

新聞主播面部動作與表情的量化分析與比較

陳宣輯

國立政治大學 資訊科學系

s9126@cs.nccu.edu.tw

摘要

近來年，新聞播報方式朝向多元化的趨勢，新聞主播出現了許多新的詮釋，從數位虛擬主播到俄羅斯的另類主播，新聞播報儼然成為了一個小型的表演舞台，然而目前仍舊沒有一個主播表情形象的參考與準則，因此本計劃「新聞主播面部表情與動作的量化分析與比較」，乃希望透過視訊分析的技術達到自動化偵測與處理，利用處理後的量化資料，進一步分析主播臉部變化與新聞內容的因果關係，並做出客觀的結論。

關鍵詞：新聞主播、臉部表情、視訊分析、自動化偵測

一、導論

台灣有線電視開放後，各家電影台競爭激烈，其中新聞更是各電視台必爭之地。為了提高收視率，新聞播報演變出許多表演化的趨勢，而對新聞主播的形象塑造上，有意無意的造成新聞主播明星化。除了外貌上總以帥哥美女為號召之外，臉部表情也更加表演化與活潑（如頻頻頓首、搖頭擺腦、眉頭深鎖、與社交性的笑容）[1]，比較國外的 CNN 或 NHK 等國際知名電視新聞，讓人總是感覺台灣新聞主播的表情動作似乎相對的豐富，如圖一。



圖一 主播各類的表情動作

此外，2000年英國的 PA New Media 電子媒體集團突發奇想地創造出全球第一位虛擬主播“安娜諾娃”[2]，這位虛擬主播的電腦系統會自動在網路上尋找突發新聞，並自動播出，即時性之高，無人可比。這種 365 天永遠不需換班，即使是緊張的實況轉播也不會唸錯稿的數位新聞主播，或許是我們未來收看新聞的趨勢潮流。但我們對於虛擬人物的面部表情往往仍然認為不夠自然與流暢，因此，大家還是習慣收看傳統的新聞播報。如果能將真實主播的五官位置加以量化，將可藉此了解臉部的表情動作，做為未來數位虛擬人物塑造與建構的參考。

本研究「新聞主播面部動作與表情的量化分析與比較」，乃針對前述的問題，希望透過視訊分析的技術，達到自動化偵測與處理，一方面可作為儀態訓練的輔助工具，另一方面也可以藉此比較不同文化、風格、類型主播的差異性。

二、研究方法

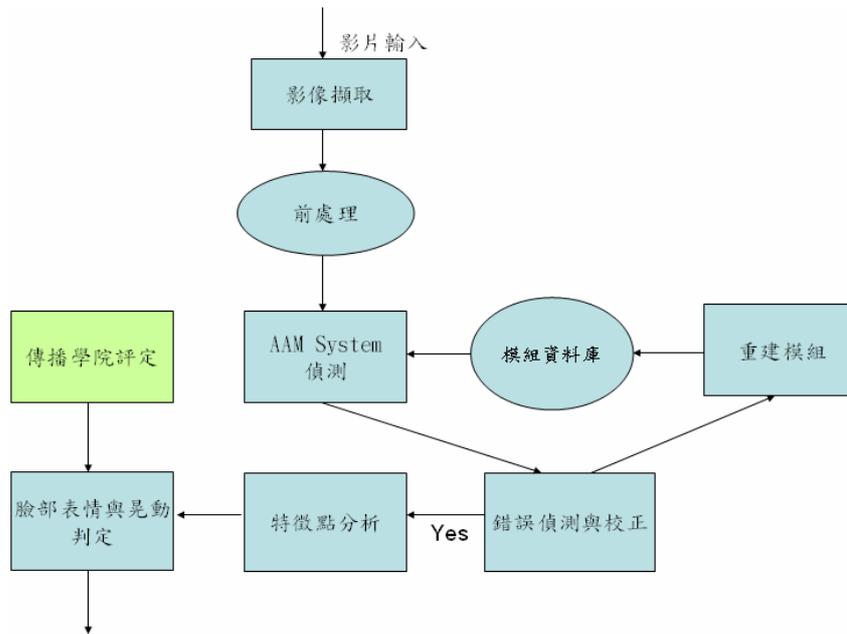
本研究的偵測範圍為臉部，但由於現在的新聞型態越來越豐富，例如：主播過多的肢體語言或臉部的傾斜角度，將影響實驗之結果，我們做出以下基本假設，進一步排除外在因素：

