

# 中華民國 第 49 屆中小學科學展覽會

## 作品說明書

---

高中組 生活與應用科學科

佳作

040816

「像」由心生-人臉門禁系統

學校名稱：國立中和高級中學

作者： 高一 吳承翰 高一 胡鎮宇	指導老師： 鄭嘉哲 李建男
-------------------------	---------------------

關鍵詞：人臉、門禁系統、圖像辨識

## 摘要

描述未來時空的科幻電影中，人類從來不帶鑰匙、卡片等額外提供機器辨識身分的工具，只要站在門前，機器就會很智慧的反應出是誰要進入，然後決定是否讓行。這種系統是透過在人類身上一些天生的特徵作為鑰匙，例如：臉、指紋、視網膜、腦波等。本次研究就是要實作用「人臉」來開門的「人臉門禁系統」。其實其他公司已有已量產化的人臉門禁系統，但是卻不普及，其中很大的原因就是「造價不斐」。但是其實此系統需要的設備很簡單，如果能大大降低此系統的造價，可以普及的裝設此系統在許多地方。這次科展所開發的系統經過測試之後，數據顯示辨識成功率最高有 96.43%。算是一個還不錯的數據，而且造價降低許多，說不定以後出門不用帶鑰匙不再是個夢。



## 壹、 研究動機

常常會看到科幻電影中，人物在未來的時空從來不帶鑰匙、卡片等額外提供機器辨識身分的工具，只要站在門前，機器就會很智慧的反應出是誰要進入，然後決定是否讓行。這種系統可能是透過在人類身上一些很難變造的部分做為天生的鑰匙，例如：臉、指紋、視網膜、腦波等。在校內科展的時候，我們研究了完整的人工智慧(Artificial Intelligence)的圖像辨識方法，後來就在思考如何將圖像辨識應用在實際的層面上。再加上學校生活與科技課程中有學到「數位家庭安全」，所以就想到每天要帶著重重的一串鑰匙很不方便，是否可以如同科幻電影免去帶鑰匙的麻煩呢？上網查詢之後，發現有種系統叫「人臉門禁系統」，只要人員站在系統裝置前，就可以準確辨認出人員是誰。看到此系統後，發現這是一個圖像辨識應用的好地方，如果可以將此系統普及化，那麼生活將會便利許多。所以就起了念頭想要研究看看「人臉門禁系統」。

## 貳、 研究目的

經過在網路上查詢關於「人臉門禁系統」的相關資訊，發現目前有幾家廠商在製作此系統。以大陸漢王生產的「人臉身份識別門禁系統 F710」來說，此系統目前並無普及化，其中很大的原因就是造價不菲，一台要價 9800 元人民幣(大約是 49660 元新台幣)，一般無特殊用途，是不會裝設的。但是其實此系統需要的設備主要只需攝影機、智慧控制單元兩種，如果能大大降低此系統的造價，可能可以普及的裝設此系統在家庭、學校、公司等。因此，本次研究的方向就是以單價不高的網路攝影機(Webcam)取代高解析度 CCD 攝影機、一般家用電腦(PC)取代晶片控制單元來實作此系統，希望可以使價格降低，但是也不失此系統的效率及其優點。



(漢王人臉通 F710)

## 參、研究設備及器材

一、硬體：網路攝影機(Webcam)、個人電腦(PC)、RS232 序列埠結合電鎖控制器(自製)。

二、軟體：Microsoft Visual C#、Microsoft Visual C++、OpenCV、Ulead PhotoImpact、MathType、Microsoft DirectShow SDK、Microsoft Speech SDK。

註 1：OpenCV 的全稱是 Open Source Computer Vision Library，是一個跨平臺的電腦視覺庫。OpenCV 是由英特爾公司發起並參與開發，以 BSD 許可證授權發行，可以在商業和研究領域中免費使用。OpenCV 可用於開發即時的圖像處理、電腦視覺以及模式識別程式。

註 2：DirectShow，開發代號 Quartz，是一種由微軟公司開發的能夠讓軟體開發者對媒體文件執行各種不同處理的應用程序設計介面。它是微軟公司對早先 Windows 影像科技的一次更新。

