ORIGINAL ARTICLE

A high-discriminative facial recognition method based on shape and grey-level appearances using landmark-points

Taner Cevik[1] · Fatih Sahin[2]

Abstract

This paper proposes a high-discriminative facial recognition method that fuses shape and grey-level features of faces. Facial landmark-points are characteristic to individuals, hence can be exploited for recognition. Spatial relationships of these landmark-points, namely their Euclidean distances between each other, are included in the feature set. Besides, the mean grey-level values calculated at the vicinity of these landmark-points are also considered as a discriminative factor and incorporated into the feature set. The results of the comprehensive simulations show the remarkable and competitive performance of the proposed method regarding recognition accuracy, as well as robustness against partial occlusion, noise, expression changes and variances in illumination.

Keywords Facial recognition · Landmark · Shape appearance · Grey-level appearance · Partial occlusion · Noise · Illumination · Pose · Expression

1 Introduction

With the development of modern science and technology, the need for authentication and identification processes to be carried out quickly, effectively and automatically, i.e., without the need for any human intervention is rapidly increasing [1]. Biometrics have received great interest due to their high performance in differentiation among people in many areas such as surveillance, identification, human–computer interaction and have gained wide application in recent years [2–4]. Individuals have some biological properties called metrics that distinguish them from others [5]. Biometry is concerned with these metrics that are categorized as physiological and behavioral [6].

Face data are one of the leading biometrics in that it can be collected and processed in real time without any discomfort and physical contact through devices such as cameras and without the need for any human intervention, as well as its high distinctive performance [7, 8]. Therefore, face recognition has been widely preferred in the security domain of commercial and law enforcement applications [9, 10]. The face recognition process, which is actually a representative of pattern recognition, consists of two main steps. The first of these is the process of inference of the features that can distinguish a face from others, also called face representation. The other is the design and application of classifiers or models that can be distinguished by the so-called facial pairing [11, 12]. Among these steps, the face representation is more crucial because face recognition is a computationally difficult task that requires a fine distinction between images of similar identity, as well as generalizing to different images of the same identity [13]. In addition, factors such as noise, partial occlusion and illumination, exposure and expression differences make the automatic recognition process much more difficult [14].

Each face has its own shape and color characteristics. The methods focus on the color characteristics of the face is concerned with the texture of the face. It is identified that each human face has its own structural properties called texture. Although there is no general consensus on its definition, texture is defined as repeated patterns in the image [15]. A descriptor, to be a good representative of a face, should carry the qualifications as a high-discriminative ratio, low com-

✉ Taner Cevik
   tanercevik@aydin.edu.tr

   Fatih Sahin
   fsahin1976@yahoo.com

[1] Department of Software Engineering, Istanbul Aydin
     University, Istanbul, Turkey

[2] Department of Computer Engineering, Istanbul Aydin
     University, Istanbul, Turkey

putational complexity and resistance to deteriorating factors such as noise, occlusion and variations in illumination and expression [16].

1.1 Related work

Plenty of face representatives have been presented in the literature that is mainly grouped under two categories: local [17, 18] and holistic descriptors [19–21]. Holistic approaches examine the entire image and consider holistic features. This holistic set of features also refers to the general characteristics of the face [22]. Principal component analysis (PCA) [19], linear discrimination analysis (LDA) [20], independent component analysis (ICA) [23] and grey-level co-occurrence matrices (GLCM) [24] are the most basic, popular and inspiring works of holistic approaches. The distinctive feature of PCA is to reduce dimensionality by the transition from high-dimensional image space to a low-dimensional orthogonal area. The transition is performed by applying a linear transformation, given the least mean square reconstruction error. LDA is interested in finding a linear transformation that minimizes the intra-class variation and maximizes the inter-class variation. First adopted by Herault and Jutten [25], ICA searches for a linear conversion to minimize the statistical dependencies of the components of a vector. GLCM is one of the basic and prominent statistical textual

feature extraction methods that has been widely used in various applications [26, 27] for texture analysis. GLCM is the matrix that holds the distribution of co-occurring intensity patterns at a given offset over a given image. The second-order statistical (Haralick) features are extracted to analyze the texture of the image which subsequently can be used for classification tasks [28]. GLCM, which is one of the primary sensitive textual descriptors, handles the visual texture of the image by evaluating the pixel spatial arrangement statistics [29].

In contrast to holistic approaches, local descriptors, such as, local binary pattern (LBP) [30], local Gabor binary pattern (LGBP) [31], center-symmetric local binary pattern (CS-LBP) [32], local directional pattern (LDP) [33], joint local binary patterns with Weber-like responses (LJBPW) [34], pyramid transform domain local binary pattern (PLBP) [35], local directional gradient pattern (LDGP) [36], local phase quantization (LPQ) [37], local directional number pattern (LDNP) [38], histogram of gradients (HoG) [39], local ternary pattern (LTP) [40], Gabor [41, 42] benefit from local appearance features. Among the local descriptors, Gabor wavelets, Radon transformation [43], Texton learning [44, 45] and LBP are the prominent ones and have inspired further studies [46, 47]. Especially, LBP has found a wide application area due to its flexibility of adaptation, high performance of discrimination and low complexity [48]. Therefore, a series

of sequential studies [49] have been proposed to develop and expand the idea of LBP.

