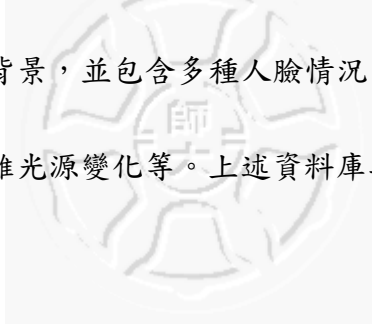


## 第四章 實驗結果與討論

本研究以兩個影像資料庫，以驗證本研究所提出人臉偵測演算法之準確度及執行效率。4.1 節為實驗影像來源；4.2 節為模型訓練；4.3 節為參數調整；4.4 節為測試資料庫之實驗結果；4.5 節與其它系統比較；4.6 節將對實驗結果進行討論。

### 4.1 實驗影像來源

本實驗之第一個實驗影像來源為 BioID 人臉資料庫(Test Set 1)[39]，此資料庫為公開且常用於不同人臉偵測演算法間之比較，共 1,521 張影像，第二個實驗影像資料庫(Test Set 2)來源為從攝影機所擷取之畫面，共 535 張影像進行偵測，上述實驗影像大部分為複雜背景，並包含多種人臉情況，如：正面、接近正面、傾斜、轉頭、臉部表情和複雜光源變化等。上述資料庫之實驗影像並不包含訓練影像。



### 4.2 模型訓練

本研究以高斯混合模型及數個粗糙分類器將輸入之視窗進行分類，梯度分佈模型中，需以人臉及非人臉之影像資料訓練高斯混合模型，人臉影像從日常生活照片蒐集而來，而非人臉影像從網路及風景照片隨機蒐集而來，為增加訓練資料量，每張人臉影像皆經過鏡射且平移後，共使用 1,966 張人臉及 4,682 張非人臉影像產生第一個訓練影像資料庫(Training Set 1, TS1)，並將每張影像正規化至 32×32 像素，用於訓練人臉及非人臉之高斯混合模型。

本研究之第二階段，串連七個梯度空間相關性模型進行精確分類，此部分之模型訓練參考 Sung 和 Poggio 之方法[15]，收集人臉及非人臉高斯混合模型之誤判影像後，加上第一個訓練影像資料庫產生第二個訓練影像資料庫(Training Set 2, TS2)，用於學習串連分類器中第一個梯度空間相關性模型(SC<sub>1</sub>)，第二個梯度空間相關性模型(SC<sub>2</sub>)也使用第二個訓練影像資料庫進行學習，其差別為 SC<sub>1</sub> 使用 10 個粗糙分類器而 SC<sub>2</sub> 使用 450 個粗糙分類器進行學習。以第 N 個梯度空間相關性模型為例(SC<sub>N</sub>, N>2)，其訓練資料庫(Training Set N, TSN)為收集 SC<sub>1</sub>, SC<sub>2</sub>, ..., SC<sub>N-1</sub> 之誤判影像而來，本實驗總計使用七個訓練影像資料庫，共使用 1,966 張人臉及 18,402 張非人臉影像訓練七個梯度空間相關性模型。



