

# 應用於學習者專注力分析的 人臉偵測法之研究

## 張循鋁

助理教授  
國立臺北商業大學資訊管理系  
E-mail: hsunli@ntub.edu.tw

## 林開榮

副教授  
育達科技大學資訊管理系暨研究所  
E-mail: linky@ydu.edu.tw

## 余誌強

育達科技大學資訊管理系暨研究所  
E-mail: ta0614@hotmail.com

## 摘要

專注力與學習成效息息相關，所以專注力的測量格外重要，在眾多的量測方式中，以結合人臉偵測技術的方式較客觀且不干擾學生學習。本文提出結合擴增式哈爾材質特徵 (Extended Haar Texture Features, EHTF) 與 EHTF-Boosting 學習演算法的人臉偵測方法，EHTF 定義了 6 種不同的矩形特徵樣版，套用於影像上可描述局部區域的邊緣、線性、方向與亮 (暗) 點等特徵，具有求算快速與不易受光照變化影響的優點，將人臉與非人臉的訓練用影像套用 EHTF 後，利用 EHTF-Boosting 學習演算法，建構數個弱分類器 (Weak Classifier) 以形成一個強分類器 (Strong Classifier)，並針對求算每一弱分類器的權重 (Weight) 求算方式進行改良以正確反映出每一弱分類器的重要性。以 LFW-Bigfoto 資料集中各 2,000 張正面人臉影像與不包含人臉的背景影像作為訓練集，實作分類器訓練並以 MIT-CMU 資料集作測試，檢測結果顯示本文所提出之 EHTF 特徵與 EHTF-Boosting 學習演算法相較於其他方法，顯著提升了人臉的偵測率 (Detection Rate) 與降低誤判率 (False Positive Rate)，並擁有使用特徵數量少、訓練速度快

與人臉偵測率較佳等優點，且可滿足即時偵測的應用需求。在教學環境應用中，本文所提出之方法可用於數位學習之學習成效分析、智慧教室即時點名、利用臉部表情診斷學習困難度、教室秩序監控等不同應用領域，讓相關分析資料的取得更加容易與客觀。

**關鍵字：**專注力、人臉偵測、AdaBoost、EHTF-Boosting。

## 壹、緒論

隨著數位時代的來臨，老師的教學方式愈來愈多元，除傳統課堂講述外，透過科技網路等數位方式進行教學亦常見，如應用媒體及網路科技的數位學習，或教師透過影音、網路、互動遊戲等引導學生學習，已是教學的常態。學生在此學習情境，或許因可接觸數位科技媒體如智慧手機或平板，提高學習動機，但也使老師較難掌握學生在學習過程是否專注。由於專注力與學習的關係非常密切，沒有專注就沒有辨識、學習與記憶（鄭昭明，2006）。所以專注力的測量就變得格外重要，在眾多量測的方式中，又以結合影像處理與辨識技術的方式，較為客觀且不影響學生的學習，如蔡沛勳（2006）首先結合影像處理及辨識技術，利用使用者臉部表情及操作行為為模式所取得的特徵，推論學生在上課時的專心程度。蘇信宏（2007）運用膚色模型進行人臉偵測，在取得人臉位置後更進一步擷取出眼睛的位置與睜閉的狀態，並據此做為評估學生專心程度的依據。王鴻勳（2010）則利用樣板比對法（Template Matching）進行人臉偵測，取得人臉與眼睛位置後，加入了偵測眼睛左顧右盼的機制，透過眼睛移動的軌跡與行為，來判斷學生的上課學習情形。龔喬琳（2012）利用膚色特徵偵測人臉，並搭配眼睛睜閉與嘴部開闔的狀態，來分析學習者的專注程度。Saneiro、Santos、Salmeron-Majadas 與 Boticario（2014）則結合人臉表情與肢體動作的變化，作為學習者專注與情緒分析評估的依據。

許多相關研究均顯示人臉偵測為專注力測量分析資料取得的重要步驟與方法，關係到後續分析的進行。人臉偵測意旨在一給定任意影像中，能夠標示出有人臉存在的位置、大小。近年來，有許多關於人臉偵測的方法被提出，其中由 Viola 與 Jones（2004）所提出運用類哈爾特徵（Haar-Like Features），結合積分影像（Integral Image）、AdaBoost 學習演算法與階層式分類器（Cascaded Classifier）架構的人臉偵測方法，實踐了第一個可滿足即時需求的人臉偵測系統，更奠定往後人臉偵測領域的發展基礎。雖然 Viola 與 Jones 所提出的方法獲得顯著的成功，但所使用的類哈爾特徵數量龐大且穩定性（robustness）不佳，容易受到外在光照變化因素的影響，而導致人臉偵測率的大幅降低，且 AdaBoost 學習演算法訓練過程花費時間甚鉅，因此目前相關的研究均著重於改善 AdaBoost 演算法的學習效率與強化用於描述人臉特徵的穩定性。

Zhang、Chu、Xiang、Laio 與 Li（2007）提出 Multi-Block Local Binary Patterns（MB-LBP）的特徵樣板，以類似九宮格的 3×3 矩形樣版為依據，利用比較相鄰矩形區塊間像素和的大小關係，進行特徵編碼，以減低光照變化所造成的影響，並利用 Gentle AdaBoost 學習演算法改善了訓練的效率，但特徵的求算仍需耗費許多時間。Guo、Yan、Zhao 與 Yang（2008）則提出了哈爾材質特徵（Haar Texture Feature, HTF），於同一影像同一位置的局部矩形區域上同時套用所定義

的 3 種不同 HTF 特徵樣板，並分別求算各特徵樣板的二元化特徵值 (binary feature values)，0 與 1，以形成 3 位元的序列，此 3 位元的序列共可有  $2^3$  種不同形式的特徵編碼。針對給定影像，逐一對訓練集中的人臉影像與非人臉影像中所有的位置進行掃描並同時套用 3 種不同的 HTF 特徵樣板，已取得 3 位元的編碼序列，掃描完畢後可形成 2 個 8-bin 的直方圖 (histogram)，用以表示 8 種不同 HTF 特徵編碼於人臉影像與非人臉影像上的機率分佈情形。給定一檢測影像，針對整張影像用不同大小的子視窗進行掃描，對每一子視窗套用 HTF 以求得特徵編碼，並於人臉與非人臉的直方圖查詢該特徵編碼的機率值，據以對影像子視窗為人臉或非人臉進行分類。HTF 雖可有效降低特徵數量及訓練時間，但所定義的 3 種 HTF 特徵樣板分別只針對水平、垂直與邊緣的特徵進行描述，描述能力有限。因此，HTF 在偵測準確率的提升上仍有很大的改善空間。

本文以 HTF 特徵樣版為基礎，提出擴增式哈爾材質特徵 (Extended Haar Texture Features, EHTF)，對 HTF 進行改善，透過加入水平線性 (horizontal linear)、垂直線性 (vertical linear) 與亮 (暗) 點 (spot) 等共 6 種矩形特徵樣板，改善 HTF 特徵樣版描述能力有限的問題，並將特徵編碼擴增為 64 ( $2^6$ ) 種，形成 64-bin 直方圖，有效提升 EHTF 特徵之鑑別度與描述能力，並透過所提出之 EHTF-Boosing 學習演算法，有效提升學習效率與偵測的準確性。同樣以 LFW-Bigfoto 資料集進行訓練，並以 MIT-CBCL 資料集進行測試的條件下，檢測結果顯示本文所提出之方法，在學習效率、特徵數量需求與人臉偵測準確率上，皆優於 Guo 等所提出之 HTF 方法，除可滿足即時偵測的應用需求外，並可供後續以影像處理與辨識技術為基礎之學習者專注力分析等相關研究使用。

