A hybrid Face Recognition Technique as an Anti-Theft Mechanism

Othman Omran Khalifa*, Mohamed K. Kashil and Aisha Hassan Abdalla Hashim

1Electrical and Computer Engineering, Faculty of Engineering, International Islamic University Malaysia, Malaysia
* khalifa@iium.edu.my

Abstract. This paper proposes an anti-theft mechanism uses biometric Face Recognition to identify thief along with alarming. This can be used as security for ATMs, airport's systems, medical records, identify customers, preventing fraud and providing VIP services as well as recognizing individuals with known shoplifting convictions and video surveillance. The proposed system based on Viola Jones algorithm, Wavelet transform and Principal Component Analysis. Experimental results are given to demonstrate the viability of the proposed face recognition system with achieved efficiency is 82%.

1. Introduction
Humans have always had the ability to recognize and distinguish between faces, but computers are able to do so only recently. Due to the increasing public availability of facial images, especially as companies like Facebook pursue “real identity” policies, may result in an immense searchable multimedia database for previously unidentified individuals. Face recognition systems are capable of matching faceprints with individuals’ names at times when consumers’ identities are known, such as when using CCTV for monitoring. Currently, Computers use face-recognition systems to automatically identify a person from a digital image or a video frame from a video source [1][2]. The traditional technique used in face recognition is Principal Component Analysis (PCA). Other popular techniques/algorithms include Linear Discriminate analysis, Elastic Bunch Graph Matching and Multilinear Subspace Learning. Other techniques also exist which are either wavelet-based or curvelet-based [3][4][5].

Another method that some facial recognition algorithms use is identifying facial features of the subject’s face. This can be called “Face Recognition Using Features”. In this method, the size, shape and position of distinguishable parts of the face, such as distance between the eyes, width of the nose, shape of the cheekbones and the length of the jaw line are analyzed. These data can then be compared with a database containing data of faces already analyzed for matching features. This can be used to recognize the subject’s face with reasonable accuracy. Face Recognition has a wide variety of applications but the field where the application has found widespread use is security. Face recognition has become popular as an anti-theft mechanism and is already being used by banks and airports. Banks use face recognition in ATMs to protect customers against identity theft and fraudulent transactions. The benefit of using Face Recognition is that there is no need for the cooperation of the participant. The analysis is effective on the frontal image of the subject’s face. Although other extremely reliable methods such as fingerprint and retinal scanning do exist, they require the laborious participation of the subject.
2. Face Recognition System

A typical Face Recognition system includes three stages. The first stage involves pre-processing. It includes image normalization, noise elimination, illumination, and normalization. The second stage is facial feature extraction from the processed image obtained in the first stage. Finally, third stage involves classification of facial images based on the results obtained in the previous stage. Figure 1 shows the typical face recognition system.

![Typical Face Recognition System Architecture](image)

Figure 1. Typical Face Recognition System Architecture

2.1. Facial feature extraction

Facial feature extraction has become an important issue in automated visual recognition of human faces [6]. It is very clear that people are good at face identification, but it is not at all obvious how faces are encoded or stored. We can use facial features such as the eyes, mouth, chin and nose as important cues for discrimination and recognition of faces [7]. Colour may also be important in the recognition of faces [8]. Some research has suggested that it confers little recognition advantage for identifying faces while other research suggests that colour is a very important cue. The clear perception of colours in the environment illustrates that colour must be important in the interpretation of complex scenes and recognizing objects in the environment. Due to the similarity of the shapes of mouth and eyes to some geometric figures, we can extract them in terms of a deformable template model [9]. Other facial features such as eyebrows and nose are so variable that we must extract them by active contour model [10].

- **Deformable template model**: The deformable templates are specified by a set of parameters which use *a priori* knowledge about the expected shape of the features to guide the contour deformation process. The templates are flexible enough to change their sizes and other parameter values, so as to match themselves to the data. The final values of these parameters can be used to describe the features.

