

總計劃名稱：類神經網路於生物測定認證技術及應用之研究
計劃期間：85年8月1日~88年7月31日

子計劃三名稱：唇語辨讀系統 Active mesh for lip-reading

計劃編號：NSC 86-2745-E-002-012

執行期限：85/08/01~86/07/31

計劃編號：NSC 87-2213-E-002-036

執行期限：86/08/01~87/07/31

計劃編號：NSC 88-2213-E-002-036

執行期限：87/08/01~88/07/31

子計劃主持人：劉長遠 台大資訊工程學系

一、中文摘要

我們以 Point Distribution Model (PDM) 作為唇形輪廓以及唇形輪廓附近顏色的模型。在影像中追蹤唇形輪廓的工作，是以上述的唇形輪廓為基礎，找到一個最適當的唇形，使得該唇形在該影像中的顏色為合法的。在唇語辨讀方面，我們使用 Time-Delay Neural Networks 來做辨識。

二、英文摘要

We model the lip contour and lip color profile by the point distribution model (PDM). Image searching task is to find a lip contour on the image based on the lip contour model which will present the best lip color profile. In lip reading, we employ Time-Delay Neural Networks to recognize the speech.

三、計劃緣由與目的

唇語辨識已經被認為是一個重要而且很有應用價值的題目。在吵雜環境中，語音辨識便會受到噪音雜訊的

干擾，以影像作為輸入的唇語辨識便沒有這方面的困擾；再者，人發出某些鼻音時，以聲音的資訊是很難辨識的，但是嘴形的變化卻是比較不同的[9]。結合影像及聲音的資訊，也可以提昇電腦對於語音的辨識[8]。

在影像處理的領域裡，要從影像中找出唇形的形狀，一直就是一個困難的而有趣的課題。這課題的主要困難度在於每個人的膚色不同，嘴唇顏色亦不同，隨著環境中不同光源的光線的照射，臉部的顏色也會改變。再者，嘴唇的顏色跟臉部顏色很接近，也就是說嘴唇和嘴唇旁邊的肌膚若要以顏色來加以區分的話，其介線並不明顯。儘管嘴唇的顏色很難加以定義，描述，但是直觀上，顏色還是辨識唇形中一個很重要的線索。

要以數學模型來描述嘴唇形狀也有其困難度，因為每個人的唇形都不一樣，就算是同一個人的嘴唇，也會因為這個人的臉轉了個角度，或者他張開了嘴而改變了他的唇形。而且嘴唇是柔軟可變形的，這也增加了描述

唇形的困難。

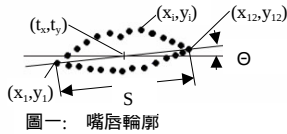
也有人認為嘴唇形狀在唇語辨識的過程中並不重要，因為人在讀唇語時顯然不是很清楚的看出對方嘴唇的形狀。倒是嘴唇運動的方式，提供了比較重要的資訊。所以 Pentland 等人就提出了以 optical flow 去估算臉部肌肉的運動情形，然後以肌肉運動的模式去辨別語者所發出的音。

在本計劃中，我們選擇利用 Point Distribution Models (PDM) 來建構嘴唇模型，此方法由 Juergen Luetttin 等人於 1996 年提出[1][2]，它可以用來描述嘴唇的輪廓，在影像中尋找嘴唇位置則是透過比對嘴唇輪廓附近的顏色來達成。

唇語辨識則是採用 Time-Delay Neural Networks. 這網路可用於辨識諸如語音辨識或者唇語辨識此類時序訊號上。

四、研究方法與成果

4.1 嘴唇形狀模型



圖一：嘴唇輪廓

我們以手動方式標示嘴唇輪廓，如圖一所示，一個嘴唇由 22 個點所組成，包括 (x_1, y_1) 及 (x_{12}, y_{12}) 兩個端點。為了讓每個標示出來的嘴唇的每一點可以相對映，我們採用一致的標示方法，上嘴唇的每點間距相等，下嘴唇的間距亦相等。