## 貳、文獻探討

過去已有許多人臉偵測相關的研究被提出，利用不同的特徵定義方式，搭配機器學習的演算法以建構出人臉／非人臉的分類器，或稱為人臉偵測器，如：Mosaic Image (Yang & Huang, 1994)、Geometrical Face Model (Jeng, Liao, Liu & Chern, 1996)、Neural Network (Rowley, Baluja & Kanade, 1996)、Template Matching (Miao, Yin, Wang, Shen & Chen, 1999)、Bayesian Discriminating Features (Liu, 2003)、AdaBoost (Viola & Jones, 2004) 和 Support Vector Machines (Waring & Liu, 2005) 等。其中由 Viola 與 Jones 所提出基於 AdaBoost 所建構之人臉偵測器，簡稱為 VJ-Detector，成為第一個可滿足即時需求的人臉偵測器，促使人臉偵測進入了一個新的里程碑。雖然 VJ-Detector 獲得了顯著的成功，但其所採用的類哈爾特徵穩定性不佳，易受到外在因素的影響致使描述能力下降，且因類哈爾特徵之整體特徵集合過於龐大，導致 AdaBoost 學習演算法於訓練

過程耗時甚鉅。因此 Guo 等所提出的 HTF-Detector 即利用 HTF 特徵，針對 VJ-Detector 的缺點進行改善，並獲得顯著的成效。本節將針對 VJ-Detector 與 HTF 特徵進行概略的介紹，並說明其優缺點。

## 一、VJ-Detector

VJ-Detector 採用了簡單的相鄰矩形區塊作為樣板定義，稱為類哈爾特徵，用以描述人臉的特徵，該方法的主要貢獻有三個：（1）提出了積分影像（Integral Image）的計算方法，加速了計算矩形特徵值的運算效率，（2）以 AdaBoost 學習演算法為基礎，提出了一個簡單而有效的分類器訓練方式，（3）提出階層式分類器（Cascaded Classifier）的架構，分層且快速地對影像中非人臉的區域進行篩選，有效縮短掃描影像的時間，以達到即時偵測所需的效能要求。

### （一）、類哈爾特徵（Haar-Like Features）

類哈爾特徵是一種簡單矩形特徵，因類似於哈爾小波（Haar Wavelet）而得名。如圖 1 所示，類哈爾特徵之特徵值的求算方式是黑色矩形和白色矩形在灰階影像子窗口中對應區域的像素總和差異值，亦即白色區塊像素總和減去黑色區塊像素總和，藉此反應出影像局部的灰階變化，這些矩形區塊的大小無固定尺寸，但是白色區塊和黑色區塊的大小必須相同。假定  $i(x,y)$  為給定之灰階影像  $I$  中於  $(x,y)$  座標處的像素值， $f_{haar}$  為矩形樣板特徵值，特徵值計算方式為：

$$f_{haar} = \sum_{x,y \in white} i(x,y) - \sum_{x,y \in black} i(x,y)$$

將圖 1 所示的 5 種類哈爾特徵樣版套用於給定的影像上，如圖 2 所示，把所有可能的樣版大小於所有可能的位置上逐一進行掃描以求算出所有對應的特徵值，依據給定影像的大小，特徵總數有所不同，可以透過表 1 所列之公式計算出影像中所有可能的特徵數量，所有可能的特徵將形成特徵池（feature pool），做為後續建構弱分類器時選取代表性特徵使用。

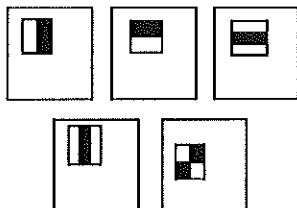


圖 1 類哈爾特徵樣板

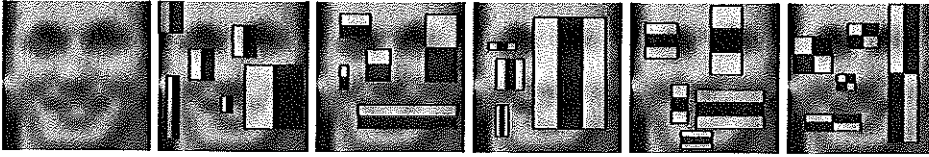


圖 2、將類哈爾特徵樣板套用於給定影像上示意圖。

表 1 各類類哈爾特徵數量計算公式

特徵樣板	公式
	$\sum_{h=1}^{H/2} \sum_{w=1}^W (H-2h+1)(W-w+1)$
	$\sum_{h=1}^{H/2} \sum_{w=1}^{W/2} (H-h+1)(W-2w+1)$
	$\sum_{w=1}^{W/3} \sum_{h=1}^H (W-3w+1)(H-h+1)$
	$\sum_{w=1}^W \sum_{h=1}^{H/3} (W-w+1)(H-3h+1)$
	$\sum_{w=1}^{W/2} \sum_{h=1}^{H/2} (W-2w+1)(H-2h+1)$

表 1 中， $W$  與  $H$  分別為給定影像的寬度與高度， $w$  與  $h$  代表矩形特徵樣板的寬度及高度。若以給定為  $24 \times 24$  的影像為例，共會得到高達 162,336 個特徵值，當給定的影像尺寸越大時，特徵數量相對也就更多，影像尺寸與特徵值數量比較如表 2 所示。

表 2 影像尺寸與特徵值數量比較

影像尺寸	18×18	24×24	30×30	36×36
特徵數量	51,705	162,336	394,725	816,264

在如此龐大的特徵數量之下，矩形特徵值的求算時間面臨著相當大的考驗，因此 Viola 與 Jones 提出了積分影像的運算方法，大幅降低了求算特徵值所需要的時間。

## (二)、積分影像 (Integral Image)

給定一灰階影像  $I$ ，在像素  $i(x,y)$  處的積分影像定義為：

$$ii(x,y) = \sum_{x \leq x, y \leq y} i(x,y)$$

如圖 3 所示，積分影像值  $ii(x,y)$  為陰影區域內所有像素灰階值的和。若要得到影像  $I$  的積分影像  $II$ ，只需對  $I$  所有像素逐一掃描一次即可求得，其公式如下：

$$ii(x,y) = ii(x-1,y) + s(x,y)$$

$$s(x,y) = s(x,y-1) + i(x,y)$$

其中  $s(x,y)=0$  且  $i(0,y)=0$ 。

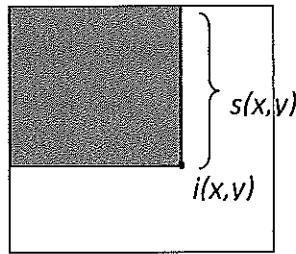


圖 3 給定影像中點  $(x,y)$  的積分影像值

藉由積分影像，可以快速求算任意矩形範圍內之像素值總和，如圖 4 所示，若要求算任一矩形範圍  $ABDC$  內的像素和，則只需將  $A$ 、 $B$ 、 $C$ 、 $D$  等四個範圍邊界於積分影像中對應的像素值取出，並帶入下列公式即可得到該矩形區塊範圍內之像素值總和。

$$s = (A + D) - (B + C)$$

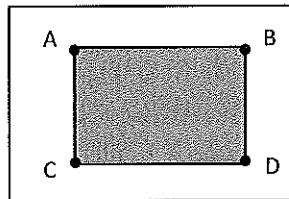


圖 4 利用積分影像快速求算矩形範圍內之像素值總和

給定一原始影像  $I$ ，求算出其對應之積分影像  $II$  後，如前述將類哈爾特徵樣板套用於  $II$  上並進行特徵值求算，基於積分影像的特性，可以快速求得特徵值，

大幅降低了求算特徵值所需要的時間。

### (三) AdaBoost 學習演算法

機器學習是一門人工智慧的科學，相關的方法主要可分為監督式學習 (supervised learning) 與無監督式學習 (unsupervised learning) 兩種。監督式學習從給定的訓練集正樣本與負樣本中，學習出一個函數，亦即分類器。當面對新的樣本時，可以根據訓練所得之分類器來進行分類。監督式學習的訓練集必須包括輸入 (特徵) 和輸出 (目標)，目標需由人為標註與給定；而無監督式學習，訓練集中則無需提供人為標註的目標。AdaBoost (Adaptive Boosting) 是一種監督式學習 (supervised learning) 的疊代演算法，主要精神是從大量的弱分類器 (weak classifier) 中，選取最具有分類意義的數個弱分類器，以組合成一個強分類器 (strong classifier)。其中弱分類器定義為比隨機分類略好一點的分類器，亦即分類正確率略大於 50%。

