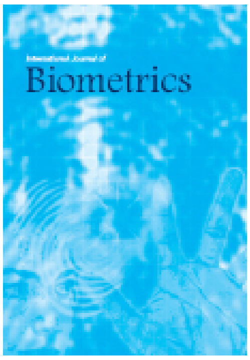


International Journal of Biometrics

ISSN online: 1755-831X - ISSN print: 1755-8301
https://www.inderscience.com/ijbm

Effect of facial expression in face biometry for a multimodal approach

Samik Chakraborty, Dipayan Chatterjee, Subrata Golui, Madhuchhanda Mitra, Saurabh Pal

DOI: 10.1504/IJBM.2023.10051693

Article History:

Received:	31 December 2020
Last revised:	25 August 2021
Accepted:	03 November 2021
Published online:	15 December 2022

Effect of facial expression in face biometry for a multimodal approach

Samik Chakraborty*

Department of AEIE,
Heritage Institute of Technology,
Kolkata, India
Email: samikasan0516@gmail.com
*Corresponding author

Dipayan Chatterjee, Subrata Golui, Madhuchhanda Mitra and Saurabh Pal

Department of Applied Physics,
University of Calcutta,
Kolkata, India
Email: dipayanchatterjee8@gmail.com
Email: subrata.golui@bsnl.co.in
Email: madhuchhanda94@rediffmail.com
Email: spal76@gmail.com

Abstract: With the advancement of technology, security has become an inseparable part of it. But many factors often influence the accuracy of the authentication system. In this current scenario, the multimodal biometric system is used where information from different modalities are fused to address the weakness of the system. In the present work, a robust biometric authentication system proposed using face and facial expression as biometric modalities. Facial recognition is the most commonly used biometric system over the years. Facial expressions of an individual are unique and it is integrated as an additional layer along with face recognition system to enhance the security of the system as the current scenario tends towards intelligent security systems for real-time surveillance. After pre-processing, eigenvalue-based and local binary pattern (LBP)-based features are extracted from the face and facial expression and the information are fused. Finally, the authentication is done using image Euclidian distance (IMED) based classifier. This proposed work evaluated using the JAFFE and Yale database and 95.71% and 88.89% authentication accuracy is achieved, respectively.

Keywords: face biometry; facial expression; local binary pattern; LBP; eigenfaces; IMED classifier.

Reference to this paper should be made as follows: Chakraborty, S., Chatterjee, D., Golui, S., Mitra, M. and Pal, S. (2023) 'Effect of facial expression in face biometry for a multimodal approach', *Int. J. Biometrics*, Vol. 15, No. 1, pp.101–119.

Biographical notes: Samik Chakraborty is currently with the Department of Applied Electronics and Instrumentation Engineering Department of Heritage Institute of Technology and holding the designation of Assistant Professor. He

has received his BTech in Applied Electronics and Instrumentation Engineering in 2007 and MTech in Instrumentation and Control Engineering in 2010. Presently, he is pursuing PhD in the Department of Applied Physics, University of Calcutta.

Dipayan Chatterjee has received his BTech in Instrumentation and Control Engineering in 2015 and MTech in Instrumentation and Control Engineering in 2019.

Subrata Golui graduated from the University of Calcutta with a BSc in Physics as a major subject in 2006. He has received his BTech in Radio Physics and Electronics and MTech in Instrumentation and Control Engineering from the University of Calcutta, India in 2009 and 2019, respectively. Since 2010, he has been working as a Sub Divisional Engineer for Bharat Sanchar Nigam Limited (BSNL), a Telecom Department of the Government of India.

Madhuchhanda Mitra is with the Department of Applied Physics, University of Calcutta over last 24 years. She received her BTech, MTech and PhD from the same department. Her field of research includes biomedical signal and image processing. She has published more than 150 research papers in national and international journals and conferences.

Saurabh Pal received his BTech, MTech and PhD (Tech) in 2000, 2006 and 2013 respectively, from the University of Calcutta, Calcutta, India. He has teaching and research experience of more than 16 years. At present, he is an Associate Professor, Department of Applied Physics, University of Calcutta, India, in the Instrumentation Engineering section. His field of interest includes digital signal processing, biomedical instrumentation and biomedical signal and image processing. He is the co-author of more than 45 research papers in biomedical signal processing and analysis domain.

1 Introduction

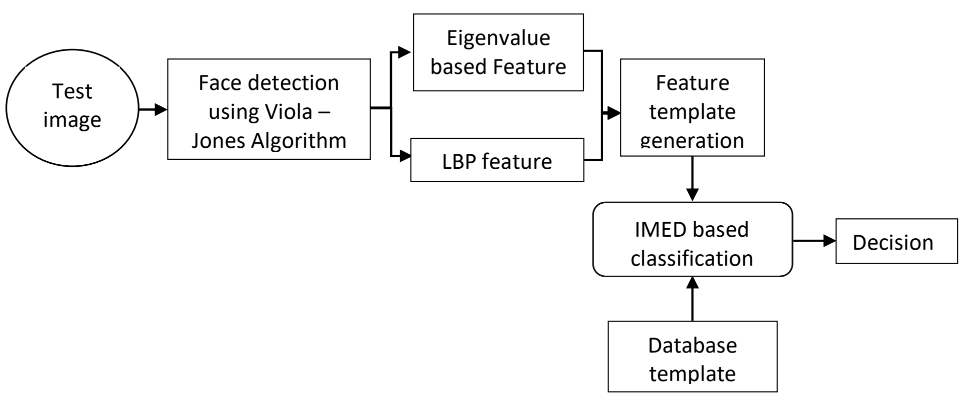
Biometric recognition is an automated method that is based on human physiological and behavioural attributes. For identification and verification, it becomes a well-known alternative over traditional password and PIN-based methods as biometric traits are unique, cannot be stolen or forgotten. In literature, several physiological and behavioural characteristics are used as biometric features which are widely known for their capability to serve promising and convincing security in the authentication system, such as, face (Turk and Pentland, 1991), fingerprint (Maltoni et al., 2003), ear (Chakraborty et al., 2020), palmprint (Kong et al., 2006), iris (Vatsa et al. 2008), ECG (Chakraborty et al., 2017), PPG (Chakraborty and Pal, 2016) and so on. Among all the biometric traits, face recognition techniques have received the most attention over the decades. In comparison with other biometrics, it is found that face biometric has distinct advantages due to its natural and non-intrusive enrolment process. Face images are gathered via a camera from a distance without having physical contact with the person himself/herself. This recognition technique is extensively used in forensic, surveillance and security authentication. During the last few decades, many approaches are developed and proposed for successful face recognition systems. Face recognition methods generally use two common approaches, i.e., appearance-based and feature-based methods. In the appearance-based method, various parameters of the pixels like intensity, histogram, etc. are represented as a 2-dimensional array while in the feature-based approach, the image is represented by locating the distinct parts of the face such as eyes, nose, ears, mouth, etc. A lot of research has been done under these categories with benchmark works are presented below.

