CrossMark

Recognition of identical twins using fusion of various facial feature extractors

Ayman Afaneh[1], Fatemeh Noroozi[2] and Önsen Toygar[1*] (iD)

Abstract

Distinguishing identical twins using their face images is a challenge in biometrics. The goal of this study is to construct a biometric system that is able to give the correct matching decision for the recognition of identical twins. We propose a method that uses feature-level fusion, score-level fusion, and decision-level fusion with principal component analysis, histogram of oriented gradients, and local binary patterns feature extractors. In the experiments, face images of identical twins from ND-TWINS-2009-2010 database were used. The results show that the proposed method is better than the state-of-the-art methods for distinguishing identical twins. Variations in illumination, expression, gender, and age of identical twins' faces were also considered in this study. The experimental results of all variation cases demonstrated that the most effective method to distinguish identical twins is the proposed method compared to the other approaches implemented in this study. The lowest equal error rates of identical twins recognition that are achieved using the proposed method are 2.07% for natural expression, 0.0% for smiling expression, and 2.2% for controlled illumination compared to 4.5, 4.2, and 4.7% equal error rates of the best state-of-the-art algorithm under the same conditions. Additionally, the proposed method is compared with the other methods for non-twins using the same database and standard FERET subsets. The results achieved by the proposed method for non-twins identification are also better than all the other methods under expression, illumination, and aging variations.

Keywords: Identical twins, Face recognition, Score fusion, Feature fusion, Decision fusion

1 Introduction

Biometrics has recently been widely used for human recognition in many different countries to identify a person under controlled or uncontrolled environments. The traditional methods for person identification such as passwords and magnetic cards have many disadvantages compared with a biometric-based method that depends on who the person is intrinsically, not what he knows or what he possesses extrinsically [1]. Biometric systems recognize the individuals based on their physical traits or behavioral characteristics; therefore, many factors must be considered when choosing any biometric trait [2, 3] to be used in a person recognition system. Universality is one of the most important factors which means that every person should have that characteristic. Uniqueness is another factor which indicates that no two person should be the same in terms of that characteristic. On the other hand,

permanence is another factor which means that the characteristic should be invariant with time. Acceptability is a factor that should be considered which indicates to what extent people are willing to accept the biometric system [1, 4].

Absence of the factors, such as universality, uniqueness, permanence, and acceptability, leads to a weak recognition system with high error rates. Therefore, all the factors must be available at the same time in order to get a good distinguishing system. In all the cases, the face trait meets the aforementioned factors perfectly which makes it a good choice as a biometric trait. However, there is a case of face recognition that represents the main challenges with one of those factors which is identical (monozygotic) twins case [5]. In identical twins case, universality, permanence, and acceptability are satisfied, but the factor that represents a serious problem is the uniqueness. It is axiomatic that the identical twins have the same face shape, size, and features, so new methods and algorithms should be studied and considered in order to deal with the high similarities in case of identical twins. It is obvious

*Correspondence: onsen.toygar@emu.edu.tr
[1]Computer Engineering Department, Faculty of Engineering, Eastern Mediterranean University, Famagusta, North Cyprus, via Mersin 10, Turkey
Full list of author information is available at the end of the article

Afaneh *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:81

Page 2 of 14

that face recognition for a population without identical twins will be more efficient and easier when constructing a system of identical twins recognition. In other words, algorithms that are able to distinguish the critical challenges such as identical twins should be more powerful in the case of non-twins recognition which is the main goal in this study.

In order to distinguish identical twins, we propose a biometric system which is mainly based on three different types of fusion, namely feature-level fusion, score-level fusion, and decision-level fusion. Additionally, principal component analysis (PCA) [6] , histograms of oriented gradients (HOG) [7], and local binary patterns (LBP) [8] are employed as feature extraction algorithms. The outputs of feature-level fusion, score-level fusion, and decision-level fusion are consolidated to form the proposed method in which the details are further explained.

This paper is organized as follows. Section 2 discusses the related studies in recognition of twins by using different biometric traits. Section 3 explains the feature extraction methods (PCA, HOG, and LBP) that were used in this study as feature extractors. Sections 4 and 5 discuss the fusion levels and the proposed technique. Section 6 presents the experimental setup, the datasets used in the experiments, and the results. Additionally, the results of recognition under different conditions of face images capturing illumination, expression, gender, and age are presented. Lastly, Section 7 presents the concluding remarks.

2 Related work

Identical twins were used in some studies in the literature especially by analyzing their faces, fingerprints, irises, palm prints, and speech. Jain et al. in 2002 [9] used the minutia-based automatic fingerprint matching and successfully distinguished the fingerprint images of identical twins. However, for non-twins matching, the accuracy was higher than the case of identical twins. In other words, the similarity between the fingerprints of identical twins was much higher than the case of non-twins. As a result, the false accept rate (FAR) of identical twins was about four times higher than that of non-twins [9].

Adapted Gaussian mixture models (GMMs) were implemented to investigate the performance of speaker verification technology for distinguishing identical twins in 2005 [10]. The tests were applied using long and short duration of speaking by GMM-UBM scoring procedure as baseline scores in the experiments [10]. Acquired scores were subjected to unconstrained cohort normalization (UCN) and labeled as UCN scores. Using UCN, EER decreased from 10.4 to 1% (short) and from 5.2 to around 0% (long) [10]. Competitive code algorithm was developed in 2006 in order to distinguish individuals who have the same genetic information such as identical twins using palm

prints as a biometric trait [11]. The authors proved that using the three principle lines of palm print is not enough to distinguish identical twins since it is genetically related. Genetically unrelated features in palm print were also used in that study, and the genuine accept rate was found to be about 97%.

Hollingsworth et al. in 2010 [12] proposed to evaluate the human ability to determine the degree of similarity between iris images and whether they belong to identical twins or not. Using 3 s to display each image, 81% accuracy was acquired using only the appearance of iris and 76% accuracy using only the appearance of periocular. Increasing the time of displaying each image of iris and periocular improved the accuracy to 92 and 93%, respectively. Demographic information such as gender and ethnicity and/or some facial marks were included to face matching algorithms in 2010 [13] with a view to enhance the accuracy of the system. When comparisons between the matching results of rank one matching accuracy of the state-of-the-art commercial face matcher (face VACS) with the proposed facial marks matcher were performed, the accuracy increased from 90.61% (face VACS) to 92.02% (proposed facial marks matcher).

Recognition experiments on identical twins in 2010 [14] showed that the multimodal biometric systems which combine different instances of the the same biometric traits lead to perfect matching compared with the unimodal systems. Using a commercial minutiae-based matcher such as VeriFinger and the iris feature representation method based on ordinal measure, the EERs of finger fusion and the fusion of right and left irises were both 0.49%. On the other hand, discriminating facial traits were determined by observation of humans in 2011 [15]. In that study, 23 people participated in the recognition experiments in which the maximum, minimum, and average success rates were 90.56, 60.56, and 78.82%, respectively. Additionally, they performed automated system matching with uncontrolled face images and obtained low success rates.