As mentioned above, the face is one of the most important biometrics because of the amount of information it offers about individuals and the fact that this information can be collected using remote, camera-like devices without any discomfort and human intervention. Many non-verbal and semantic information such as a person's identity, intent, emotion can be obtained by looking at an individual's face. On the face, there are some important key points, namely landmark-points, on which the face shape analysis is made. Some leading computer vision applications, such as head-pose-estimation [50, 51] and facial expression recognition [52, 53], exploit the data belonging to the landmark-points. Moreover, the landmark-points around the eye can provide the first estimate of the central position of the pupils, to be used in eye detection and eye tracking [54]. Data retrieved from landmark-points can be an important source of information for human and computer interaction, entertainment, security surveillance and medical applications. Due to some reasons, the detection of the landmark-points becomes struggling. Firstly, the change in facial appearance differs from person to person in different facial expression insertions and head position changes. Secondly, environmental conditions such as lighting significantly affect the appearance of faces in images. And, finally, the self-occlusion due to extreme changes in head poses, or the occlusion caused by other objects leads to missing face appearance information [55].

1.2 Our contributions

Although facial landmark-points have been heavily exploited for facial expression recognition and head-pose-estimation, their discriminative performance on personal identification has been poorly explored. In this study, the discriminative performance of the landmark-points in terms of facial recognition by proposing a method that exploits both the facial appearance and face shape features belonging to them are analyzed. In the first step, all sample images are put into a standard form by pre-processing to provide uniformity. Following the completion of the uniformization phase of the samples, the landmark-points detection process starts. By the end of this process, sixty-six landmark-points on each face image are defined with their spatial coordinates. Subsequently, by exploiting the spatial coordinates of these landmark-points, some shape-based features encompassing their Euclidean distances between each other as well as grey-level-based appearance features are extracted and included in the feature set. The classification is performed according to this compound feature set that comprises both the shape and appearance-based relationships between the landmark-points. Extensive face recognition experiments are conducted on four widely used face datasets.
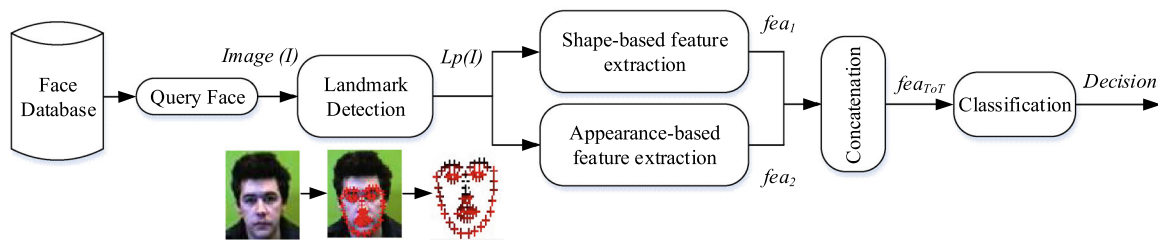
Fig. 1 The block diagram of the entire process

As clearly identified in the simulation results, the strengths of the proposed method can be listed as follows:

- Remains stable under varying illumination.
- Maintains its high individual discrimination despite increasing noise.
- Retains its ability to recognize faces at a satisfactory rate, even when exposed to partial obstacles that make it very difficult to distinguish individuals.

The rest of this paper is organized as follows. Section 2 explains the proposed method briefly. Section 3 reports the experimental results and discussions among them. Finally, Sect. 4 concludes the paper.

2 Methodology

The method proposed in this study is comprised of three major steps as landmark-points detection, shape and grey-level appearance-based feature extraction and classification. Figure 1 illustrates the entire operational block diagram of the proposed method.

2.1 Landmark detection

It is the purpose of the facial landmark detection algorithms to automatically detect the positions of the landmark-points of faces in images or videos. These landmark-points define the positions of some dominant components of the face, such as corners of lips and eyes, the tip of the nose or the interpolated points that connect these fiducial staff on a spline or a face contour [55]. The methods for facial landmark detection are generally classified under three headings, as the holistic [56, 57], part-based [58, 59] and regression-based [60] methods [61].

Part-based methods build the idea of local facial appearances and holistic shape appearances, which makes these methods robust to occlusion and illumination. This idea, which is suggested firstly by the study named active shape models (ASM) [62] and following improved by another

Fig. 2 Sample image and its annotated landmarks

study, constrained local models (CLM) [63] has pioneered many subsequent research activities [64, 65].

In contrast to the part-based ones, holistic methods exploit the global facial shape patterns and holistic facial appearance information. The active appearance model (AAM) [56] is indeed a statistical model that uses a few coefficients to fit the face images to control both the facial appearance and the shape changes. AAM relies on principal component analysis (PCA) while building the holistic facial appearance model and global shape model subsequently, during the model construction. The determination of the landmark-points is done by adapting the learned appearance and shape models to the test images. Model coefficients are estimated in the conventional AAM, by repeated calculations based on the model coefficient update estimation that relies on the current model coefficients and error image [55].

Regression-based landmark detection is a method of interest in recent research studies. The regression-based methods, rather than putting forward a global face model as done by holistic and part-based methods, intend to learn directly a mapping from the appearance of an image to its landmark locations. The regression-based methods roughly are categorized as direct regression methods, cascaded regression methods and deep-learning-based regression methods. Direct regression methods, which are further sub-categorized as local and holistic approaches, exert to predict the landmark locations at once without the need for an initialization. In contrast, cascaded regression methods, require subsequent, cascaded iterations to correctly localize the landmark-points with the requirement of pre-initialization of the landmark locations [55].