VJ-Detector 透過積分影像快速求算類哈爾特徵之特徵值，不同大小與位置的矩型特徵，都可構成一個弱分類器。如表 2 所示，以給定為大小為  $24 \times 24$  的影像為例，共會得到高達 162,336 個特徵值，亦即將建構出 162,336 個弱分類器，藉由 AdaBoost 每次疊代的訓練過程中，可從如此龐大數量的弱分器中，選取一個最具有分類意義的弱分類器，亦即分類誤差最小，透過多次的疊代，最後整合選出的數個弱分類器建構成一個強分類器。

給定訓練用影像樣本集  $X$ ， $X = \{x_1, x_2, \dots, x_n\}$ ，與對應之類別標籤  $Y$  (目標)， $Y = \{y_1, y_2, \dots, y_n\}$ ，且  $y_i = \{0, 1 \mid i = 1, \dots, n\}$ ， $y_i = 0$  表示為非人臉，反之則為人臉。對  $X$  中所有影像於同一位置套用某一類哈爾特徵  $f$ ，並求算特徵值  $f(x)$ ， $x \in X$ ，則依據  $f$  所建構之弱分類器  $h$  定義如下：

$$h(x, f, p, \theta) = \begin{cases} 1 & pf(x) < p\theta \\ 0 & \text{otherwise} \end{cases}$$

其中  $\theta$  為閾值， $p \in \{-1, +1\}$  用以指示不等式  $pf(x) < p\theta$  的方向。 $h(x, f, p, \theta) = 1$  表示針對  $x$ ，以  $f$  所求算出之特徵值依據，利用所建構出的弱分類器  $h$  進行分類，分類結果為人臉，反之，亦即  $h(x, f, p, \theta) = 0$ ，則分類為非人臉。對於  $x$  套用  $f$  後共可求算出  $n$  個對應的特徵值，並將其依照大小排序，藉由掃描一遍排序後的特徵值，以滿足最小分類誤差為目標，可以為  $f$  決定一個最佳的閾值  $\theta$  與不等式的方向  $p$ ，用以建構出基於  $f$  的弱分類器  $h$ 。在每一回合的訓練過程中，套用所有可能的特徵樣版於  $X$ ，因為每一個可能的特徵樣版皆可建構出一個弱分類器，以大小為  $24 \times 24$  的影像為例，共會建構出 162,336 個弱分類器，將每一個弱分類器套用於後求算分類誤差 (classification error)，並選取分類誤差



最低的弱分類器作為該回合中所選取之最佳弱分類器。假定共需選取  $T$  個弱分類器，於第  $t$  回合疊代過程中， $t = 1, 2, \dots, T$ ，所選取的最佳弱分類器為  $h_t$ ，並依據  $h_t$  對  $X$  分類之誤差調整每一個訓練樣本的權重值，並求算  $h_t$  的權重值  $\alpha_t$  以反應  $h_t$  最後於強分類器  $C(x)$  中的重要性。在經過  $T$  次疊代後，將可建構出由  $T$  個弱分類器組合而成的一個強分類器  $C(x)$ ，完整 AdaBoost 演算法如下所述。

---

AdaBoost Algorithm :

- Given example images  $(x_1, y_1), \dots, (x_n, y_n)$  where  $y_i = 0, 1$  for negative and positive examples respectively.

- Initialize weights  $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$   $y_i = 0, 1$  respectively, where  $m$  and  $l$  are the number of negatives and positives respectively.

- For  $t = 1, \dots, T$  :

1. Normalize the weights,  $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$

2. Select the best weak classifier with respect to the weighted error

$$\varepsilon_t = \min_{f,p,\theta} \sum_i w_i |h(x_i, f, p, \theta) - y_i|$$

3. Define  $h_t(x) = h(x, f_t, p_t, \theta_t)$  where  $f_t, p_t$ , and  $\theta_t$  are the minimizers of  $\varepsilon_t$

4. Update the weights:

$$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$

where  $e_i = 0$  if example  $x_i$  is classified correctly,  $e_i = 1$

otherwise, and  $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$

- The final strong classifier is:

$$C(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases}$$

where  $\alpha_t = \log \frac{1}{\beta_t}$

---

#### (四) 階層式分類器 (Cascaded Classifiers)

VJ-Detector 可滿足即時的需求，階層式分類器的架構扮演了關鍵的角色，透過層層快速過濾掉明確非人臉的影像子視窗，加速人臉偵測的速度。在說明階層式分類器架構之前需先定義何為誤判率 (False Positive Rate, FPR) 以及檢測率 (Detection Rate, DR)。誤判率代表將檢測影像之子視窗影像誤判為人臉的機率，檢測率為找到人臉之正確率。換言之，誤判率表示將非人臉的子視窗影像誤判為人臉的比率，檢測率則為是人臉的子視窗影像且判別為人臉的比率。當希望達到較高的檢測率時，誤判率也會連帶升高，反之，若希望達到較低的誤判率，則檢測率將會下降。

階層式分類器最主要的用途就是使得計算速度增快，利用選擇各個階層的誤判率以及檢測率，透過層層的篩選最終達到低於所指定整體誤判率的要求，來使得 AdaBoost 能應用滿足即時的需求。如圖 5 所示，階層式分類器的架構為將數個強分類器結合運用，最一開始的分類器所能夠接受的誤判率最高，其主要目的在於快速過濾掉明確非人臉的影像子視窗，往後的每一層分類器將負責篩選在前一回合通過篩選所保留下來的影像子視窗，而且所能接受的誤判率隨著層級愈高也愈來愈低，亦即愈高層級的分類器將負責處理愈複雜且不易正確分類的影像子視窗，且通過篩選的條件也愈嚴苛，直到通過最後一個分類器，則所留下來的影像子視窗將會分類為人臉影像，並於原始影像中標記出位置與大小，達到人臉偵測的目的。

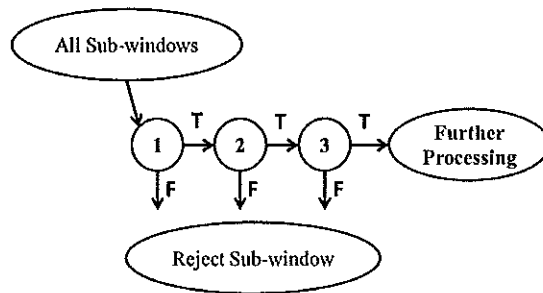


圖 5 階層式分類器架構圖

階層式分類器總共需多少個階層，將根據誤判率以及檢測率來決定。階層式分類器演算法概述如下，首先決定每回合可接受的最大誤判率  $f$ 、最小檢測率  $d$  與整體誤判率  $F_{target}$ 。接著使用人臉樣本集  $P$  與非人臉樣本集  $N$  套用 AdaBoost 進行訓練，在 AdaBoost 訓練過程中，將依據  $f$ 、 $d$  與  $F_{target}$  決定弱分類器選取的數量，以建構出一強分類器，做為階層式分類器第一層分類使用。將訓練好之第一層分類器套用於非人臉樣本集  $N$  上，將誤判為人臉 (非人臉誤判為人臉) 的樣本集合

作為訓練下一階層分類器之非人臉樣本集使用，重複此過程直到整體誤判率小於  $F_{target}$  為止，至此，階層式分類器訓練完成。完整階層式分類器訓練演算法如下所示。