Three different commercial face matchers in addition to local region principle component analysis (LR-PCA) were used in 2011 [16] for distinguishing identical twins. Experiments were run under several conditions such as expression, light control, and presence of glasses. The best performance with a minimum EER (from 0.01 to 0.12%) was acquired by Cognitec matcher under ideal conditions. On the other hand, the accuracy of identical twins' matching was increased by cascading of appearance-based verifier and motion-based verifier in 2012 [17] compared with the results of using both of them separately. Six face expressions were examined using motion-based matchers, Simple Spare Displacement Algorithm (SDA) and Dense Displacement Algorithm (DDA). The best performance was acquired by motion-based matcher which was

Afaneh *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:81

Page 3 of 14

increased from 93.3 to 96% after applying a cascading approach.

Paone et al. in 2014 performed some experiments that were implemented with different conditions on face images of identical twins [18]. The primary goal of these experiments is to measure the ability of some algorithms to distinguish two different faces that have a large similarity such as identical twins (monozygotic). Three of the top submissions to Multiple Biometric Evaluation (MBE) 2010 face track algorithms [19] were used in addition to four commercially available algorithms. Measuring the performance of all algorithms and comparing the results in order to determine the best algorithm with the lowest error rate were done. The experiments were only applied on frontal faces without wearing glasses, and all EER results were demonstrated in that study. Consequently, these results are used in our experiments for comparison purposes in Section 6.

3 Feature extraction methods

In this study, two different categories of feature extraction techniques are used, namely appearance-based and texture-based techniques. Appearance-based techniques are based on mapping the high-dimensional face image into a lower dimensional sub-space in order to generate a compact representation of the entire face region in the acquired image. This sub-space is defined by a set of representative basis vectors, which are learned using a training set of images. The most commonly used appearance-based technique for facial feature extraction is PCA [1].

The appearance-based technique which is implemented in this work to extract features is PCA, which is the earliest automated method proposed for face recognition. PCA uses the training data to learn a subspace that accounts for as much variability in the training data as possible. This is achieved by performing an eigenvalue decomposition of the covariance matrix of the data [1].

The goal of PCA is to obtain eigenvectors of the covariance matrix (C) as $C\mathrm{w} = \lambda\mathrm{w}$ where

$$C = XX^T = \frac{1}{N}\sum_i\sum_j(\overline{x}_{i_j} - \overline{m})(x_{i_j} - \overline{m})^T, \qquad (1)$$

$$X = [X_1 - m, \; X_2 - m, \; \ldots, \; X_N - m] \qquad (2)$$

with X_i representing the image vector of the ith image and

$$m = \frac{1}{N}\sum_{i=1}^{N} X_i.$$

where m is the average of the training set and N is the number of training samples.

On the other hand, texture-based approaches try to find robust local features that are invariant to pose or lighting variations. LBP with 5×5 segments and HOG are implemented as texture-based approaches in this study, and these methods are also used in many recognition/classification problems [20–23].

LBP face analysis algorithm has been one of the most commonly used applications in recent years. Facial image analysis is an active research topic in computer vision with a wide range of important applications, e.g., human-computer interaction, biometric identification, surveillance, and security [24]. The original LBP operator labels the pixels of an image with dec imal numbers, called LBP codes, which encode the local structure around each pixel [8, 25].

LBP divides the image into several nonoverlapped blocks with equal size. In order to extract the local features, LBP texture descriptors are performed on each block separately. Then, for each block, a histogram is extracted to hold information related to the patterns on a set of pixels. Finally, the extracted features of each block will be directly concatenated to produce a single global feature vector. LBP is checking a local neighborhood surrounding a central point R which is sampled at P points and tests whether the surrounding points are greater than or less than the central point to classify textures. The LBP value of the center pixel in the P neighborhood on a circle of radius R is calculated by

$$\mathrm{LBP}_{(P,R)} = \sum_{p=0}^{P-1} S(g_p - g_c)2^p,$$

$$S(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \qquad (3)$$

where g_p and g_c are the gray value of the surrounding points and the center pixel, respectively.

HOG descriptors [26] used in computer vision and image processing for the purpose of object detection count occurrences of gradient orientation in localized portions of an image. Calculation of the classic HOG descriptor begins by dividing an image under the detection window into a dense grid of rectangular cells. For each cell, a separate orientation of gradients is calculated. The gradient magnitude $|G|$ and the orientation of the gradient θ for an image $I_{x,y}$ are calculated as follows:

$$|G| = \sqrt{I_X^2 + I_Y^2}, \quad \text{where}$$

$$I_X = I * D_X, \quad I_Y = I * D_Y,$$

$$D_X = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}, \quad D_Y = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix},$$

$$\qquad (4)$$

Afaneh *et al. EURASIP Journal on Image and Video Processing*   (2017) 2017:81

Page 4 of 14

where $*$ is the convolution operator and $\theta = \text{atan2}(I_Y, I_X)$ radians that returns a value in the interval $(-\pi, \pi]$.

The angle transformed into degrees is $\alpha = \theta * 180/\pi$ that gives values in the range $(-180, 180]$ degrees. For the "signed" gradient, it is needed to translate the range of the gradient from $(-180, 180]$ to $[0, 360)$ degrees. This is performed as follows:

$$\alpha_{\text{signed}} = \begin{cases} \alpha, & \text{if } \alpha \geq 0 \\ \alpha + 360, & \text{if } \alpha < 0 \end{cases}$$

The histogram consists of evenly spaced orientation bins accumulating the weighted votes of gradient magnitude of each pixel belonging to the cell. Additionally, the cells are grouped into blocks, and for each block, all cell histograms are normalized. The blocks are overlapping, so the same cell can be differently normalized in several blocks. The descriptor is calculated using all overlapping blocks from the image detection window.

4   Fusion of facial data in different levels

Biometric fusion can be implemented in two different modes, either prior to matching process or after matching process. In this study, fusion techniques from each biometric fusion mode were used such as feature-level, score-level, and decision-level fusion techniques. Feature-level fusion represents biometric fusion prior to matching. However, score-level and decision-level fusion are methods of biometric fusion techniques that are implemented after a matching process. There are many biometric systems employing fusion of different levels [21, 27–29].

4.1   Feature-level fusion

Consolidating two or more different biometric feature sets of the same user in order to form them as one feature set is a definition of feature- or representation-level fusion. Feature-level fusion can be classified into two different classes such as homogenous and heterogeneous feature fusion. A homogeneous feature fusion scheme is used when the feature sets to be combined are obtained by applying the same feature extraction algorithm to multiple samples of the same biometric trait (e.g., minutia sets from two impressions of the same finger). This approach is applicable to multi-sample and multi-sensor systems. Heterogeneous feature fusion techniques are required if the component feature sets originate from different feature extraction algorithms or from samples of different biometric traits (or different instances of the same trait).