1. 新聞主播的臉幾乎都是以正面為主，所以我們對於臉部偵測的範圍將以正面為主。目前正面臉部偵測(frontal face detection)的技術已發展成熟，正確率高達 95% 以上[3]。
2. 新聞畫面都是動態的連續畫面，而且主播的移動範圍有限，因此我們可以利用動態連續擷取多張影像，藉由影像資訊得知移動物體狀況和位置。靜態影像的人臉偵測由於取樣少，相形之下，動態的影像更方便偵測。

基於以上兩點的假設，並參考過許多研究報告及文獻後，我們認為應該有能力將新聞主播面部五官變化加以量化分析。例如我們可以使用光流追蹤(optical flow)與 Active Appearance Models(AAM)或其他特徵匹配技術來追蹤影像中的特定區域。透過量化後的資料，將可較客觀地對新聞主播的臉部變化做深入的研究與比較，觀察主播是否受到播報內容，是否參雜個人喜好與情緒，影響表情的變化。並且藉此觀察世界各主播的面部表情，是否因為文化與教育的不同，產生明顯的差異性。

三、實作步驟與結果

本計劃是運用 Active Appearance Models (AAM) [4]進行五官定位，再依據各定位點(feature points)的座標位置，進而得到新聞記者的表情變化與身體晃動情形。我們提出的系統架構如圖二所示：

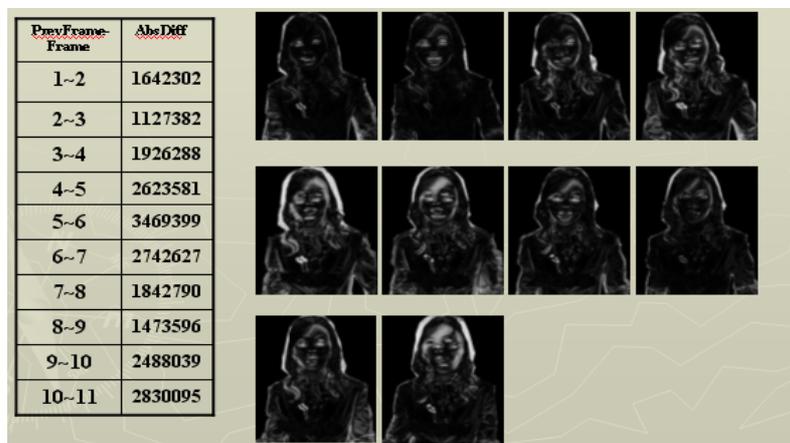


圖二、系統架構

1. 影像擷取

為了估測視訊影片中臉部變化速度與大小，用來決定最終的擷取頻率決定，做了以下的實驗（如圖三）：

收集新聞影片資料後，以每秒 30 張的頻率進行擷取，然後將後張減前張，得到前後變化量的大小。觀測以下的結果後，我們發現雖然 frame 與 frame 之間只相差了 1/30 秒，但新聞主播的表情變化與晃動量都已經很大，而且有些變化情形需要由一連串的座標變化才能精確感測，因此在本系統中我們決定以每秒 25~30 張做為擷取的畫格數。



圖三、連續畫格之變化量

2. 前處理

在做過多次 AAM System 後，發現背景常影響偵測的結果，造成偵測失敗(圖四(a))，連續畫面中單張偵測的失敗將造成後續一連串的錯誤。因此雖然在「錯誤偵測」過程中可把這些失敗的例子挑出，不做下一部分的分析，也不以此結果做為重建模組的參考；然而假使連續的錯誤省略，將可能無法判定出那段時間的表情變化。因此為了提高 AAM System 的正確率，做了去背景的前處理。為了解此步驟的效果，做出以下的實驗。