- 
- $F_0 = 1.0$ ;  $D_0 = 1.0$ ;  $F_{target} = 0.001$
  - $i=0$
  - While  $F_i > F_{target}$ 
    - $i \leftarrow +1$
    - $n_i=0$ ;  $F_i=F_{i-1}$
    - §§§§§  $F_i = f \times F_{i-1}$
    - \*  $n_i \leftarrow n_i + 1$
    - \* Use  $P$  and  $N$  to train a classifier with  $n_i$  features using AdaBoost
    - \* Evaluate current cascaded classifier on validation set to determine  $F_i$  and  $D_i$
    - \* Decrease threshold for the  $i$ th classifier until the current cascaded classifier has a detection rate of at least  $d \times D_{i-1}$  (this also affects  $F_i$ )
    - $N \leftarrow \emptyset$ ,  $\emptyset$  means empty set.
    - If  $F_i > F_{target}$  then evaluate the current cascaded detector on the set of non-face images and put any false detections into the set  $N$
- 

## 二、哈爾材質特徵 (Haar Texture Feature)

Guo 等學者所提出的哈爾材質特徵 (Haar Texture Feature, HTF)，針對 VJ-Detector 所使用的類哈爾特徵數量過於龐大，致使 AdaBoost 若分類器訓練耗時的問題，透過降低整體特徵數量，獲得了改善。HTF 使用了 3 種類哈爾特徵樣版，如圖 6 所示，因此 HTF 可視為類哈爾特徵的一個子集合。給定一張訓練影像，相較於 VJ-Detector，HTF 最大的不同之處，在於對同一位置且同樣大小的影像子視窗中，同時套用 3 種特徵樣版，並求算出對應的特徵值。每一個特徵樣版所對應的特徵值求算方式，一樣是先將影像轉換為積分影像後，於積分影像對應之子視窗中，求算白色區塊的像素總和後，再減去黑色區塊的像素總和，並將白色與黑色區塊的差異值二元化 (binary)，表示為 0 或 1 的特徵值，若白色區塊像素和大於黑色區塊像素和，則特徵值為 1，否則為 0。

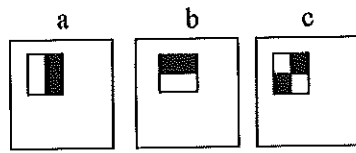
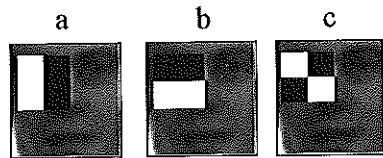


圖 6 HTF 所採用的 3 種類哈爾特徵樣版

HTF 使用的 3 種特徵樣版，可用以描述影像子視窗中，矩型區域之水平 (Horizontal)、垂直 (Vertical) 與對角 (Diagonal) 的灰階值差異。依序將如圖 6 所示之 a、b、c 三種特徵樣版套用於影像上，所求得之二元化特徵值可形成一個特徵向量，依序為  $[a,b,c]$ ，利用二進制向量編碼機制，透過公式：

$$E = 4a + 2b + c + 1$$

轉換為範圍 1 到 8 的 10 進位編碼值，如圖 7 所示。



特徵值： 1 1 0

$$E = 4a + 2b + c + 1$$



$$7 = 4 \times 1 + 2 \times 1 + 0 + 1$$

圖 7 HTF 特徵值編碼計算方式說明

HTF 的特徵計算結果為 1 到 8 的特徵編碼值，分別表示 8 種不同的特徵組合情形，因此，分別對訓練集中所有人臉影像以及所有非人臉影像，同時套用 3 種特徵樣版於所有可能的位置，所有特徵編碼值將會形成 2 個 8-bin 的直方圖，分別表示於人臉與非人臉影像中 HTF 特徵組合機率分佈的情形，如圖 8 所示。由圖 8 可知，HTF 特徵分佈於人臉影像集中於少數幾個特定的 bin 上，以圖 8(a) 中所示為例，HTF 特徵集中於第 1 個 bin 與第 5 個 bin，亦即編碼為第 1 與第 5 的 HTF 特徵組合，而非人臉影像則較為平均分佈，據此，可作為分類人臉影像與非人臉影像的依據。

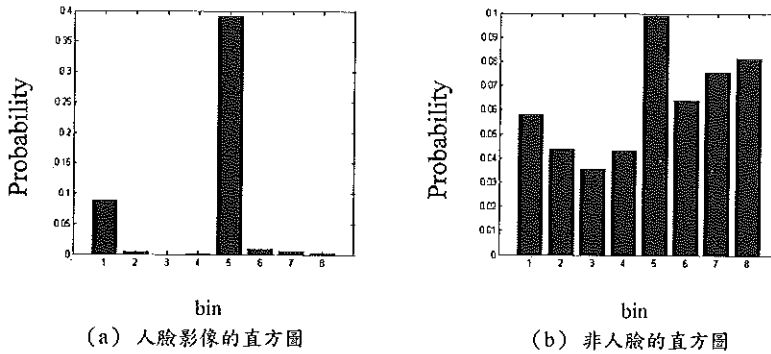


圖 8 HTF 套用於人臉與非人臉訓練集所形成的直方圖

HTF 只用了 3 種的類哈爾特特徵樣版，因此在特徵數量明顯較 VJ-Detector 所使用的類哈爾特特徵少。以  $18 \times 18$  的訓練影像為例，使用 HTF 所求算出的可能特徵總數約 6,561 個，但使用類哈爾特特徵的可能特徵總數為 51,705 個，兩者在特徵總數有顯著的差距，HTF 減低了約 8 倍的特徵數量，可有效縮短 AdaBoost 學習演算法中訓練弱分類器所花費的時間，且同時考量了水平、垂直與對角的灰階值差異，所以較不易受到外在光照變化的影響。

## 參、研究方法

VJ-Detector 提出了使用類哈爾特特徵，結合積分影像、AdaBoost 演算法與階層式分類器架構的人臉偵測方法，滿足了即時偵測的需求，但因類哈爾特特徵所產生龐大的特徵數量而導致 AdaBoost 訓練過程耗時甚鉅，仍有改進的空間。Guo 等學者提出的 HTF 雖改善了 VJ-Detector 的缺失，但只能對影像子視窗中，水平、垂直與對角的灰度差異進行描述，對較複雜之人臉材質區塊有著描述與鑑別能力不佳的缺點。本研究所提出了擴增式哈爾材質特徵，透過加入新定義的特徵樣版，補強原本 HTF 不足之處，並以 AdaBoost 為基礎提出 EHTF-Boosting 學習演算法，用以建構出最後可用於判別是否為人臉的強分類器。

### 一、擴充式哈爾材質特徵 (Extended Haar Texture Feature, EHTF)

HTF 使用了 3 種特徵樣版描述影像子視窗中水平、垂直與對角的灰階值差異，有著特徵描述與鑑別能力不佳的缺點。本研究提出了基於 HTF 的 EHTF 特徵表示方法，透過增加水平邊緣、垂直邊緣與亮(暗)點的 3 種特徵樣版，強化特徵描述能力，進而提高其人臉偵測正確率，EHTF 所採用的 6 種類哈爾特徵樣版如圖 9 所示，透過 EHTF 所定義的 6 種特徵樣版，可用以描述影像子視窗中，

矩型區域之水平 (a)、垂直 (b)、對角 (c)、邊緣 (d、e) 與亮 (暗) 點 (f) 的灰階值差異，大幅提升了特徵的描述能力。

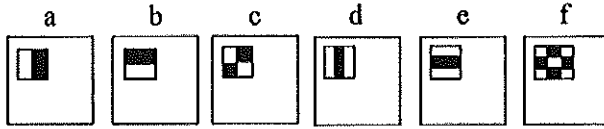


圖 9 EHTF 所採用的 6 種類哈爾特徵樣版

給定一張訓練影像，於同一位置且同樣大小的影像子視窗中（子視窗大小需為 6 的倍數），同時套用 EHTF 所定義的 6 種特徵樣版，並求出對應的特徵值。每一個特徵樣版所對應的特徵值求算方式如下，先將影像轉換為積分影像後，於積分影像對應之子視窗中，求算白色區塊的像素總和後，再減去黑色區塊的像素總和，並將白色與黑色區塊的差異值二元化，表示為 0 或 1 的特徵值，若白色區塊像素和大於黑色區塊像素和，則特徵值為 1，否則為 0。依序將如圖 9 所示之 a、b、c、d、e 與 f 共 6 種特徵樣版套用於同一影像子視窗上，所求得之二元化特徵值可形成一個特徵向量，依序為  $[a, b, c, d, e, f]$ ，利用二進制向量編碼機制，透過公式：