The initial framework for the detection of the face was proposed by Lucas and Kanade (1981) and then developed further by Tomasi and Kanade (1991). Later, Viola and Jones (2004) developed a method using the AdaBoost learning algorithm that was very fast and could rapidly detect frontal view faces. Turk and Pentland (1991) described the eigenfaces method which is the most well-known approach for face recognition. For encoding texture and shape description for digital images, Ojala et al. (1996) introduced the local binary patterns (LBP) operator. Later, Ahonen et al. (2004) successfully applied the LBP operator to face recognition by dividing an image into regions from which LBP features were extracted. LBP analyses the relation between a pixel and its neighbours and encodes this relation into a binary word. Khan et al. (2018) proposed PCA eigenface algorithm for face recognition system and achieved 86% and 80% accuracy for constrained and unconstrained environments respectively. Abdullah et al. (2017) suggested a PCA based face recognition method for criminal detection with 80% recognition accuracy. Dhamija et al. (2017) used a combination of PCA, Fisher face and SVD techniques for face recognition on a data set that consists of images of 40 subjects and achieved 99.5% recognition rate. This study concluded that a combination of these techniques enhances the efficiency and accuracy of the system. Bhavani et al. (2017) used a sparse representation technique for recognition of faces in a video where the Viola-Jones algorithm was used to detect the faces in an image. Kakkar and Sharma (2018) provided an adequate face recognition system for criminal identification by using Haar feature-based cascade classifier and LBP histogram. Maheswari et al. (2020) concluded that the PCA approach can still recognise faces successfully, despite the fact that several factors influence face recognition results. Bah and Ming (2020) proposed a new method that combines the LBP algorithm with modern image processing techniques to address some of the challenges that impede face recognition accuracy. To improve the system's accuracy, Sahan and Al-Itabi (2021) merged local and global features. The local descriptor is based on the Radon Transform, which captures the image's local variation. Chebyshev– Fourier moments are used to extract global features.

Although face biometrics is one of the most effective and widely used biometric systems, different factors can affect the reliability of the system. Major challenges encountered by the face recognition system lies in the difficulties of handling varying poses, illuminations and different expressions. Facial expression interprets emotional and mental states of a human being. During facial expression, movement of facial muscles results in a shape change in facial features (Jafri and Arabnia, 2009). A facial image registered with neutral or a specific expression can be misinterpreted when compared with the same face with a different expression. Many studies have been done to analyse the reliability of face biometric system under the influence of facial expression but a worthwhile outcome has not been achieved yet. Also, over the past few years, researchers showed their interest to utilise facial expressions for human identification. Different facial expressions contribute to the facial behaviour of a person, however, these changes in expressions due to changes in behaviour, cannot serve as distinct factors independently while identifying a subject. Rather it can be considered as a soft biometric. Extensive research has found that six fundamental facial expressions such as anger, disgust, fear, happiness, sadness, surprise shared by human reflect crucial information about the mental, emotional and even intentions (Ekman and Keltner, 1997; Matsumoto and Kupperbusch, 2001) Few studies have been made on the relevance of facial expression as a biometric trait. Tulyakov et al. (2007) reported that facial expressions such as smile, etc. are useful for identification system. Faundez-Zanuy et al. (2008) reveal that some facial expressions produce a low recognition rate. As it has been observed that facial expression does not provide enough discriminating power; it can be used as a modality of a multimodal biometric system in order to make it more reliable. Using this concept Mei Yin et al. (2018) proposed a match score level fusion-based scheme using face and facial expression for authentication and achieved a significantly high accuracy rate over the unimodal system.

Motivated from the existing literature, for enhancing the security applications, a robust biometric system based on the human face and facial expression is proposed here. The major aim of this study is to generate a typical feature template-based automatic authentication system regardless of expression to maintain the reliability of the biometric system. The basic advantage of using face and facial expressions as biometric modalities is that it requires a common sensor to acquire data and also these two modalities act as physiological and behavioural traits of the face respectively. Moreover, with the fusion of the above-mentioned parameters verification is done in a more versatile manner encompassing possible physiological and behavioural variations that may occur. With the advancement of cybercrimes day by day, face biometry will become vital for authentication in everyday life. Integration of facial expression can act as an additional layer to identify the genuine user, thereby increasing the security of the system.

Figure 1 Block diagram of the proposed biometric recognition system



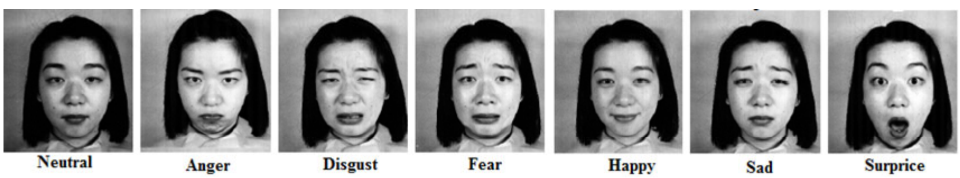
In the present work, a complete biometric authentication framework is proposed where face image is detected using the Viola-Jones detection algorithm and from the detected image eigenface based features and geometrical features are extracted for face and facial expression recognition respectively. A biometric feature template is generated by fusing information from both the attributes. Finally, authentication is claimed by the image Euclidean distance (IMED) algorithm (Wang et al., 2005). The general block diagram of the proposed recognition system is shown in Figure 1. The remainder of this paper is organised as follows: Section 2 describes the databases which are used to develop and evaluate the proposed technique. Section 3 describes the methodology in detail. Experimental results are shown in Section 4 and Section 5 concludes the paper.

2 Database used

Effectiveness of the proposed recognition system is evaluated using two databases – JAFFE database (Lyons et al., 1998) and Yale database (Belhumeur et al., 1997).

Set A JAFFE database contains images from ten subjects. Each subject posed at least three times for each of the seven facial expressions: neutral, happy, sad, surprise, anger, disgust, fear. In Figure 2 an example of a subject with seven different expressions are shown.