- **Active contour model (snake)**: Active contour is an energy-minimizing spline guided by the external constraint forces and influenced by image forces that pull it towards features such as lines and edges. Snakes lock onto nearby edges, localizing them accurately [11].

To use colour features in face recognition, we do the following:

1. **Training**: The colour image was split into three different images. The first image contains the information about the red colour, the second image contains the information about the green colour and the last image contains the information about the blue colour. Each pixel in each image has the value which represents the contrast for each colour. Then, each separate image (in the training set) is fed into a face recognition algorithm to calculate all the required variables which will be later stored.

2. **Recognition**: The same procedure was applied as above to the image recognition, where each separate image is fed into the recognition stage. After recognition is completed, three distance computations (one for Red, one for Green and one for Blue) are made. The result of distance
computation is then fed into a decision-making component. The global distance between the test and the training image is the weighted sum of the different distances [12].

2.2. Deformable template model

Conventional edge detectors cannot find facial features accurately from local evidence of edges because they cannot organize local information into a sensible global perception. For this reason, we can use templates. A template is specified by a set of parameters which uses a priori knowledge about the expected shapes of the features to guide the detection process. These templates are flexible enough to be able to change their size and shapes by changing the analytic parameter values of the template, and to match themselves to the data. The final values of these parameters can be used to describe the features.

2.3. Eye template

The deformable template acts on three representations of the image, as well as on the image itself. The first two representations are peak and valleys in the image intensity and the third is the place where the image intensity changes quickly. A simplified model is shown as in figure 2 below:

![Figure 2. simplified eye template [13]](image)

We assume that the possible ranges of length, height₁, height₂ and orientation are: a) length $\varepsilon[l-\alpha_1, l+\alpha_2]$, b) height₁ $\varepsilon[h_1-\beta_1, h_1+\beta_2]$, c) height₂ $\varepsilon[h_2-\gamma_1, h_2+\gamma_2]$ and d) orientation $\varepsilon[\theta-\theta_1, \theta+\theta_2]$. The tolerance $\alpha_1, \beta_1, \gamma_1$ and $\theta_1, i=1, 2$ are determined by Rough Contour Estimation Routine. ($P_x, P_y$) represents the centroid point. ($x', y'$) is the translated and rotated version of (x, y) and their relationships are given by:

$$x - P_x = \dot{x} \cos \theta + \dot{y} \sin \theta$$
$$y - P_y = -\dot{x} \sin \theta + \dot{y} \cos \theta$$

The total energy function is defined as:

$$E_{total} = E_{edge} + E_{white} + E_{black}$$

The $E_{edge}$, $E_{white}$ and $E_{black}$ are defined in the following:

- The edge potentials are given by the integral over the curves of upper and lower parabola divided by length:

$$E_{edge} = \frac{w_1}{\text{upper-length}} \int_{\text{upper-bound}} \phi_{edge}(x,y) dS - \frac{w_2}{\text{lower-length}} \int_{\text{lower-bound}} \phi_{edge}(x,y) dS$$

where Upper-bound and Lower-bound represent upper part and lower part of the eye, $\phi_{edge}(x,y)$ represents the edge response of point (x,y).

- The potentials of white and black points are defined as the integral over the area bounded by the upper and lower parabola divided by the area:
\[ E_{w-b} = \frac{1}{\text{Area}} \int_{\text{Para-area}} \left(-w_b N_{\text{black}}(x,y) + w_w N_{\text{white}}(x,y)\right) dA \]  

(4)

where \( N_{\text{black}}(x,y) \) and \( N_{\text{white}}(x,y) \) represent the numbers of black and white points. \( W_b \) and \( W_w \) are weights related with black and white points. In order not to be affected by improper threshold \((\theta)\), we define the tolerance \((\epsilon)\) for determining black and white points as:

- \( P(x,y) \) is a Black point if \( I(x,y) \leq \theta - \epsilon \)
- \( P(x,y) \) is a white point if \( I(x,y) \geq \theta + \epsilon \)
- \( P(x,y) \) is an ambiguous point if \( \theta - \epsilon \leq I(x,y) \leq \theta + \epsilon \)

Using the energy functions defined above, we can calculate the energy in the range of little modulations of \( \text{length, height}_1, \text{height}_2 \) and \( \text{orientation} \). When the minimum energy value takes place, the precise contour is extracted.