A heterogeneous feature fusion technique is used in this paper by combining different feature sets which are extracted by PCA, HOG, and LBP methods. The first case is done by consolidating the extracted feature sets of PCA, HOG, and LBP, as one feature set, while the fusion in the second case is implemented by using only the feature sets which are extracted by HOG and LBP.

4.2   Score-level fusion

When a final recognition decision can be acquired by combining two or more match scores of different biometric matchers, fusion is said to be done at the score-level. After capturing the raw data from sensors and extracting feature vectors, the next level of fusion is based on match scores. It is relatively easy to access and combine the scores generated by different biometric matchers; as a result, score-level fusion is the most commonly used methods in multibiometric systems. There are many types of score-level fusion such as likelihood-ratio-based fusion and transformation-based fusion. In this paper, transformation-based fusion (sum rule) was used.

4.3   Decision-level fusion

In a multibiometric system, fusion is carried out at decision level when only the decision outputs by the individual biometric matchers are available. The decision-level fusion rules such as "AND" and "OR" rules, majority voting, weighted majority voting, Bayesian decision fusion, the Dempster-Shafer theory of evidence, and behavior knowledge space are used to integrate the multiple decisions to produce the final decision. In this study, we used a hybrid decision-level fusion strategy which is explained in the next section.

5   Proposed method

A novel method for the recognition of identical twins is proposed and implemented in this study. The proposed method is based on the output of feature fusion and score fusion of HOG and LBP methods beside the output of the decision fusion of LBP, HOG, and PCA approaches as shown in Figs. 1 and 2. The proposed method works under verification mode; therefore, the user must claim his/her identity in order to check if he/she is genuine or impostor. On the other hand, if the user is recognized as impostor in any partial decision, the recognized ID will be used where the system checks not only the template of the claimed ID but also all the stored templates of all the users that are stored in the database. In the case that the user is not recognized and it is not included in the database, the partial decision becomes "unrecognized."

The main steps of the proposed method are presented below:

1- Apply feature-level and score-level fusion using HOG and LBP in addition to decision-level fusion using PCA, HOG, and LBP.
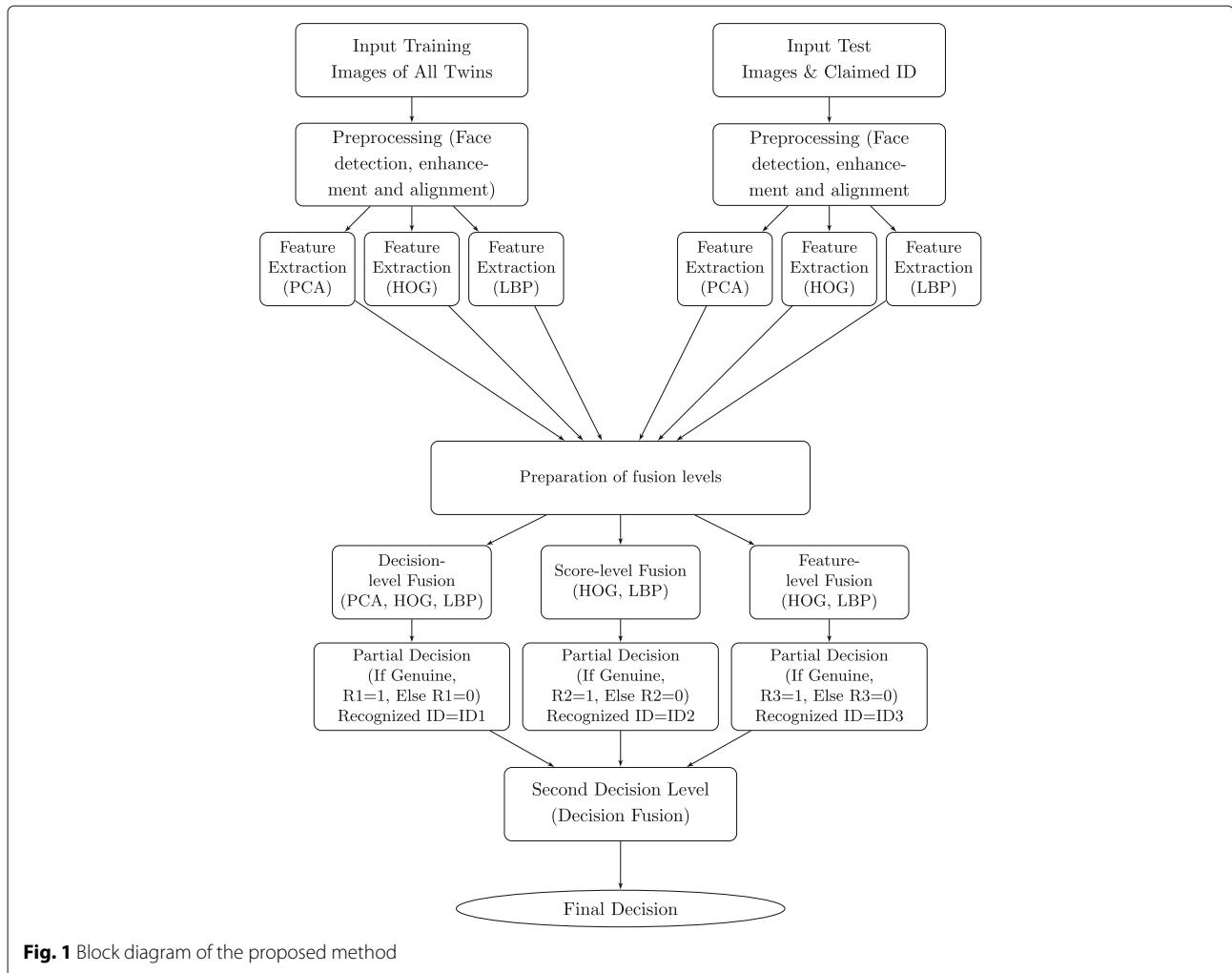
Fig. 1 Block diagram of the proposed method

2- Partial decisions from each level of fusion will be acquired as follows: If (**Partial Decision** = Genuine), then Ri = 1, else (**Partial Decision** = Impostor) Ri = 0.

3- In both decision cases, either genuine or impostor, the partial decision will present the recognized ID of the individual.

4- If two or more of the fusion levels recognize the input image as genuine based on the claimed ID, the whole system will recognize the user as genuine.

5- In the case of only one fusion level recognizes the input image as genuine, the system will check the recognized IDs of other algorithms. If they are not the same, the whole system will recognize the user as genuine; otherwise, the system will recognize the user as impostor. Table 1 clarifies this step.

Figure 1 shows the general block diagram of the proposed method while the details related to the second decision level are presented in Fig. 2.