註 2：Speech SDK 是微軟開發的語音系統，能使開發者免費使用此函式庫，快速的開發出語音辨識、合成等系統。

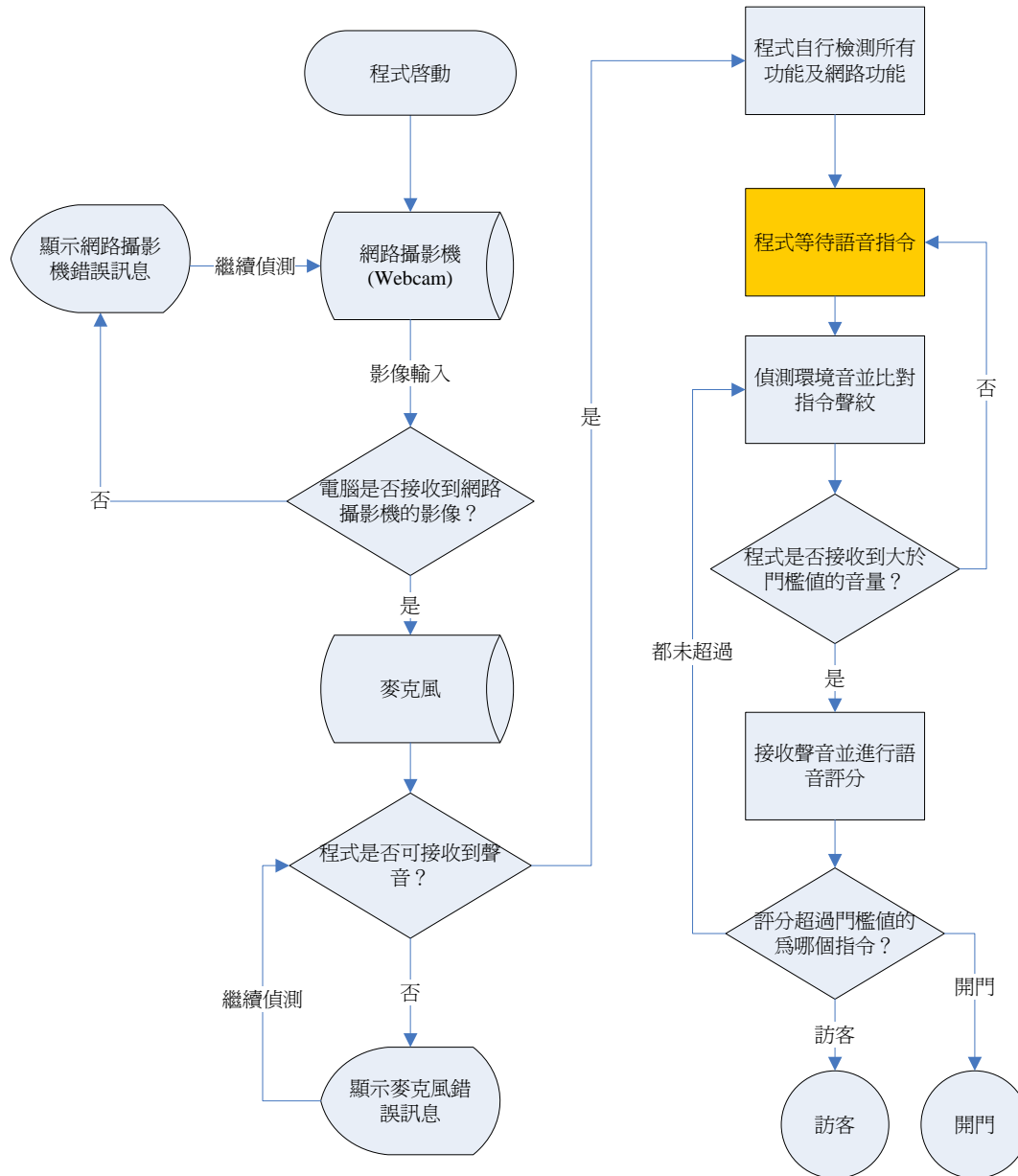


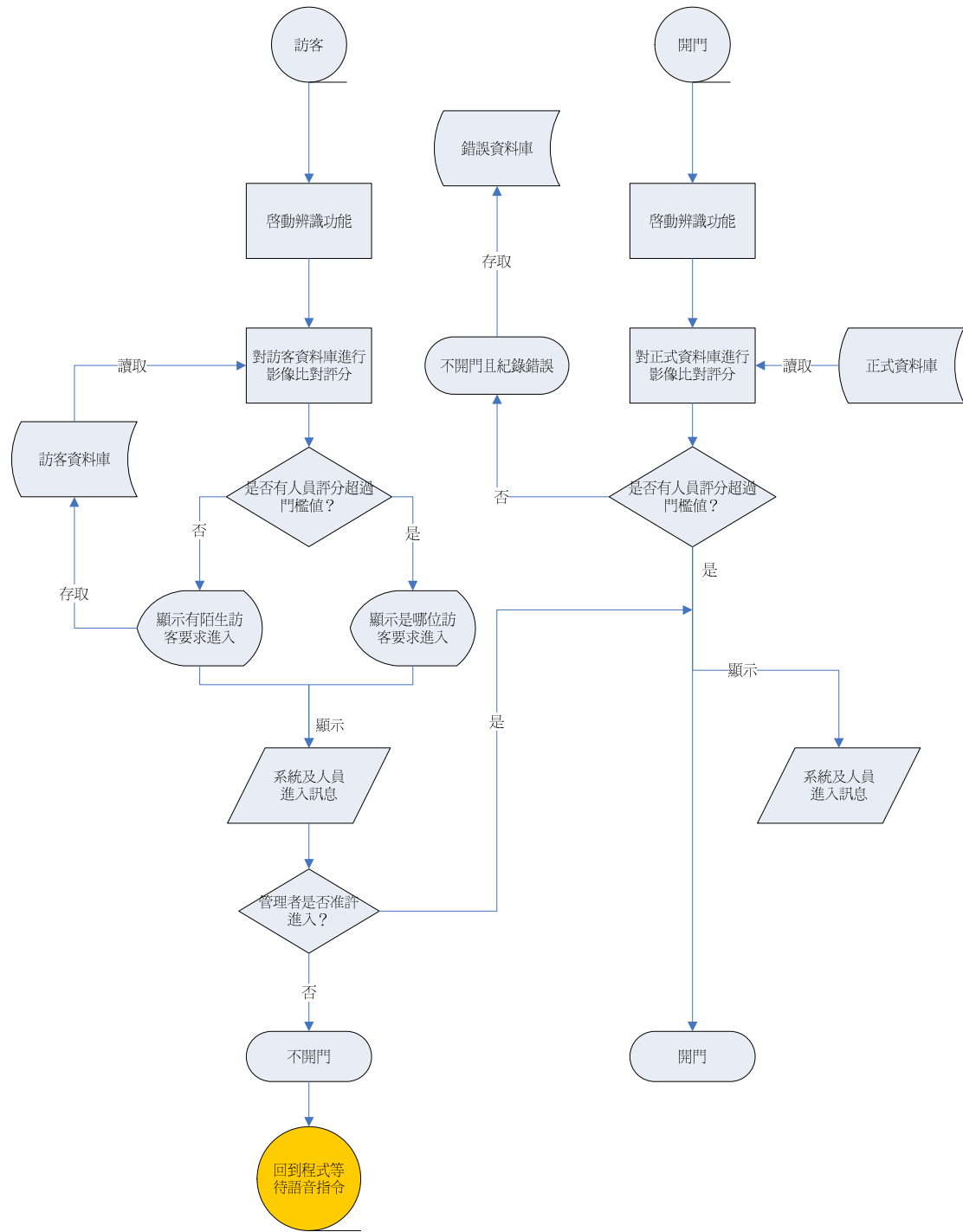
## 肆、研究方法

### 一、人臉門禁系統架構介紹

架構上基礎可分為：輸入、處理、輸出。輸入部分有：網路攝影機(含麥克風)，處理部分有：電腦，輸出有：RS232 序列埠結合電鎖控制器。

工作流程圖如下：

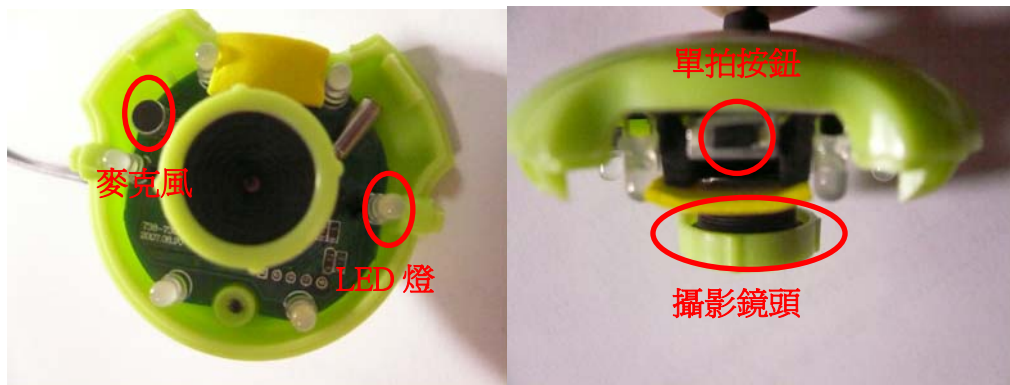




## 二、裝置

### 〈一〉 輸入

目前在市面上最常見的，也是價格最低廉的攝影器材就是網路攝影機了。介於此因，本次科展我們使用了一具 524 萬畫素的網路攝影機來作為影像輸入，這個網路攝影機上內建麥克風，因此語音輸入裝置也就不必另外加裝了。Webcam 與電腦通訊的方式為 USB，幾乎市面上所有電腦都支援 USB，相容性也不錯。攝影機上有一個單拍按鈕，按下即可拍下一張數位照片，但是此動作可透過程式完成，因此此按鈕將不使用。仔細看看發現還有 LED 燈在上面，那些 LED 燈(六個)是可變換亮度的，但是無法透過程式控制(需要自行用手轉換)，另外，我們在門上也有加裝燈具，所以不會使用攝影機內建的 LED 照明設備。觀察一下，攝影鏡頭後面有幾圈黑黑的紋路，那是用來調節焦距的，本攝影機預設焦距是 3.85MM，我們調整到 50CM，才能使鏡頭前的影像是清晰的。