Indeed, a landmark detection algorithm specifies the locations of N landmarks, $Lp = \{x_1, y_1, x_2, y_2, \ldots, x_N, y_N\}$ on a given image f. In this study, a CLM-based method, that is, the discriminative response map fitting (DRMF) [65] is applied to define the landmarks of the face images. Sixty-
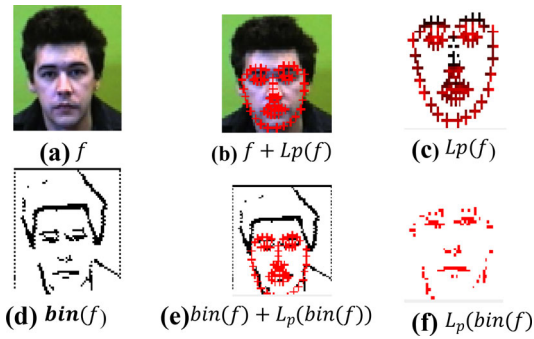
(a) f

(b) $f + Lp(f)$

(c) $Lp(f)$

(d) $bin(f)$

(e) $bin(f) + L_p(bin(f))$

(f) $L_p(bin(f))$

Fig. 3 A sample image and its binary version with their annotated landmarks

Table 1 Image coordinates (X, Y) of the landmark-points of a sample image and its binary version

$X(L_p(f))$	$Y(L_p(f))$	$X(L_p(bin(f)))$	$Y(L_p(bin(f)))$
12.12987	36.30967	12.67734	38.81291
12.68323	40.61763	13.15086	43.56321
13.57034	44.86027	13.90462	48.28583
14.85724	49.18057	14.93438	52.92806
16.81583	53.43989	16.53872	57.30252
19.72069	57.07365	19.17155	61.04213
23.36007	59.82478	22.64357	63.99347
27.40592	61.82429	26.68399	66.08498
31.78441	62.43856	31.22114	66.62399
36.1457	61.25646	35.85715	65.63002
40.00452	58.78397	39.82174	63.08491
43.37626	55.6071	43.14045	59.75187
45.86989	51.60234	45.62617	55.7707
47.24051	47.01898	47.10655	51.35436
47.87201	42.43008	47.86383	46.78536
48.09931	37.96361	48.23014	42.17451
48.01737	33.48796	48.28696	37.5547
14.5989	32.79396	15.28539	33.60506
16.70609	30.91407	17.48864	31.73869
19.59609	30.20413	20.35555	31.0144
22.63233	30.43724	23.33207	31.18806
25.57247	31.2455	26.21372	31.95197
32.93123	30.60969	33.46875	31.75856
35.83196	29.32294	36.23001	30.82089
38.91826	28.6139	39.14148	30.43596
42.00653	28.9136	42.02274	30.89196
44.49064	30.51393	44.37334	32.50591
29.54023	35.05933	30.07173	35.74509
29.74289	38.00162	30.11559	38.73692
29.94943	40.94511	30.15553	41.71238
30.16003	43.86658	30.19773	44.67409
27.61102	45.93014	27.35574	47.17923
28.9889	46.30545	28.86667	47.67258
30.40842	46.41401	30.46833	47.83739
31.85814	46.07692	32.04978	47.56523
33.21936	45.48471	33.51087	46.96988
18.58342	36.41866	19.12932	36.95469
20.41554	35.01748	20.99095	35.70374
22.79524	34.87084	23.28537	35.61264
24.78259	35.98658	25.20425	36.7432
22.77831	36.77687	23.1979	37.31306
20.60786	37.02824	21.08796	37.47068
34.67149	35.17974	34.94474	36.42996
36.46659	33.74709	36.74474	35.19575
38.86335	33.53635	39.0245	35.10579

six landmark-points are identified when DRMF applied on a face image as illustrated in Fig. 2.

DRMF shows promising performance regarding landmark-points detection, even on binary images. The location values of the landmark-points of an image and its binary version are almost the same, with a mean square error (MSE) value of *2.8315*, as clarified in Fig. 3 and Table 1.

2.2 Feature extraction

The descriptor in this study is comprised of two types of feature sets that are extracted relying on the facial landmark-points, as appearance-based features and shape-based features. The following section describes each feature set and the way of their retrieval in detail.

2.2.1 Shape-based feature set

DRMF identifies sixty-six landmarks on each face image. It is examined whether the spatial relationships of these landmark-points are unique for individuals or not. Let f be an image with N landmarks, $Lp = \{x_1, y_1, x_2, y_2, \ldots, x_N, y_N\}$, where $N = 66$. The Euclidean distances between these points are calculated. Totally $N \times (N-1)$ distance values are calculated as:

$$d(Lp_i, Lp_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (1)$$

$$fea_1 = [d(Lp_1, Lp_2)d(Lp_1, Lp_3)\ldots.d(Lp_{N-1}, Lp_N)] \quad (2)$$

where fea_1 denotes the feature set that is comprised of the Euclidean distances between the landmark-points.

2.2.2 Appearance-based feature set

The second feature set fea_2 includes the mean pixel intensity value differences of the landmark-points. To mitigate the effect of the single-pixel intensity changes due to noise

Table 1 continued

$X(L_p(f))$	$Y(L_p(f))$	$X(L_p(bin(f)))$	$Y(L_p(bin(f)))$
40.92885	34.62152	40.95582	36.23431
39.00332	35.54088	39.04976	36.88558
36.79986	35.64094	36.95332	36.86539
25.15109	52.79591	24.26228	54.00413
26.59238	51.22011	25.97602	52.42645
28.47418	49.98056	28.13225	51.34361
30.75566	50.17207	30.54656	51.49857
33.05967	49.62748	32.99945	51.16811
35.18908	50.55731	35.26755	52.08658
36.97161	51.85436	37.12641	53.56337
35.44784	53.48653	35.54884	55.41116
33.45097	54.62027	33.37958	56.63587
31.16589	55.14614	30.81342	57.05753
28.90241	54.97643	28.2527	56.81596
26.82357	54.16791	26.00239	55.74238
28.61613	51.80899	28.02457	53.16331
30.88918	51.7607	30.63792	53.08958
33.22669	51.45047	33.32162	52.97792
33.21794	52.01534	33.41687	54.06619
30.95401	52.51079	30.73026	54.49448
28.71534	52.36075	28.07193	54.25958

or illuminative changes, the mean value of the pixels at the *k-hop* neighborhood, as well as the pixel at each landmark, is considered.