$$E = 32a + 16b + 8c + 4d + 2e + f + 1$$

轉換為範圍 1 到 64 的 10 進位的編碼值，如圖 10 所示。

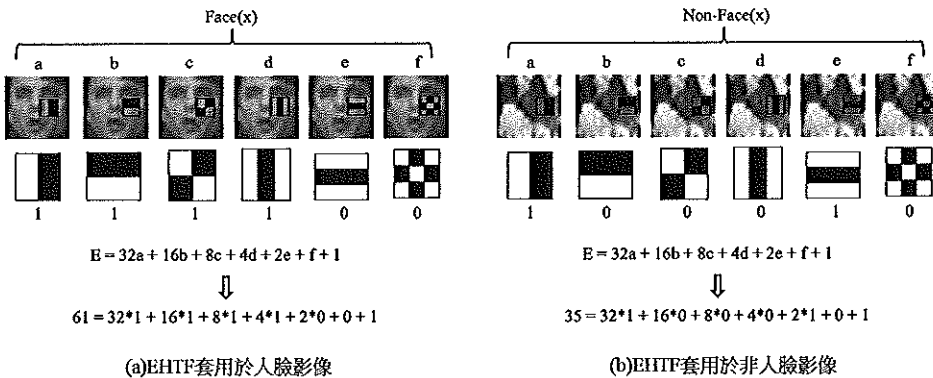


圖 10 EHTF 特徵值求算方式說明

EHTF 的特徵求算結果為 1 到 64 的特徵編碼值，分別表示共 64 種不同的特徵組合情形，對所有訓練集中人臉影像以及非人臉影像，同時套用 EHTF 所定義的 6 種特徵樣版於所有可能的位置，所有特徵編碼值將會形成 2 個 64-bin 的直方圖，分別表示於人臉與非人臉影像中 EHTF 特徵組合機率分佈的情形，如圖 11 所示。由圖 11 可知，EHTF 特徵於人臉影像的分佈集中於少數幾個特定的 bin 上，

而於非人臉影像則分佈較為平均。給定一測試影像的子視窗，以套用 EHTF 所定義的 6 種特徵樣版所得之編碼值為依據，於人臉影像與非人臉影像的直方圖上，比較查詢編碼值所對應 bin 的機率值，作為將子視窗分類為人臉或非人臉的依據。

以  $18 \times 18$  的訓練影像為例，使用類哈爾特徵的特徵總數則為 51,705 個，HTF 則為 6,561 個，而 EHTF 為 441 個。與類哈爾特徵相比，EHTF 減少了約 100 倍的特徵數量，而與 HTF 相比則減少了約 15 倍的特徵量，因此可大幅降低 AdaBoost 訓練弱分類器所需的時間。EHTF 同時使用了 6 種特徵樣版，對於外在光照變化影響的穩定性較佳，且可表示共 64 種不同的特徵組合情形，相較於 HTF 只能表示出 8 種特徵組合，因此具備了更佳的特徵描述能力。

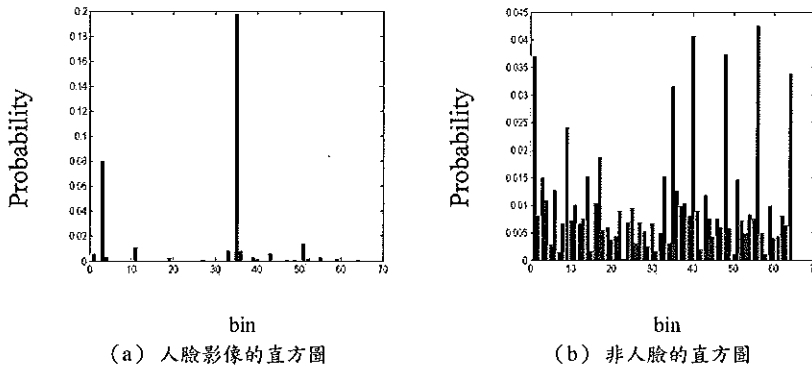


圖 11. EHTF 套用於人臉與非人臉訓練集所形成的直方圖

## 二、EHTF-Boosting 學習演算法

本研究所提出的 EHTF-Boosting 演算法是以疊代的方式，於每回合訓練過程中，首先對所有人臉與非人臉的訓練樣本進行權重 (weight) 的標準化 (normalized)，並對所有訓練樣本逐一套用所有可能的 EHTF 特徵，每一個特徵將形成一個弱分類器，以大小為  $18 \times 18$  的訓練影像為例，所有可能的 EHTF 特徵為 441 個，共可形成 441 個弱分類器。EHTF-Boosting 以最小加權錯誤 (minimize weighted error) 為目標函式，從眾多弱分類器中選取一個滿足最小加權錯誤的弱分類器為該回合最佳的弱分類器，並以所選取的分類器對訓練樣本進行檢測，依檢測的結果重新調整每一樣本的權重值，將分類錯誤樣本的權重值放大，以供下一回合訓練時著重於那些分類錯誤的樣本，最後求算出最佳弱分類器的權重值，用以表示該弱分類器的重要性。經過  $T$  回合訓練後，將得到由  $T$  個弱分類器所組成的強分類器，做為人臉檢測使用。

給定共  $n$  張大小相同的訓練用人臉與非人臉影像集  $X$ ， $X = \{(x_1, y_1), \dots, (x_n, y_n)\}$

且  $y_i = \{-1, 1\}$ ,  $i = 1, \dots, n$ ,  $y_i = 1$  表示  $x_i$  為人臉影像, 反之則為非人臉影像。其中  $X$  包含共  $l$  張人臉影像與共  $m$  張非人臉影像, 即  $n = l + m$ , 將每一張人臉訓練樣本的權重值初始化為  $1/2l$ , 每一張非人臉訓練樣本的權重值初始化為  $1/2m$ 。假定共需經過  $T$  個回合的訓練以選取  $T$  個弱分類器以建構一個強分類器, 則於每一回合訓練過程  $t$ ,  $t = 1, \dots, T$ , 執行下述步驟:

(一) 令  $w_{t,i}$  表示於第  $t$  回合訓練中第  $i$  個訓練樣本之權重值, 將所有樣本的權重值以公式  $w_{t,i} \leftarrow w_{t,i} / \sum_{j=1}^n w_{t,j}$  進行標準化以形成機率分佈, 即  $\sum_{i=1}^n w_{t,i} = 1$ , 用以表示每一訓練樣本於第  $t$  回合訓練中所受的關注程度。

(二) 對所有訓練樣本  $X$ , 逐一套用所有可能的 EHTF 特徵, 每一個特徵將形成一個弱分類器  $h$ , 以最小加權錯誤 (minimize weighted error) 為目標函式, 從眾多弱分類器中選取一個滿足最小加權錯誤的弱分類器為該回合最佳的弱分類器  $h_t$ 。假定  $Z$  表示於第  $t+1$  回合開始時所有訓練樣本的權重總和,  $Z$  之求算方式可表為下式:

$$Z = \sum_{i=1}^n D_{t+1}(x_i) = \sum_{i=1}^n D_t(x_i) \exp(-y_i h_t(x_i))$$

$D_{t+1}(x_i)$  為第  $t+1$  回合中樣本  $x_i$  的權重值, 而  $D_t(x_i)$  為第  $t$  回合中樣本  $x_i$  的權重值。由  $Z$  可知, 第  $t+1$  回合開始時所有訓練樣本的權重是由第  $t$  回合中所選取的最佳的弱分類器  $h_t$  對訓練樣本進行分類後之結果所決定, 如樣本  $x_i$  為人臉, 即  $y_i = 1$ , 最佳的弱分類器  $h_t$  對訓練樣本進行分類後之結果  $h_t(x_i)$  如為 1 則表示分類正確, 則  $-y_i h_t(x_i) = -1$ , 依據指數函數  $\exp$  的特性,  $\exp(-y_i h_t(x_i)) < 1$ , 因此, 於第  $t+1$  回合開始時, 樣本  $x_i$  的權重值相較於第  $t$  回合時將被降低, 因為在第  $t$  回合所建構出最佳的弱分類器  $h_t$  已可分類正確, 反之  $h_t(x_i)$  如為  $-1$  則表示分類錯誤, 則  $-y_i h_t(x_i) = 1$ , 所以  $\exp(-y_i h_t(x_i)) > 1$ , 於第  $t+1$  回合開始時, 樣本  $x_i$  的權重值相較於第  $t$  回合時將被放大, 因為在第  $t$  回合所建構出最佳的弱分類器  $h_t$  無法分類正確, 所以在第  $t+1$  回合時, 將著重於第  $t$  回合中分類錯誤的樣本。因每一個 EHTF 特徵套用於  $X$  可得到兩張 64-bin 的直方圖, 分別表示一個 EHTF 特徵於人臉與非人臉影像中特徵組合機率分佈的情形, 因此  $Z$  可改寫為

$$\begin{aligned} Z &= \sum_{i=1}^n D_{t+1}(x_i) = \sum_{i=1}^n D_t(x_i) \exp(-y_i h_t(x_i)) \\ &= \sum_{i: x_i \in X_f \cap y_i = +1}^{64} W_f^+ \exp(-h_t(x_i)) + \sum_{i: x_i \in X_f \cap y_i = -1}^{64} W_f^- \exp(h_t(x_i)) \end{aligned}$$



$w_j^+$  代表套用一個 EHTF 特徵於人臉訓練樣本  $x_i$  所得之特徵編碼在人臉影像直方圖中（如圖 11(a) 所示）對應的 bin 查詢所得權重值， $w_j^-$  代表套用同一個 EHTF 特徵於非人臉訓練樣本  $x_i$  所得之特徵編碼在非人臉影像直方圖中（如圖 11(b) 所示）對應的 bin 查詢所得權重值。因以最小加權錯誤（minimize weighted error）為目標函式來選取最佳的弱分類器  $h_i$ ，所以可將  $Z$  對  $h_i$  進行一次微分以求得滿足目標函式的最佳的弱分類器  $h_i$ ， $h_i(x_i)$  求算方式為：

$$-w_j^+ \exp(-h_i(x_i)) + w_j^- \exp(h_i(x_i)) = 0$$

經整理後可得

$$h_i(x_i) = 0.5 * \ln \left( \frac{w_j^+}{w_j^-} \right)$$

將  $h_i(x_i)$  帶入  $Z$  可得

$$Z = 2 \sum_{j=1}^{64} \sqrt{w_j^+ w_j^-}$$

因此，從眾多弱分類器中選取一個滿足

$$h_i = \arg \min_h Z$$

的弱分類器則為該回合所選取用於建構最佳的弱分類器  $h_i$  的 EHTF 特徵。

（三）依據滿足最小  $Z$  的 EHTF 特徵建立弱分類器

$$h_i(x_i) = 0.5 * \ln \left( \frac{w_j^+}{w_j^-} \right)$$

（四）將  $h_i$  套用於  $X$  進行分類，依據分類結果使用

$$w_{t+1,i} = w_{t,i} \beta^{1-e_i}$$

來更新每一樣本於第  $t+1$  回合的權重值，其中

$$\beta_i = \frac{Z}{1-Z}$$

且  $x_i$  使用  $h_i(x_i)$  分類結果為 1 時， $e_i = 1$ ，反之， $e_i = 0$ 。因弱分類器定義為比隨機分類略好一點的分類器，亦即分類錯誤率略小於 50%，即  $Z < 0.5$ ，所以  $\beta_i < 1$ ，如於  $x_i$  於第  $t$  回合時分類錯誤，其權重值將維持不變，如分類正確，其權重值將變小。

（五）經過  $T$  回合的訓練過程後，將選取出  $T$  個弱分類器建構出一個強分類器：

$$C(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ -1 & \text{otherwise} \end{cases}$$

其中

$$\alpha_t = \log \frac{1}{\beta_t}$$

完整 EHTF-Boosting 演算法如下所示：

EHTF-Boost Algorithm :

- Given example images  $(x_1, y_1), \dots, (x_n, y_n)$  where  $y_i = -1, 1$  for negative and positive examples respectively.

- Initialize weights  $w_{1,j} = \frac{1}{2m}, \frac{1}{2l}$ , where  $m$  and  $l$  are the number of negatives and positives respectively.

- For  $t = 1, \dots, T$  :

1. Normalize the weights,  $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$

2. Select the best weak classifier with respect to the min weighted error

$$\arg \min Z = \sum_{i: x_i \in X_j \cap y_i = +1}^{64} W_j^+ \exp(-h_t(x_i)) + \sum_{i: x_i \in X_j \cap y_i = -1}^{64} W_j^- \exp(h_t(x_i))$$

3. Define  $h_t(x_i) = 0.5 * \ln \left( \frac{W_j^+}{W_j^-} \right)$

4. Update the weights:

$$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$

where  $e_i = 0$  if example  $x_i$  is classified correctly,

otherwise  $e_i = 1$ , and  $\beta_t = \frac{Z}{1-Z}$

- The final strong classifier is:

$$C(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ -1 & \text{otherwise} \end{cases}$$

where  $\alpha_t = \log \frac{1}{\beta_t}$

## 肆、研究結果

### 一、實驗環境設定

為將本研究所提出之 EHTF 方法與 HTF 在特徵數量、訓練效能與人臉偵測率進行比較，本研究實作了基於 HTF 與基於 EHTF 的人臉偵測器，實驗硬體設備規格為 Intel® Core™ i7 處理器與 8G 記憶體，作業系統採用微軟 Windows 7 專業版，程式開發工具為 MATLAB 2012。使用 LFW-Bigfoto 資料集進行訓練，LFW-Bigfoto 訓練集為結合了 LFW 人臉影像資料庫 (Huang, Jain & Learned-Miller, 2007) 與 Bigfoto 背景影像 (不包含任何人臉) 資料庫 (<http://www.bigfoto.com/>)，手動於 LFW 資料集中擷取出 2,000 張正面人臉影像，並統一將影像尺寸調整為  $18 \times 18$ ，如圖 12(a) 所示，於 Bigfoto 影像資料庫中選取 2,000 張背景影像，如圖 12(b) 所示，大小統一縮放為  $320 \times 240$  或  $240 \times 320$ ，並於前 500 張背景影像中以任意矩形大小 (矩形寬與高需一樣) 隨機於不同位置擷取出 10,000 個子影像，並統一將影像尺寸調整為  $18 \times 18$  後，再隨機選取 2,000 張做為非人臉影像訓練集。

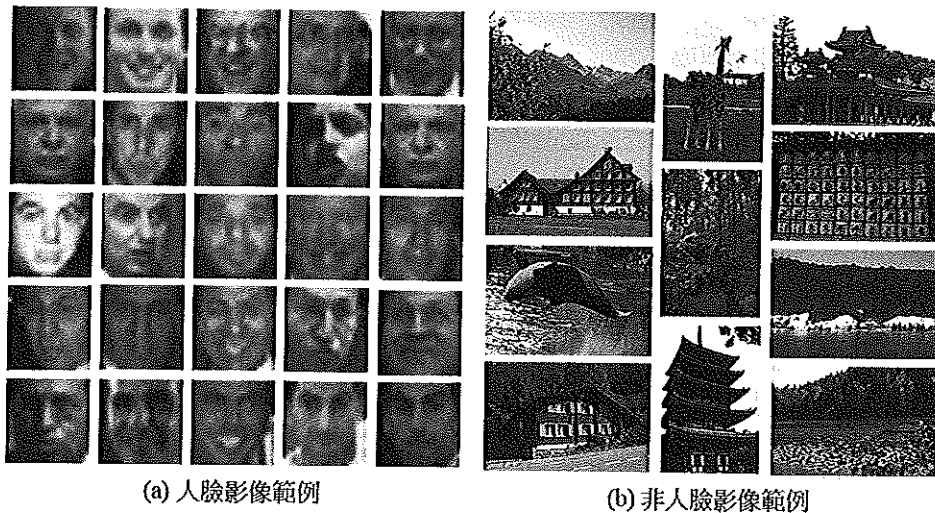


圖 12 部分人臉與非人臉影像範例

### 二、階層式分類器訓練結果比較與分析

在利用階層式分類器的訓練過程中，首先使用 LFW-Bigfoto 中的前 1,000 張人臉影像與非人臉訓練集中的前 1,000 張影像，共 2,000 張做為第一層分類器訓練用的樣本集其餘 2,000 張則作為驗證集 (validation set)，檢測率目標為 0.99，整體誤判率目標為 0.001，將訓練完成的第一層分類器針對驗證集中背景影像前 500 張影像進行掃描，掃描子視窗的最小尺寸為  $18 \times 18$ ，且需為 6 倍數，故尺寸

