing juggling. In this case, the trajectory of a hand gesture can be easily and spontaneously detected by tracking the fastest motion. The feature vectors extracted from this hand trajectory are recognized in the classification (ART) module. In the experiment, we performed the hand gesture recognition of Arabic numbers from '0' to '9' (all programs were coded in Java language).

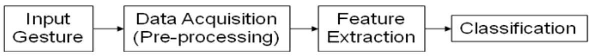

Figure 1. Structure of gesture recognition systems

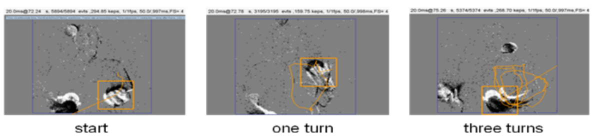

Figure 2. Hand gestures captured by the dynamic vision sensor

2.1. Feature Extraction

To improve the accuracy of gesture recognition, the feature extraction becomes critically important. In the experiment, we considered 10 categories of hand gestures from '0' to '9' as shown in Fig. 3. For each category of hand gesture, we extracted 11 feature vectors. For example, the feature vector x represents the position of the hand motion which lies along the x axis in Cartesian coordinate system. Similarly, the feature vector y represents the y-axis position. The feature vectors dx and dy represent the relative variations in the x and y positions. In addition, we derived various feature vectors (for example, angle, angular speed, linear speed, etc.).

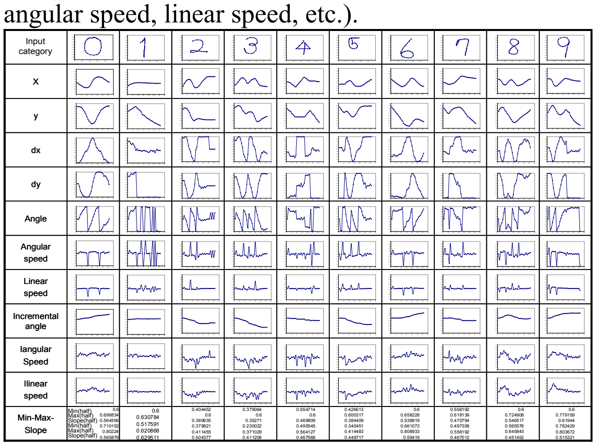

Figure 3. Various extracted feature vectors for the input gestures ('0'~'9')

Figure 4 shows the recognition accuracy of 100 test gestures measured while changing the feature vectors. In this case, 100 gestures were trained previously before the test procedure. The committed nodes represent the output categories learned by the ART algorithm. Our results show that, among 11 feature vectors, only 5 feature vectors (i.e., y, dx, dy, angle, and incremental angle) exceed 90 % recognition accuracy. However, the performance of the feature vector y can deteriorate when small or large patterns are utilized (for example, the accuracy was reduced to be less than 60 % in small-sized gestures). In addition, the recognition accuracy of the feature vector angle was measured to be less than 90 % in a large-scale test procedure (800 test results). Thus, we found out that the feature vectors dx, dy, and incremental angle had superior performance to others for dynamic gesture recognition. In Figure 4, the feature vector dx+dy appeared to have the best performance in gesture recognition since the accuracy is maximum (100 %) while the active committed nodes are minimum (10). Thus we chose the feature vector dx+dy for recognizing dynamic hand gestures.

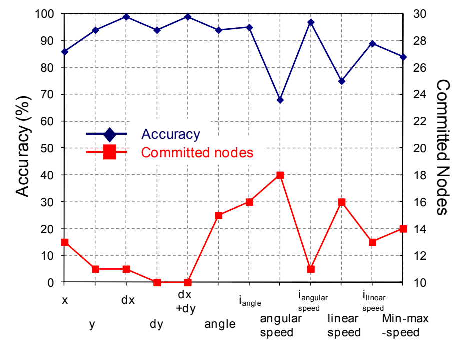

Figure 4. Recognition accuracy measured while changing the feature vectors

2.2. Recognition of Continuously Moving Gestures

Basically, moving gesture recognition has one critical problem, known as "gesture spotting", to detect the start and end points of a continuous gesture. This problem can be solved by using either HMM or recurrent neural networks. However, the previous method used in HMM-based classification requires a heavy computational load due to its large number of states and frame-based calculation [4]-[5]. To overcome these problems, in this paper, we applied the default ARTMAP algorithm based on the decision threshold adjustment of distributed prediction as shown in Fig. 5. Default ARTMAP employs winner-take-all coding during training and distributed coding during testing [8]. This algorithm produces continuous-valued test set predictions distributed across output classes. Then, as shown in Fig. 5 (b), the start and end points of the continuous gesture can be recognized simply by detecting when the distributed prediction exceeds the decision threshold. Thus, the proposed technique requires the pre-evaluation process to obtain a proper decision threshold before testing.