2.4. Mouth template
If the front view of the face is taken, the mouth has a relatively important role to play in face recognition. The mouth template is shown as in figure 3.

- The equation of middle lips (i.e. crack between upper and lower lips) is given as:
  \[ \dot{y} = \text{height}_s - \frac{\text{height}_s}{\text{length}_2} \dot{x}^2, \dot{x} \in [M_{xc} - \frac{\text{length}}{2}, M_{xc} + \frac{\text{length}}{2}] \]  
  (5)

- The equation of the lower lip is:
  \[ \dot{y} = \text{height}_d - \frac{\text{height}_d}{\text{length}_2} \dot{x}^2, \dot{x} \in [M_{xc} - \frac{\text{length}}{2}, M_{xc} + \frac{\text{length}}{2}] \]  
  (6)

- The equation of the upper lip consists of two parts:
Mouth contour energy function consists of edge terms $E_{\text{edge}}$ and black terms $E_{\text{black}}$. The black term is dominant at the edge area whereas the black term encloses as many black points belonging to the mouth as possible. They are described as:

$$E_{\text{total}} = E_{\text{edge}} + E_{\text{black}}$$

The equation for $E_{\text{edge}}$ consists of three parts: middle lip, lower lip and upper lip.

The equation for $E_{\text{edge}}$, is:

$$E_{\text{edge}} = -\frac{w_{\text{lower}}}{\text{Lower} - \text{length}} \int_{\text{Lower}} \phi_{\text{edge}}(x,y) \, dS - \frac{w_{\text{lef}}}{\text{Lef} - \text{length}} \int_{\text{Lef}} \phi_{\text{edge}}(x,y) \, dS - \frac{w_{\text{righ}}}{\text{Righ} - \text{length}} \int_{\text{Righ}} \phi_{\text{edge}}(x,y) \, dS$$

where $\text{Lower}$ represents the lower boundary of mouth, $\text{Lef}$ represents the left part of upper lip, $\text{Righ}$ represents the right part of upper lip and $\phi_{\text{edge}}(x,y)$ represents edge response of point $(x,y)$.

The black energy function is designed to enclose as many black points belonging to the mouth as possible. It is defined as:

$$E_{\text{black}} = \frac{1}{\text{Area}} \int_{\text{Lbound}} w_{\text{black}} N_{\text{black}}(x,y) \, dA + \frac{1}{\text{Mid} - \text{length}} \int_{\text{Mid}} -w_{\text{Mid}} N_{\text{black}}(x,y) \, dS$$

where $\text{Lbound}$ represents lower lips, $\text{Ubound}$ represents upper lip, $\text{Mid}$ represents gap between lips.

2.5. Active contour

The shapes of eyebrows, nostrils and face, unlike the eye and the mouth, varies for different people and their contours cannot be captured by using deformable template. In this case, we can apply a model called “snake” or “active contour models” to extract the contours of eyebrows, nostrils and face. A snake is an energy minimizing spline guided by external constraint forces and influenced by image forces that pull it towards features such as lines and edges. Snakes are active contour models. They lock onto nearby edges and localize them accurately. The energy function of the active contour model is defined as:

$$E_{\text{snake}} = \int_0^1 E_{\text{snake}}(v(s)) \, ds = \int_0^1 [E_{\text{internal}}(v(s)) + E_{\text{images}}(v(s)) + E_{\text{constraint}}(v(s))] \, ds$$

where $v(s)$ represents the position of snake, $E_{\text{internal}}$ represents the internal energy of the contour due to bending, $E_{\text{images}}$ gives rise to the image forces and $E_{\text{constraint}}$ represents the external energy. The original active contour model is user-interactive. The advantage here is that the final form of the snake can be influenced by feedback from a higher level process.