為  $18 \times 18$ 、 $24 \times 24$ 、 $30 \times 30$ 、 $36 \times 36 \dots$ 等，且掃描時每次移動 2 個像素，最後將掃描背景影像（不包含人臉）所得前 4,000 張誤判為人臉之子視窗影像統一調整大小為  $18 \times 18$  後，做為下一層弱分類器訓練的非人臉訓練集，當不足 4,000 張時，將再給予 500 張背景影像進行掃描。最後，所訓練出 10 個階層的階層式分類器，共選出 54 個不同的特徵，其中第 1 層分類器所選出之 7 個特徵樣式，如圖 13 所示，皆為重要人臉特徵之所在，可知 EHTF 特徵具備良好的特徵描述能力與鑑別能力。



圖 13 階層式分類器之第 1 層分類器所選出之 7 個 EHTF 特徵

圖 14 為 EHTF 與 HTF 訓練階層式分類器之人臉偵測率比較結果，從圖中可知，在相同誤判率的條件下，本研究所提出 EHTF 的偵測率皆優於 HTF，由此可知相較於 HTF，基於 EHTF 的人臉偵測器於偵測率有明顯的提升，且誤判率較低。

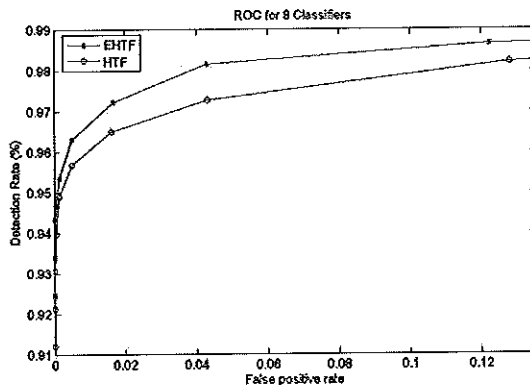


圖 14 Receiver operating characteristic (ROC) curve for the 9 classifier

圖 15 與圖 16，皆以訓練 10 層階層式分類器為例，將 EHTF 與 HTF 在每一層分類器所使用特徵數量與每一層訓練所需時間做比較。EHTF 與 HTF 在前 4 層的分類器建構所使用的特徵數量差異性不大，於第 5 到 8 層開始有差異，於第 9 與第 10 層時差異顯著提升，階層式分類器越後面的層級為處理越難分類的樣本，因此需要使用較多的特徵（較多的弱分類器），而 EHTF 於第 9 與第 10 層時所使用的特徵數較 HTF 所使用特徵數少約 1.5 倍，如圖 15 所示，由實驗結果可知 EHTF 特徵描述能力優於 HTF。在訓練所需時間上，因 EHTF 整體特徵數量少於 HTF，所以訓練所需時間明顯少於 HTF，且於第 10 層分類器訓練所花時間較 HTF 少約 6 倍，如圖 16 所示。

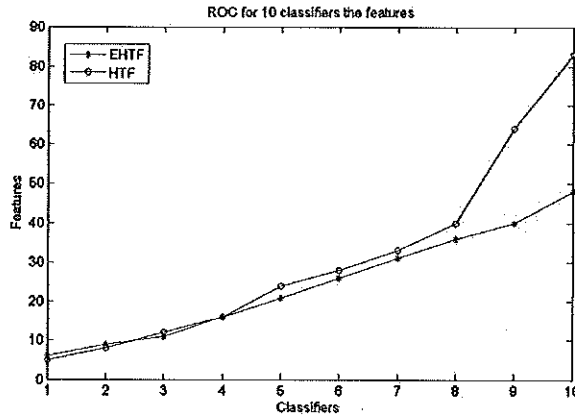


圖 15 EHTF 與 HTF 特徵數量比較

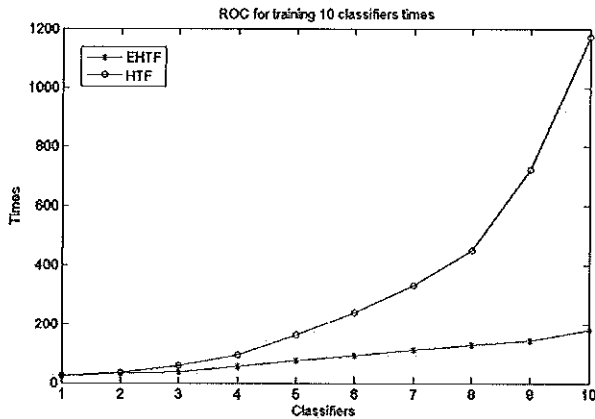


圖 16 EHTF 與 HTF 弱分類器訓練時間比較

### 三、階層式分類器訓練結果應用於 MIT-CMU 測試集實驗結果

利用本研究所提出的 EHTF 方法，實作了 19 層階層式分類器，用於掃描多張 MIT-CMU 測試集 ([http://vasc.ri.cmu.edu/idb/html/face/frontal\\_images/index.html](http://vasc.ri.cmu.edu/idb/html/face/frontal_images/index.html)) 影像之實驗結果如圖 17 所示，其中側臉以及頭部較歪的影像無法順利偵測到人臉的位置，因為在訓練時均使用正面且角度為正的樣本進行訓練，而導致部分影像中的人臉無法順利偵測並正確標示。

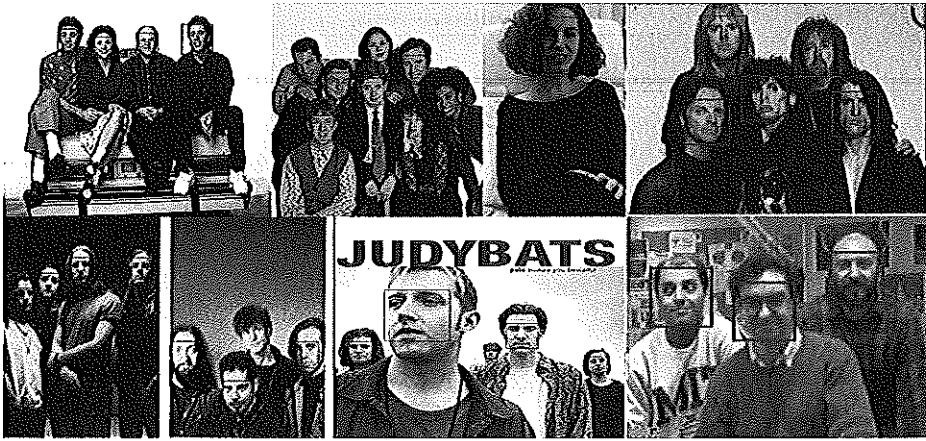


圖 17 EHTF 階層式分類器掃描 MIT-CMU 測試集實驗結果

## 伍、結論

隨著科技的進步，資訊科技正快速地應用於教學環境中，學生的學習也不僅限於課堂，在如此多元化的教學與學習環境中，如何能更客觀地評估學習者的學習成效是一個需要審慎面對的問題。近年來，許多相關研究利用人臉偵測作為學習者專注力分析資料取得的重要方法，後續分析等皆須基於成功偵測並定位出人臉的前提上方可進行，此種方式具有直覺且不影響學習者學習過程的優點。然而，如何能快速且準確地於影像中偵測出人臉，仍然是一個值得研究的議題。