〈上面兩張圖片是本次使用的 Webcam，鑲嵌在門中〉

### 〈二〉 處理

在處理影像、聲音的訊號時，需要強大的計算單元以及可儲存的資料庫。原先有考慮使用單晶片來時作此系統，但是介於它計算及儲存功能都沒有電腦來的好，所以使用了電腦。電腦收到影像及聲音後，會進行分析，再與已儲存的資料庫進行比對，這些功能都必須透過撰寫一套專用的程式才能達成。

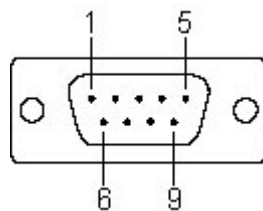


在程式已執行後，會對聲音進行檢測，把聽到的聲音進行 FFT(快速傅利葉轉換)，然後用迴歸方程式比對指令，得到指令相似度，如相似度高於門檻值，即程式收到指令，會立即從攝影鏡頭上擷取一張照片，進行濾波去除雜訊，進行下列比對程序：

1. 從影像找出特徵點(哈里斯角點偵測 Harris Corner Detector)。
2. 與已儲存的影像進行特徵點比對(灰度相似性)，產生數據。
3. 利用灰度相似性的數據進行極線幾何(Epipolar Geometry)產生限制極線，用 RANSAC 最佳化求解。
4. 利用灰度數據做單應(Homography)限制，並用 RANSAC 求解。
5. 將灰度相似性、限制極線、單應限制三者引導交集，剔除在交集區以外比對相似的點，得到最終相似點。
6. 分析數據，給出相似度，再判斷是為哪一人員，是否通行。

### 〈三〉 輸出

程式主要判斷的結果有兩種：1.開門 2.錯誤(也就是不開門)，但是如何控制電鎖，且表現給使用者了解現在的狀態呢？我們在設計了一個通訊電路板，此電路板是透過 RS232 序列埠(俗稱 COM Port)來通訊，C#可以利用指令(SerialPort 一種序列埠元件)來控制 RS232 上的 DTR、RTS 腳位的電位(高低)，電位不同，會造成電路板上的繼電器(Relay)切換，可由此控制電鎖和訊號燈。RS232 腳位如下：



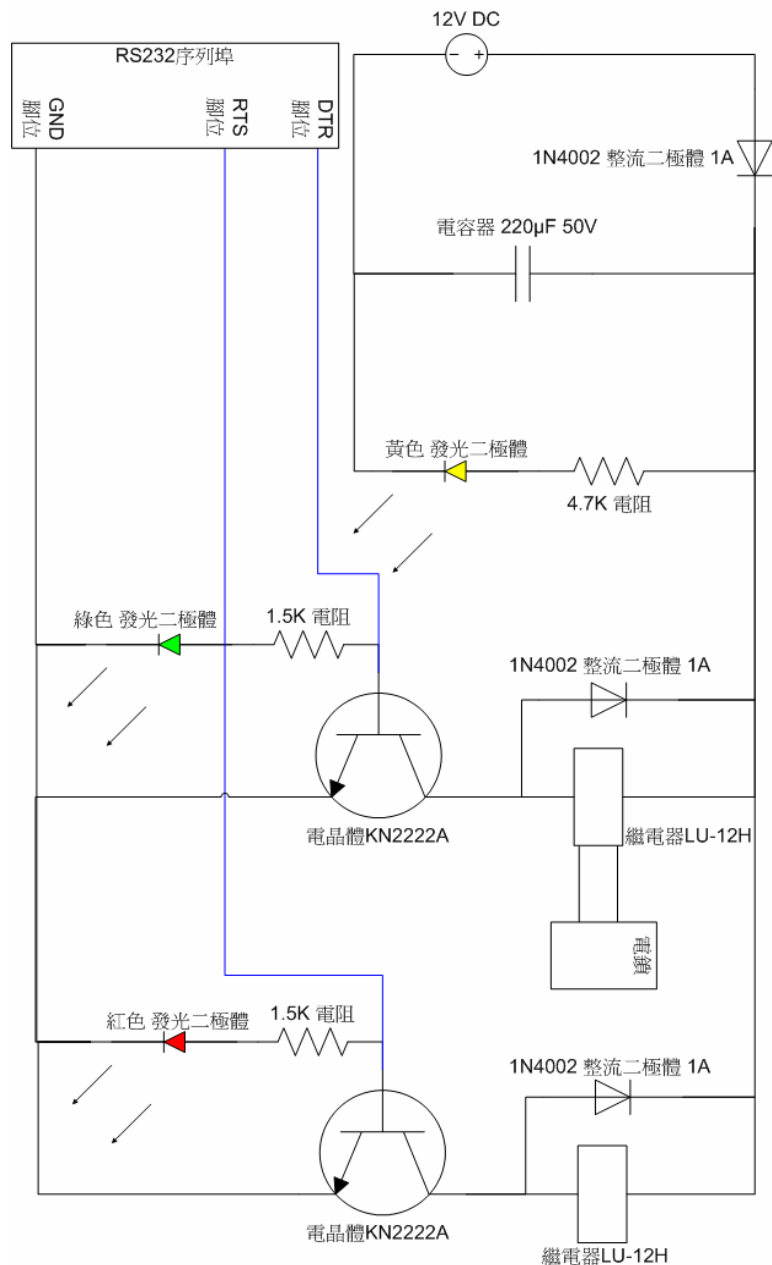
RS-232-DB9 公頭

1	2	3	4	5	6	7	8	9
DCD	RxD	TxD	DTR	GND	DSR	RTS	CTS	---

電路板上的訊號燈有三種：

- 1.黃燈(電源燈)：黃燈只要有電源連接就會亮起，表示電路板可使用。
- 2.綠燈(通行燈)：辨識通過的話，程式控制 DTR 腳位到高電位，造成繼電器切換，亮起綠燈 2 秒(2 秒後切換回低電位)並開啓電鎖。
- 3.紅燈(錯誤燈)：紅燈閃一下為程式皆收到語音指令。紅燈閃三下為程式判斷錯誤(主人不允許進入、辨識錯誤、語音指令不明確等)，控制 RTS 腳位到高電位，繼電器切換，且不通行。