Figure 2 A typical subject from JAFFE database with different facial expressions



Set B The Yale facial expression database consists of the faces of 15 individuals. Each individual contains 11 grayscale images. Images were captured from various angles, with different light conditions and various facial expressions. Out of 11 images, six images are of different expressions, i.e., happy, neutral, sad, sleepy, surprise and wink and the other three images are of different lighting positions: centre light, right and left light and the rest two images are with and without glass images.

Figure 3 A typical subject from Yale database with different facial expressions

3 Methodology

3.1 Face detection using Viola-Jones algorithm

Detecting faces from the backdrop is the first stage in facial recognition systems. Also, face detection can be used for picture capture, video coding, video conferencing, and crowd control in public places. The face detection method is useful for searching and storing data about facial features in images and videos that comprise faces of different shapes and sizes. The Viola-Jones method, also known as the first object detection system, is a rapid, accurate, and efficient method for finding a face in an image. As it is fast and adaptable in providing a framework with a high detection rate, this approach is suitable for real-time use (Viola and Jones, 2004). It is Paul Viola and Michael Jones presented this fast and robust method for face detection in the year 2001. The algorithm follows four steps: Haar features selection, creating an integral image, Adaboost training algorithm and classification using cascaded classifiers. The basic principle is to scan a sub-window for detecting significant sections across the image. This algorithm is a cascaded Adaboost classifier based on image integral and rectangular features reminiscent of Haar wavelets.

The Viola-Jones face detector analyses a given sub-window using features consisting of two or more rectangles. The different types of features are shown in Figure 4. All human faces share some similar properties. These regularities may be matched using Haar Features. Figure 5 shows examples of Haar features used for nose and eye detection

Figure 4 Different types of Haar features

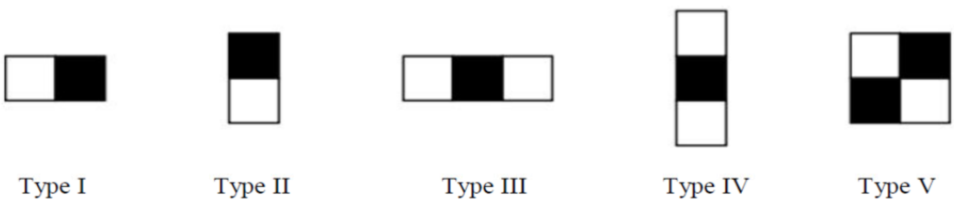
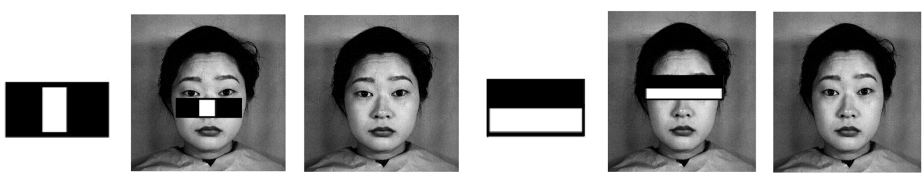


Figure 5 (a) Haar feature used for nose detection and (b) Haar feature used for eye detection



(a) (b)

An integral image is an intermediate representation of an image where the value of the location of an integral image is computed from an input image by making each pixel equal to the entire sum of all pixels above and to the left of the concerned pixel as in Figure 6.

Figure 6 (a) Input image (b) Integral image

1	1	1
1	1	1
1	1	1

(a)

1	2	3
2	4	6
3	6	9

(b)

AdaBoost algorithm is a machine learning-based boosting algorithm which combines weak classification functions and forms a strong classifier through a weighted combination of weak classifiers.

The cascade classifier is more efficient and it rejects most of the negative sub-windows and detects the positive sub-windows. The cascaded classifier is composed of stages each containing a strong classifier. Each stage classified a sub-window passed to the next stage or discarded depending on the sub-window consists of face or not respectively. The more stages a given sub-window passes, the higher the chance the sub-window contains a face. Viola-Jones also refers to the cascaded classifier as an attention-based cascade as more attention is directed towards the regions of the image suspected to contain faces. When reducing some computation in identifying the face area, it allows the background area of the image to be swiftly deleted.

For a trained cascade classifier, the rate of detection is:

$$D = \prod_{i=1}^{K} d_i \tag{1}$$

where K is denoted as the number of classifier and d_i is the detection rate of the i^{th} classifier.

3.2 Preprocessing

Before use as an input image to the LBP operator, skin region of the face image is detected first. Garcia and Tziritas (1999) proposed a reliable skin colour model that is adaptable to people of different skin colour s and to different lighting conditions. The common *RGB* representation of colour images are not suitable for characterising skin colour as the *R*, *G*, *B* component represents not only colour but also luminance, which varies due to ambient lighting, makes an unreliable measure in the skin from non-skin regions. *YCbCr* chromatic colour space separates luminance in *Y* component from the colour information and hence, provides a way to use only colour information for segmenting skin and non-skin regions. Due to this fact, skin colour model is developed in chromatic colour space (*YCbCr* chromatic colour space) and only chrominance components (*Cb* and *Cr*) are used for modelling the skin pixels. The values of *Cb* and *Cr* are selected from Owusu et al. (2014). The skin classifier is represented in terms of 0 and 1. The values of 0 and 1 represent non-skin and skin pixel respectively.

3.3 LBP based facial expression feature extraction

In facial expression analysis position of facial features plays very essential role as emotion is more often communicated by the facial movement which will change the textural structure of facial features and other regions in a face. LBP is used to extract the local texture description of a grayscale image. It is widely used in face recognition system. Ojala et al (1996) first introduced LBP operators for encoding image information of texture and description. LBP operator is calculated in a 3 × 3 neighbourhood as shown in Figure 7. For a grayscale image I, p_c (i_c, j_c) is any pixel position within the local area of the image. To extract LBP features LBP operator is calculated in a 3 × 3 neighbourhood where p_c as centre of a 3 × 3 window and the other points are $p_0 \dots p_7$. So, the local area texture T can be defined as

$$T_{LBP} = t\left(p_c, p_0, p_1, \cdots p_7\right) \tag{2}$$

Now, a binary code is produced by comparing its neighbourhood with the value of the central pixel where the value of the centre pixel used as threshold and it can be rewritten as

$$T_{LBP} = t\left(p_c, \left(p_0 - p_c\right), \left(p_1 - p_c\right) \cdots \left(p_7 - p_c\right)\right) \tag{3}$$