VJ-Detector 雖然獲得顯著的成功，確立了人臉偵測領域的重要里程碑，且已廣泛應用於日常生活中，如：相機、社群網路、手機與行動裝置等，但其仍有著特徵數量龐大與人臉偵測器訓練過程耗時甚鉅的缺點。HTF 的提出雖然有效地改善了 VJ-Detector 的缺點，但又因 HTF 特徵過於簡單而致使偵測效果不佳。本研究提出了一個應用於學習者專注力分析的人臉偵測方法，針對 HTF 的缺失提出了 EHTF 特徵與 EHTF-Boosting 學習演算法，可分為三大重點：

(一) 加強了 HTF 的描述能力，加入了邊緣與亮(暗)點的描述，共使用 6 種不同的特徵樣版，強化了 EHTF 的特徵描述能力與鑑別力，有效提升整體的偵測率。

(二) 大幅降低特徵數量，相較於 VJ-Detector 降低了約 100 倍的特徵數量，而相較於 HTF 則降低了約 15 倍的特徵數量，因此可大幅縮短訓練分類器所需的時間。

(三) 利用所提出的 EHTF-Boosting 演算法，使用不同於 AdaBoost 的最佳弱分類器選擇機制，透過疊代的方式選取多個弱分類器以建構出強分類器。

由檢測結果可知，基於 EHTF-Boosting 所訓練出的階層式分類器相較於 HTF

只需少量的特徵即可達成目標，以建構 10 層的階層式分類器為例，EHTF 相較於 HTF 所使用特徵數量少約 1.5 倍，所以相對也降低了偵測時間，使得人臉偵測的過程更快速，可滿足應用於即時環境的需求。

本文所提出之 EHTF 人臉偵測相關技術，在教學環境應用中可用於數位學習之學習成效分析、智慧教室即時點名、利用臉部表情診斷學習困難度、教室秩序監控等不同領域，利用 EHTF 建構出人臉偵測器，讓相關分析資料的取得更加容易與客觀。而未來針對人臉偵測技術領域，可往三個研究方向探討：

(一) 特徵的變化或延伸，以提高人臉鑑別度，設計不同的特徵針對側臉、俯視、五官遮蔽物，如：眼鏡等，甚至表情等，上述容易影響偵測結果之原因進行改善。

(二) 針對 AdaBoost 演算法進行改善，降低計算時間，以達到訓練時間縮短的目的。

(三) 因光影變化會影響偵測結果，故若能針對光影問題加以處理，方可提高偵測率並降低誤判率。

## 參考文獻

- 王鴻勳 (2010)。結合影像偵測與模糊推論來分類學生上課學習行為。未出版之碩士論文。臺北市立教育大學，臺北市。
- 蔡沛勳 (2006)。遠距教學之全時精神狀態偵測系統之研製。未出版之碩士論文，私立朝陽科技大學，臺中市。
- 鄭昭明 (2006)。認知心理學。臺北市：桂冠。
- 蘇信宏 (2007)。數位學習情意偵測專心程度之影像處理。未出版之碩士論文，私立北臺灣科學技術學院，臺北市。
- 龔喬琳 (2012)。人臉偵測應用於學習專注力與群組參與度之分析。未出版之碩士論文，國立臺中教育大學，臺中市。
- Huang, G. B., Jain, V., & Learned-Miller E. (2007). Unsupervised joint alignment of complex images. *IEEE 11th International Conference on Computer Vision* (pp.1-8). Brazil.
- Guo, Z. B., Yan, Y. Y., Zhao, C. X., & Yang, J. Y. (2008). HTF-Boosting Learning and Face Detection. *Computational Intelligence and Industrial Application, 2008. PACIIA '08. Pacific-Asia Workshop on* (pp.376-380). China: Wuhan.
- Jeng, S. H., Liao, H. -Y. M., Liu, Y. -T, & Chern, M. -Y. (1996). An Efficient Approach for Facial Feature Detection using Geometrical Face Model. *Pattern Recognition, 1996., Proceedings of the 13th International Conference on* (pp.426-430). Vienna.
- Liu, C. (2003, June). A Bayesian Discriminating Features Method for Face Detection.

- Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25, 725-740.
- Miao, J., Yin, B., Wang, K., Shen, L., & Chen, X. (1999, July) A Hierarchical Multi-scale and Multi-angle System for Human Face Detection in a Complex Background using Gravity-center Template. *Pattern Recognition*, 32, 1237-1248.
- Ojala, T., Pietikäinen, M., & Harwood, D. (1996, March). A Comparative Study of Texture Measures with Classification Based on Feature Distributions. *Pattern Recognition*, 29, 51-59.
- Rowley, H. A., Baluja, S., & Kanade, T. (1996). Neural Network based Face Detection. *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on* (pp.23- 38). San Francisco, CA.
- Saneiro, M., Santos, O. C., Salmeron-Majadas, S. & Boticario, J. G. (2014, April). Towards Emotion Detection in Educational Scenarios from Facial Expressions and Body Movements through Multimodal Approaches. *The Scientific World Journal*, 2014, 1-14.
- Waring, C. A., & Liu, X. (2005, June). Face Detection using Spectral Histograms and SVMs. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 35, 467-476.
- Viola, P., & Jones, M. J. (2004, May). Robust Real-Time Face Detection. *International Journal of Computer Vision*, 57, 137-154.
- Yang, G., & Huang, T. S. (1994, January). Human Face Detection in a Complex Background. *Pattern Recognition*, 27, 53-63.
- Yan, S. Y., Shan, S. G., Chen, X. L., & Gao, W. (2008). Locally Assembled Binary (LAB) feature with feature-centric cascade for fast and accurate face detection. *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp.1-7). Anchorage, AK.
- Zhang, L., Chu, R., Xiang, S., Liao, S., & Li, S. Z. (2007). Face Detection Based on Multi-Block LBP Representation. *Advances in Biometrics Lecture Notes in Computer Science*, 4642, 11-18.



# Face Detection Approach Applied on Learners' Attention Analysis

Hsun-Li Chang

Associate Professor  
Department of Information Management  
National Taipei University of Business  
E-mail: hsunli@ntub.edu.tw

Kai-Yung Lin

Associate Professor  
Department and Graduate Institute of Information Management  
Yu Da University  
E-mail: linky@ydu.edu.tw

Chih-Chiang Yu

Department and Graduate Institute of Information Management  
Yu Da University  
E-mail: ta0614@hotmail.com

## *Abstract*

*Attention is closely related to learning. How to measure students' attention levels is a very important topic. Among a lot of measuring approaches, face detection is more objective and has less effect on students. In this paper, we proposed a novel face detection approach based on the Extended Haar Texture Features (EHTF) and the EHTF-Boosting learning algorithm. EHTF uses 6 differential rectangular feature templates to describe characteristics of edge, linearity, orientation, and spot in the images' local area. EHTF has benefits to evaluation faster and illumination invariant. Applying EHTF on the training face/non-face images, EHTF-Boosting is used to train several weak classifiers to form a strong classifier. The proposed EHTF-Boosting learning algorithm can correctly show the importance, i.e. the weight value, of each weak classifier. By using 2,000 frontal face images and 2,000 background im-*

*ages from LFW-Bigfoto dataset as the training set, we implemented a face detector based on the proposed methods and applied it on the MIT-CMU testing dataset. Comparing EHTF with other methods, the experimental results show that our approach can improve detection rate and reduce the false positive rate obviously. It has several advantages: using fewer features, less training time and higher detection rate. The proposed method can also satisfy the requirements for real-time applications and can be applied on different instructional environments as the data acquiring method, for example online learning, attendance analysis, learning difficulty analysis, and keeping order in course. Analysis data acquiring for related researches can be more convenient and objective.*

***Key words: Attention, Face Detection, AdaBoost, EHTF-Boosting***