電路圖請見下：



## 伍、研究結果

### 一、人臉比對

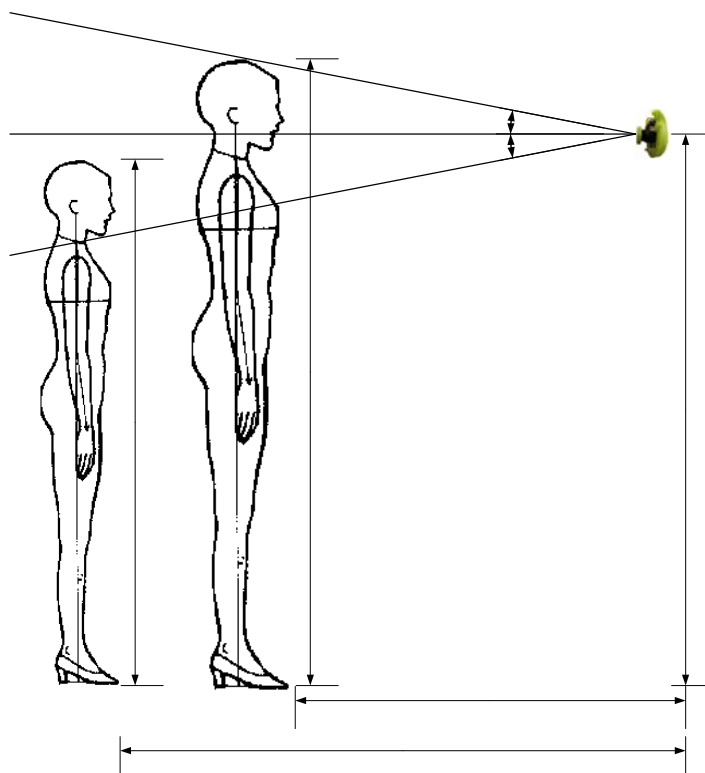
因為我們的攝影鏡頭沒有辦法改變攝影距離，所以我們要讓固定攝影機與使用者的距離。經過測試，如要得到幾乎全畫面都是臉部的話，大約距離是 50~60CM。如下圖，120CM 過遠，拍入了一些不必要的景物，造成辨識的多餘。50CM 則為適當距離，臉部幾乎可全照到，也沒有其他景物。



(120CM 距離)

(50CM 距離)

接下來，攝影鏡頭的高度、角度也是很重要的，不同身高的人，要站不同的距離。其中關係請見下圖：



假設右邊是高於攝影機的人，他與鏡頭的距離是  $L1$ ，身高是  $H1$ 。左邊是矮於攝影機的人，他與鏡頭的距離是  $L2$ ，身高是  $H2$ 。 $\theta$  為鏡頭垂直可視角度的  $1/2$ ， $H$  為攝影鏡頭高度， $F$  是人臉的垂直大小。

身高高於攝影鏡頭的人， $H1 > H$ ，為了得到較好的影像，人頭頂必須在影像的最上方，由此可推得一公式：
$$\frac{H1-H}{L1} = \tan \theta \Rightarrow L1 = \frac{H1-H}{\tan \theta}$$

身高矮於攝影鏡頭的人， $H1 < H$ ，為了得到較好的影像，人頭下巴必須在影像的最下方，由此可推得一公式：

$$\frac{H-(H2-F)}{L2} = \tan \theta \Rightarrow L2 = \frac{H-(H2-F)}{\tan \theta}$$

為了要求出人員與鏡頭的攝影距離，必須要知道人員的身高、攝影機的高度、人臉長度以及  $\theta$  鏡頭垂直可視角度。其中攝影機的高度、人臉長度以及  $\theta$  鏡頭垂直可視角度可先輸入，使用者只需知道自己的身高，按照地上所畫的線，以腳尖對齊，即可獲得最佳攝影距離。

拍了照之後，就要進行圖像比對了，其中開發所使用的理論及演算法請參閱附錄〈一〉，下面是我們的比對圖像的步驟：

1. 將圖像先過濾雜訊，再找出特徵角點(哈里斯角點偵測)

2. 用灰度相似性  $C = \sum_{u,v \in patch}^n (I_1(u,v) - I_2(u,v))^2$  找出相似的點，其中

$I_n(u,v)(n=1,2)$  是第  $n$  幅圖像在  $(u,v)$  座標處的灰度， $patch$  是比

對搜尋視窗， $C$  越小表示相似度越高。如果某個特徵點在兩個圖像中都表現高相似性，其點為相似點。

3. 找出兩幅圖像的極線幾何限制，要先估計出一個基礎矩陣：

$m'^T F m = 0$ ，其中  $F$  為基礎矩陣， $m$  與  $m'$  為一對相似點。

4. 估計出基礎矩陣後，用 RANSAC 求解，並且在在計算時加上秩 2 限制，去掉可能重疊或不相交的極線：

$$\begin{cases} \min(\sum_i d(m_i, \hat{m}_i)^2 + d(m'_i, \hat{m}'_i)^2) \\ \text{rank}(F) = 2 \end{cases} \quad \text{其中, } m_i \text{ 與 } m'_i \text{ 是通過灰}$$

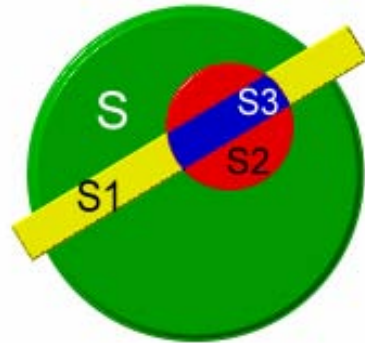
度相似性比對的一對相似點。 $\hat{m}_i$  與  $\hat{m}'_i$  是待估計的真相似點，它們準確的滿足  $\hat{m}'_i{}^T F \hat{m}_i = 0$

5. 找出單應限制。令  $m, m'$  為兩幅圖像對應的相似點，其單應矩陣  $H$  滿足： $m' = Hm$ 。

6. 求解過程使用 RANSAC 剔除錯誤比對點

$$\begin{cases} \min(\sum_i d(\hat{m}'_i - m'_i)^2) \\ \hat{m}'_i = H\hat{m}_i \end{cases}$$

7. 由極線幾何可知， $m'_i$  的相似點  $\hat{m}_i$  必然在對應的極線附近，又由單應限制可知，相似點  $m'_i$  也必然在估計點  $\hat{m}'_i$



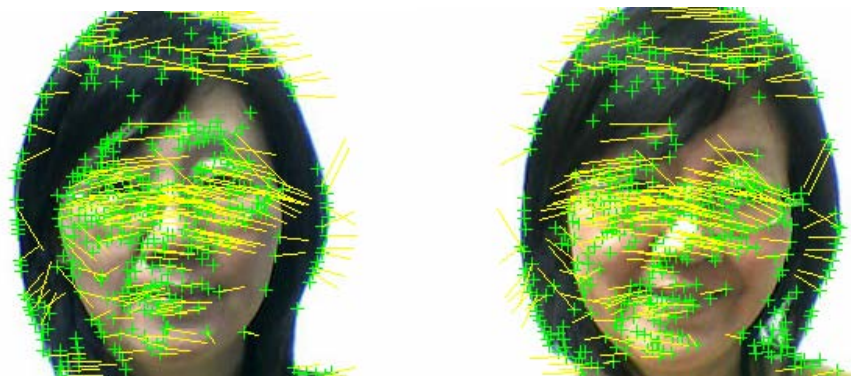
附近，上述兩區域的交集就是相似點所在的大致區域。S 為灰度相似性搜尋的區域，S1 為極線幾何限制的區域，S2 為單應限制的區域，綜合上述限制可知，相似點必然落在 S1 與 S2 的交集 S3 上，比較 S 搜尋區域，選擇 S3 定義的區域進行比對就大大的縮小相似點的搜尋範圍，也減少了錯誤比對的可能性。