Assuming the differences is independent of p_c, it is further described as

$$T_{LBP} = t\left(p_c\right) t\left(\left(p_0 - p_c\right), \left(p_1 - p_c\right) \cdots \left(p_7 - p_c\right)\right) \tag{4}$$

Since luminance of an image is described by $t(p_c)$, local image texture can be described by

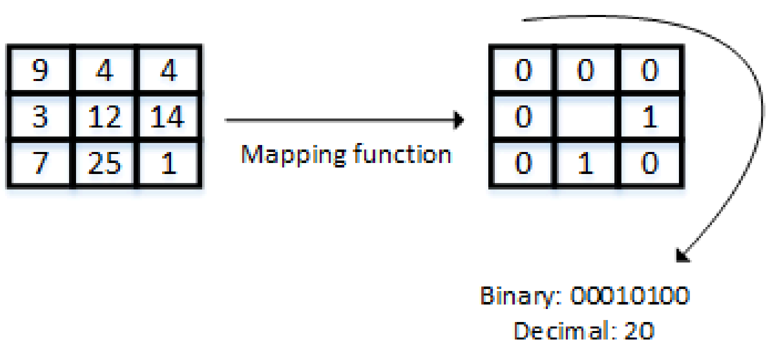
$$T_{LBP} \approx t\left(s\left(p_0 - p_c\right), \left(p_1 - p_c\right) \cdots \left(p_7 - p_c\right)\right) \tag{5}$$

where, $s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$

Now the unique LBP code is achieved by assigning a binomial weight to each sign $s(p_n - p_c)$ where $n = 0, 1, \ldots, 7$.

$$T_{LBP}\left(i_c, j_c\right) = \sum\nolimits_{n=0}^{7} s\left(p_n - p_c\right) \cdot 2^n \tag{6}$$

Figure 7 LBP operator on the centre pixel as a threshold



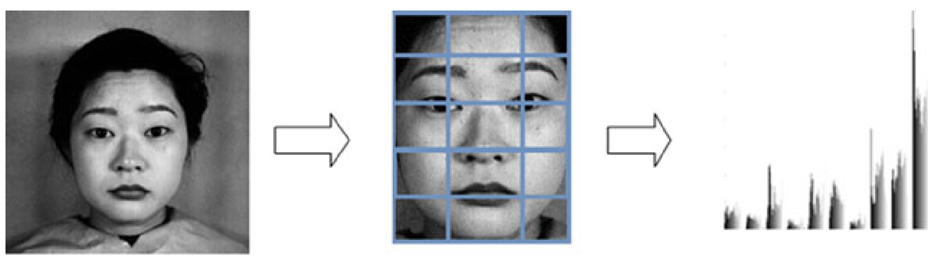
After scanning a facial expression image by the LBP operator, the LBP coding image of the original image is obtained. Then the texture feature of the image can be described by counting the facial expression image histogram. The process of LBP feature extraction is shown in Figure 8. The circular neighbourhood, defined by the number of neighbours (N) and radius (R), rotation invariance, and uniformity, are the three parameters of the general LBP operator. These characteristics must be explored in order to find the optimal combination for a particular application. The distribution of spots and edges throughout the entire image is captured in the LBP histogram in this approach. There are 256 unique labels in the (8, 1) neighbourhood where N = 8 and R = 1 and the LBP histogram descriptor has a 256-dimensional depth. As this LBP histogram descriptor generates histogram for the entire image, the total image is divided into block by block. There are

LBP variants with a uniform mapping, no mapping, and number of neighbours is 8 and radius of the circle is 1. Each of these LBP blocks is then used to determine the contribution of local features to the total local feature vector using 2D-DCT. Only the most significant 2D-DCT components are chosen after zigzagging scanning each of the generated DCT coefficients. Then by concatenating the block-wise local feature vectors, the overall local feature vector is generated. In the experiment, blocks of various sizes are used for this local extraction part and it has been found that block sizes of 12×12 gave the best results.

An LBP histogram in his approach contains information about facial micro-patterns like the distribution of edges, spots and flat areas over the whole image. In case of (N, R) = (8, 1) neighbourhood, there are 256 unique label and the dimension of LBP histogram descriptor is 256.

The LBP coding image includes local micro-mode information of the original image, such as edge, feature points and spot, etc. So, the local texture feature of a facial expression image can be described by a histogram which is formed by LBP codes.

Figure 8 LBP based face representation (see online version for colours)



3.4 Eigenvalue based facial feature extraction

Eigenvalues play an important role in image processing applications. A well-known approach for face recognition is the eigenfaces method described by Turk and Pentland (1991). Eigenfaces find a set of subspaces based on Principal Component Analysis (PCA) from unlabelled face image data, where grey-level images are reduced down to the most variant feature by projecting them to a lower dimension subspace. PCA is an important statistical procedure. It determines the data mean and principal components. With this technique, the data is analysed for strong patterns and variations in the data. In most cases, this approach is used to maximise variance and capture strong patterns in a dataset. PCA performs effectively for correlated data. Data with high correlation exists in images as well. PCA performs better at extracting features from images because of this. Image matrix is turned into a lower-dimensional eigen subspace by performing several operations on it. Afterwards, calculate the covariance matrix from the matrix of smaller dimension. Covariance matrix represents the relative variance between pixels in an image. This covariance matrix is then used to create eigenvectors. Principal components are those eigenvectors with the highest eigenvalues. Eigenface is a set of eigenvectors used for human face identification. Image features are extracted from the image and the most prominent eigenface is picked for facial recognition. Principal component analysis is used in face recognition to represent the face image efficiently using eigenface.

To extract the features first mean (μ) across all training images (M) are calculated as,

$$\mu = \frac{1}{M}\sum_{i=1}^{M} x_i \tag{7}$$

where x_i is one of the vectors in the training set. Then the covariance matrix S is calculated as,

$$S = \frac{1}{M}\sum_{i=1}^{M} (x_i - \mu)(x_i - \mu)^{T_i} \tag{8}$$

where T_i is the transposed vector.

A positive value of covariance shows that dimensions are directly proportional to each other, i.e., increase or decrease together. A negative value indicates an increase in one dimension while a decrease in the other and a value of zero shows that the dimensions are independent of each other. Eigenvalues (λ_i) and eigenvectors (v_i) of covariance matrix S are computed using equation (9).

$$SV_i = \lambda_i v_i \tag{9}$$

Eigenvectors are organised by their eigenvalues. Eigenvectors with small eigenvalue are less significant than those with higher eigenvalue and can be simply ignored. Eigenvectors with higher eigenvalues are the principal component of data. The group of selected eigenvectors is called the eigenfaces. After obtaining the eigenfaces all images are projected into the eigenface space and the weights of the image in that space are stored.