Active contour energy is defined as:
\[ E_{\text{total}} = \int_0^1 (\alpha(s)E_{\text{continuity}}(v(s)) + \beta(s)E_{\text{curvature}}(v(s)) + \gamma(s)E_{\text{images}}(v(s)))ds \]  
(12)

Fast iteration method is used for the minimal energy iteration process of the active contour model but this method has its disadvantages where contour points will accumulate at certain strong portions of the active contour. The approximation used is shown:

\[ \left| \frac{dv_i}{ds} \right| = \left| v_i - v_{i-1} \right|^2, \left| \frac{d^2v_i}{ds^2} \right| = \left| v_{i-1} - 2v_i - v_{i-1} \right|^2 \]  
(13)

3. Wavelet Transform

Wavelet transform technique is a new field in face recognition and it has an impact on some old and new disciplines. Compared with Fourier transform and Gabor transform, it can extract useful information effectively. Thus, it can solve some problems that Fourier transform cannot resolve. Wavelet transform provides a powerful and versatile framework for image processing. It is widely used in the fields of image de-noising, compression, fusion, and so on. The changes of expressions in the sample images of an individual result in the differences of higher frequency band of the images. The basic functions of wavelet transform are obtained from a single prototype wavelet (or mother wavelet) by translation and dilation.\(^6\)

\[ \psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi(\frac{t-b}{a}) \]  
(14)

Where “a” and “b” are both real numbers which quantify the scaling and translation operations. Substitute “a” and “b” with “2m” and “n\times2m” respectively, the basic functions become:

\[ \psi_{a,b}(t) = 2^{-m/2} \psi(2^{-m}t-n) \]  
(15)

The process of one-dimensional discrete wavelet transform and reconstruction a signal \( x(t) \) is defined as follows:

\[ W(m,n) = \langle x(t), \psi_{m,n}(t) \rangle \]

\[ \hat{x}(t) = \sum_{m,n} W(m,n) \hat{\psi}_{m,n}(t) \]  
(16)

The two-dimensional wavelet transform is got by applying one-dimensional wavelet transform to the rows and columns of two-dimensional data. An approximation image is derived from 1-level wavelet decomposition of an image and three detail images in horizontal, vertical and diagonal directions respectively. The approximation image is used for the next level of decomposition.

Figure 4. The process of decomposing an image
3.1. Face Recognition

Once the face has been cropped out, we must work on it to complete the recognition process. We do a wavelet transform on this image. The coding for wavelet transform is available in the Appendix. The wavelet transform [15] is a well-known signal processing tool which has been successful to many image processing applications, particularly for image compression, where the transient features of the image are represented using few scaling function coefficients whereas the sharp nature of the image are represented using wavelet function coefficients.

Once wavelet transform is complete, we have to do a Principle Component Analysis [16]. This is a standard tool in modern data analysis in diverse fields including computer graphics because it is a simple method for extracting relevant information from confusing data sets. Principle Component Analysis (PCA) transforms data in the image to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate, the second greatest variance on the second coordinate and so on. The code for Principle Component Analysis is shown in the Appendix. By performing PCA on the image after wavelet, we can use this information to compare images in the database with the input image. Wavelet transform and PCA have been already done on the images in the database. Now, the system will compare this input’s result with the database and attempt to find a match. If a match is found, the system will allow access to the user. If no match is found, access is denied.

4. Principle Component Analysis
Principal Component Analysis (PCA) is used to find a low dimensional representation of data. Some important details of PCA are highlighted as follows:

Let \( X = \{X_n R^d | n = 1, \ldots, N\} \) be an ensemble of vectors. In imaging applications, they are formed by row concatenation of the image data, with \( d \) being the product of the width and the height of an image.