(不同時間，距離 50CM 所拍的兩幅照片)



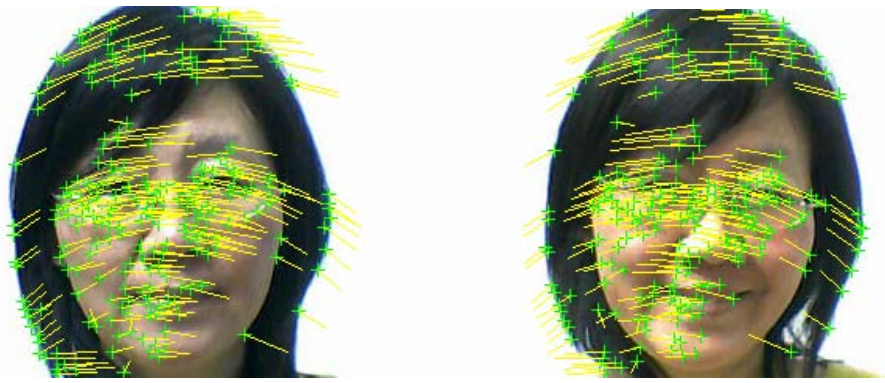
(進行雜訊濾波，哈里斯角點偵測 1000 個特徵點)



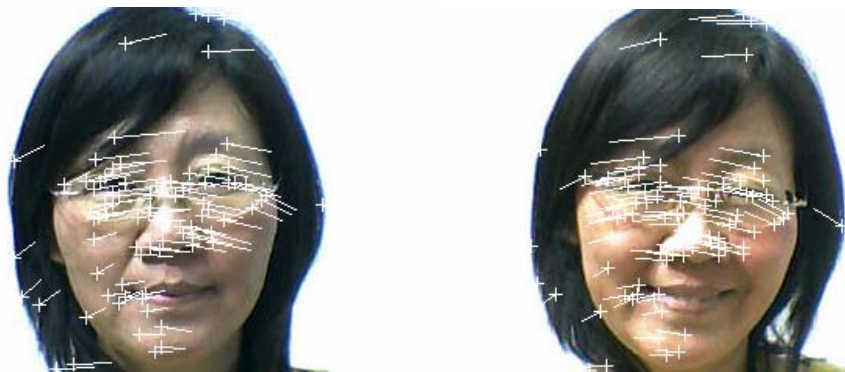
(進行灰度相似性比對，得到綠色相似點，黃線連接左右角點)



(進行極線幾何限制，白色極線，與極線相交的是紅圈，不相交的為藍點)



(進行單應限制，少了許多不正確的相似點)



(交集極線幾何限制、單應限制與灰度相似性的最終相似點)

現在我們拍了 320X240(像素)不同解析度的照片 4 組，640X480 大小的 4 組，分別編號 320X240：1a、1b、2a、2b、3a、3b、4a、4b，640X480：1a、1b、2a、2b、3a、3b、4a、4b，a b 為同一人不同時間所拍的照片。如需察閱，請見附錄〈二〉。320X240 與 640X480 分析的數據請參閱附錄〈三〉。

這裡設定特徵角點搜尋 1000 個、限制相鄰角點距離 3(像素)、灰度相似性搜尋範圍 50(像素)、灰度相似性搜尋視窗大小 20(像素)，經過測試，以上設定可以獲得比較好的結果及效率。

而解析度高低的影響層面有：比對速度、準確度。320X240 大小的圖像，比對速度大約是 4.1(秒/組)、準確度大約是 92.86%，640X480 大小的圖像，比對速度大約是 2.3(秒/組)、準確度大約是 96.43%。

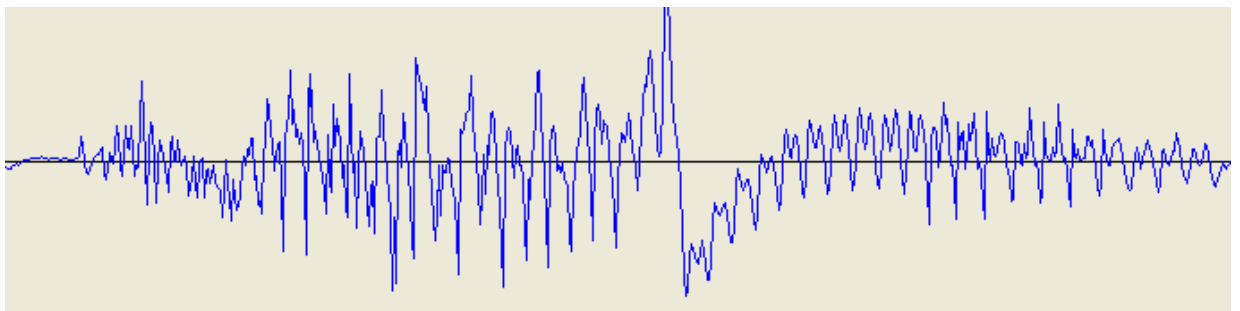
分析以上結果與檢視錯誤辨識的圖像，我們可知道影像較大的圖像擁有較好速度及準確度。

## 二、語音辨識

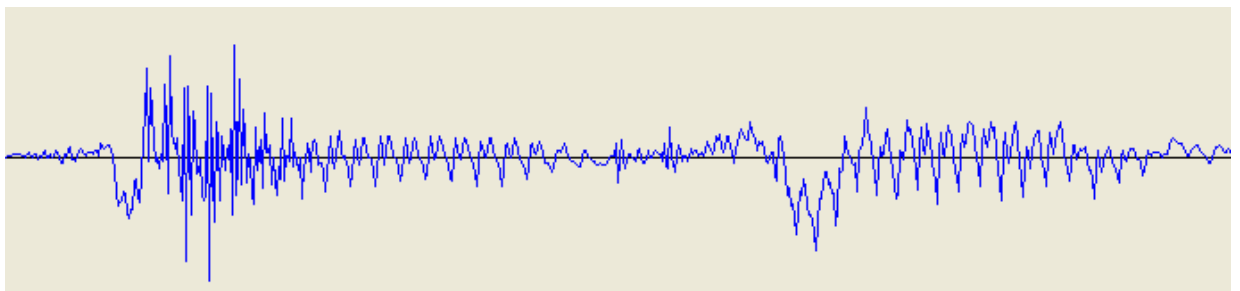
爲甚麼要使用語音辨識呢？因爲我們的攝影鏡頭沒有辦法每分每秒都在進行辨識，就算能捕捉到人臉，也不能確定他的目的是不是要進入這扇門。所以這裡就要使用語音辨識了。