3.5 Feature template generation by concatenation of feature sets

LBP is one of the most powerful descriptors, which summarise the local structure of an image. Due to its computational simplicity, it has been successfully used for many different image analysis tasks (Ahonen et al., 2004; Zhang et al., 2004). From the review of previous research, it has been found that mouth, eyebrow areas contribute most to facial expressions. Eigenvector based method extracts the global grayscale feature of the whole facial image and reduced the data size at the same time. To enhance the performance of the biometric recognition method fusion of feature is introduced. To obtain the global feature vector, PCA is applied to the entire image in this study. The eigenvectors are sorted in the PCA method according to the eigenvalues' descending values. However, because PCA is also used for dimension reduction, after sorting the original global feature vector, only the most significant features vector values are selected. LBP histogram shows the distribution of edges, spots, and flat areas across the entire image, as well as other facial micro-patterns. It's worth noting that the basic LBP histogram descriptor obtained in this case is for the entire image and can be considered a global descriptor. The descriptor represents facial patterns, but due to the global histogram operation, the spatial location information is lost. However, the block-by-block version of the LBP image used in this study results in a local representation of the image's facial features. This local and global feature template of each subject is concatenated into a new feature vector template that works as a person's composite identity template, enriching the feature set with the invariant properties of both the feature sets. Serial concatenation, parallel fusion employing a complex vector, and algorithms that extract correlation features from several modalities are all common feature fusion techniques (Ma et al., 2020). The serial concatenation approach, in contrast to the other two ways, is straightforward to use, effective, and extensible to combine more than two modes. In this study, the final template is created by combining the features of each of the templates as shown in equation (10) and the lengths of the feature vectors were discovered to be 489 following fusions.

If, F_i^L and F_j^G are two different feature set where F^L and F^G refers to the LBP and eigenvector feature sets and i, j corresponds to the size of the LBP and eigenvalue respectively, then the final template is written as,

$$F_{(i+j)}^{fusion} = F_i^L + F_j^G \tag{10}$$

3.6 IMED based classification

To evaluate the recognition rate of the proposed model IMED based minimum distance classifier is used. Among all the image metrics, Euclidean distance is the most commonly used due to its simplicity. It is used to classify unknown image data by measuring the minimum distance between the test image data and the stored data in the database. Here, distance is defined as an index of similarity so that the minimum distance corresponds to the maximum similarity. The conventional Euclidean distance is susceptible to distortion and translation since it does not take into account pixels' spatial relationship. To some extent, IMED compensates for this flaw. Wang et al. (2005) proposed an IMED that takes into consideration the spatial relationship of image pixels and is resistant to noise and slight deformation. The two-step procedure of calculating the IMED of images has been demonstrated. It was calculated in two steps. The original images are first transformed linearly, and then the Euclidean distance between them is calculated. As a result, it may be readily integrated into a number of current pattern classifiers, including PCA and SVMs. IMED is based on the assumption that pixels next to one another have little variation in grey levels, and it simply considers the distance between pixels on the image lattice when determining the relationship between pixels.

Let x, y be two M by N images, $x = (x^1, x^2, \cdots, x^{MN})$, $y = \left(y^1, y^2, \cdots, y^{MN}\right)$, where x^{kN+1}, y^{kN+1} are the grey levels at location (k, l). The Euclidean distance d^E (x, y) is given by,

$$d_E^2(x, y) = \sum_{k=1}^{MN} \left(x^k - y^k\right)^2 \tag{11}$$

4 Result and discussion

The total experiment is divided into three parts. In the first and second part of the experiment, the proposed recognition system is used as a unimodal system using face and facial expression respectively as an individual modality. In the third part, the proposed system is used as a multimodal biometric system by combining both attributes. Performance of the proposed methods is measured by calculating recognition accuracy, sensitivity and specificity which are defined as follows:

$$Recognition\ Accuracy(\%) = \frac{No.\ of\ correctly\ classified\ subject}{Total\ number\ of\ subject} \times 100\% \tag{12}$$

$$Sensitivity(\%) = \frac{No.\ of\ true\ positive}{Total\ number\ of\ true\ positive\ and\ false\ negative} \times 100\% \tag{13}$$

$$Specificity(\%) = \frac{No.\ of\ true\ negative}{Total\ number\ of\ true\ negative\ and\ false\ positive} \times 100\% \tag{14}$$

4.1 Experiment on face based unimodal system

As mentioned earlier, in the first part, the proposed system is evaluated as a unimodal system where the face and facial expression act as individual biometric modality respectively. For eigenfeature-based face recognition system, features from images with the neutral expression of every subject are extracted and stored as a database. Then, for new entry similar features are extracted and a comparison is done with the new entry against the stored database. It is obvious that for test image only image with neutral expression is considered.

First, the proposed system is evaluated using the JAFFE database and Yale database separately. In the JAFFE database, each subject has at least three images with neutral expressions. Out of them one image of the neutral expression of every subject is used as a training image whereas the rest of the neutral images are used as testing images for performing the recognition process. It has been found that using IMED classifier for a random testing image, 90% accuracy is obtained while the face is using as a single modality for biometric authentication. Both sensitivity and specificity as calculated using this technique are 90% (Table 6).

Secondly, this proposed system again re-evaluates using the Yale database. Yale database consists of face images of 15 subjects and each subject contains neutral expression images of different configurations: normal, with glasses, with no glasses, centre-light, left-light and right-light. Out of these, features from normal expression images of each subject are stored as a database. Images of a subject with or without glass along with the different lighting effects have been considered as a new entry in the field of study. It is to be noted that in reality the subject to wear glass or not depends on normal circumstances. The Yale database has images of 15 subjects, out of which two (subject 8 and 13) are found to be wearing glasses in normal condition with a normal expression. Therefore, for these two subjects' images with glasses are considered within the training database for the study. Rest 60 images are used as testing images and 86.67% accuracy, has been obtained using IMED classifier. The sensitivity and specificity as derived in this method are 87.50% and 88.89%, respectively (Table 6). Table 2 shows the percentage of accuracy achieved by the system when images of different configurations are used as a new entry. From this experiment, it has been seen that images with no glasses of all the subjects classified correctly but the recognition rate decreased when images of different lighting conditions are used.