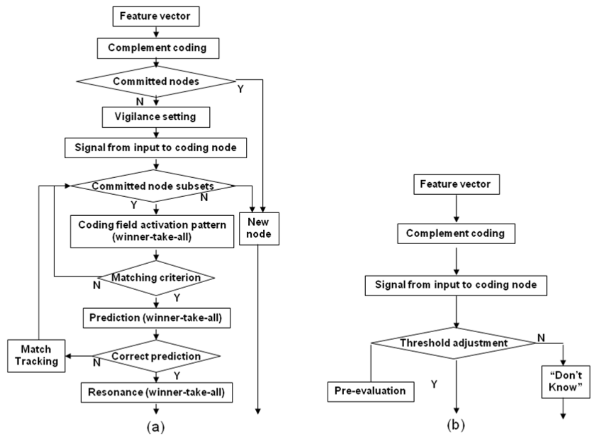

Figure 5. Default ARTMAP flowchart (a) training and (b) testing

Figure 6 (a) shows the signal from input to coding node (defined in [8]) calculated when the feature vector of the gesture '0' was entered into the classification module sequentially over time. Basically, this value increases with the similarity between the input and the committed node. As expected, the signal from input to coding node of the committed node (output class) '0' is higher than the others and increases as the starting time of the input gesture (feature vector) becomes synchronized with the learned gesture. Thus, this phenomenon shows that we can discriminate the desirable committed node from the others by adjusting a threshold. In this case, two techniques based on the threshold can be utilized as shown in Fig. 6 (b). One is to determine the output class as soon as the signal from input to coding node exceeds the threshold (threshold 1). The other is to calculate the maximum value during a certain period of time after when it becomes higher than the threshold (threshold 2). The results show that the performance of the threshold 2 is superior to the threshold 1 technique. For example, the maximum accuracy was measured to be 99 % when the threshold 2 and calculating period were set to be 61.5 and 150 ms, respectively.

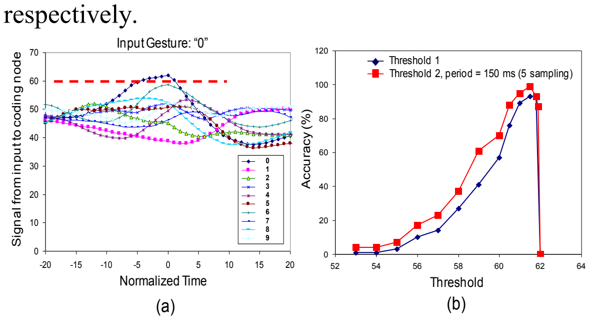

Figure 6. (a) The signal from input to coding node calculated when the gesture "0" was entered sequentially over time (b) Recognition accuracy measured while varying the threshold

However, the performance of the gesture recognition system can be degraded when the speed of the testing gesture is different to the trained gesture as shown in Fig. 7. Our results show that the measured accuracy of the fast-speed gesture (100 test sets) (i.e. two times faster than the normal-speed trained gesture) was less than 20 % in case of without training. In this case, the measured accuracy can be improved dramatically by training the fast-speed gesture in advance. It should be noted that the performance measured when the fast-speed gesture is trained separately against the normal-speed gesture (i.e., trained in a different class) is superior to the case using same-class training (Fig. 7 shows that the measured accuracy is increased up to 84 %). To evaluate the performance of various gestures, we measured the recognition accuracy while varying the threshold and period simultaneously as shown in Fig. 8. In this experiment, various gestures with different speeds were simulated. A total of 700 gestures (normal speed = 500, fast speed = 100, medium speed (between normal and fast) = 100) were trained and 1000 gestures (normal speed = 800, fast speed = 100, medium speed = 100) were tested. The results show that the maximum recognition accuracy was measured to be 95 % when the decision threshold and calculating period are 60.7 and 11 (/samples, i.e., 330 ms), respectively. This decision threshold depends on the application, but can be determined automatically because the parameters used in the choice by difference rule (for example, adaptive weights and dimension of an input pattern) are set before testing [9]. It should be noted that, the computation of default ARTMAP was performed within 3 ms per one feature input (@2-GHz CPU). Thus, we expect that this technique based on threshold adjustment can be utilized usefully for real-time recognition of various dynamic patterns.

