



# 模糊化類神經網路於動態手勢辨認之應用

## Application of Fuzzy Neural Networks in Dynamic Hand Gesture Recognition

4中9

計畫編號：NSC 88-2213-E-032-014

執行期限：87年8月1日至88年7月31日

主持人：蘇木春 副教授 淡江大學電機系

e-mail: muchun@ee.tku.edu.tw

### 一、摘要

根據資料顯示，臺灣所使用的手語，基本手型的有五十一種，但會因為手語打出時的位置不同而有不一樣的意思；而手或手臂所面對或指向的方向，也會影響手語所代表的意義。除此之外，手型的變化與組合也代表不同的手語。然而，手語中與手部移動軌跡有關的動態手勢佔有比率不低，因此，一個完整且實用的手語發聲系統，勢必要完成所有動態手勢的辨識。要能達成上述的目標，必須要有三度空間軌跡追蹤器(3D-Tracker)的配合才能獲得手部移動的資訊，為了能讓聾啞者能真正利用此套系統以改善他們的生活，我們決定在此一年的計劃中，自行開發出低價位的3D-Tracker，然後發展出能利用模糊化類神經網路來辨識此種動態手勢的技術。希望此種系統的建立能對我國之聾啞同胞能有所助益。

關鍵詞：類神經網路、模糊系統、決定樹、手勢辨識

### Abstract

According to literatures, there are about 51 basic hand shapes in the hand gestures used by the deaf in Taiwan. However, different positions and orientations of the hands result in different interpretations of hand gestures. Besides, different combinations of basic hand shapes also result in different had gestures. However, dynamic hand gestures which involve the movements of hands play an important role in hand-gesture to speech systems. If we want to implement a complete hand-gesture to speech system we have to develop a technique for recognizing the whole dynamic hand gestures. In order to achieve the goal, we need a 3D-tracker to capture the trajectory information of hands. The commercial products of 3D-Tracker are very expensive, therefore, we decide to develop a low-cost 3D-Tracker. In addition, we also want to develop a technique of recognizing dynamic hand gestures by using fuzzy neural networks. We really hope to implement such a low-cost system to make the deaf benefit from it as soon as possible.

Keywords: neural networks, fuzzy systems, decision

tree, gesture recognition.

### 二、計畫緣由與目的

對於身體部分功能有障礙者來說，這個世界充滿了許多的阻礙，如何藉助科技來幫助他們，完成一個無障礙空間，實在是刻不容緩的任務。目前工程界也的確有了不錯的成績，譬如說，助聽器可以幫助聽力受損的人，重新感受到外界的聲音刺激，至於身體不便於行的患者，在操控方便的電動輪椅的協助下，可以有較大的自由度來擴展他們的生活空間；可是對於聾啞同胞，我們卻始終未能完成一種實用的復健裝置來幫助他們。

由於先天性疾病或後天意外的發生，導致聾啞人士的聽覺或語言能力受損，因此手語變成他們的溝通工具，但是對大多數的人來說，如何懂得使用手語卻是一門艱深的學問，所以聾啞者與一般大眾只能透過口譯人員或書寫來完成雙向溝通，這種溝通上的障礙，很可能導致聾啞者的求職困難，伴隨著經濟能力較差，甚至於會影響他們的親人或子女的成长，如何幫助他們克服溝通障礙，是個十分迫切且富有意義的研究方向，而手語發聲系統就是個能幫助聾啞者復健的裝置。

根據資料的收集，目前已知的，有人從事美國手語[1]-[10]及日本手語[11]-[17]的研究，而本人投入臺灣手語的辨識研究也已經有一些初步的成果[19]-[21]。手語會因國度的不同而有極大的差異性，甚至因地區的差異也會導致同一國家的聾啞者，彼此之間也有些微的溝通障礙，因此，即使國外有發展成功的手語系統，也無法全然移植於國內使用，除此之外，價格因素也是重要的考量因素之一，所以臺灣勢必自行發展本身的手語發聲系統。

手語不像一般口語一樣，由所謂的母音與子音所組合而成，手語是由一些視覺因素所組成，三個基本構成要素為(1)手型(hand shape)、(2)位置(location)、以及(3)手部的運動軌跡及指向方向(hand trajectory and orientation)。而手語又大致分為靜態手語及動態手語兩種。靜態手語指的是那些只要知道手型(即手指彎曲度)及手部位置，即代表一個字或詞的手語而言，而動態手語就