Table 1 Result of IMED classifier for a random test image of subject 3 of JAFFE database

Subject	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Min. distance	19.57	10.54	*0.57*	6.93	8.21	15.53	7.23	11.86	7.43	11.32

Table 2 Accuracy achieved for testing images of different configurations

Images considered as testing image (nos)	*Accuracy achieved (%)*
W/glasses or w/no glasses images (15)	100
Image with centre-light (15)	86.67
Image with left-light (15)	80
Image with right-light (15)	80
Average	*86.67*

4.2 Experiment on facial expression based unimodal system

To estimate the effectiveness of using the facial expression as a modality of a biometric authentication system, features from the neutral expression of faces per subject are used as training images and one image of the different class of expressions of the same subject is used for testing purposes. This experiment also evaluated using JAFFE and Yale databases.

In the JAFFE database, each subject contains more than one image for seven basic expressions, i.e., neutral, happy, sad, surprise, anger, disgust and fear. Feature template for the stored signature database is prepared by averaging the LBP features of any two neutral expression images. While features from the images of the different classes of expressions of the same subject are considered as a new entry. JAFFE database contains at least three images for each basic expression of ten subjects and features from the neutral image stored as database and images of all other expressions considered as testing images. It has been observed that facial expression as biometric modality 78.33% accuracy is achieved including all expression for the entire database. Table 3 presents the results of the IMED classifier where neutral expression of a subject used as training and the other class of expressions the same subject as testing images whereas Table 4 shows the expression specific percentage of accuracy achieved by the classifier.

Similarly, the effectiveness of using the facial expression as a modality of a biometric authentication system, the proposed system once again evaluated using Yale database. In this study, a feature template for the database is prepared by averaging LBP features of the normal image, an image with no glasses or with glass depending on the subject's habit of generally wearing glass or not. Yale database contains only one image per configuration or expression per subject and from Section 4.1 it has been seen while the classification of the facial image that images with no glasses or with glasses of all subjects have been classified correctly with a hundred percent accuracy. So, for obvious reason, an image with no glasses of each subject is used with a normal image for the averaging feature set except for subject 8 and subject 13. In the case of these two subjects, with glass images are used for the averaging feature set. Yale database contains images of 15 subjects with normal expression and five other basic expressions, i.e., happy, sad, sleepy, surprise and wink. Therefore, a total of 75 images were used as testing images and it has been observed that facial expression as biometric modality

74.67% accuracy is achieved. An overview of the classifier result of the Yale database is shown in Table 5.

Table 3 Minimum distance value between neutral image (training image) and image with different expressions (testing image) of the same subject of JAFFE database

Neutral expression of a subject as training image	*Different expression of the same subject as testing image*					
	Happy	*Sad*	*Surprise*	*Anger*	*Disgust*	*Fear*
Subject 1	2.32	1.34	5.94	3.02	7.83	3.23
Subject 2	3.24	1.43	5.53	3.33	8.67	4.56
Subject 3	2.41	1.87	6.34	3.46	7.54	3.44
Subject 4	2.75	1.63	5.65	2.95	10.54	5.63
Subject 5	3.5	2.31	5.32	4.01	9.32	5.42
Subject 6	2.91	2.03	4.61	3.06	7.53	5.39
Subject 7	2.95	2.63	7.64	4.65	12.54	6.53
Subject 8	1.89	2.85	7.91	3.78	11.73	4.54
Subject 9	3.72	1.56	6.04	3.07	10.23	6.37
Subject 10	3.27	3.94	5.98	3.64	8.67	6.84

Table 4 Result of IMED classification on JAFFE database with facial expression as test image

Expression specific accuracy (%)						*Accuracy achieved (%)*	*Sensitivity (%)*	*Specificity (%)*
Happy	*Sad*	*Surprise*	*Anger*	*Disgust*	*Fear*	(*including all expression*)		
93.33	93.33	70.00	76.67	56.67	80.00	*78.33*	*78.57*	*78.95*

Table 5 Result of IMED classification on Yale database with facial expression as test image

Expression specific accuracy (%)					*Accuracy achieved (%)*	*Sensitivity (%)*	*Specificity (%)*
Happy	*Sad*	*Sleepy*	*Surprise*	*Wink*	*(including all expression)*		
86.67	86.67	80.00	60.00	60.00	74.67	74.47	75

The results from both the databases show that expressions like ‘happy’ and ‘sad’ are more reliable as a biometric trait with respect to other common ones available in the test databases. This indicates less variation in face pattern due to sudden expression in these two emotions compared to others. It is observed from the results of the JAFFE database that poor accuracy is achieved in the case of disgust expression. If disgust expression is eliminated from the testing images, then including all other expressions achieved accuracy, sensitivity and specificity increased up to 82.67%, 84.75% and 85% respectively. From this, it can be concluded that during authentication extreme care should be taken with disgust expression.

4.3 Experiment on multimodal system

In the last part of the experiment, both face and facial expression attributes combined by fusion of features (Section 3.5) and achieved higher accuracy. To act the system as a multimodal system, the final feature template is generated by concatenation of eigenface feature and LBP feature as shown in equation (10). For the JAFFE database, by averaging the final feature template of two neutral images of each subject, the database for classification is created and images of other basic expressions are used as a new entry to the system. The accuracy of the system is measured as 95.71%, whereas face and facial expression as individual modality provide 90% and 78.33% respectively. The system's sensitivity and specificity improved by 97.06% and 94.44%, respectively. As discussed in Section 4.2, by eliminating the images with the expression of disgust, the accuracy of the multimodal system increased up to 96.67%and the sensitivity and specificity achieved in this case are 97.73% and 95.65%.