$$mp_{Lp_i} = \left(\sum_{n=1}^{k^2-1} p_{Lp_n} + p_{Lp_i} \right)/k^2 \qquad (3)$$

$$fea_2 = \left[d\left(mp_{Lp_1}, mp_{Lp_2}\right) \dots d\left(mp_{Lp_{N-1}}, mp_{Lp_N}\right) \right] \qquad (4)$$

where p_{Lp_i}, mp_{Lp_i} and fea_2 refer to the pixel intensity value of the landmark-point at Lp_i, mean intensity value belonging to its *k-hop* neighborhood, and the feature set containing these mean intensity values, respectively.

After the extraction stage of the two feature sets, they are concatenated to form the overall feature set, fea_{ToT}, as:

$$fea_{ToT} = fea_1 | fea_2 \qquad (5)$$

3 Simulation results and discussions

Various experiments are conducted to measure and analyze the performance of the proposed framework under different circumstances. The evaluation of the proposed framework is performed mainly on the CAS-PEAL-R1 dataset [67]. The CAS-PEAL-R1 dataset is a subset of the CAS-PEAL dataset

(a) Images of individuals wearing different accessories

(b) Images of individuals having varying expressions

(c) Face images exposed to varying illumination

Fig. 4 Exemplary face images retrieved from different categories in the CAS-PEAL-R1 dataset

that contains several tens of thousands of images of 1040 subjects. The CAS-PEAL-R1 dataset is preferred because it contains a large number of images taken under various conditions such as varying lighting, expression and accessories. These variable factors make the recognition process much more difficult.

Sample images retrieved from different categories in the CAS-PEAL-R1 dataset are presented in Fig. 4.

The performance analysis of the proposed method is carried out in two steps. In the first step, the resistance of the method to noise effects, partial occlusion and illumination changes is explored. Then, by performing comprehensive simulations, the facial recognition accuracy performance of the proposed method is analyzed and discussed. The performance of the proposed method is compared to a number of the-state-of-the-art methods (LBP, Gabor, local tetra patterns (LTetP) [68], local monotonic pattern (LMP) [69], local phase quantization (LPQ), Weber local descriptor (WLD) [70], local gradient pattern (LGP) [71], median binary pattern (MBP) [72], local arc pattern (LAP) [73], monogenic binary coding (MBC) [74]) proposed in the literature.

Following the feature extraction stage, the classification task is implemented. The classification is done by means of supervised training. *k*-nearest neighbor (*k*-NN) method is used as the training algorithm. In the *k*-NN method, *k* value is taken as 1. The reason for this is that individuals show very close features regarding facial characteristics. Therefore, if you increase the radius of the circumference during

classification, you may accidentally assign an incorrect label to a subject. On the outset, the data are split into two parts (train and test) randomly to train the model. The train-part forms the %80 of the dataset while test includes the rest %20. Besides, the train-part of the dataset is also partitioned in two as the train (%80) and validation (%20).

Experiments are performed on MATLAB 2017b running on the Intel CORE i7-5500U 2.4 GHz processor and 16 GB RAM computer system.

3.1 Analysis of stability

Noise, occlusion and changes in illumination can significantly affect recognition performance. Therefore, the proposed method should not fail under exerting conditions and should remain stable to achieve satisfactory recognition performance. First, the robustness of the proposed method against noise is analyzed. Secondly, it is presented the behavior of the proposed method in the occasion of exposure to variances in illumination. Lastly, it is explored the resistance of the proposed descriptor under variable partial occlusions.

3.1.1 Noise resistance analysis

An important consideration to be taken into account during the performance analysis of a face identifier is how it resists against a challenging factor such as noise without any improvement, i.e., filtration. Therefore, the reaction of the proposed method is examined by applying artificially produced noise. Two types of noise, i.e., salt-pepper and Gaussian, are artificially applied to each image in the dataset.

At the outset, salt-pepper noise is considered. Images may sometimes suffer to noise, namely impulse noise, during acquisition, transmission or recording operations. Impulse noise is generally classified as random-valued impulse noise (RVIN) and fixed-valued impulse noise (FVIN). These two types of noise models differ in the intensity value change occurring on the noise pixels. In FVIN model, each noise-suffered pixel takes the value 0 or 255, that is, the pixel turns to either black or white. The modeling of FVIN is ordinarily is done as follows:

$$x'_{ij} = \begin{cases} \{0, 255\} \text{ with probability } p \\ x_{ij} \text{ with probablity } 1 - p \end{cases} \quad (6)$$

where x_{ij}, x'_{ij}, p refer to the original, noisy pixel intensity values at image coordinate (i, j), and the noise density, respectively [75].