複雜多了，又可再細分為：

- (1) 手部位置不變但手形有變。
- (2) 手形不變但手部位置或方向有變。
- (3) 手形及手部位置皆變。

從以上的分類，我們很清楚地知道，要能成功的辨識手語，實在不是件容易的工作。

至於動態手勢的辨識方法，有些研究者是提出隱藏式馬可夫模型(Hidden Markov Model)的方法來解決，但是此種方法需要大量訓練資料以便估測出所需的參數，如此一來才能有較高的辨識率。本計畫希望能以模糊化類神經網路為架構，發展出一種較有效率且辨識率高的動態手勢辨識單元。目前，我們已有辨識靜態手語及動態手勢中的第(1)部分的手語的能力，希望能藉助這個計劃來完成其餘動態手勢的辨認工作，以便早日為臺灣聾啞人士建立一個無障礙空間。

### 三、結果與討論

本計劃已開發出如圖一所示的低價位的 3D-Tracker，並提出兩種辨識手部運動軌跡的方法：

#### 1. 方法

(1) 決定樹方式：

- Step 1 取樣：利用開發出低價位的 3D-Tracker 取樣。
- Step 2 平滑化：利用平滑濾波器將原始資料平滑化。圖二顯示了平滑效果。
- Step 3 特徵點粹取：找到端點和轉折點。
- Step 4 加標：根據曲度特性將每個線段分別標示為如圖三所示的區段。
- Step 5 辨識：利用圖四的決定樹來辨識手部運動軌跡。

(2) 模糊化類神經網路方式：

- Step 1 取樣：利用開發出低價位的 3D-Tracker 取樣。
- Step 2 平滑化：利用平滑濾波器將原始資料平滑化。
- Step 3 投影：將平滑化後之資料投影至具代表性平面上。
- Step 4 樣本產生：利用八方向的 chain coding 技術將資料予以編碼，以便產生樣本。
- Step 5 辨識：利用下式計算輸入對第 k 個手部運動軌跡的相似度：

$$S_k = \sum_{i=1}^{N_{d_1}^k} Out_{d_1^k}(x_i) + \sum_{i=N_{d_1}^k+1}^{N_{d_1}^k+N_{d_2}^k} Out_{d_2^k}(x_i) + \dots + \sum_{i=N_{d_1}^k+N_{d_2}^k+\dots+N_{d_{D-1}}^k+1}^N Out_{d_{D-1}^k}(x_i) \quad (1)$$

其中  $Out_j(x)$  代表第 j 個模糊化類神經元的輸出， $N_{d_j}^k$  代表與第  $d_j$  個方向相似的圖樣數目。此種模

糊化類神經元的輸入/輸出的關係定義如下：

$$net_j(x) = \sum_{i=1}^n w_{ji} x_i = w_j^T x \quad (2)$$

和

$$Out_j(x) = f(net_j(x)) = \frac{1}{2} (net_j(x) + 1) \quad (3)$$

其中  $x = (x_1, x_2, \dots, x_n)^T$  是輸入向量， $w_j = (w_{j1}, w_{j2}, \dots, w_{jn})^T$  是鍵節值向量。如果  $\|x\| = \|w_j\| = 1$  那麼

$$Out_j(x) = \frac{1}{2} (\|x\| \cdot \|w_j\| \cos(\theta) + 1) = \frac{1}{2} (\cos(\theta) + 1) \quad (4)$$

此時  $Out_j(x)$  代表  $x$  和  $w_j$  的相似度

Step 6 決定單元：最後利用下式將此輸入判定為具有最大相似度之手部運動軌跡：

$$S_k \geq S_{k'} \quad \text{for } k \neq k' \quad (5)$$

#### 2. 結果

本計劃要求九個人比十種於臺灣手語常見的手部運動軌跡(如表一)，每人每種手部運動軌跡各比十次。然後利用上述兩種辨識手部運動軌跡的方法來予以辨識，其辨識率分別為 94%(決定樹方式)和 98.9%(模糊化類神經網路方式)。

#### 3. 結論

本計劃已開發出低價位的 3D-Tracker，可申請專利；除此之外，所提出的兩種辨識手部運動軌跡的方法，不僅簡單且其辨識率亦高，有學術和實際應用的價值。

#### 四. 計劃成果自評：

所提出之辨識方法，已發表在 1998 年之 IEEE Workshop on Knowledge Engineering and Data Engineering [22] 和 1999 年之 IEEE International Joint Conference on Neural Networks [23]。我們亦著手進行將計劃結果做進一步的整理，以便以期刊論文的形式發表。因此，已達成本計劃之目標。