6 Experiments and results

In order to demonstrate the validity of the proposed method in distinguishing identical twins, several experiments have been conducted on ND-TWINS-2009-2010 Dataset [5, 30]. The following subsections present the details about the dataset used, the experimental setup, and the results of different types of experiments such as expression-based, illumination-based, gender-based, and age-based experiments. Additionally, experiments related to the recognition of non-twins are also presented in the following subsections using ND-TWINS-2009-2010 and FERET Dataset [31, 32].

6.1 ND-TWINS-2009-2010 Dataset

ND-TWINS-2009-2010 Dataset contains 24,050 color photographs of the faces of 435 attendees of the Twins Days Festivals in Twinsburg, OH, performed in 2009 and 2010. All images were captured by Nikon D90 SLR cameras. Images were captured under natural light in "indoor" and "outdoor" configurations ("indoor" was a tent). Facial
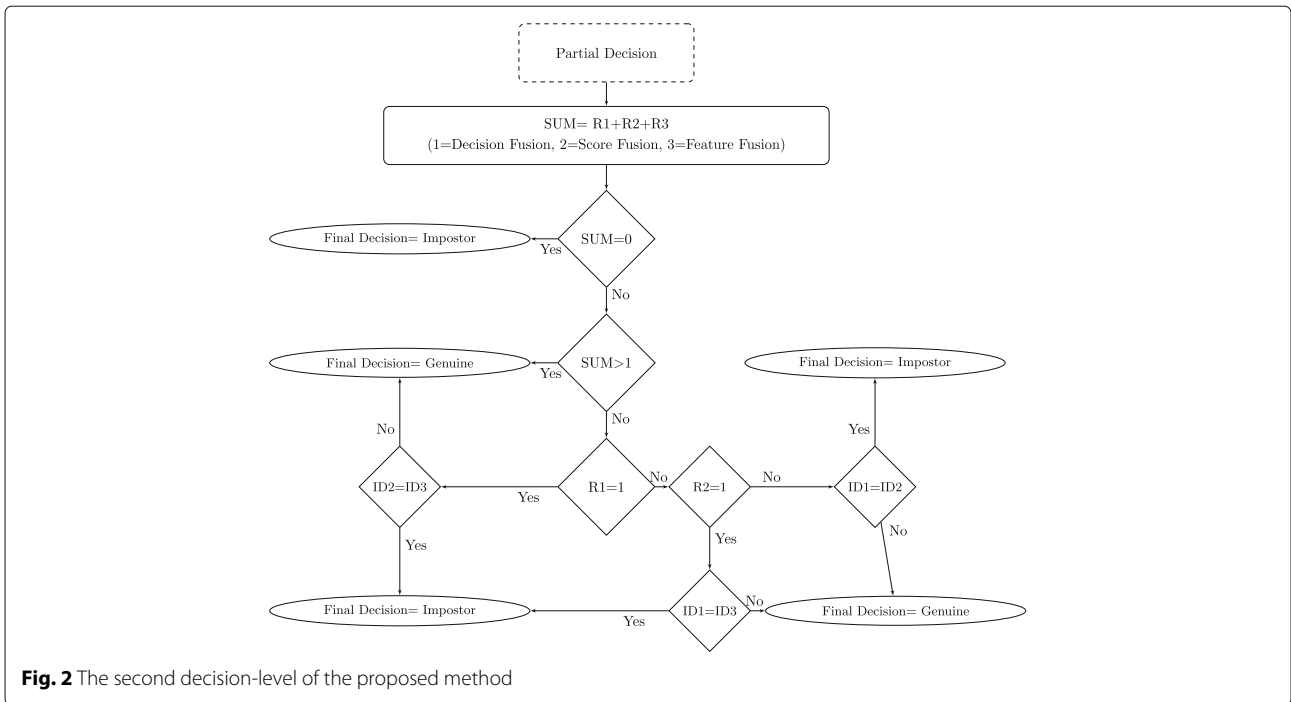
Afaneh *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:81

Page 6 of 14

Fig. 2 The second decision-level of the proposed method

capturing angle varied from -90 to $+90°$ in steps of $45°$ ($0°$ was frontal). Additionally, images were captured under natural and smiling expression. Example images can be seen in Fig. 3 for two different people (identical twins) where each image shows two different samples of the same person. Figure 4 also demonstrates two different images for twins of more than 40-year-old women.

6.2 Standard FERET Dataset

The standard FERET Dataset is a subset of FERET database that contains 1196 gallery images for training and four different subsets of FERET database images under various challenges. The training images that are in category "fa" (1196 images) are used as gallery images for four probe sets namely "fb," "fc," "duplicate I," and "duplicate II." The subset fb includes 1195 images with variations in expressions. The subset fc includes 194 images with illumination variations. On the other hand, images with

aging variations are in duplicate I and duplicate II subsets. Duplicate I subset consists of 722 facial images which are recorded at different times compared to fa subset images. Duplicate II is a subset of duplicate I (234 images) which includes images taken at least 18 months later after the gallery image was taken. Duplicate I and duplicate II subsets are useful for aging experiments using face recognition methods. The standard FERET subsets are used in this study to compare various face recognition algorithms and the proposed method under different challenges for non-twins.

6.3 Experimental setup

A set of experiments is conducted for identical twins based on their face images by using 352 users (176 identical twins) and 1512 image samples from ND-TWINS-2009-2010 Dataset. Four algorithms, namely convolutional neural networks (CNN) [33–35], PCA,

Table 1 Combination possibilities of partial decisions

First partial decision	Second partial decision	Third partial decision	Final decision
Genuine	Genuine	Genuine	Genuine
Impostor (ID:A)	Genuine	Genuine	Genuine
Genuine	Not recognized	Genuine	Genuine
Genuine	Not recognized	Impostor (ID:A)	Genuine
Impostor (ID:A)	Genuine	Impostor (ID:B)	Genuine
Not recognized	Not recognized	Genuine	Impostor
Impostor (ID:B)	Genuine	Impostor (ID:B)	Impostor

Afaneh *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:81

Page 7 of 14

Fig. 3 Example frontal faces. Images in **a** are of the first twin under different illumination and expression while images in **b** are of the second twin under different illumination and expression

HOG, and LBP are implemented for comparison purposes. Additionally, three fusion methods namely feature-level, score-level, and decision-level fusion and the proposed method are implemented in order to find the most reliable system that is able to correctly match identical twins by face recognition. The effect of the four conditions (illumination, expression, gender, and age) is also examined. All the selected images in the experiments were frontal face images without glasses. Manhattan distance measure is used to measure the similarity between test and train images.