For RVIN, two types of models have been proposed. In the first of these models [76], a noisy pixel can take one of the values in a fixed interval of the length m, rather than two fixed values as in FVIN. This model is called as fixed range impulse noise (FRIN) and formulated as:

$$x'_{ij} = \begin{cases} [0, m) \text{ with probability } p_1 \\ x_{ij} \text{ with probability } 1 - p \\ (255 - m, 255] \text{ with probability } p_2 \end{cases} \quad (7)$$

The second proposition [75] for RVIN is called as general fixed-valued impulse noise (GFN) or multi-valued impulse noise (MVIN) and formulated as follows:

$$x'_{ij} = \begin{cases} \{0, 255\} \text{ with probability } p \\ x_{ij} \text{ with probablity } 1 - p \end{cases} \quad (8)$$

where S is the set of pulse noise values consisting of k elements selected from values in the range [0, 255].

Every image in the dataset is exposed to salt-pepper noise artificially and the recognition accuracy performance analysis is conducted on the salt-pepper noisy images. Figure 5 demonstrates the recognition accuracy values of the proposed and other methods. d, $(fea_{ToT}, k = 1)$, $(fea_{ToT}, k = 2)$ denote the noise density, the proposed method considering 1-hop and 2-hop neighboring pixels while calculating the mean pixel intensity value, respectively. Clearly, the proposed method remains stable even in case of increasing salt-pepper noise exposure. Since the method proposed in this study fully relies on the facial landmark-points, the salt-pepper noise does not affect the extraction of these points and the same or nearly similar feature set is continued to be captured. Thus, while the recognition accuracy performances of most of the other methods seriously degrade, our method can distinguish individuals despite the increasing salt-pepper noise rate.

In the next stage of noise durability performance analysis, another type of noise, Gaussian noise, is taken into consideration. The two dominant noise sources for digital image acquisition are the stochastic quality of photon counting in detectors and the internal and electronic fluctuations of the collection devices [77]. This most common noise arising due to the image acquisition system can be modeled as Gaussian random noise generally [78]. Gaussian noise is statistical noise with a probability density function (PDF) equal to that of the normal distribution, also known as the Gaussian distribution, named after Carl Friedrich Gauss. In other words, the positions of the pixels exposed to noise and scattering of the values are subject to Gaussian distribution. The PDF of a Gaussian random variable is formulated as in the following:

$$p_G(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \quad (9)$$

where z, μ, σ denote the grey-level, mean value and standard deviation, respectively.

Figure 6 presents the recognition accuracy values of the proposed method and other competitors in the case of Gaussian noise exposure. The data in the table show the

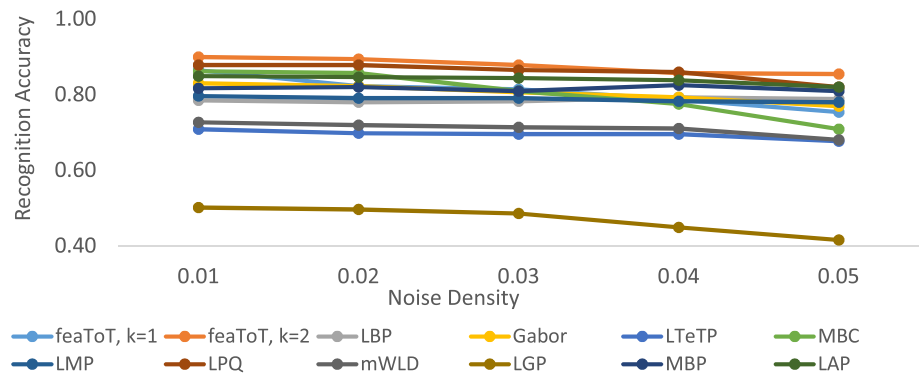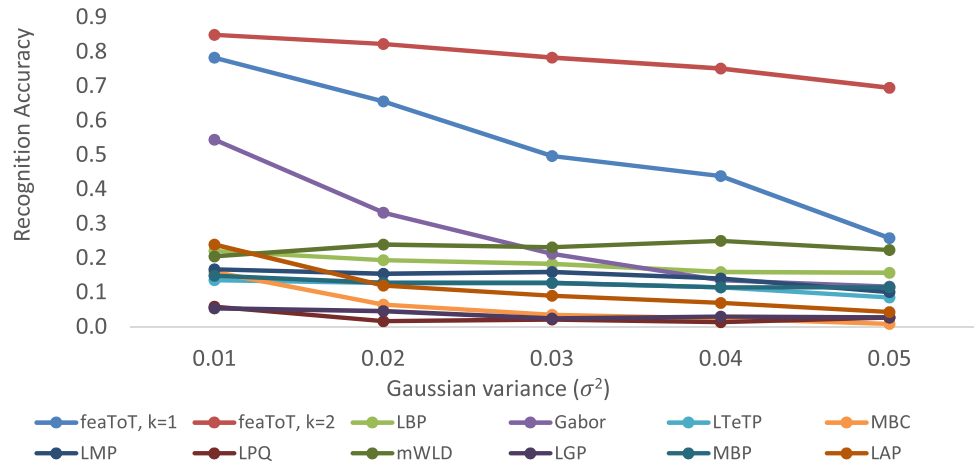**Fig. 5** The effect of salt-pepper noise on the recognition accuracy

Fig. 6 The effect of Gaussian noise on the recognition accuracy

recognition accuracy values for the increasing Gaussian variance (σ^2) values by taking the Gaussian mean constant $\mu = 0.001$. Inherently, the recognition accuracy performance of a method degrades due to the increase in the variance of the noise. However, a robust descriptor should be as resistant as possible despite the increasing and changing noise parameters and be able to withstand its disturbing effects. Obviously, our method keeps its robustness even under varying and increasing noise values, while the recognition accuracy performances of other methods degrade seriously.