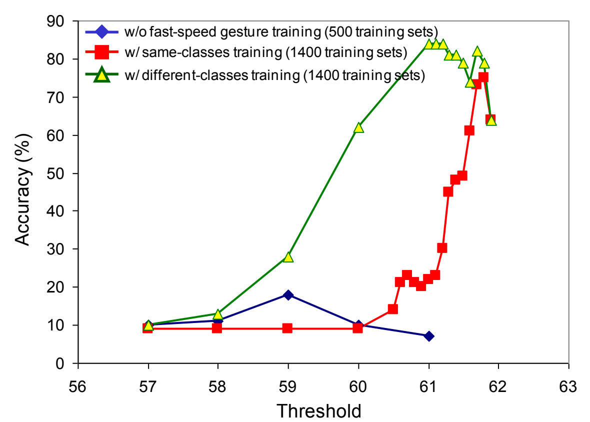

Figure 7. Recognition accuracy measured when a fast-speed gesture is applied

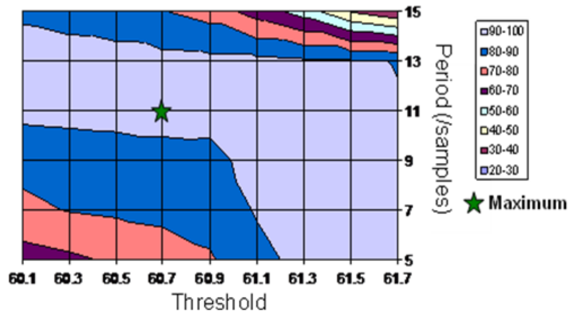

Figure 8. The recognition accuracy measured while varying the threshold and period (number of training gestures = 700, number of testing trials = 1000)

In addition, the performance comparison between ART and HMM is shown in Fig. 9. For this comparison, we utilized Left-Right (LR) and Left-Right Bended (LRB) models as the HMM classifier (features: angle, number of states: 5) [3]. The results show that the learning times of HMM and ART are increased with the number of training gestures as shown in Fig. 9 (a). In particular, we found out that the learning time of HMM is about 2000 times greater than ART (~30 ms when the number of training gestures is 400). This is mainly because the conventional Baum-Welch algorithm calculates every state transition probability during learning while the synaptic weight in ART is updated only when the input is close enough to internal expectations (match-based learning) [8]-[10]. Fig. 9 (b) shows the measured accuracy while varying the number of training gestures. There was not much difference in measured accuracies between ART and HMM (> 90 %).

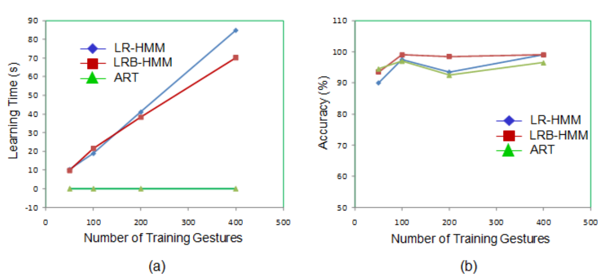

Figure 9. The performance comparison between ART and HMM (a) learning time (number of feature elements = 30) and (b) accuracy (number of testing trials = 200)

3. Summary

We have reported on learned gesture categorization and recognition using the default ARTMAP model. Unlike conventional non-recurrent neural networks, the gesture spotting problem can be easily solved by using the adaptive decision threshold technique in the distributed prediction process. The results showed that 95 % recognition accuracy could be obtained by optimizing the decision threshold and the calculating period and that learning is about 2000 times faster than with alternative methods. ARTMAP models also have a great advantage of solving the stability-plasticity dilemma and thereby self-stabilizing their learned memories under fast learning conditions. Thus, this technique holds promise for use in reliable real-time learning and recognition of various dynamic patterns (for example, gesture recognition, hand writing recognition, and speech recognition).

Acknowledgements

The authors are grateful to Prof. Tobi Delbruck (INI-ETH) for providing the dynamic vision sensor and software framework for the neuromorphic interface.

References

[1] D. C. Wunsch II. ART properties of interest in engineering applications. *International Joint Conference on Neural Networks*, 3380-3383, 2009.

[2] S. Grossberg. The complementary brain: Unifying brain dynamics and modularity. *Trends in Cognitive Sciences*, 4:233-246, 2000.

[3] M. Elmezain and et al. A Hidden Markov Model-based isolated and meaningful hand gesture recognition. *World Academy of Science, Engineering and Technology*, 393-400, 2008.

[4] H. K. Lee and J. H. Kim. An HMM-based threshold model approach for gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:961-973, 1999.

[5] H. S. Park and et al. HMM-based gesture recognition for robot control. *Lectures Notes in Computer Science*, 3522:607-614, 2005.

[6] P. Lichtsteiner and et al. An 128x128 120dB 15us-latency temporal contrast vision sensor. *IEEE J. Solid State Circuits*, 566-576, 2007.

[7] J. H. Lee and et al. Real-time Gesture Interface based on Neuromorphic Silicon Retina. submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[8] G. A. Carpenter. Default ARTMAP. *International Joint Conference on Neural Networks*, 1396-1401, 2003.

[9] G. P. Amis and G. A. Carpenter. Default ARTMAP 2. *International Joint Conference on Neural Networks*, 777-782, 2007.

[10] G. A. Carpenter and S. Grossberg. *Pattern Recognition by Self-Organizing Neural Networks*. Cambridge, MA: MIT Press, 1991.