Table 6 Comparison of accuracy rate of unimodal and multimodal approach on JAFFE and Yale database

Database / *Used approach*	*JAFFE database*			*Yale database*		
	Accuracy	*Sensitivity*	*Specificity*	*Accuracy*	*Sensitivity*	*Specificity*
Eigenvalue based	90%	90%	90%	86.67%	87.50%	88.89%
LBP based	78.33%	78.57%	78.95%	74.67%	74.47%	75%
Fusion of Eigenvalue with LBP (including all expression)	*95.71%*	*97.06%*	*94.44%*	*88.89%*	*90.70%*	*89.36%*

Table 7 Comparison between different approaches

Study	*Database used*	*Feature used*	*Accuracy*
Shih et al. (2008)	JAFFE database	DWT	94.13%
Ramesha and Raja (2011)	JAFFE database	DT-CWT	88.6%
Li and Yahya (2014)	JAFFE database	Gabor wavelets and SVD	75.2%
Rahulamathavan and Rajarajan (2017)	JAFFE database	PCA	94.37%
Fatihah et al. (2018)	Yale database	LBP	82%
Wang et al. (2018)	Yale database	2D DWD	84.3%
Maw et al. (2020)	JAFFE database	Eigenfaces	80%
Shinwariet al. (2019)	JAFFE database	LDA	99.53%
Ayeche et al. (2020)	JAFFE database	LGN descriptor	84.28%
Present work	Yale database	Eigenvalue + LBP features	88.89%
	JAFFE database		95.71%

Notes: DWT: Discrete wavelet transformation, DT-CWT: Dual tree complex wavelet transform, SVD: Singular value decomposition, PCA: Principal component analysis, DWD: Discrete wavelet domain, LDA: Linear discriminant analysis, LGN: Local gradient neighbourhood.

Similarly, in the Yale database, computation of final stored data is done by averaging the final feature template of normal and no glass image of each subject except subjects 8 and 13. Similarly, like previous sections, these two subjects with glass images are used for the generation of the final feature template. Happy, sad, sleepy, surprise, wink and centre light configuration images of each subject are used as a new entry to the system. The accuracy of the system has been measured as 88.89% which is slightly higher than previous results obtained from the face and facial expression based unimodal system. The sensitivity and specificity of the system rises correspondingly up to 90.70% and 89.36%. Table 6 shows the accuracy rate, sensitivity and specificity of the suggested approaches applied to both databases. It is observed that combining with face the accuracy of the system increased and also the fusion of facial expression with face acts as an additional layer of security. The result obtained is compared with some previously reported works as in Table 7. The comparison is done on the basis of accuracy only as the other performance parameters are not reported in previous articles. It is noticed that only LDA based classification performs better than the present work. However, other statistical indicators might have been more useful to compare the performances.

5 Conclusions

This paper proposes a novel biometric system applying a fusion approach for face and facial expression as biometric modalities. Face and facial expression are the physiological and behavioural traits of human beings and both are fused for authentication thus the proposed method presents a relatively more robust, accurate recognition rate and cost-effective model. From the performance evaluation, it has been concluded that the fusion of face and facial expressions significantly increases the accuracy rate if a multimodal recognition system is employed. This study found that during the expression of surprise and disgust the accuracy of the system decreases whereas during happy and sad expressions better accuracy achieved. This is because the facial pattern mostly changes during surprise and disgust than the latter two. From this study, it can be concluded that utmost care must be taken during the course of authentication with surprise, disgust and fear expression if the normal facial image without any expression is stored. Feature expression has a great potential to be used in human-computer interaction such as for security and surveillance. It is found to be effective as an offender's behaviour can be predicted by analysing the images of their face. This kind of system can be used to preserve the security of the user's identity and its data by authorising the user using his emotions as behavioural traits where the intensity of his emotion will be used as a verification.

References

Abdullah, N.A., Saidi, M.J., Wen, C.C. et al. (2017) 'Face recognition for criminal identification: an implementation of principal component analysis for face recognition', in *Proceedings of 2nd International Conference on Applied Science and Technology 2017* (*ICAST'17*), Langkawi, Kedah, Malaysia, pp.1–6.

Ahonen, T., Hadid, A. and Pietikäinen, M. (2004) 'Face recognition with local binary patterns face recognition with local binary patterns', in Pajdla, T. and Matas, J. (Eds.): *Computer Vision – ECCV 2004, Lecture Notes in Computer Science*, Vol. 3021, pp.469–481, Springer, Berlin, Heidelberg.

Ayeche, F., Alti, A. and Boukerram, A. (2020) 'Improved face and facial expression recognition based on a novel local gradient neighbourhood', *Journal of Digital Information Management*, Vol. 18, No. 1, pp.33–42.

Bah, S.M. and Ming, F. (2020) 'An improved face recognition algorithm and its application in attendance management system', *Array*, Vol. 5, No. 1, pp.1–7.

Belhumeur, P.N., Hespanha, J.P. and Kriegman, D.J. (1997) 'Eigenfaces vs. fisherfaces: recognition using class specific linear projection', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp.711–720.

Bhavani, K., Dhanaraj, V., Siddesh, N.V. et al. (2017) 'Real time face detection and recognition in video surveillance', *International Research Journal of Engineering and Technology*, Vol. 4, No. 6, pp.1562–1565.

Chakraborty, S. and Pal, S. (2016), 'Photoplethysmogram signal based biometric recognition using linear discriminant classifier', in *Proceedings of 2nd International Conference on Control, Instrumentation, Energy & Communication* (*CIEC*), Kolkata, India, pp.183–187.

Chakraborty, S., Mitra, M. and Pal, S. (2017) 'Biometric analysis using fused feature set from side face texture and electrocardiogram', *IET Science, Measurement & Technology*, Vol. 11, No. 2, pp.226–233.

Chakraborty, S., Mitra.M. and Pal, S. (2020) 'Geometric retrieval algorithm-based ear biometry with occluded images', International Journal of Biometrics, Vol. 12, No. 3, pp.247–268.

Dhamija, J., Choudhary, T., Kumar, P et al. (2017) 'An advancement towards efficient face recognition using live video feed', Proceedings of International Conference on Computational Intelligence and Networks (CINE), Odisha, India, pp.53–56.

Ekman, P. and Keltner, D. (1997) 'Universal facial expressions of emotion: an old controversy and new findings', in Segerstråle, U.C. and Molnár, P. (Eds.): *Nonverbal communication: Where Nature Meets Culture*, pp.27–46, Lawrence Erlbaum Associates, Mahwah, NJ.

Fatihah, N.N., Ariyanto, G., Latipah, A.J. and Pangestuty, D.M. (2018) 'Face recognition using local binary pattern and nearest neighbour classification', *International Symposium on Advanced Intelligent Informatics* (*SAIN*), Yogyakarta, Indonesia, pp.142–147.

Faundez-Zanuy, M. and Fabregas, J. (2008) 'On the relevance of facial expressions for biometric recognition', in Esposito, A., Bourbakis, N.G., Avouris, N. et al. (Eds.): *Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction, Lecture Notes in Computer Science*, Vol. 5042, pp.32–43, Springer, Berlin, Heidelberg.

Garcia, C. and Tziritas, G. (1999) 'Face detection using quantized skin colour regions merging and wavelet packet analysis', *IEEE Transactions on Multimedia*, Vol. 1, No. 3, pp.264–277.