圖四(a)與圖四(b)為同一時刻的畫面，但在下列實驗中圖四(a)因為受到背景中有類似臉頰的形狀，因此造成偵測上的錯誤，如果能先切割(crop)出此畫面中臉部範圍，再進行偵測，發現準確率將提升許多(圖四(b))。



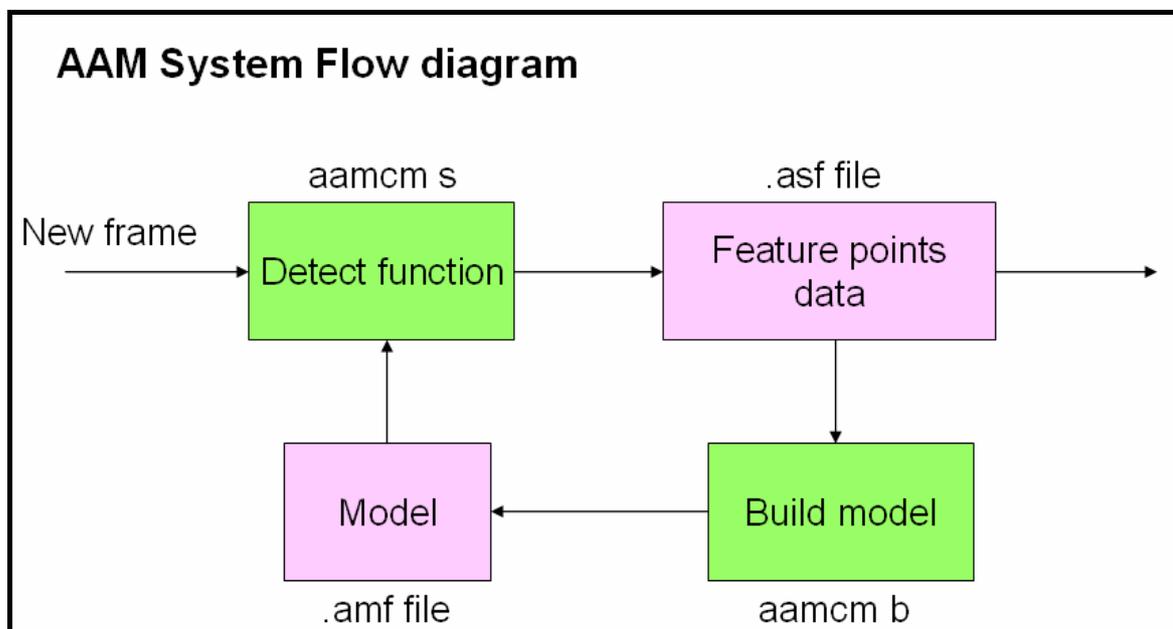
圖四 (a)雜亂背景影響偵測結果 (b) 使用人臉偵測技術先切出臉部後再套用 AAM 結果

因此在前處理上將採用 [3] 所提及的過程與演算法。偵測到臉部，再進行切割(crop)，此種方法在正面臉部偵測(frontal face detection)有很高的準確率。

3. AAM System 偵測

現有的 AAM 函式庫每次只處理一個畫面，也因此我們使用 MATLAB 外部呼叫 AAM System 以達到系統的自動化分析功能，偵測流程如(圖五)，利用新偵測出來的結果，重做出最新的模組，而每次所套用的模組將由「模組資料庫」所提供，模組資料庫會從「錯誤偵測」的結果考慮是否套用最新產生的模組，因為由以下實驗得知，偵測錯誤會導致產生的新模組正確率降低。如(圖六)是同一個畫面，左圖採用系統最初的模組偵測，而右圖則是套用最新產生的模組，但從結果看來，反而是套用一開始的模組產生的結果比較好，在此次經驗後，我們往回查看，原來是因為前一張受到背景所影響，造成偵測結果不佳，使得接下來重建的模組正確度降低。然而一般認為越新的模組所偵測的結果應該會較準確，但事實上由於新模組的一點誤差將會影響接下來的偵測，進而使接下來所產生的