The unimodal biometric systems that are implemented in this study use PCA, HOG, LBP, and CNN. For PCA, we use the maximum number of non-zero eigenvectors. HOG algorithm uses 64×128 image size and divides the facial image into 16×16 blocks with 50% overlapping. The images are also processed using LBP by dividing it to 5×5 partitions (segments). Finally, we trained a CNN to perform recognition based on image samples. The output of the CNN is recognition rate. Every value corresponds to one of the conditions such as age, gender, illumination, and expression. In order to train and test by using more than 1000 images, we chose a medium size CNN, namely, GoogLeNet [36]. The model follows the concept of "a

network in the network," which is based on the repetition of the inception module. In the GoogLeNet, the module is repeated nine times. The first level includes 1×1 convolutions and a 3×3 max pooling. The second level contains 1×1, 3×3, and 5×5 convolutions. The third level consists of an inception module with filter concatenations, which merges the results that have been obtained in the previous steps. From early to top inception modules, the number of filters varies from 256 to 1024. In order to add the capability of back-propagating the gradients, auxiliary classifiers are connected to the intermediary layers. They are fed by the outputs of the inception modules. While training, their losses are multiplied by 0.3 and added to the overall loss, but they do not count in making the inference. Initially, the images are resized to 256×256 pixels. Next, during training, a crop of 224×244 pixels is randomly taken from every image.

The performance of the proposed method is also measured in the case of non-twins using ND-TWINS-2009-2010 Dataset. These set of experiments are conducted by dividing 176 identical twins into two equal groups. The first group contains the first brother/sister of each twin, while the second group contains the second brother/sister of each twin. In that case, each group contains 88 of

Fig. 4 Example frontal faces. Images in **a** are of the first twin under different expression and controlled illumination while images in **b** are of the second twin under different expression and uncontrolled illumination

Afaneh *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:81

Page 8 of 14

Table 2 EER results of expression-based experiments for identical twins

	Expression (training-test)	EER results (%) of		
		Natural-natural	Natural-smiling	Smiling-smiling
Implemented algorithms	CNN	13.41	24.44	0
	PCA	12.65	24.54	20.83
	HOG	5.60	11.11	4.17
	LBP	2.76	8.80	0
	Score fusion (PCA, HOG, LBP)	3.76	11.11	3.13
	Score fusion (HOG, LBP)	2.30	4.63	2.09
	Feature fusion (PCA, HOG, LBP)	3.53	6.02	4.17
	Feature fusion (HOG, LBP)	2.38	4.17	0
	Decision fusion (PCA, HOG, LBP)	3.37	6.94	0
	Proposed method	*2.07*	*3.24*	0
Algorithms implemented in [18]	A	4.50	7.00	4.20
	B	39.40	39.20	40.00
	C	6.70	37.60	7.40
	D	22.20	22.90	19.90
	E	14.40	13.50	13.50
	F	9.40	10.80	9.30
	G	7.70	8.80	6.80

The italicized data reflect the highest equal error rates of each experiments

users who are not twins. By implementing the same type of experiments on these two groups separately, the face recognition performance on non-twins is measured. Using the same database, same users, and same samples in the recognition experiments on twins and non-twins, the comparison is more realistic than using different database, since the capturing conditions of images such as illumination, expression, distance to camera, etc. are the same.

On the other hand, standard FERET subsets are also used to evaluate the proposed method in the absence of identical twins. In this study, five different subsets of FERET Database are used namely "fa," "fb," "fc," "duplicate I," and "duplicate II" subsets. The first subset which is named as fa contains frontal face images with ideal conditions (natural expression and controlled lighting), and it is used for training (gallery) purposes. On the other hand, fb subset includes frontal face images with alternative face expression. In fc subset, the included frontal face images were captured under uncontrolled illumination. Duplicate I subset contains probe frontal face images that

Table 3 EER results of expression-based experiments for non-twins

	Expression (training-test)	EER results (%) of		
		Natural-natural	Natural-smiling	Smiling-smiling
Algorithms	CNN	5	3.6	6
	PCA	4.4	6.4	3.6
	HOG	0.8	3.6	0.0
	LBP	0.0	2.8	0.0
	Score fusion (PCA, HOG, LBP)	0.0	2.8	1.2
	Score fusion (HOG, LBP)	0.0	1.2	0.0
	Feature fusion (PCA, HOG, LBP)	0.0	1.6	0.0
	Feature fusion (HOG, LBP)	0.0	0.4	0.0
	Decision fusion (PCA, HOG, LBP)	0.0	2.0	0.0
	Proposed method	0.0	*0.0*	0.0

The italicized data reflect the highest equal error rates of each experiments

Afaneh *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:81
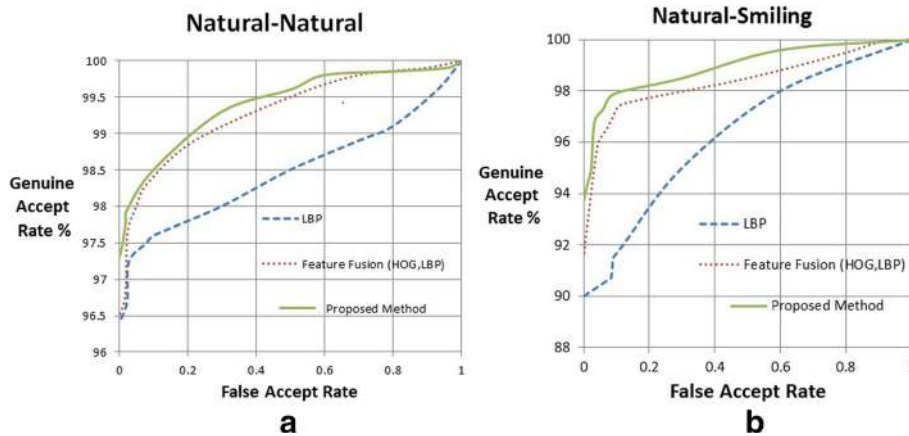
Page 9 of 14

Fig. 5 ROC curves for **a** natural-natural expression/**b** natural-smiling expression

were obtained anywhere between 1 min and 1031 days after their respective gallery matches. Additionally, duplicate II subset includes probe frontal face images that are a strict subset of the duplicate I images, and they are those taken only at least 18 months after their gallery entries. fb, fc, duplicate I, and duplicate II subsets are used for testing operations.

The performance of all algorithms is measured and reported by equal error rate (EER). EER is defined as the point that false reject rate (FRR) and false accept rate (FAR) have the same value. EER is also used to compare the efficiency of the implemented methods under different conditions.

6.4 Experiments on ND-TWINS-2009-2010 Dataset

We conducted four sets of experiments using ND-TWINS-2009-2010 Dataset. These are expression-based, illumination-based, gender-based, and age-based experiments. The following subsections present the details of these experiments for the recognition of identical twins and non-twins separately.

6.4.1 Expression-based experiments

The first set of experiments aim to measure the efficiency of face recognition for identical twins and non-twins under the condition of expression variation. In these experiments, both smiling and natural expressions of the face image that were captured under controlled lighting were used. Tables 2 and 3 show the EER of natural-natural (N-N), natural-smiling (N-S), and smiling-smiling (S-S) as training-test combination for identical twins and non-twins, respectively.