又，S 為兩端點的距離，用以定義此嘴唇的 Scale。Θ 是兩端點方向和水平線的角度。

嘴唇的平均輪廓，以及變異的主要模式是由一組訓練範例求得。為了得到訓練範例，必須先以人工的方式，在影像中標示出嘴唇的輪廓。訓練範例

的總個數為 N，第 i 個訓練範例使用向量 v_i 表示：

$$V_i = (x_{i0}, y_{i0}, x_{i1}, y_{i1}, \dots, x_{in-1}, y_{in-1})^T$$

其中 (x_{ij}, y_{ij}) 是第 i 個訓練範例嘴唇形狀的第 j 個點的座標。

為了讓標準化嘴形為單一寬度以及零位移及旋轉角度為零，使用下列式子得到標準化的嘴形：

$$X_i = M \left(\frac{1}{s}, -q \right) [v_i - t_c]$$

其中

$$t_c = (t_x, t_y, t_x, t_y, \dots, t_x, t_y)^T$$

t_x, t_y 為嘴唇的中心點；M(a,b) 轉化矩陣可將向量縮放 a 倍，並且旋轉 b 角度。

由訓練範例我們求出 N*N 的共變矩陣，再求出此變矩陣的 Eigenvectors. 我們知道訓練範例在 Eigenvalue 最大的 Eigenvector 方向的投影，有最大的變異，Eigenvalue 次大的 Eigenvector 方向投影之變異次之，以此類推。

於是標準化過的唇形可以使用下列式子趨近：

$$x = \bar{x} + Pb$$

其中 $P = (p_1 p_2 \dots)$ p_1 為 Eigenvalue 最大之 Eigenvector, p_2 之 Eigenvalue 次之。

4.2 嘴唇顏色模型

在嘴唇輪廓的每一點附近取 9 點，每個嘴唇共可取 198 個點，以這些點的顏色代表嘴唇顏色。每個訓練範例的嘴唇顏色可表示成：

$$h_i = (g_{i0}, g_{i2}, g_{i3}, \dots, g_{i197})^T$$

於是我們由 $h_1 \dots h_x$ 求出共變矩陣 S_g ，在求出共變矩陣的 Eigenvectors 及 Eigenvalues.

又，每個嘴唇顏色可以下列式子來趨

近:

$$h = \bar{h} + P_g b_g$$

其中 $P_g = (P_{g1} P_{g2} \dots P_{gt})$, $P_{g1} \dots P_{gt}$ 分別為其對應的 Eigenvalue 最大至第 t 大的 Eigenvectors.

4.3 嘴唇的定位

我們使用下列式子降低嘴唇顏色的維度:

$$b_g = P_g^T (h - \bar{h})$$

以下列式子來評估所找到的嘴唇顏色和嘴唇顏色模型的類似程度:

$$E_p = (h - \bar{h})^T (h - \bar{h}) - b_g^T b_g$$

定位嘴唇的方法是在影像中找出一嘴唇形輪廓, 使得 E_p 得以最小化.

4. 唇語的辨識

我們選擇以 Time-Delay Neural Networks 來辨識唇語. 網路由三層神經元構成, 以定位得到的嘴唇輪廓序列作為輸入值, 使用 Error Backpropagation 來學習.

五、總結

Point Distribution Modeling 可用在嘴唇輪廓的定位上, Tim-Delay Neural Networks 可以使用於在唇語辨識上.

六、參考文獻

1. Juergen Luetttin, Neil A. Thacker, "Speechreading using Probabilistic Models", *Computer Vision and Image Understanding*, Vol 65, No. 2, February, pp.163-178, 1998
2. Juergen Luetttin, Neil A. Thacker and Steve W. Beet, "Visual Speech Recognition Using Active Shape Models and Hidden Markov Models",
3. Juergen Luetttin, "Towards Speaker Independent Continuous Speechreading", *Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH'97)*, 1997
4. J. Yang and A. Waibel, "A real-time face tracker," *Proceedings of WACV'96*, pp. 142-147 (Sarasota, Florida, USA)
5. Chung-Lin Huang, Wen-Yi Huang, "Sign language recognition using model-based tracking and a 3D Hopfield neural network", *Machine Vision and Applications* (1998), 10, pp. 292-307
6. Christoph Bregler, Hermann Hild, Stefan Manke, and Alex Waibel, "Improving connected letter recognition by Lipreading",
7. Greg I. Chiou and Jenq-Neng Hwang, "Lipreading from Color Video", *IEEE Transactions on Image Processing*, Vol. 6, No.8, August 1997, pp.1192-1195
8. Keren Yu, Xiaoyi Jiang, Horst Bunke, "Lipreading: A classifier combination approach", *Pattern Recognition Letters* 18 (1997), pp. 1421-1426.
9. K. Mase and A. Pentland, "Automatic lipreading by optical-flow analysis," *Syst. Comput. Jpn.*, vol 22, pp.67-76, 1991
10. T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, "Active shape models - Their training and application", *Computer Vision and Image Understanding* 61, 1995, pp.38-59.