模組錯誤率放大，連續的錯誤將造成系統整個失敗（圖七）。因此接下來的「錯誤偵測」將扮演很重要的角色，決定是否採用新的模組與評估五官特徵點(feature points)是否準確而值得拿來做後續的分析工作。



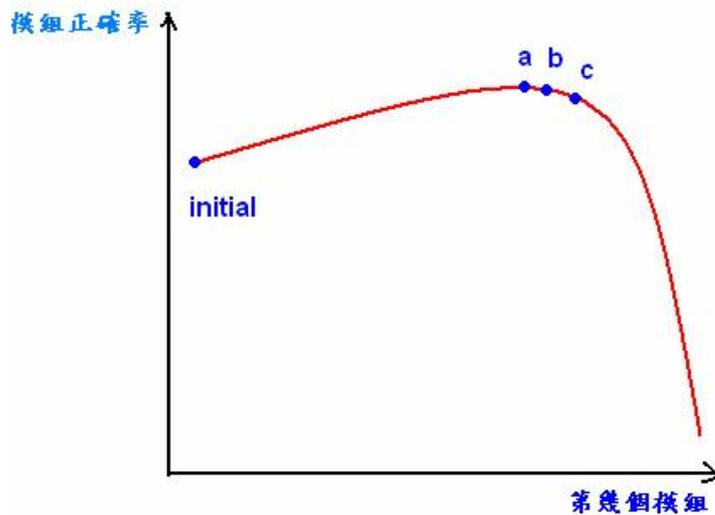
圖五、AAM 偵測流程圖



(a)套用最初的模組

(b)套用最新的模組

圖六、(b)套用新模組的偵測準確度還不及(a)套用原模組

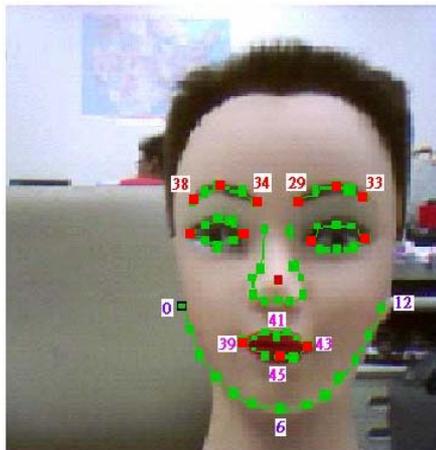


圖七、當模組正確率緩慢提升到點 a 時，此時可能因單張偵測錯誤造成接下來模組正確率急速下降

4. 校正與錯誤偵測

4-1. 校正

為了提升「臉部表情與晃動判定」的準確度，我們將幾個關鍵的表情判定點進行校正工作，因為這些點對於之後的表情分析有較重要的指標性作用，例如：嘴角對於未來嘴型的判斷是相對重要。當此張圖經過「AAM System 偵測」後，我們取用（圖八）紅色各點與其周圍八點進行 corner detection [6]，選擇這九點 corner 強度最強為此點的座標位置。此座標也影響之後的「錯誤偵測」。