3.1.2 Varying-illumination-resistance analysis

Face recognition becomes a strongly challenging paradigm especially in unconstrained environments [79]. Varying illumination is one of the most decisive factors that make facial recognition difficult. Image variations because of the varying illumination that induce changes on the images such as cast or attached shadows can be larger than that due to the innate differences between individuals [80]. These challenges stemming from varying illumination have attracted considerable attention from researchers, and many studies have been conducted to overcome these challenges. These

studies are broadly classified into three categories as, normalization and pre-processing, illumination invariant feature extraction and modeling [81].

The method proposed in this study falls into the category illumination invariant feature extraction. If a method exploits the feature set that heavily relies on pixel intensity values, it is inevitable to be affected by changes in illumination. Therefore, when designing our face descriptor, it is intended not to produce a feature set based purely on pixel density values that would reduce immunity to the damaging effects of varying illumination.

The recognition accuracy performance of our method and other state-of-the-art methods are explored by conducting extensive simulations on the CAS-PEAL-R1 dataset. The reason for selecting the CAS-PEAL-R1 data set is that instead of being artificially produced, it contains a subset of images that are exposed to varying illumination that occurs naturally in the indoor environment. Figure 4c presents sample face images that are exposed to varying natural illumination. Obviously, without any pre-processing or normalization process, the distinguishing of the individuals is highly struggling for any descriptor that descriptor-feature-set heavily relies on pixel intensity values. As illustrated in Fig. 7, our method

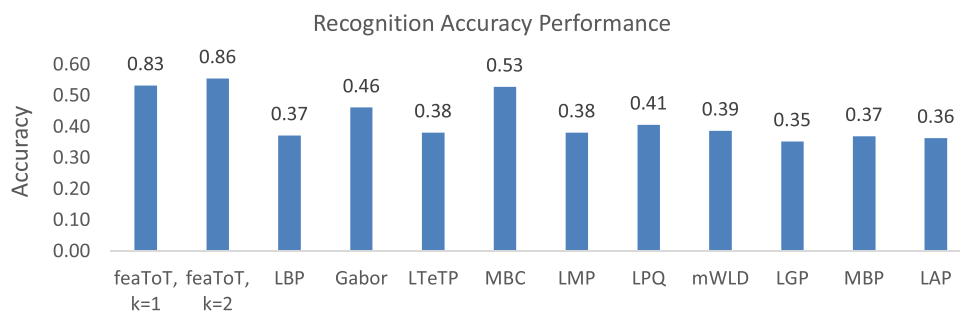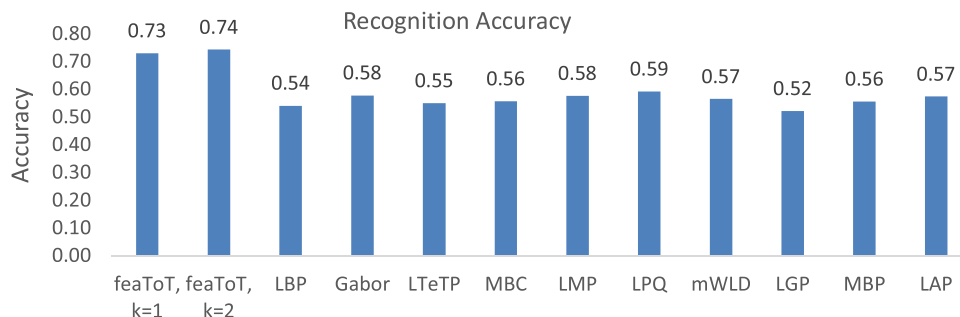**Fig. 7** The effect of illumination variation on the recognition accuracy

Recognition Accuracy Performance

Fig. 8 The effect of partial occlusions on the recognition accuracy

Recognition Accuracy

does the best regarding recognition accuracy even under the diminishing effects of shadows as a result of changing illumination. The performances of the other descriptors sharply fall due to their feature sets' reliance purely on pixel intensity values.

3.1.3 Recognition performance analysis under facial-accessory-caused partial occlusion

Although there is much research on face recognition to mitigate the distorting effects of pose and illumination changes, problems caused by occlusions are often overlooked. However, face occlusion is quite common and may occur due to intentional or unintentional spontaneous reasons. For example, football hooligans and ATM criminals may wear scarves and/or sunglasses to prevent their faces from being recognized. Other than that, some people wear veils because of religious beliefs or cultural habits. Other sources of facial occlusion include medical masks, beards, hats, facial hair, mustaches, make-up and so on. Of course, facial occlusion can significantly affect the performance of the most complex face recognition systems, unless occlusion analysis is specifically considered. The robustness of face recognition systems against partial occlusion is therefore very important nowadays [82].

Local feature-based methods have been recognized to be robust to partial occlusions and less susceptible to such problems, unlike traditional holistic approaches such as PCA, LDA and ICA. Some of the local feature-based methods

Fig. 9 Exemplary face images retrieved from the ExtendedYaleB dataset

focus solely on illumination and/or expression changes, while others [83, 84] aim to overcome the problems caused by partial occlusions.

As abovementioned, the method in this study does not solely rely on the pixel intensity values, rather, it is based on facial landmark features that are stable even under partial occlusions. Therefore, it can be seen from the results in Fig. 8 that our method has a high performance even on the data set of partially obstructed images, such as glasses and hats, which seriously impair facial recognition (Fig. 4a). Compared to the performance of other methods, our method provides a great advantage.