Figures 5 and 6 demonstrate the ROC curves for natural-natural expression and natural-smiling expression, controlled-controlled illumination, and controlled-uncontrolled illumination, respectively.
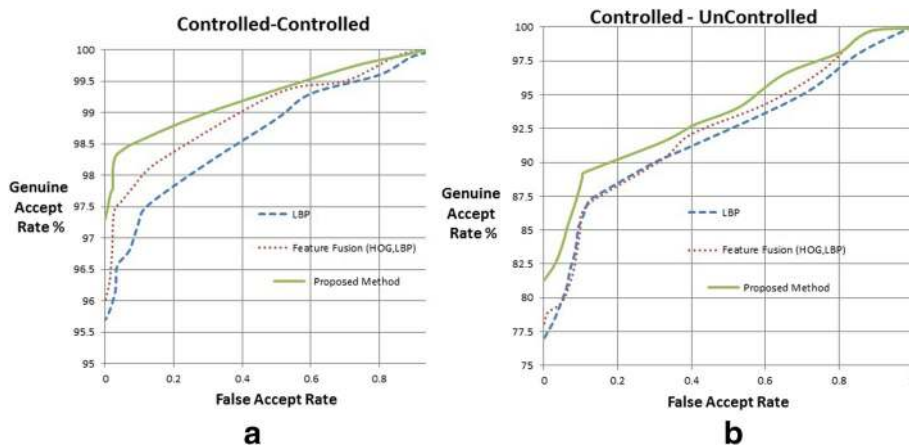

Fig. 6 ROC curves for **a** controlled-controlled illumination/**b** controlled-uncontrolled illumination

Afaneh *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:81

Page 10 of 14

Table 4 EER results of illumination-based experiments for identical twins (*Cont* controlled condition, *Uncont* uncontrolled condition)

		EER results (%) of		
	Illumination (Training-Test)	Cont-Cont	Cont-Uncont	Uncont-Uncont
Implemented algorithms	CNN	25.00	18.84	20.09
	PCA	12.01	31.15	14.04
	HOG	6.09	25.77	9.23
	LBP	3.45	12.69	4.08
	Score fusion (PCA, HOG, LBP)	4.40	20.19	9.04
	Score fusion (HOG, LBP)	2.58	13.65	5.77
	Feature fusion (PCA, HOG, LBP)	3.89	21.92	8.27
	Feature fusion (HOG, LBP)	2.58	12.88	*4.04*
	Decision fusion (PCA, HOG, LBP)	3.58	12.50	4.81
	Proposed Method	*2.20*	*10.77*	*4.04*
Implemented algorithms in [18]	A	4.70	5.90	11.50
	B	35.90	40.70	41.40
	C	9.00	34.10	32.30
	D	14.50	20.90	26.50
	E	10.20	13.80	24.00
	F	7.30	12.40	19.40
	G	8.00	7.80	16.20

The italicized data reflect the highest equal error rates of each experiments

6.4.2 Illumination-based experiments

Various face images that were captured under the same and different lighting conditions are used in the second set of experiments. For these experiments, there are two possibilities: controlled illumination (image acquired under the tent) and uncontrolled illumination (images acquired outdoor in rainy or sunny weather). Using face images that were captured under controlled and uncontrolled illumination, the tests were conducted in three different cases, namely controlled-controlled (C-C), controlled-uncontrolled (C-U), and uncontrolled-uncontrolled (U-U) as training-test combinations. Tables 4 and 5 show the EER results of these experiments performed under illumination conditions for identical twins and non-twins, respectively.

6.4.3 Gender-based experiments

In the next set of experiments, we separated the subjects used in the previous experiments (expression-based and illumination-based) to male and female face images. The

Table 5 EER results of illumination-based experiments for non-twins (*Cont* controlled condition, *Uncont* uncontrolled condition)

Algorithms	Illumination (training-test)	EER results (%) of		
		Cont-Cont	Cont-Uncont	Uncont-Uncont
	CNN	9	24	30
	PCA	8.3	17.7	10.8
	HOG	3.7	11.4	5.7
	LBP	1.7	8.5	2.8
	Score fusion (PCA, HOG, LBP)	2.5	9.4	5.7
	Score fusion (HOG, LBP)	0.6	6.3	2.8
	Feature fusion (PCA, HOG, LBP)	1.1	8.2	4.6
	Feature fusion (HOG, LBP)	0.6	8.0	1.7
	Decision fusion (PCA, HOG, LBP)	1.1	6.3	2.3
	Proposed method	*0.0*	*5.7*	*1.1*

The italicized data reflect the highest equal error rates of each experiments

Afaneh *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:81

Page 11 of 14

Table 6 EER results of gender-based experiments for identical twins

	Gender	EER results (%) of	
		Male	Female
Implemented algorithms	CNN	10.08	9.92
	PCA	14.94	17.56
	HOG	11.42	10.81
	LBP	4.91	5.64
	Score fusion (PCA, HOG, LBP)	6.74	9.12
	Score fusion (HOG, LBP)	6.12	5.54
	Feature fusion (PCA, HOG, LBP)	8.88	8.70
	Feature fusion (HOG, LBP)	4.04	5.38
	Decision fusion (PCA, HOG, LBP)	4.55	6.07
	Proposed method	*3.90*	*4.64*
Implemented algorithms in [18]	A	4.10	8.10
	B	39.40	39.10
	C	7.30	35.10
	D	22.30	21.30
	E	14.10	16.70
	F	9.80	13.10
	G	6.70	11.50

The italicized data reflect the highest equal error rates of each experiments

experiments were performed based on gender as female and male in which the facial images are grouped separately. The results based on EER values of identical twins and non-twins are shown on Tables 6 and 7, respectively.

6.4.4 Age-based experiments

The goal of the last experiment set is to study the effect of age using several algorithms for distinguishing identical twins and non-twins. Therefore, the images are divided into two categories based on age: "over 40 years old" and

"40 years old and younger." The results of these experiments are demonstrated on Tables 8 and 9 for identical twins and non-twins, respectively.