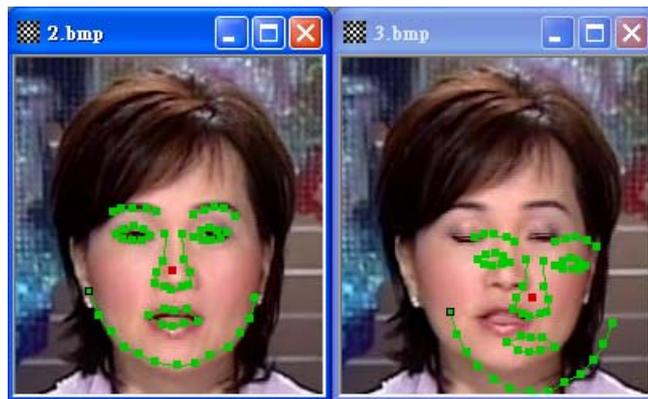
圖八、校正時，取用紅點進行 corner detection

4-2. 錯誤偵測

在測試許多資料後，我們發現當 AAM System 偵測有問題時，大部分都錯得

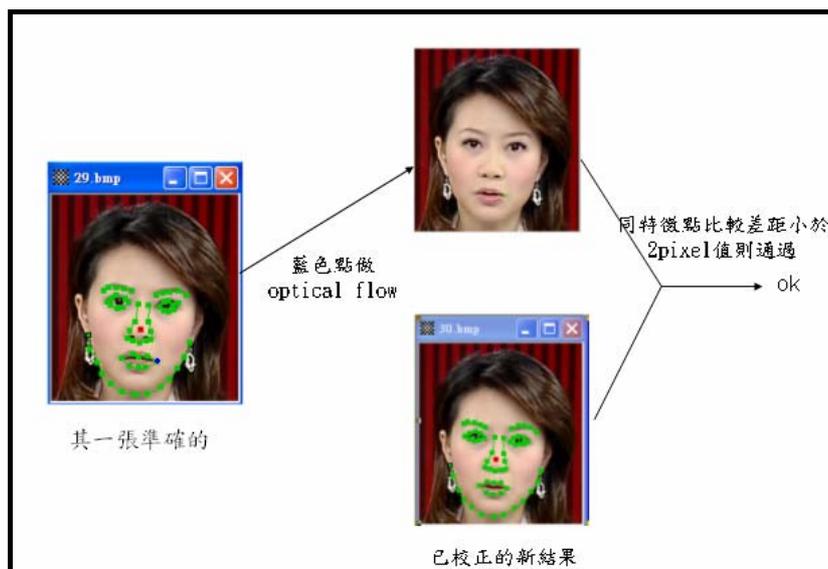
相當離譜¹，而且新聞記者的臉部位置在不到一秒的時間內並不會改變很多，因此我們可以借由五官特徵點的座標判定此張的錯誤率。當通過錯誤偵測後再送往「重建模組」與「特徵點分析」。

方法一（圖九），是拿最簡當的重心座標當做依據，因為每張 frame 之間的時間極短，臉部位置並不會離開很多。依據此張的重心位置是否已經遠離前一張的重心，利用方法一判決此類的錯誤，已得到不錯的偵測率。



圖九、重心遠離情形，選擇方法一來偵測錯誤

方法二（圖十），一些誤差不大的錯誤無法以方法一偵測，此時可以 optical flow 來解決這種情況，由於將整張圖去做 optical flow [5]會花較多的時間，而且在錯誤偵測上也不需要，所以選取前一張的一些邊緣點進行 sparse optical flow，例如；眉角、嘴角、眼角等邊緣點，將前張此點的變化情形與此張校正後得到的座標的相減，AAM System 對這些點的偵測已較為準確，如果這些點差距均大於幾個 pixel 值時，就可視為錯誤發生。



圖十、方法二之流程示範

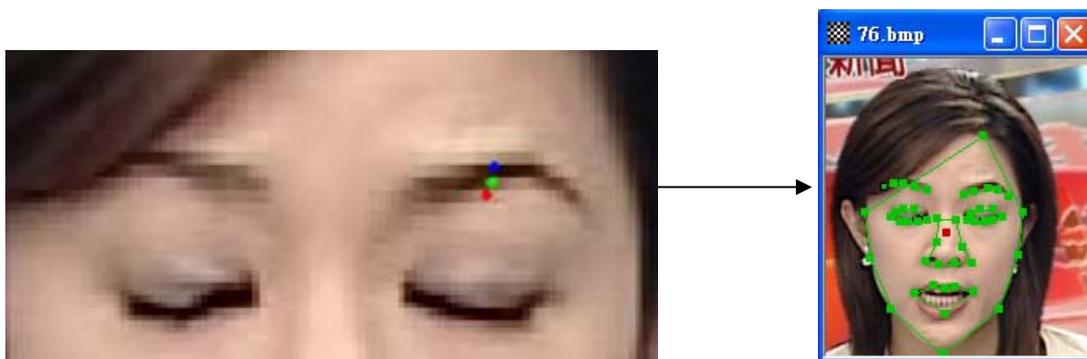
¹ 由於 AAM 使用 local search，因此起始點若差距太多，就有可能收斂到錯誤的結果。