這裡使用了 Microsoft Speech SDK 作爲我們的語音系統，因爲開發語音系統必須要經過長久的資料蒐集，才可能做出比較準確的引擎(Engine)。

進入門的指令有兩種，分別爲「開門」和「訪客」波形如下：



(開門)



(訪客)

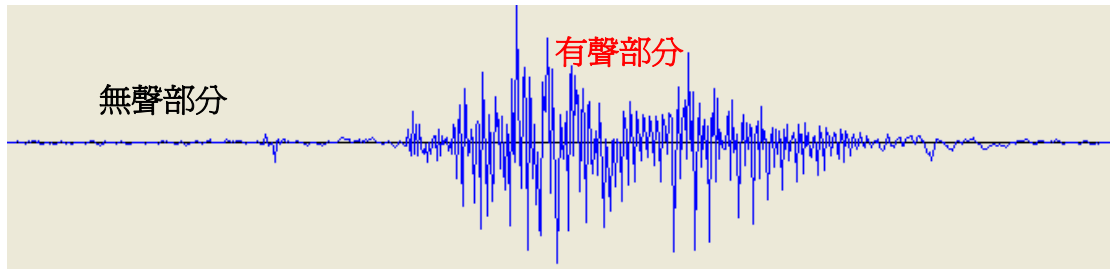
語音辨識的主要步驟如下：

1. 開啓麥克風，偵測目前的收到的聲音的音量
2. 偵測到聲音大於門檻值，開始錄音
3. 偵測到聲音小於門檻值，結束錄音
4. 將錄下的聲音進行特徵提取
5. 把特徵向量比對，確認結果(開門、訪客、錯誤)

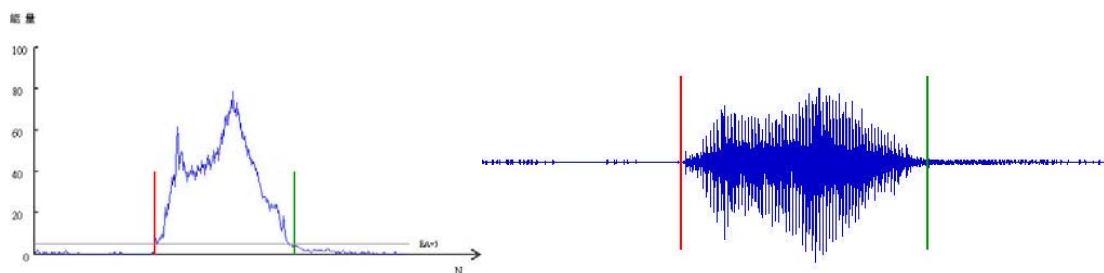
爲甚麼還要偵測聲音是否大於門檻值呢？當錄製一個語音時，如使用手動錄音的話前後一定會有靜音的部分，但這些靜音的部分我們並不需



要，所以我們必須偵測語音的端點，只錄取有聲部分。我們使用來偵測端點的方法是能量量測，能量量測法是在一段語音資料中，判斷有聲、無聲區域的方法。因為靜音部分的能量一定比有聲部分為低，因此只要能找出一個門檻值來區分有聲或無聲區域，然後判斷從哪一個音框開始平均能量超過預設的門檻，就能夠判斷出有聲部分是從哪一個音框開始了。



每一音框數為 256 點，將每一音框的能量值算出並畫出其能量圖如下面左圖，並設一門檻值，這裡的門檻值為 5(dB)，可以找出主要能量的起點與終點。

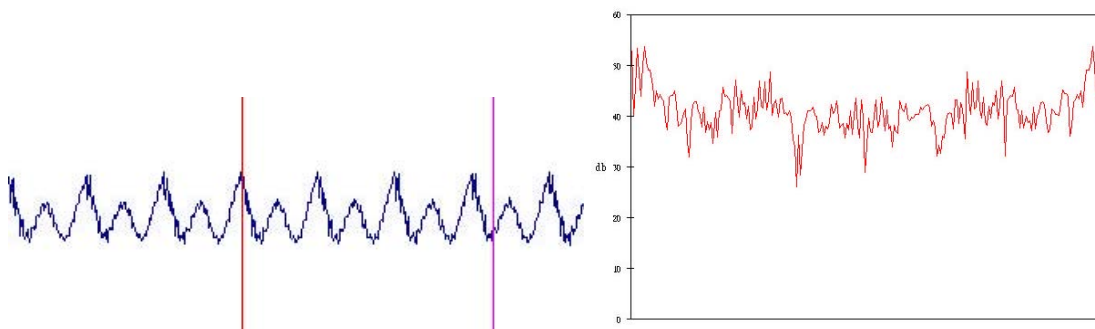


(紅線是錄取起點，綠線是結束點)

錄下了一旦聲音以後，就要進行比對，來確認是不是使用者的指令，或者只是環境聲音而已。

1. 我先將這段聲音去雜訊(35dB 以下)，然後標準化(當錄製語音時，會因為每個人說話的大小聲而影響語音信號取樣值的大小，而造成比對錯誤，因此我們需將語音信號的取樣值標準化。標準化的目的是將原始語音信號做等比例的放大或縮小，使其取樣值都落在同一範圍中)。

2. 將一語音信號切成數個音框，每個音框 256 個點。
3. 提取特徵向量，先對每個音框做 FFT(快速傅利葉轉換 Fast Fourier Transform)轉成頻域。下面左圖為其中一音框的在時域上的信號。經由 FFT 運算後，轉成下面右圖頻域上的信號。



4. 將一個音框的頻域切成 20 個頻帶，然後計算每個頻帶的能量，算出每個頻帶的能量佔全部頻帶能量的比例，得到每個音框的特徵向量。
5. 重複第四步驟，分析出所有音框，與之前所儲存的指令模型用樣本相關係數進行比對。

樣本相關係數的定理如下：

設有兩組樣本  $X_1, \dots, X_n$  及  $Y_1, \dots, Y_n$ ，其樣本平均數分別為  $\bar{X}, \bar{Y}$ ， $R$  表兩組樣本之間的相關程度，其值介於 -1 與 1 之間，越靠近 -1 與 1 表示越相似，越靠近零越不相似，可以把數據取絕對值，得到正數。

$$R = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\left[ \sum_{i=1}^n (X_i - \bar{X})^2 \right]} \sqrt{\left[ \sum_{i=1}^n (Y_i - \bar{Y})^2 \right]}}$$