Let \( E(X) = \frac{1}{N} \sum_{n=1}^{N} X_n \)

be the average vector in the ensemble. After subtracting the average from each element of \( X \), we get a modified ensemble of vectors,

\[
\bar{X} = \{\bar{X}_n, n = 1, \ldots, N\} \text{with } \bar{X}_n = X_n - E(X)
\]

The auto-covariance matrix \( M \) for the ensemble \( X \) is defined by

\[
M = \cos(\bar{X}) = E(X \circ X)
\]

where \( M \) is a \( d \times d \) matrix, with elements

\[
M(i, j) = \frac{1}{N} \sum_{n=1}^{N} \bar{X}_n(i) \bar{X}_n(j), 1 \leq i, j \leq d
\]

It is well known from matrix theory that the matrix \( M \) is positively definite (or semi-definite) and has only real non-negative eigenvalues. The eigenvectors of the matrix \( M \) form an orthonormal basis for \( R^d \). This basis is called the K-L basis. Since the auto-covariance matrix for the K-L eigenvectors are diagonal, it follows that the coordinates of the vectors in the sample space \( X \) with respect to the K-L basis are un-correlated random variables. The PCA of a vector \( y \) related to the ensemble \( X \) is obtained by projecting vector \( y \) onto the subspaces spanned by \( d' \) eigenvectors corresponding to the top \( d' \) eigenvalues of the autocorrelation matrix \( M \) in descending order, where \( d' \) is smaller than \( d \). This projection results in a vector containing \( d' \) coefficients \( a_1, \ldots, a_{d'} \). The vector \( y \) is then represented by a linear combination of the eigenvectors with weights \( a_1, \ldots, a_{d'} \).

PCA has been widely adopted in human face recognition and face detection since 1987. However, in spite of PCA's popularity, it suffers from two major limitations: poor discriminatory power and large computational load. It is well known that PCA gives a very good approximation in face image. However, in eigen space, each class is closely packed.

5. Proposed method
Traditionally, to represent the human face, PCA is performed on the whole facial image. However, this approach suffers from the limitations mentioned earlier. To resolve these limitations, we proposed a new method to use PCA - applying PCA on the wavelet subband 4. Daubechies wavelet D4 is adopted for image decomposition. Proposed system consists of two stages, namely, training and recognition stages as shown in figure 9. Training stage computes the representational bases for images in the domain of interest (that is reference images) and converts them into training image representations. The training image representations of each image are stored into the library.
6. Results analysis

The face detection process was very fast and efficient during simulation. Viola Jones algorithm was very effective in detecting multiple faces in images. But this algorithm wasn’t perfect. There were instances where it could have many bounding boxes- some with the faces cropped out, the others with just the background.

We can see in the above image that Viola Jones algorithm detects all the four faces that are visible in the image but it also has detected two extra “faces”. One of them is on the right bottom corner of the image which is actually part of a shirt. The other one is in the middle bottom part of the image and the detected “face” is very small and also part of a shirt.

This occurs in solo images (images with one face) as well, as shown above. In this image, the face is detected along with another “face” which is actually part of a design in the shirt. We can see that the falsely detected second “face” is actually a pair of headphones but the shape of it confused the system resulting in a face being detected. This is fatal to the system. When two “faces” are detected, the system crashes resulting in halting of the face recognition process. The system is supposed to crop the face out, but since two faces have been detected, it is confused as to which face to crop out. This results in the process stopping. After face detection is complete, it was important that the cropped-out
face be resized to a standard resolution of 92×112. All the images saved in the database are of this particular resolution but the resolution of the face that will be detected varies for each image. If the resolution of the test image did not match the resolution of the images in the database, the program retrieves an error saying, “Matrix dimensions do not match”. Therefore, I had to edit the code so that once the face is detected, the program will automatically resize the cropped image to the standard size. Table 1 shows the results of a series of tests was done on the system to evaluate its performance. Figure 10 shows an image with a face that is angled towards the left. Viola Jones algorithm detects the inclined face.