6.5 Experiments on standard FERET Datasets for non-twins recognition

In these experiments, the proposed method is evaluated using non-twins face images. Table 10 shows the EER results of non-twins recognition using standard FERET subsets. These set of experiments are conducted under

Table 7 EER results of gender-based experiments for non-twins

	Gender	EER results (%) of	
		Male	Female
Algorithms	CNN	6	15.7
	PCA	8.1	8.7
	HOG	4.0	4.3
	LBP	2.2	2.7
	Score fusion (PCA, HOG, LBP)	3.3	3.7
	Score fusion (HOG, LBP)	1.5	1.8
	Feature fusion (PCA, HOG, LBP)	2.1	2.5
	Feature fusion (HOG, LBP)	1.5	1.8
	Decision fusion (PCA, HOG, LBP)	1.7	2.1
	Proposed Method	*0.6*	*0.9*

The italicized data reflect the highest equal error rates of each experiments

Afaneh *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:81

Page 12 of 14

Table 8 EER results of age-based experiments for identical twins

	Age	EER results (%) of	
		40 and younger	Over 40
Implemented algorithms	CNN	18.8	4.74
	PCA	14.54	17.10
	HOG	11.89	10.14
	LBP	4.00	5.59
	Score fusion (PCA, HOG, LBP)	7.55	8.51
	Score fusion (HOG, LBP)	6.36	5.16
	Feature fusion (PCA, HOG, LBP)	8.79	8.53
	Feature fusion (HOG, LBP)	3.92	5.36
	Decision fusion (PCA, HOG, LBP)	5.79	5.30
	Proposed Method	*3.85*	*4.50*
Implemented algorithms in [18]	A	9.60	7.40
	B	38.40	39.00
	C	15.50	34.10
	D	24.30	21.60
	E	19.40	21.60
	F	14.50	12.50
	G	13.50	11.00

The italicized data reflect the highest equal error rates of each experiments

three different challenges, namely expression, illumination, and aging variations.

6.6 Results and discussion

All the experimental results demonstrate that the decision fusion (PCA, HOG, LBP) is better than or comparable with the state-of-the-art methods. However, the proposed method is better than the decision fusion (PCA, HOG, LBP), and it shows superior performance compared to the state-of-the-art methods in this field for all types of experimental conditions including expression, illumination, gender, and age variations. The high performance of the proposed method is caused by the usage of a combination of feature-level, score-level, and decision-level fusion in one method in addition to the usage of different voting techniques in the second-decision level.

Distinguishing identical twins under standard conditions is possible as shown in the experimental results. However, when conditions of the captured images are not ideal, distinguishing identical twins is a hard challenge.

Table 9 EER results of age-based experiments for non-twins

	Age	EER results (%) of	
		40 and younger	Over 40
Algorithms	CNN	6	16
	PCA	10.3	8.0
	HOG	5.1	3.3
	LBP	3.1	2.9
	Score fusion (PCA, HOG, LBP)	3.9	3.1
	Score fusion (HOG, LBP)	1.7	1.7
	Feature fusion (PCA, HOG, LBP)	2.5	2.7
	Feature fusion (HOG, LBP)	2.1	1.9
	Decision fusion (PCA, HOG, LBP)	1.9	1.7
	Proposed method	*1.0*	*0.8*

The italicized data reflect the highest equal error rates of each experiments

Afaneh *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:81

Page 13 of 14

Table 10 EER results for non-twins using standard FERET subsets under expression, illumination and age variations

| | Challenge subset | EER results (%) of | | | |
		Expression (fb)	Illumination (fc)	Aging (duplicate l)	Aging (duplicate ll)
Algorithms	CNN	6.10	26.44	9.88	13.54
	PCA	11.4	34.3	23.2	30.7
	HOG	9.5	31.9	13	14
	LBP	3.43	24.9	11.15	12
	Score fusion (PCA, HOG, LBP)	6.2	22.7	12.6	14.67
	Score fusion (HOG, LBP)	5.9	24.2	10.25	10.7
	Feature fusion (PCA, HOG, LBP)	8.7	28.1	11.15	14
	Feature fusion (HOG, LBP)	3.9	24.5	10.75	12
	Decision fusion (PCA, HOG, LBP)	2.9	21.8	8.15	8.7
	Proposed method	*2.5*	*20.6*	*7.5*	*8*

The italicized data reflect the highest equal error rates of each experiments

Identical twins represent a very difficult recognition problem, and the results achieved for the recognition of identical twins are worse than the results obtained to recognize non-twins.

7 Conclusions

A novel method is proposed for the solution of distinguishing identical twins by using facial images. The proposed method uses feature-level fusion, score-level fusion, and decision-level fusion with three feature extraction approaches. PCA, HOG, and LBP are implemented as feature extractors and matching is performed using KNN. Various experiments are conducted using ND-TWINS-2009-2010 and standard FERET Datasets. The experiments that use ND-Twins-2009-2010 database are performed under different illumination, expression, age, and gender conditions using samples of identical twins and non-twins separately. Additionally, the performance of the proposed method is measured using standard FERET Dataset of non-twins' faces under different expression, illumination, and aging conditions. Experiments show that the recognition of identical twins is harder when the conditions of capturing samples are different. Consequently, the degree of difference between images is lower when both training and test samples are acquired under the same conditions such as uniform lighting and natural expression. Results are not significantly affected by variation in age and gender. In addition, the high similarity between identical twins significantly affects the performance of any recognition system compared with the non-twins case. The proposed method is compared with four unimodal and five multimodal systems that are conducted in this work in addition to seven state-of-the-art algorithms. The lowest equal error rates of identical twins recognition that are achieved using the proposed method are 2.07% for natural expression, 0.0%

for smiling expression, and 2.2% for controlled illumination compared to 4.5, 4.2, and 4.7% equal error rates of the best state-of-the-art algorithm under the same conditions. Consequently, all the experimental results demonstrate that the proposed method outperforms all aforementioned techniques under different expression, illumination, gender, and aging conditions for both identical twins and non-twins recognition.

Availability of data and materials
The web links to the sources of the data (namely, images) used for our experiments and comparisons in this work have been provided in this article.

Authors' contributions
All authors read and approved the final manuscript.

Ethical approval and consent to participate
Not applicable

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Computer Engineering Department, Faculty of Engineering, Eastern Mediterranean University, Famagusta, North Cyprus, via Mersin 10, Turkey. [2]Institute of Technology, University of Tartu, Nooruse 1, Tartu, Estonia.