#### 五、參考文獻

- [1] J. Kramer and L. Leifer, "The talking glove: an expressive and receptive verbal communication aid for the deaf, deaf-blind, and nonvocal," Proc. of the Third Annual Conf. On Computer Technology / Special Education / Rehabilitation, Northridge, pp. 335-340 CA, Oct. 1987.
- [2] J. Kramer and L. Leifer, "The talking glove: a speaking aid for nonvocal deaf and deaf-blind

- individuals," Proc. of the RESNA 12<sup>th</sup> Annual Conf. New Orleans, Louisiana, pp. 471-472 1989.
- [3] S. S. Fels and G. E. Hinton, "Glove-Talk: a neural network interface between a Data-Glove and a speech synthesizer," *IEEE Trans. on Neural Networks*, vol 1, no.4, pp4-8, 1993.
- [4] C. Vogler and D. Metaxas, "Adapting Hidden Markov Model," *IEEE Int. Conf. on Systems, Man, and Cybernetics*, pp. 156-161,1997.
- [5] K. Grobel and M. Assan, "Isolated sign Language Recognition using Hidden Markov Models", *IEEE Int. Conf. on Systems, Man, and Cybernetics*, pp. 162-167,1997.
- [6] A Braffort, "ARGO: An architecture for sign language recognition and interpretation," In P. A. Harling, A. D. N. Edward, editors, *Progress in gestural interaction. Proceedings of Gesture Workshop '96*, pp. 17-30. Springer, Berlin, New York 1997.
- [7] R Erenshsteyn and P. Laskov, "A multi-stage approach to fingerspelling and gesture recognition," *Proceedings of the Workshop in the Integration of Gesture in Language and Speech*, pp. 185-194, Wilmington, DE, 1996.
- [8] M.B. Waldron and S. Kim, "Isolated ASL sign recognition system for deaf persons," *IEEE Transactions on Rehabilitation Engineering*, 3(3):261-271, September 1995.
- [9] H. Liang and M. Ouhyoung, "A sign language recognition system using Hidden Markov Model and Contest Sensitive Search", *ACM VRST*, 1996.
- [10] J. Schlenzig, E. Hunter and R. Jain. "Video Based Hand Gesture Interpretation Using Recursive Estimation," *Proc of ACSSC, Pacific Grove, NOV. 1995*, pp. 1267-1271.
- [11] K. Kamata, T. Yoshida, M. Watanabe and Y. Usui, "An Approach to Japanese-sign language translation system," *IEEE International Conf. on SYSTEM, MAN, AND CYBERNETICS*, pp. 1089-1090, 1989.
- [12] K. Murakami and H. Taguchi, "Gesture recognition using recurrent neural networks," *CHI'91 Proc.*, pp. 237-242, 1991
- [13] T. Kurokawa, "Gesture coding and a gesture dictionary for a nonverbal interface," *IEICE trans. Fundamentals*, vol E75-A, no.2, pp.112-121, Feb, 1992.
- [14] Y. Nagashima, T. Onodera, H. Ngashima, M. Terauchi, and G. Ohwa, "A study of recognition method of Japanese finger spelling," *Human Interface 44-15*, pp. 109-115, 1992.
- [15] S. Okamura and A. Kumamoto, "Noncontact hand-shape recognition and its application," *Human Interface*, vol. 8, pp.169-176,1993.
- [16] H. Sagawa, H. Sakou and M. Abe, "Sign language translation system using continuous DP matching," *Human Interface 44-12*, pp. 87-94, 1992.
- [17] K. Morimoto, T. Izuchi, E. Fufishige, S. Watanabe, T. Morichi and T. Kurokawa, "Analysis of spatio-temporal structure of sign language for its machine translation," *8<sup>th</sup> Symposium on Human Interface*, pp. 621-626,1992.
- [18] J. Yamato, J. Ohya and K. Ishii, "Recognizing Human Action in Time-Sequential Images Using Hidden Markov Model," *Proc. of the IEEE Conf. on Comp. Vision and Pattern Recognition*, 1992, pp. 379-385.
- [19] M. C. Su, W. F. Jean, and H.T. Chang, "A static hand gesture recognition system using a composite neural network," *Proc. of Fifth IEEE International Conf. On Fuzzy Systems*, vol 2, pp.786-792, 1996.
- [20] M. C. Su, "A speaking aid using neural networks for the deaf.", *Biomedical Engineering Applications, Basis & Communications*, vol. 8,no.4, pp.33-39,1996.
- [21] M. C. Su, H. Huang, C. H. Lin, C. L. Huang and L. D. Lin, "Application of Neural Networks to Spatio-Temporal Hand Gesture Recognition," *IEEE int. Conf. On Neural Networks*, pp. 2116-2121, 1998.
- [22] M. C. Su, Y. Y. Chen, K. H. wang, C. Y. Tew, and H. Huang, "A syntactic approach to 3D arm movement recognition," *IEEE Knowledge and Data Engineering Extranage Workshop*, pp. 68-73, 1999.
- [23] M. C. Su, H. Huang, P. S. Chiu, Y. X. Zhao, Y. Y. Chen, and H. F. Chen, "Recognition of 3D movements using neural networks," to be published in *IEEE Int. Joint conf. On Neural Networks*, 1999.