然而如果有一連串的此類錯誤發生，可能代表這時期的某個五官位置移動過大，與下一張過大的差距將造成 AAM 這類 Model-Based 的偵測方法產生較大誤差。例如：當主播的眉型在短時間內起伏過大，會造成眉毛單獨區域偵測的不準，如圖十一。



圖十一、連續兩張眉型變化過大，造成 AAM 在眉毛的偵測上產生誤

如果我們依然以方法二將上述（圖十一）僅因單獨區域的偵測誤差判斷為錯誤發生而省略分析，將無法得到之後的挑眉結果。因此我們試著利用剛剛做過 optical flow 的結果，進行修正，再採用修正之後的五官定位點（圖十二）重建造成新的模組，經過修正產生的模組將更能有效地抓出接下來的眉型變化。



圖十二、綠點為圖十一左圖的偵測位置。紅點為圖十一右圖的偵測的位置，藍點為透過 optical flow 的修正結果，將眉毛各點以此方式修正後得到箭頭右方的五官定位，明顯改善了圖十一右圖的偵測結果，再採用這些定位點重新建造出新的模組。

5. 臉部表情與晃動判定

「臉部表情與晃動判定」部分，乃與本校傳播學院新聞系合作，由新聞系專業的教授們評判出那一些屬於不專業的表情動作，例如：有參雜主觀意見和情緒於其中，或者過份誇張等；當然如果資料收集更多時，可以使用機器學習理論找出一個適當的門檻值，做為判斷那些屬於自然表現的範圍。

5-1. 單張的判定

在觀察主播臉部畫面時，有些表情是可由單張的資料就可感覺的到，因此在

定義此類表情時，我們只需由取用單張的特徵點就可以。這種判定的類型如下表：

原因	偵測方式
一、眉形的角度	由特徵點測得
二、不過份誇張	測量笑容大小與臉部的比例
三、臉部是否偏向	由特徵點偵測臉部方向

★原因一、眉型角度

AAM System 通常在眉稍本來就較具準，加上校正的過程，因此準確度應足夠從透過 29、31、33 三點，將眉型找到（圖十三、表十三）。



圖十三、眉形變化的偵測

	29 至 31			31 至 33		
	ΔX	ΔY	彎曲的角度	ΔX	ΔY	彎曲的角度
左圖	11.69	3.61	15.88	10.94	-4.54	-22.57
右圖	12.86	3.36	14.66	12.80	-2.27	-10.06