Afaneh *et al. EURASIP Journal on Image and Video Processing*   (2017) 2017:81

Page 14 of 14

References

1. AK Jain, AA Ross, K Nandakumarr, *Introduction to biometrics*. (Springer Science Business Media, New York, 2011), pp. 978–0387773254
2. A Bolotnikova, H Demirel, G Anbarjafari, Real-time ensemble based face recognition system for NAO humanoids using local binary pattern. Analog Integr. Circ. Sig. Process. **92**(3), 1–9 (2017)
3. I Lüsi, JCJ Junior, J Gorbova, Baró, S Escalera, H Demirel, J Allik, C Ozcinar, G Anbarjafari, in *Joint challenge on dominant and complementary emotion recognition using micro emotion features and head-pose estimation: databases*. Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference On (IEEE, Washington, 2017), pp. 809–813
4. G Anbarjafari, Face recognition using color local binary pattern from mutually independent color channels. EURASIP J. Image Video Process. **2013**(1), 6 (2013)
5. P Phillips, P Flynn, K Bowyer, R Bruegge, P Grother, G Quinn, M Pruitt, in *Proc. IEEE Conf. Autom. Face Gesture Recognit. Workshops*. Distinguishing identical twins by face recognition (IEEE, Santa Barbara, 2011), pp. 185–192. doi: 10.1109/FG.2011.5771395
6. KI Kim, K Jung, HJ Kim, Face recognition using kernel principal component analysis. IEEE Signal Proc. Lett. **9**(2), 40–42 (2002)
7. O Déniz, G Bueno, J Salido, F De la Torre, Face recognition using histograms of oriented gradients. Pattern Recogn. Lett. **32**(12), 1598–1603 (2011)
8. T Ahonen, A Hadid, M Pietikäinen, Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI). **28**(12), 2037–2041 (2006). doi:10.1109/TPAMI.2006.244
9. A Jain, S Prabhakar, S Pankanti, On the similarity of identical twin fingerprints. Pattern Recog. **35**(11), 2653–2663 (2002). doi:10.1016/S0031-3203(01)00218-7
10. A Ariyaeeiniaa, C Morrison, A Malegaonkara, B Black, A test of the effectiveness of speaker verification for differentiating between identical twins. Sci. Justice. **48**(4), 182–186. (2008). doi:10.1016/j.scijus.2008.02.002
11. AW Kong, D Zhang, G Lu, A study of identical twins' palmprints for personal verification. Pattern Recognit. **39**(11), 2149–2156 (2006). doi:10.1016/j.patcog.2006.04.035
12. K Hollingsworth, K Bowyer, P Flynn, in *Proc. IEEE Comput. Soc. Conf. CVPRW*. Similarity of iris texture between identical twins, (2010), pp. 22–29. doi:10.1109/CVPRW.2010.5543237
13. U Park, A Jain, Face matching and retrieval using soft biometrics. IEEE Trans. Inf. Forensic Secur. **5**(3), 406–415 (2010). doi:10.1109/TIFS.2010.2049842
14. Z Sun, AA Paulino, J Feng, Z Chai, T Tan, AK Jain, A study of multibiometric traits of identical twins. Proc. SPIE Biom. Technol. Hum. Identif. VII. **7667**, 1–12 (2010). doi:10.1117/12.851369
15. S Biswas, KW Bowyer, PJ Flynn, A study of face recognition of identical twins by humans. Proc. IEEE WIFS, 1–6 (2011). doi:10.1109/WIFS.2011.6123126
16. MT Pruitt, JM Grant, JR Paone, PJ Flynn, RWV Bruegge, in *Proc. 1st Int. Joint Conf. Biometrics*. Facial recognition of identical twins, (2011), pp. 185–192. doi:10.1109/IJCB.2011.6117476
17. L Zhang, N Ye, EM Marroquin, D Guo, T Sim, New hope for recognizing twins by using facial motion. WACV IEEE, 209–214 (2012). doi:10.1109/WACV.2012.6163026
18. JR Paone, PJ Flynn, PJ Philips, KW Bowyer, RWV Bruegge, PJ Grother, GW Quinn, MT Pruitt, JM Grant, Double trouble: differentiating identical twins by face recognition. IEEE Trans. Inf. Forensic Secur. **9**(2), 285–295 (2014). doi:10.1109/TIFS.2013.2296373
19. PJ Grother, GW Quinn, PJ Phillips, *Mbe 2010: Report on the evaluation of 2d still-image face recognition algorithms*. NIST, Gaithersburg, MD, USA, Tech. Rep (NISTIR 7709) (2010)
20. C Kalyoncu, Ö Toygar, GTCLC: a novel leaf classification method using multiple descriptors. IET Comput. Vis. **10**(7), 700–708 (2016). doi:10.1049/iet-cvi.2015.0414
21. M Eskandari, Ö Toygar, Selection of optimized features and weights on face-iris fusion using distance images. Comp. Vision Image Underst. **137**, 63–75. (2015). doi:10.1016/j.cviu.2015.02.011
22. M Farmanbar, Ö Toygar, Spoof detection on face and palmprint biometrics. SIViP, 1–8 (2017). doi:10.1007/s11760-017-1082-y
23. M Farmanbar, Ö Toygar, A hybrid approach for person identification using palmprint and face biometrics. Int. J. Pattern Recognit. Artif. Intell. **29**(06), 1556009 (2015)
24. D Huang, C Shan, M Ardebilian, Y Wang, L Chen, Local binary patterns and its application to facial image analysis: A survey. IEEE Trans. Syst. Man Cybern. **41**(6), 765–781 (2011). doi:10.1109/TSMCC.2011.2118750
25. T Ojala, M Pietikäinen, T Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Patt. Anal. Mach. Intell. (PAMI). **24**(7), 971–987 (2002). doi:10.1109/TPAMI.2002.1017623
26. N Dalal, B Triggs, Histograms of oriented gradients for human detection. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn (2005). doi:10.1109/CVPR.2005.177
27. M Eskandari, Ö Toygar, Fusion of face and iris biometrics using local and global feature extraction methods. SIViP. **8**(6), 995–1006 (2014). doi:10.1007/s11760-012-0411-4
28. K Delac, M Grgic, in *Electronics in Marine, 2004. Proceedings Elmar 2004. 46th International Symposium*. A survey of biometric recognition methods (IEEE, Zadar, 2004), pp. 184–193
29. A Dantcheva, P Elia, A Ross, What else does your biometric data reveal? A survey on soft biometrics. IEEE Trans. Inf. Forensic Secur. **11**(3), 441–467 (2016)
30. (2013). CVRL data sets [online] available: https://sites.google.com/a/nd.edu/public-cvrl/data-sets. Accessed Jan 2017
31. PJ Phillips, H Wechsler, J Huang, PJ Rauss, The feret database and evaluation procedure for face-recognition algorithms. Image Vis. Comput. **16**(5), 295–306 (1998)
32. PJ Phillips, H Moon, SA Rizvi, PJ Rauss, The feret evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell. **22**(10), 1090–1104 (2000)
33. F Noroozi, M Marjanovic, A Njegus, S Escalera, G Anbarjafari, Audio-visual emotion recognition in video clips. IEEE Trans. Affect. Comput (2017). doi:10.1109/TAFFC.2017.2713783
34. F Noroozi, M Marjanovic, A Njegus, S Escalera, G Anbarjafari, in *Pattern Recognition (ICPR), 2016 23rd International Conference On*. Fusion of classifier predictions for audio-visual emotion recognition (IEEE, Cancun, 2016), pp. 61–66
35. J Grobova, M Colovic, M Marjanovic, A Njegus, H Demire, G Anbarjafari, in *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference On*. Automatic hidden sadness detection using micro-expressions (IEEE, Washington, 2017), pp. 828–832
36. C Szegedy, W Liu, Y Jia, P Sermanet, S Reed, D Anguelov, D Erhan, V Vanhoucke, A Rabinovich, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Going deeper with convolutions (IEEE, Boston, 2015), pp. 1–9