3.2 Recognition accuracy performance analysis

In the previous section, the analysis of the stability and robustness of the proposed method is performed on the challenging dataset CAS-PEAL-R1, which comprises recognition difficult, occluded and varying illuminated images. Besides, the resistance of the method is explored by exposing these images to artificially generated salt-pepper and

Fig. 10 Recognition accuracy performance analysis on the ExtendedYaleB dataset
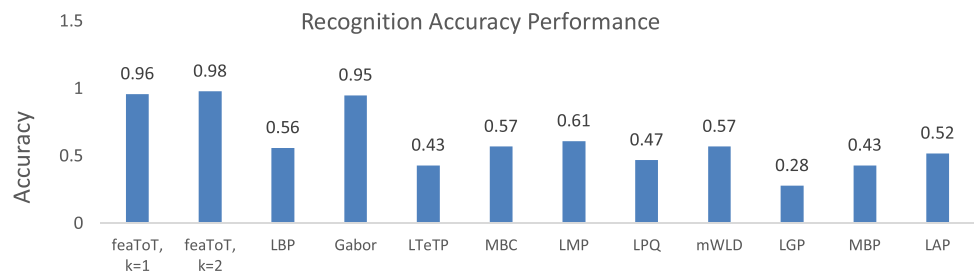

Recognition Accuracy Performance

Fig. 11 Exemplary face images retrieved from the Face94 dataset
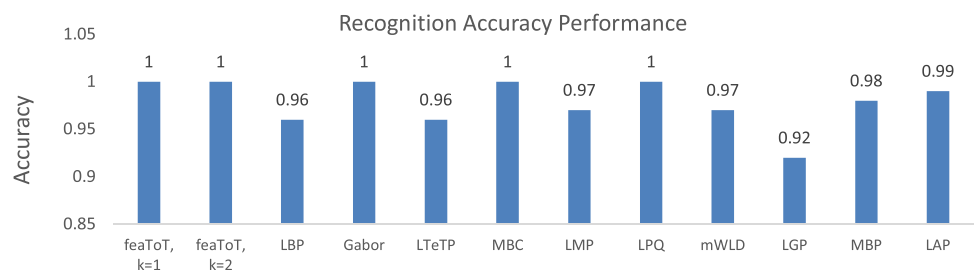

Fig. 13 Exemplary face images retrieved from the JAFFE dataset

Gaussian noise. This section clarifies the recognition accuracy performance of the proposed method on different datasets, namely ExtendedYaleB [85], Face94 [86] and JAFFE [87–89]. The Extended Yale B dataset contains 16352 images of 28 different individuals, 640×480 pixel size, exposed to 9 different poses and 64 different illuminations. Although images in the Extended Yale B dataset do not include any expression variation, they are exposed to significant pose and illumination changes. Figure 9 presents sample images of ExtendedYaleB dataset.

Figure 10 gives the recognition accuracy performance of the proposed and other state-of-the-art methods on the ExtendedYaleB dataset. Similarly, a result consistent with the reality of the previous results is obtained, i.e., in terms of recognition performance, the method we propose gives the second-best performance following Gabor.

Simulations are continued to be conducted on another dataset, namely Face94, which is composed of a total of 1860 images of which 20 belong to each individual. Despite the fact that it is not as challenging as CAS-PEAL-R1 and ExtendedYaleB datasets, individuals pose varying expressions (Fig. 11). Not surprisingly, as presented in Fig. 12, the proposed method performs the best among the others.

Following, the face recognition performance is analyzed on the JAFFE dataset. The database contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects (Fig. 13). As clarified in Fig. 14, despite the high-varied expressions of the models, the proposed method achieves one of the best discrimination performances.

4 Conclusion

In this paper, a high-discriminative facial recognition method, which fuses shape and grey-level features of faces, is proposed. Facial markings are used as shape properties because they present the characteristics of individuals that are at least unchanged or variable even in the event of exposure to destructive factors such as occlusion, varying illumination and noise. A number of features are produced of these facial landmark-points. These features include both some spatial values and pixel intensity values that are calculated considering only those landmark-points and their vicinities. As clearly presented in the simulations results, the

Fig. 12 Recognition accuracy performance analysis on the Face94 dataset

Recognition Accuracy Performance

Fig. 14 Recognition accuracy performance analysis on the JAFFE dataset
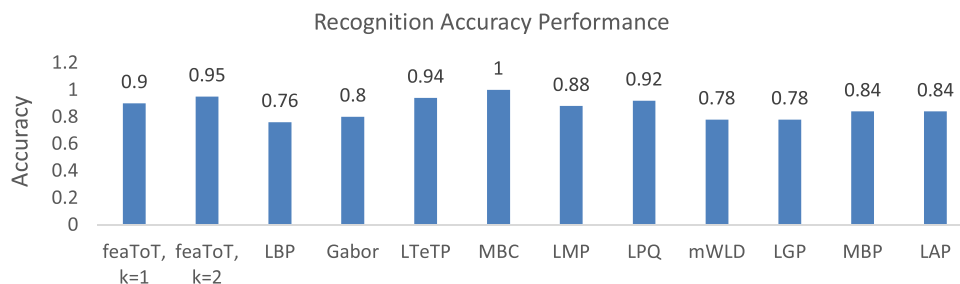
proposed method remains stable even under the challenging factors such as varying illumination, noise and partial occlusion, which diminishes of a facial recognition method significantly. The recognition and robustness performance of a number of the-state-of-the-art methods are also analyzed. The proposed method competes promisingly with others in terms of recognition accuracy and at the same time makes a clear difference to them when exposed to challenging factors.