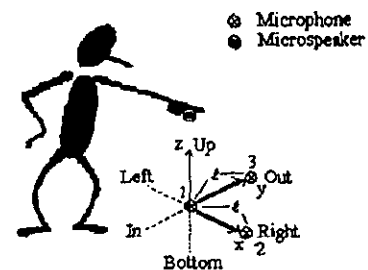
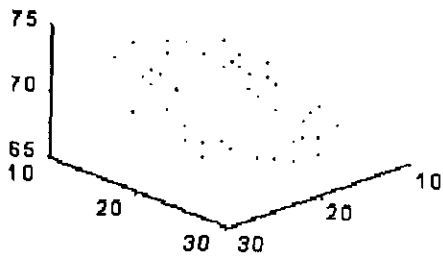
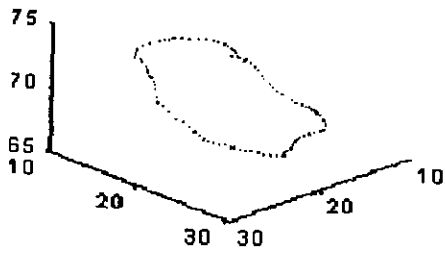


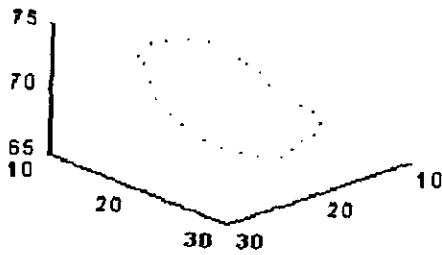
Figure 1. The 3D ultrasonic tracker.



(a)



(b)



(c)

Figure 2 An example of the smoothing effect: (a) the original raw data. (b) the smoothed data. (c) the chained data.

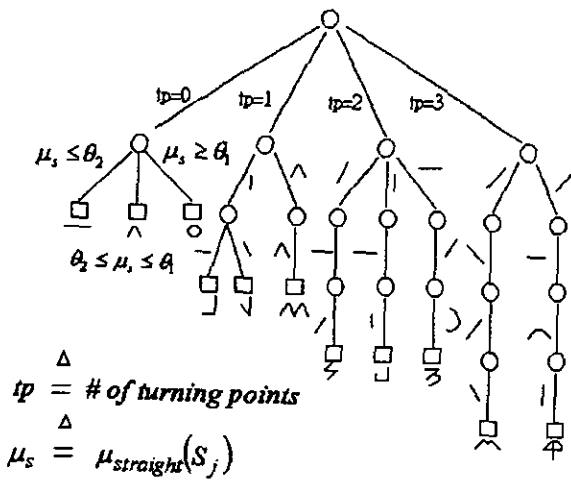


Figure 3 The ideal decision tree.

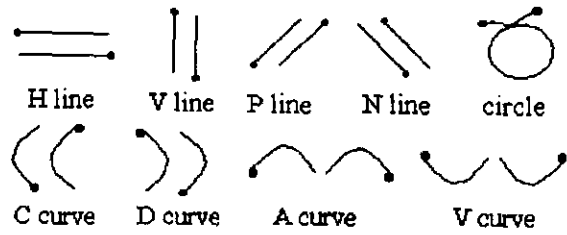


Figure 4 Nine types of segments (• represents a starting point).

Table.1 The 11 typical arm movements involved in TSL.

	Movement primes	Taiwanese Sign Language
1	—	Left, right, heavy, ...
2	∩	Newspaper, sky, evening, ...
3	○	Look for, inspect, empty, ...
4	└	Chou, ...
5	↓	Enough, ...
6	∩	Grandfather, frog, ...
7	⚡	Electricity, sound, ...
8	┘	Then, ...
9	∧	Mountain, ...
10	ㄨ	Ten thousand, ...