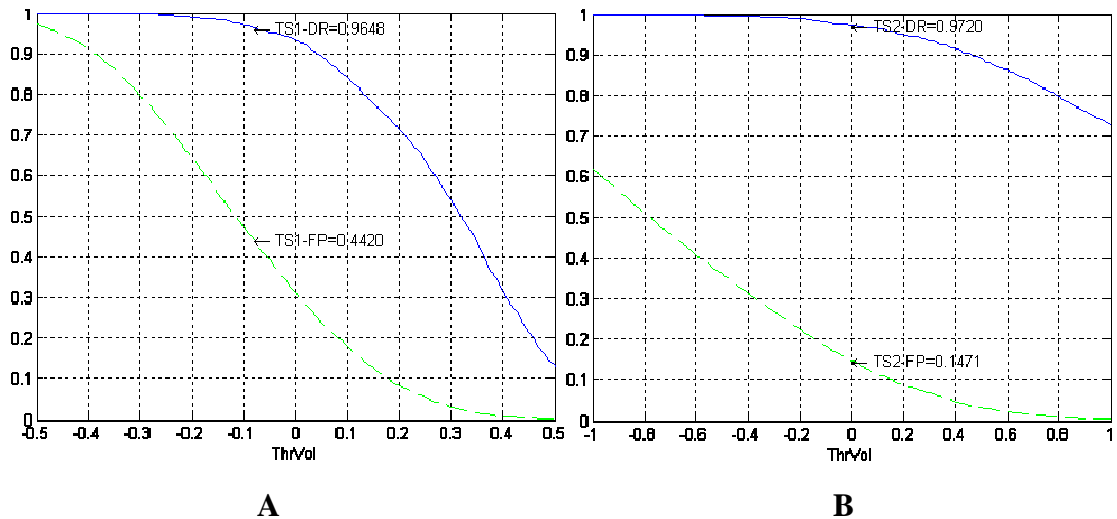
## 4.3 參數調整

### 4.3.1 高斯分佈數目

人臉及非人臉之梯度分佈以兩個高斯混合模型組成，每個高斯混合模型由數個高斯分佈之線性組合而成，高斯分佈的數目決定第一階段之偵測率及誤判率，

表二：高斯分佈數目與偵測率和誤判率之關係

#Gaussian	DR	FP	
2	0.7396	0.2879	
4	0.8174	0.1633	
8	0.8578	0.1208	#Gaussian：高斯分佈數目
12	0.8632	0.1178	DR：偵測率
16	0.8758	0.0951	FP：誤判率



圖十四：實線為偵測率與門檻值曲線，虛線為誤判率和門檻值曲線，當  $SC_1$  之門檻值  $ThrVol=-0.08$  時，在訓練影像資料庫(TS1)偵測率(TS1-DR)和誤判率(TS1-FP)各為 0.9648 和 0.4429，而當  $SC_2$  之門檻值  $ThrVol=0.00$  時，在訓練影像資料庫(TS2)偵測率(TS2-DR)和誤判率(TS2-FP)各為 0.9720 和 0.1471。

由表二可知，若以 TS1 作為高斯分佈數目調整之依據，超過 8 個高斯分佈對偵測率並無顯著提升，因此本實驗以 8 個高斯分佈(M=8)訓練人臉高斯混合模型。而超過 12 個高斯分佈對誤判率下降並無顯著幫助，因此本實驗以 12 個高斯分佈訓練高斯混合模型(M=12)，3.2 節為高斯混合模型之訓練方法。

#### 4.3.2 門檻值與誤判率之關係

本研究串連七個梯度空間相關性模型進行人臉偵測，如節 3.5 所述，每個模型需選擇適合之門檻值，以保留人臉視窗並移除誤判視窗。圖十四 A 為  $SC_1$  以不同的門檻值將 TS1 進行分類之結果，由此可知  $SC_1$  之門檻值( $ThrVol$ )為-0.08 時，偵測率(TS1-DR)和誤判率(TS1-FP)各為 0.9648 及 0.4420，此門檻值之意義為在 TS1 中，約 96% 之人臉訓練影像經過  $SC_1$  分類仍可正確判斷，圖十四 B 為  $SC_2$  以門檻值 0.00 將 TS2 進行分類時，約 97% 之人臉影像經過  $SC_2$  分類仍可正確判斷，

表三：SC<sub>1</sub> 至 SC<sub>7</sub> 的分類器數目、門檻值、偵測率、誤判率和整體誤判率表

	#WC	ThrVol	TSn-DR	TSn-FP	Overall FP
SC <sub>1</sub>	10	-0.08	0.9648	0.4420	0.5413
SC <sub>2</sub>	450	0.00	0.9720	0.1471	0.2355
SC <sub>3</sub>	474	-0.03	0.9771	0.1342	0.1273
SC <sub>4</sub>	410	-0.13	0.9725	0.3815	0.0919
SC <sub>5</sub>	411	-0.14	0.9908	0.5569	0.0744
SC <sub>6</sub>	267	-0.17	0.9913	0.4617	0.0542
SC <sub>7</sub>	250	-1.21	0.9903	0.5846	0.0454

#WC：粗糙分類器數目

ThrVol：SC<sub>1</sub> 至 SC<sub>7</sub> 之門檻值

TSn-DR：SC<sub>n</sub> 於訓練影像資料庫 TSn 之偵測率

TSn-FP：SC<sub>n</sub> 於訓練影像資料庫 TSn 之誤判率

Overall FP：串連梯度空間相關性模型之整體誤判率

如表三所示，SC<sub>1</sub> 至 SC<sub>7</sub> 門檻值之選擇以偵測率高為主，目的為提高系統偵測率。

表三中 TSn-DR 和 TSn-FP 欄位為單一梯度空間性模型(SC<sub>n</sub>)在其訓練影像資料庫(TSn)之偵測率與誤判率。以 SC<sub>1</sub> 為例，SC<sub>1</sub> 僅有 10 個粗糙分類器，且在 SC<sub>1</sub> 之訓練影像資料庫 TS1 達到 0.9648 的偵測率和 0.4420 的誤判率，意義為 SC<sub>1</sub> 僅以 10 個粗糙分類器即可降低誤判率至 0.4420，大幅降低後續分類器之運算量。