| Total number of images tested | No face detected | False face detected | Single correct face detected | Face detection percentage |
|------------------------------|------------------|---------------------|-----------------------------|--------------------------|
| 50                           | 1                | 8                   | 41                          | 82%                      |

Figure 11. Testing face detection at different face angle

7. Conclusion
In this paper, an enhance face detection and recognition as an Anti-Theft Mechanism was proposed, the system we came up with was to use facial features of the human for face recognition. The wavelet transform was used in the pre-processing stage, but they were not the main method used for the actual recognition. The distinctive features of the human face such as the eyes and mouth to come up with a system to recognize the human face. The system build that can be used in ATMs, or PCs to prevent theft of financial valuables such as cash or credit or valuable information. The Proposed Anti-theft system can be made more efficient by improving precision of Face Recognition and adding other biometric applications such as finger print recognition along with Face Recognition feature for more secureness also by incorporating neural networks, a machine-learning AI technique that may increase the efficiency of these systems.

REFERENCES

[1]. Yugang J, Ping G 2005 Face Recognition by Combining Wavelet Transform and k-Nearest Neighbor, Journal of Communication and Computer, Vol. 2, No.9, pp 50-53.
[2]. Haiyang Z 2011 Image Preprocessing Methods in Face Recognition”, 2011 Symposium on Photonics and Optoelectronics pp 16-18.
[3]. Samadani R, Sundararajan A and Said A 2004 Deringing and deblocking DCT compression artifacts with efficient shifted transforms, In Proceedings of International Conference on Image Processing Volume 3, pp 1799–1802.

[4]. Adini Y M 1997 Face Recognition: the Problem of Compensating for Changes in Illumination Direction. IEEE Trans. on Pattern Analysis and Machine Intelligence pp 721-732.

[5]. Finlayson G D, Schiele B and Crowley J 1998 Comprehensive colour image normalization. In: Burkhardt H., Neumann B. (eds) Computer Vision — ECCV’98. ECCV 1998. Lecture Notes in Computer Science, vol 1406. Springer, Berlin, Heidelberg.

[6]. Govindaraju V, Sripati S N and Sher D B 1990 A computational Model for Face Location, Proceedings Third International Conference on Computer Vision.

[7]. Waters K, Terzopoulos D 1991 Modeling and animating Faces using scanned data”, Journal of Visualization Computer Animation, Vol. 2, no. 4, pp 123-123.

[8]. Karimi B and Krzyżak A 2007 A Study on Significance of Color in Face Recognition using Several Eigenface Algorithms, Canadian Conference on Electrical and Computer Engineering.

[9]. Yuille A I, Cohen D S and Hallinan P W 1989 Feature extraction from faces using deformable templates, in proceedings of CVPR, pp 104-109.

[10]. Berger M O and Mohr R 1990 Towards Autonomy In Active Contour Models, In Proceedings of the 10th ICPR (I), pp. 847-85.

[11]. Kass M, Witkin A and Terzopoulos D 1988 Snakes: Active contour models, International Journal of Computer Vision 1, pp 321-31.31.

[12]. Jacobs I S and Bean C P 1963 Fine particles, thin films and exchange anisotropy, in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, pp 271–350.

[13]. Chung-Lin Huang and Ching-Wen Chen, Human Facial Feature Extraction for Face Interpretation and Recognition, Institute of Electrical Engineering, National Tsing Hua University

[14]. Viola P and Jones M 2001 Robust Real-time Object Detection, Cambridge.

[15]. Malmurugan N, Shanmugam A, Jayaraman S and Chander V V 2005 Academic Open Internet Journal, Vol 14.

[16]. Jonathan S 2009 A Tutorial on Principal Component Analysis, (April 22, 2009), Center for Neural Science, New York University.