### 三、資料庫

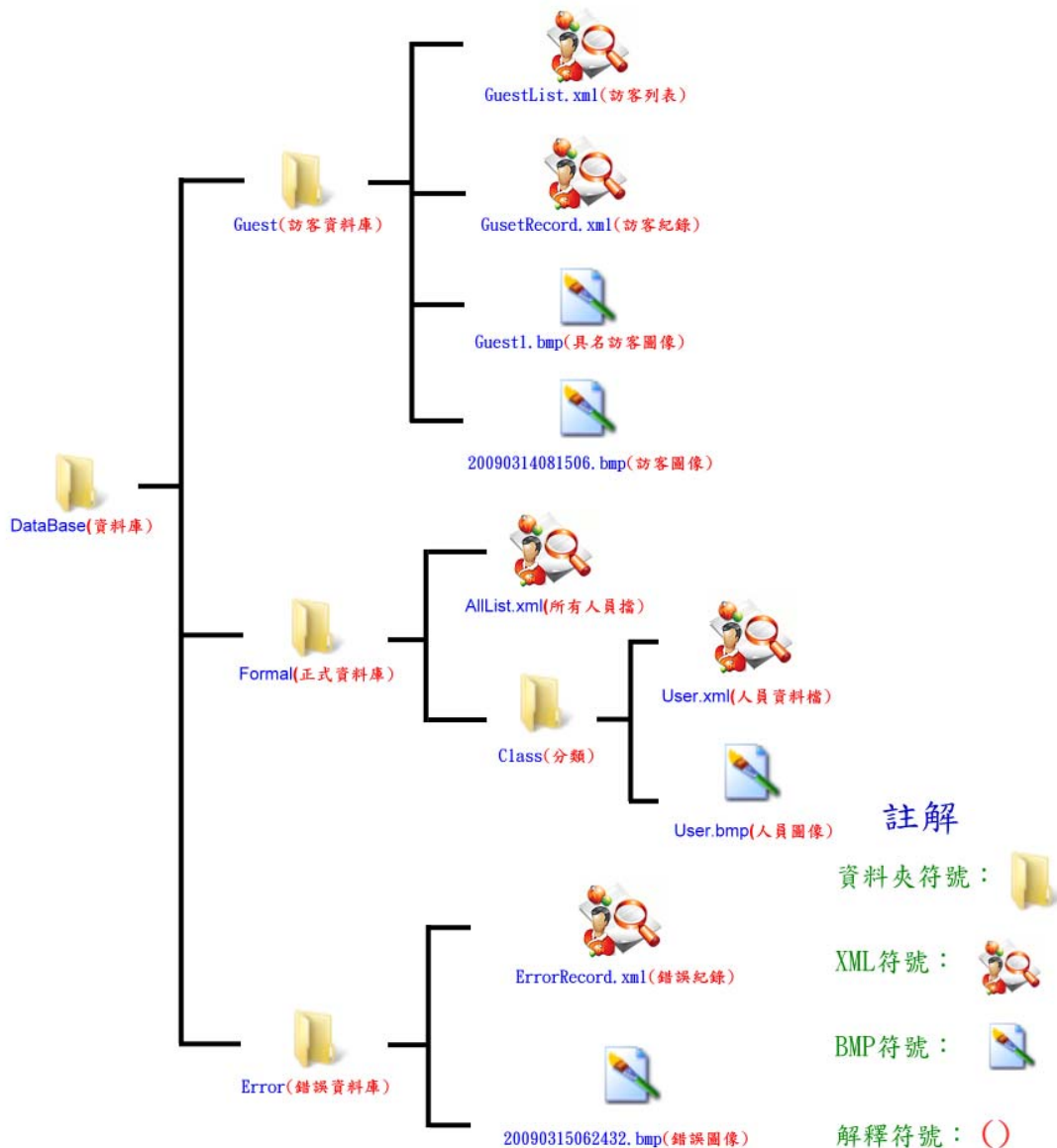
資料庫是程式很重要的一部分，因為我們需要把人員的資料進行儲存，而且以後還會常常讀取出來進行比對，所以希望資料庫能有結實的結構，也能很有效率的被讀出。

目前能跨越最多平台，相容性最好的資料格式便是 XML(eXtensible Markup Language 可擴展標示語言)。

XML 結構非常簡單，`<XXX>`左邊的兩個符號代表一個標籤，XXX 則代表這個標籤的意義，標籤結束只要在 XXX 前加上/即可，如右`</XXX>`，標籤中間則夾著你所要寫入的內容。

有些標籤是沒有開始與結束的，因為他只是給程式的訊息，而非主要要被顯示的內容。每個 XML 文檔都由 XML 序言開始，在前面的程式碼中的第一行便是 XML 序言：`<?xml version="1.0" encoding="BIG5"?>`，xml version 是描述 XML 的解析器和瀏覽器，encoding 代表本 XML 的解碼方式(Big5 是繁體中文)。還有`<![CDATA[XXX]]>`是描述 XML 文件中，不需要被解讀的符號(XXX)。另外還有一種標籤，`<?XXX?>`這行程式碼中的 XXX 代表要給解析器的指令，沒有任何規定。`<!--XXX-->`中的 XXX 則代表額外訊息。

整個資料庫的基本結構如下圖：



### 〈一〉 正式資料庫

我們資料庫中有分成兩種，一種是正式資料庫，另一種則是訪客及錯誤資料庫，這兩種資料庫的不同在於他們的資料內容是不一樣的。正式資料庫代表是通過這個門禁系統的人是這個空間的主要人員，系統如果辨識通過，會直接放行。

所有人員描述檔檔名為 AllList.xml，內容如下

```
<?xml version="1.0" encoding="BIG5"?>
```

```
<Class>
<ClassName>家人</ClassName>
<User1>ChougSan.xml</User1>…(依此類推)
</Class>
```

所有人員描述檔的內容包括<Class>分類、<ClassName>分類名稱、<User1>人員檔名。程式會透過這些訊息，往分類目錄讀取人員檔。

人員資料檔檔名由電腦產生，在使用者將人員加入資料庫後，電腦會將使用者給的人員名稱轉換成英文編碼，檔名也由此產生，內容如下：

```
<?xml version="1.0" encoding="BIG5"?>
<UserCode>ChougSan</UserCode>
<UserName>張三</UserName>
<AddTime>2009/03/14/23:12:03</AddTime>
<ImageName>ChougSan.bmp</ImageName>
<Rights>Allow</Rights>…
```

人員資料檔內容包括<UserCode>人員編碼、<UserName>人員名稱、<AddTime>加入時間、<ImageName>圖像名稱、<Rights>權限。程式會透過以上訊息讀入此使用者的圖像名稱，讀出圖像，為之後的比對工作做準備。

## 〈二〉 訪客及錯誤資料庫

訪客及錯誤資料庫與正式資料庫的資料不同在於：訪客及錯誤資料庫需要記錄的資料較少。訪客資料庫要記錄：訪客姓名(可有可無)、訪客編碼、造訪時間、訪客圖像，錯誤資料庫要記錄：錯誤時間、錯誤圖像。

訪客列表(GuestList.xml)結構如下：

```
<?xml version="1.0" encoding="BIG5"?>
<GusetCode>Guset1</GusetCode>
<GuestName></GusetName>
```

```
<GuestImage>Guest1.bmp</GuestImage>...
```