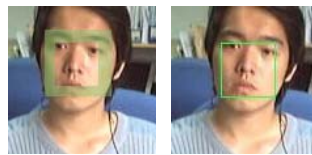
表三之 Overall FP 欄位為串連數個梯度空間相關性模型之整體誤判率。以 SC<sub>2</sub> 為例，整體誤判率 0.2355 意為串連 SC<sub>1</sub> 及 SC<sub>2</sub> 兩個模型在七個訓練影像資料庫  $\bigcup_{i=1...7} \text{TS}_i$  的誤判率，而 SC<sub>3</sub> 之整體誤判率為 0.1273，意義為串連 SC<sub>1</sub>、SC<sub>2</sub> 及 SC<sub>3</sub>

三個模型在七個訓練影像資料庫的誤判率，由表三可知每多串連一個模型可有效降低整體誤判率，並維持偵測率在 96% 以上。4.4 節之實驗參數皆以表三為準。

## 4.4 實驗驗證

### 4.4.1 成功偵測區域

本研究定義若某人臉被成功偵測，則此偵測視窗最大不能超過從眉毛至嘴巴所形成的正方形內，且不能小於從眼睛至嘴巴所形成的正方形內，如圖十五 A 所示，綠色區域為本研究可接受之成功偵測區域，圖十五 B 為成功偵測範例。



A

B

圖十五：圖 A 綠色區域為本研究可接受之成功偵測區域，圖 B 為成功偵測範例。

### 4.4.1 Test Set 1 實驗結果

Test Set 1 為 BioID 影像資料庫，共 1521 張影像總計 1522 張人臉，姿勢皆為正面，影像背景單純且較少光源變化，此資料庫測試重點在不同人臉表情、人種及遮蔽物(眼睛、髮型及鬍鬚等)對系統偵測率誤判率之影響。

本實驗之偵測率為 0.9146，1521 張影像經第一階段處理後總計 12,777,921 個



圖十六：以 Test Set 1 進行人臉偵測結果範例。



圖十七：以 Test Set 2 進行人臉偵測結果範例。

視窗中誤判 25 個視窗，誤判率約為  $1.9564 \times 10^{-6}$ ，偵測結果範例於圖十六。

#### 4.4.2 Test Set 2 實驗結果

Test Set 2 共 534 張影像，每張影像人臉數目不固定，背景較 BioID 複雜且光源方向不穩定，如正光、側光和逆光等多種變化，同時臉部可能為正面、接近正面、傾斜、部分遮蔽(眼鏡和手指等)及不同表情，總計 574 張人臉，此實驗之測試重點為不同光源和人臉姿勢下，對系統偵測率及誤判率影響。

本實驗之偵測率為 0.9512，534 張影像經第一階段處理後總計 3,929,040 個視窗中僅誤判 4 個視窗，誤判率約為  $1.0180 \times 10^{-6}$ ，偵測結果範例於圖十七。

#### 4.4.3 偵測速度實驗

本研究以 Test Set 2 資料庫測試系統速度，此資料庫中每張影像大小皆為  $320 \times 240$ ，在 Pentium M 1.5GHz 筆記型電腦上，平均每張影像處理時間為 102ms，約每秒可處理十張影像，因此本研究所提出之方法可即時偵測人臉。

### 4.5 與其它系統比較

表四：不同系統間之比較

	Test Set 1 (BioID)		Test Set 2		FPS (Test on 320x240 image)
	偵測率	誤判數目	偵測率	誤判數目	
H. A. Rowely et al.[4]	0.9835	102	0.9721	111	0.13 FPS (200MHz R4400 SGI Indigo 2)
BuFaLo[38]	0.9454	47	0.6324	15	18 FPS (Pentium M 1.5GHz)
B. Fröba and C. Küblbeck[12]	0.8450	16	0.9703	17	16 FPS (Pentium IV 2.0GHz)
本系統	0.9146	25	0.9512	4	10 FPS (Pentium M 1.5GHz)