Jafri, R. and Arabnia, H.R. (2009) 'A survey of face recognition techniques', *Journal of Information Processing Systems*, Vol. 5, No. 2, pp.41–68.

Kakkar, P. and Sharma, V. (2018) 'Criminal identification system using face detection and recognition', *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 7, No. 3, pp.238–243.

Khan, A., Rehman, S., Waleed, M. et al. (2018) 'Forensic video analysis: passive tracking system for automated person of interest (POI) localization', *IEEE Access*, Vol. 6, No. 1, pp.43392–43403.

Kong, A., Zhang, D. and Lu, G. (2006) 'A study of identical twins palm print for personal verification', *Pattern Recognition*, Vol. 39, No. 11, pp.2149–2156.

Li, L.S. and Yahya, N. (2014) 'Face recognition technique using Gabor Wavelets and singular value decomposition (SVD)', in *IEEE International Conference on Control System, Computing and Engineering*, Penang, Malaysia, pp.455–459.

Lucas, B.D. and Kanade, T. (1981) 'An iterative image registration technique with an application to stereo vision', in *Proceedings of 7thInternational Joint Conference on Artificial Intelligence* (*IJCAI*), Canada, August 1981, pp.674–679.

Lyons, M., Akamatsu, S., Kamachi, M. et al. (1998) 'Coding facial expressions with Gabor Wavelets', in *Proceedings of 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, pp.200–205.

Ma, Y., Huang, Z., Wang, X. and Huang, K. (2020) 'An overview of multimodal biometrics using the face and ear', *Mathematical Problems in Engineering*, Article ID. 6802905, Vol. 2020, No. 1, pp.1–17.

Maheswari, V., Sari, C.A., Setiadi, D.R.I.M. and Rachmawanto, E.H. (2020) 'Study analysis of human face recognition using principal component analysis', *International Seminar on Application for Technology of Information and Communication* (*iSemantic*), Semarang, Indonesia, pp.55–60.

Maltoni, D., Maio, D., Jain, A.K. et al. (2003) *Handbook of Fingerprint Recognition*, 2nd ed., Springer-Verlag, London.

Matsumoto, D. and Kupperbusch, C. (2001) 'Idiocentric and allocentric differences in emotional expression, experience, and the coherence between expression and experience', *Asian Journal of Social Psychology*, Vol. 4, No. 2, pp.113–131.

Maw, H.M., Thu, S.M. and Mon, M.T. (2020) 'Vision based facial expression recognition using Eigenfaces and multi-SVM classifier', in Hernes, M., Wojtkiewicz, K. and Szczerbicki, E. (Eds.): *Advances in Computational Collective Intelligence*, *ICCCI 2020*, *Communications in Computer and Information Science*, Springer, Cham, Vol. 1287, pp.662–673.

Mei Yin, D.B., Mukhlas, A., Wan-Chik, R.Z. et al. (2018) 'A proposed approach for biometric-based authentication using of face and facial expression recognition', in *Proceedings of 2018 IEEE 3rd International Conference on Communication and Information Systems* (*ICCIS*), Singapore, pp.28–33.

Ojala, T., Pietikaien, M. and Harwood, D. (1996) 'A comparative study of texture measures with classification based on feature distribution', *Pattern Recognition*, Vol. 29, No. 1, pp.51–59.

Owusu, E., Zhan, Y. and Mao, Q.R. (2014) 'An SVM-AdaBoost facial expression recognition system', *Applied Intelligence*, Vol. 40, No. 3, pp.536–545.

Rahulamathavan, Y. and Rajarajan, M. (2017) 'Efficient privacy-preserving facial expression classification', in *IEEE Transactions on Dependable and Secure Computing*, Vol. 14, No. 3, pp.326–338.

Ramesha, K. and Raja, K.B. (2011) 'Dual transform based feature extraction for face recognition', *International Journal of Computer Science*, Vol. 8, No. 2, pp.115–121.

Sahan, A.M. and Al-Itbi, A.S. (2021) 'The fusion of local and global descriptors in face recognition application', in Hura, G., Singh, A. and Siong Hoe, L. (Eds.): *Advances in Communication and Computational Technology*, *Lecture Notes in Electrical Engineering*, Vol. 668, pp.1397–1408, Springer, Singapore.

Shih, F.Y., Chuang, C-F. and Wang, P.S.P. (2008) 'Performance comparisons of facial expression recognition in Jaffe database', *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 22, No. 3, pp.445–459.

Shinwari, A.R., Balooch, A.J., Alariki, A.A. et al. (2019) 'A comparative study of face recognition algorithms under facial expression and illumination', in *Proceedings of 21st International Conference on Advanced Communication Technology* (*ICACT*), Pyeong Chang Kwangwoon_Do, Korea (South), pp.390–394.

Tomasi, C. and Kanade, T. (1991) *Detection and Tracking of Point Features*, Carnegie Mellon University Technical Report, CMU-CS-91-132.

Tulyakov, S., Slowe, T., Zhang, Z. et al. (2007) 'Facial expression biometrics using tracker displacement features', in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA 2007, pp.1–5.

Turk, M. and Pentland, A. (1991) 'Eigenfaces for recognition', *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp.71–86.

Vatsa, M., Singh, R. and Noore, A. (2008) 'Improving iris recognition performance using segmentation, quality enhancement, match score fusion, and indexing', *IEEE Transactions on Systems, Man, and Cybernetics*, *Part B* (*Cybernetics*), Vol. 38, No. 4, pp.1021–1035.

Viola, P. and Jones, M.J. (2004) 'Robust real-time object detection', *International Journal of Computer Vision*, Vol. 57, No. 2, pp.137–154.

Wang, J-W., Le, N.T., Lee, J-S. and Wang, C-C. (2018) 'Illumination compensation for face recognition using adaptive singular value decomposition in the wavelet domain', *Information Sciences*, Vol. 435, No. 1, pp.69–93.

Wang, L., Zhang, Y. and Feng, J. (2005) 'On the Euclidean distance of images', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 8, pp.1334–1339.

Zhang, G., Huang, X., Li, S.Z., Wang, Y. and Wu, X. (2004) 'Boosting local binary pattern (LBP)-based face recognition', in Li, S.Z., Lai, J., Tan, T., Feng, G. and Wang, Y. (Eds.): *Advances in Biometric Person Authentication. SINOBIOMETRICS 2004. Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, Vol. 3338, pp.179–186.