程式會透過以上訊息讀入訪客列表，為之後的比對做預先讀入。

當有客人造訪，不管是否之前有來過，相關訊息都會寫入訪客紀錄 (GusetRecord.xml)，結構如下：

```
<?xml version="1.0" encoding="BIG5"?>  
<VisitTime>2009/03/14/08:15:06</VisitTime>  
<GuestImage>20090314081506.bmp</GuestImage>...
```

訪客的圖像有兩種：訪客圖像、具名訪客圖像，訪客圖像每次有客人造訪時，都會擷取一張下來儲存。但是如果是之前有造訪過的客人，程式不但會擷取一張照片下來儲存，還會比對前的圖像，了解是哪位客人造訪。但是訪客要如何具名呢？這時候就要靠使用者在客人造訪後自行輸入了。

錯誤資料庫需要記錄的資料更少了，以下是錯誤紀錄 (ErrorRecord.xml)的結構：

```
<?xml version="1.0" encoding="BIG5"?>  
<ErrorTime>2009/03/15/06:24:32</ErrorTime>  
<ErrorImage>20090315062432.bmp</ErrorImage>...
```

程式會記錄錯誤的時間，跟當時錯誤所拍到圖像，讓使用者可以檢視。

## 陸、討論

### 一、問題討論

我們已經完成了兩組數據，證明系統之可行性，但如能解決下列問題，將可使系統之效能大幅提昇，並可減少使用限制：

1. 使用者的臉部問題：如果使用者的臉上出現重大改變，辨識系統將不能使用。
2. 使用者面對鏡頭的角度：如果用側面進行比對，會造成比對失敗。
3. 比對的速度：比對最快的速度是 2.3(秒/組)，但如果需要比對的組很多，可能速度就不夠了，但是此值易受電腦硬體速度影響。
4. 語音辨識：使用者口齒不清、有奇怪的腔調、環境音過大，很可能會造成無法辨識指令。

### 二、解決方法

本研究提出了相對應的解決方法：

- 〈一〉除了臉部辨識系統之外，可以加上說話人識別、密碼、RFID 感應卡、指紋、視網膜等，也是很方便辨識身分的系統，來輔助解決此問題。
- 〈二〉提醒使用者要使用正面面對鏡頭，也可以加上側面的鏡頭。
- 〈三〉改變比對的演算法(SIFT)或是使用硬體較好的電腦，一般現在市面上買到的新電腦，都可以縮短辨識速度。
- 〈四〉建立多方的語音模型並多次比對，減少辨識錯誤的機率。程式加入過濾環境雜音的功能，減少程式受環境雜音影響。

### 三、未來展望

未來可再往下列功能發展，以精進系統之方便性：

- 〈一〉本系統可以加裝另外一鏡頭，辨識臉的側面，但是需考慮減少雙鏡頭系統辨識的時間。

- 〈二〉有兩個鏡頭之後，讓兩個鏡頭所擷取到的圖像進行 3D 模型建立，產生一個臉部的 3D 模型，然後與之前產生的模型比對，減少失誤率。
- 〈三〉攝影裝置可以改為紅外線偵測、熱感應裝置，讓拍攝出來的圖像，比較不會受光線影響。
- 〈四〉程式加入臉部偵測(Face Detection)的功能。
- 〈五〉鏡頭加裝步進馬達來達到臉部追蹤(Face Tracking)、自動焦距調整(Zoom In/Out)。



## 柒、結論

### 一、辨識系統

經過開發設計測試，我們確實開發出了一套可使用的門禁系統，而準確度有 96.43%，圖像需要辨識的時間為 2.3(秒/組)，顯示此系統之實用性及顯著降低成本的可能性。

### 二、成本估算

本次科展研究之另一目的是要降低人臉門禁系統的設備成本。本研究所使用的個人電腦最低需求配備為：

中央處理器	記憶體	硬碟機	液晶螢幕
P4 2.66GHz 處理器	1G DDR	IDE 40G	19 吋

台灣的電腦普及率為 57.5%，本研究可使用一般家用電腦，只需購買簡單的設備，1,000 元以內即可完成。以下列出使用的設備、零件之費用：

設備	網路攝影機	通訊板	線材	燈具	總價
價格(元)	490	200	150	100	940

經過搜尋了解目前已商業化之人臉門禁系統，大陸一套要價(新台幣)49,660 元，本研究開發的系統成本大幅降低。

### 三、實用性討論

目前此系統已經達到相當的準確度，而且因為本研究大幅降低製作成本，所以普及性可以大大增加。如果經過商業化後將系統裝在每戶住家、公司或學校等地方，將能提升生活之便利性，以後出門不用帶鑰匙已不再是個夢。

## 捌、參考資料及其他

1. 梁棟, 童強, 屈磊, 王年, 韋穗,。一種基於極幾何和單應約束的圖像匹配演算法。系統仿真學報 2006 年 第 18 卷 第 01 期
2. 李健, 梁琨。基於小波多解析度分析的快速圖像匹配。陝西科技大學學報(自然科學版) 2006 年 第 24 卷 第 06 期
3. 楊茗惠。用隱藏式馬可夫方法於頻域特徵之國語數字辨識。2003 年
4. Epipolar Geometry, from the World Wide web:  
[http://en.wikipedia.org/wiki/Epipolar\\_geometry](http://en.wikipedia.org/wiki/Epipolar_geometry)
5. Homography, from the World Wide Web: <http://en.wikipedia.org/wiki/Homography>
6. 唐政元,吳怡樂,劉啓宗,陳加珍,林顯昌。對有遮蔽物的建築物影像作影像修補。2006 年
7. XML, from the World Wide web: <http://zh.wikipedia.org/wiki/XML>
8. Chinese XML Now, from the World Wide web: <http://xml.ascc.net/zh/big5/index.html>
9. 單位矩陣, from the World Wide web:  
<http://zh.wikipedia.org/wiki/%E5%96%AE%E4%BD%8D%E7%9F%A9%E9%99%A3>
10. OpenCV China, from the World Wide web: <http://www.opencv.org.cn/index.php/>
11. 唐政元,吳怡樂,許子彥,王文宏。使用 SIFT 與 PCA 做人臉辨識。2008 年
12. 林芸。非線性加權式演算法應用於 FDG-PET 動態參數評估。2006 年 7 月
13. 鄧寶松,高宇,楊冰,吳玲達。序列圖像中三維空間點的精確重建方法。2007 年 6 月
14. 徐偉,王朔中。基於視頻圖像 Harris 角點檢測的車輛測速。2006 年
15. DirectShow, from the World Wide web: <http://zh.wikipedia.org/wiki/DirectShow>
16. 趙輝。SIFT 特徵匹配技術講義。2006 年 6 月

## **【評語】 040816**

此作品強調成本的降低，創新的探討較少著墨，甚為可惜，系統的實用性，因諸多的限制，使用上並不是很方便，能夠將系統整合起來，誠屬不易。