FPS：Frame Per Second，每秒處理張數

H. A. Rowley 等人提出以類神經網路建立人臉偵測系統[4]，對於較差的影像品質仍能保持高偵測率，但卻產生較高的誤判率，如表四所示，其方法在 Test Set 2 中有高達 111 個錯誤偵測，如圖十八所示，錯誤偵測偏高原因為類神經網路對複雜圖形較為敏感，本系統於 Test Set 2 中雖偵測率雖較低，但偵測速度卻約快數倍，較適合應用於即時需求的系統中。

BuFaLo(Base-unit for Face Localization)系統實做 P. Viola 和 M. Jones 提出的人臉偵測演算法[38]，其方法主要以區塊間強度差異進行偵測，如表四所示，其方法在 Test Set 1 中偵測率可達 0.9454，偵測速度上皆優於表四中所有系統，但在光源變化較大的測試資料庫 Test Set 2 中之偵測率僅有 0.6324，如圖十八所示，其偵測率極低之原因為人臉區塊強度差異對光線變化較為敏感，於複雜光源環境較不適合。

B. Fröbac 和 C. Küblbeck 提出以邊緣方向樣版進行人臉偵測[12]，其方法可快速且能應用於複雜光源環境，於 Test Set 2 中偵測率高達 0.9703，但也產生較多的錯誤偵測，於 Test Set 2 中共有 17 個錯誤偵測之處，圖十八為誤判範例。本系統之偵測率於 Test Set 2 為 0.9512，偵測率略低但僅有 4 個錯誤偵測。



A

B

C

圖十八：圖 A 說明類神經網路對影像中複雜圖形較為敏感[4]，易造成誤判。圖 B 說明 BaFaLo 系統未能偵測影像中光源變化較大的人臉[38]。圖 C 說明邊緣方向樣版所造成的誤判範例[12]。

## 4.6 討論

本研究將人臉偵測分為兩階段，第一階段以人臉及非人臉梯度分佈之高斯混合模型，並使用動態間隔偵測法，減少影像中所需偵測的視窗數目，如圖十八所示，以  $32 \times 32$  之視窗等比放大 13 次掃瞄一張  $458 \times 342$  影像，約有  $1.8804 \times 10^6$  個視窗需進行偵測，動態間隔偵測法和高斯混合模型分類判斷後，將所需掃瞄視窗數目降至 18,530 個，約為原始影像總視窗數目的 0.99%。

初步偵測後，第二階段串連七個以梯度空間相關性模型( $SC_1, SC_2, \dots, SC_7$ )，總計 2,272 個粗糙分類器，將上階段可能為人臉之視窗進一步分類， $SC_1$  需掃瞄上階段被分類為人臉共 18,530 個視窗，而  $SC_2$  需掃瞄被  $SC_1$  分類為人臉之視窗共



	需掃瞄 視窗數目	估原始影像 視窗數目比例
原始影像	$1.8804 \times 10^6$	1.000
動態間隔法+高 斯混合模型	18,530	$\frac{18530}{1.8804 \times 10^6} = 0.0099$
SC <sub>1</sub>	18,530	$\frac{18530}{1.8804 \times 10^6} = 0.0099$
SC <sub>2</sub>	646	$\frac{646}{1.8804 \times 10^6} = 3.4354 \times 10^{-4}$
SC <sub>3</sub>	22	$\frac{22}{1.8804 \times 10^6} = 1.1700 \times 10^{-5}$
SC <sub>4</sub>	5	$\frac{5}{1.8804 \times 10^6} = 2.6590 \times 10^{-6}$
SC <sub>5</sub>	3	$\frac{3}{1.8804 \times 10^6} = 1.5954 \times 10^{-6}$
SC <sub>6</sub>	2	$\frac{2}{1.8804 \times 10^6} = 1.0636 \times 10^{-6}$
SC <sub>7</sub>	1	$\frac{1}{1.8804 \times 10^6} = 5.3180 \times 10^{-7}$

圖十八：每階段分類器所需掃瞄視窗數目及比例

646 個視窗，由此可知 SC<sub>1</sub> 僅需 10 個粗糙分類器即可移除上階段

96% ( $\frac{18530 - 646}{18530} = 0.9651$ ) 以上的誤判視窗，有效減少後續 SC<sub>2</sub> 至 SC<sub>7</sub> 所需掃瞄之視

窗數目。圖十九所示，當臉部極度黑暗及遮蔽過大，或畫面中包含與人臉梯度分

佈相似之物體時，本系統可能會產生未偵測到人臉或誤判之情形。



圖十九：本系統未偵測到人臉及誤判範例影像。