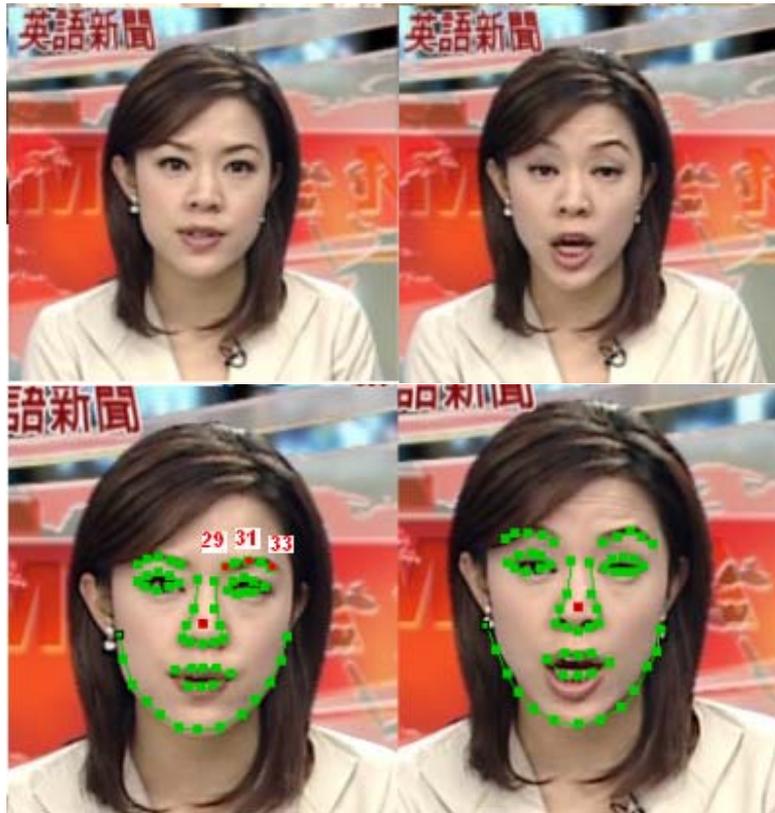
表十三

★原因二、不過份誇張

為了從五官的座標點偵察出五官的變化情形，將採行下列三項標準作為自動判斷的依據。

- (1) 笑容的大小
- (2) 五官位置的比例
- (3) 眉毛的角度

請參考下列範例（圖十四與表十四）



圖十四、右邊表情明顯比左邊誇張

標準	圖例	左圖		右圖	
		(1) 笑容的大小/平均笑容大小	0.9 倍		1.02 倍
(2) 眉毛與眼睛差距/平均差	0.87 倍		1.32 倍		
(3) 眉毛的角度(度)	左半	右半	左半	右半	
	17.56	20.47	22.95	31.16	

表十四

因此由這些資料我們可以了解，右圖的表情誇張度應該是較左圖強烈，在這裡的 training data 的標準就交由新聞系老師再進行判定，之後我們在由此標準定出偵測的判定值(threshold)。

★原因三、臉部是否偏向

臉部是否有偏向，可檢查一些比較不隨臉部表情變化的特徵點而發現，另外最好取用一些準確性較高的點，作為偵測對象，因此做了以下的實驗。取用已經校正過的眉梢兩點，當做依據點(圖十五)，而得到以下的結果(表十五):

由此實驗得知，只要特徵點誤差不高，對於水平轉動偵測的正確率應該能夠得到滿足。



圖十五、左邊兩圖為未偵測前，第二張有明顯的偏向，第三與第四張為偵測的結果，數據如下

node frame	29		34		29-34		臉轉動 的角度
	X	y	x	y	x	y	
左圖	218.28	193.089	133.36	133.09	25.19	0.26	0.60
右圖	245.98	163.93	223.07	159.38	23.91	4.56	10.79

表十五

5-2. 連續多張的判定

有些臉部變化是無法從單張畫面中判定出來的，例如：大範圍的臉部移動，就單一 frame 只能知道目前各特徵點的位置，所以這類的改變需要由連續幾張的偵測才可辨識出來。這種判定的類型如下表:

原因	偵測方法
1.表情刻意加強	由相同方向移動累積量與正常的累積量做比較
2.臉部無故晃動	

連續的例子如下(圖十六、表十六)：



圖十六、連續的晃動情形，偵測情形如下表

node frame	29		34		29-34		臉轉動 的角度
	x	y	x	y	x	y	
1	379.113	214.958	458.973	215.538	79.860	0.580	0.417
2	351.778	216.767	432.162	213.372	80.385	-3.395	-2.420
3	335.098	217.622	414.508	211.002	79.411	-6.620	-4.768
4	323.776	224.640	402.036	216.009	78.260	-8.630	-6.296
5	374.242	217.972	452.693	218.344	78.452	0.372	0.272
6	406.744	199.898	485.762	205.674	79.018	5.776	4.183
7	420.622	198.186	499.585	206.976	78.963	8.790	6.355
8	440.724	188.590	519.806	200.990	79.082	12.401	8.916

表十六

5-3. 變化程度由量化顏色表示法

變化程度往往是一件很抽象感覺，例如：常常有人認為某位記者笑的很誇張，但是到底以多誇張，我們仍然無法以一個具體的量化資料來判斷；在此研究中我們可以利用已偵測坐標進行變化程度的量化顏色表示法。在進行之前，我們要先選取一張標準圖(standard image)，此標準圖應選擇無表情或自然的臉部表現，再將所有的 frame 與它比較，經由比較後的倍數數字，做為量化表示的依據。如果由不同的新聞畫面選擇標準圖，需要先標準化成將被偵測的大小，才不會造成誤差。量化顏色表示法的演算法如下：

```

void 量化顏色表示法( standard_image 採用的標準圖 ,threshold1 ,threshold2 )
{
FOR i= 第一個 frame : 最後一個 frame
case 1 : 眉毛的量化比較
times(倍數)= i.眉毛高度 / standard_image. 眉毛高度 ;
case 2 : 嘴巴張大的量化比較
times(倍數)= i.嘴型大小 / standard_image.嘴型大小 ;
case 3 : .....其他的五官比較

if |times| < 誤差值
no-op; //此狀態視為沒有變化，此點的顏色依然為綠色
else if times >= threshold1
{ 放大，擴張狀態
視放大的程度將此點由粉紅->深紅 }
else if times <= threshold2
{ 縮小，降低狀態
視所小的程度將此點由淺藍->深藍 }
}

```

量化顏色表示法

圖十七為量化顏色表示法的 case 1 的範例：



圖十七、 量化顏色表示法

四、討論

此研究主要使用 AAM system 做為臉部偵測的工具，我們將 AAM 原來單張的偵測以迴圈的操作方式，讓它可以追蹤連續的臉部畫面，但是 AAM 仍然有許多特性與缺失需要去改進，因此在此提出討論，並說明我們提出的解決方案。

- 4.1. AAM 容易受背景的影響：使用人臉偵測技術，在前處理階段將之解決。
- 4.2. 模組的製作會有誤差累積放大的情形：由錯誤偵測的方式，及時發現誤差

的出現，再由校正的方式減少誤差或者乾脆採用更前面的模組重新偵測，其目的就是為了解決誤差放大與模組準確度急速崩解的情況。

- 4.3. 對於 AAM 這種 model-based 的偵測法，由於是屬於 local search，只能偵測局部的範圍，但如過對於過大的變化，或者由小擴張到大的五官變化仍無法追蹤準確：我們在此情況中，採用 optical flow 進行輔助修正，解決了 AAM 此類的缺失，並得到不錯的成效。

五、結語

在這次的專題中最常遇到的困難就是五官特徵點(feature points)定位不夠準確，造成接下來自動化的判定的誤差很大，但是在測一些臉部影片由自己所拍攝的結果就相對準確許多，所以很有可能是由於現在的新聞主播臉部影像過小造成像數過低影響了偵測的準確度，常常變化只在於幾個 pixel 之內，但與新聞系經過討論後，認為此系統可應用於輔助主播表情訓練上，由於訓練時，我們擴大臉部的影像範圍，當面部解析度增大後，系統的準確度將足夠自動挑出一串影片的某一片段，此時主播臉部出現特別的變化，此功能對於輔助主播訓練將有很大的實用價值。

在未來來的工作中，會為五官變化定出一些標準的判斷範圍，例如：眉毛彎曲角度超過幾度就屬於表情變化。在單獨定義這些五官變化後，會試著與表情做出連結[7]。例如：在實驗中觀察到挑眉的動作常伴隨著眼睛睜大，這樣的臉部動作是否屬於驚訝的表情，因此希望能透過多種五官的變化發現某些情緒表情。

六、參考文獻

- [1] 鍾汝德, "當新聞主播遇上虛擬主播－數位化時代新聞主播的定位與未來", 政治大學碩士論文
- [2] "不用口紅的數位主播"《數位時代》
http://www.bnext.com.tw/mag/2000_03/2000_03_562.html
- [3] Paul Viola and Michael Jones, "Robust Real-Time Face Detection", International Journal of Computer Vision 57(2), 137-154, 2004.
- [4] T.F. Cootes, G.J. Edwards, C.J. Taylor, "Active Appearance Models (1998)", Lecture Notes in Computer Science, <http://www2.imm.dtu.dk/~aam/>
- [5] B. K. P. Horn and B. G. Schunck. Determining optical flow. Artificial Intelligence, 17:185-203, 1981
- [6] C. Harris and M. Stephens. A combined corner and edge detector. In Alvey Vision Conference, pp. 147-151, 1988
- [7] Facial Action Coding System (FACS).
http://face-and-emotion.com/dataface/facs/new_version.jsp