Compliance with ethical standards

References

1. Liu, X., Lu, L., Shen, Z., Lu, K.: A novel face recognition algorithm via weighted kernel sparse representation. Future Gener. Comput. Syst. **80**, 653–663 (2018)
2. Zhong, F., Zhang, J.: Face recognition with enhanced local directional patterns. Neurocomputing **119**, 375–384 (2013)
3. Guan, Z., Wang, C., Chen, Z., Bu, J., Chen, C.: Efficient face recognition using tensor subspace regression. Neurocomputing **73**, 2744–2753 (2010)
4. Cevik, N., Cevik, T., Gurhanli, A.: Novel multispectral face descriptor using orthogonal walsh codes. IET Image Proc. (2019). https://doi.org/10.1049/iet-ipr.2018.6423
5. Jain, A.K., Hong, L., Pankanti, S.: Biometric identification. Commun. ACM **43**(2), 90–98 (2000)
6. Jain, A.K., Ross, A.: Introduction to biometrics. In: Jain, Anil K., Ross, Arun (eds.) Handbook of Biometrics, pp. 1–22. Springer, New York (2008)
7. Dubey, S.R.: Local directional relation pattern for unconstrained and robust face retrieval (2017). arXiv:1709.09518 [cs.CV]
8. Jafri, R., Arabnia, H.R.: A survey of face recognition techniques. J. Inf. Process. Syst. **5**(2), 41–68 (2009)
9. Huang, H., Li, J., Liu, J.: Enhanced semi-supervised local Fisher discriminant analysis for face recognition. Future Gener. Comput. Syst. **28**(1), 244–253 (2012)
10. Chen, X., Zhang, J.: A novel maximum margin neighborhood preserving embedding for face recognition. Future Gener. Comput. Syst. **28**(1), 212–217 (2012)
11. Duan, Y., Lu, J., Feng, J., Zhou, J.: Context-aware local binary feature learning for face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **40**(5), 1139–1153 (2018)
12. Lu, J., Liong, V.E., Zhou, J.: Simultaneous local binary feature learning and encoding for homogeneous and heterogeneous face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **40**(8), 1979–1993 (2018)
13. Abudarham, N., Shkiller, L., Yovel, G.: Critical features for face recognition. Cognition **182**, 73–83 (2019)
14. Cevik, N., Cevik, T.: DLGBD: a directional local gradient based descriptor for face recognition. Multimed. Tools Appl. **78**(12), 15909–15928 (2018). https://doi.org/10.1007/s11042-018-6967-4
15. Nanni, L., Brahnam, S., Ghidoni, S., Menegatti, E., Barrier, T.: Different approaches for extracting information from the co-occurrence matrix. PLoS ONE **8**(12), 1–9 (2013)
16. Liu, L., Fieguth, P., Guo, Y., Wang, X., Pietikainen, M.: Local binary features for texture classification: taxonomy and experimental study. Pattern Recognit. **62**, 135–160 (2017)
17. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
18. Liu, C., Wechsler, H.: Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition. IEEE Trans. Image Process. **11**(4), 467–476 (2002)
19. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cogn. Neurosci. **3**(1), 71–86 (1991)
20. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces versus Fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720 (1997)
21. Lei, Z., Liao, S., Pietikäinen, M., Li, S.Z.: Face recognition by exploring information jointly in space, scale and orientation. IEEE Trans. Image Process. **20**(1), 247–256 (2011)
22. Çevik, N., Çevik, T.: A novel high-performance holistic descriptor for face retrieval. Pattern Anal. Appl. (2019). https://doi.org/10.1007/s10044-019-00803-5
23. Comon, P.: Independent component analysis—a new concept? Signal Process. **36**, 287–314 (1994)
24. Haralick, R., Shanmugan, K., Dinstein, I.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. **3**, 610–621 (1973)
25. Jutten, C., Herault, J.: Blind separation of sources, part I: an adaptive algorithm based on neuromimatic architecture. Signal Process. **24**(1), 1–10 (1991)
26. Champion, I., Germain, C., Da Costa, J.-P., Alborini, A., Dubois-Fernandez, P.: Retrieval of forest stand age from SAR image texture for varying distance and orientation values of the grey level co-occurrence matrix. IEEE Geosci. Remote Sens. Lett. **11**(1), 5–9 (2014)
27. Fan, D.P., et al.: Scoot: a perceptual metric for facial sketches. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, Korea (2019)
28. Ou, X., Pan, W., Xiao, P.: In vivo skin capacitive imaging analysis by using grey level co-occurrence matrix (GLCM). Int. J. Pharm. **460**(1–2), 28–32 (2014)

29. Adur, J., Carvalho, H.F., Cesar, C.L.: Nonlinear optical microscopy signal processing strategies in cancer. Cancer Inform. **13**(13), 67–76 (2014)

30. Ahonen, T., Hadid, A., Pietikainen, M.: Face recognition with local binary patterns. In: Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic, pp. 469–481 (2004)

31. Zhang, W.C., Shan, S.G., Gao, W., et al.: Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition. In: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, pp. 786–791 (2005)

32. Heikkilä, M., Pietikäinen, M., Schmid, C.: Description of interest regions with local binary patterns. Pattern Recognit. **42**(3), 425–436 (2009)

33. Jabid, T., Kabir, M.H., Chae, O.: Robust facial expression recognition based on local directional pattern. ETRI J. **32**(5), 784–794 (2010)

34. Dan, Z., Chen, Y., Yang, Z., et al.: An improved local binary pattern for texture classification. Optik **125**, 6320–6324 (2014)

35. Qian, X., Hua, X.-S., Chen, P., et al.: PLBP: an effective local binary patterns texture descriptor with pyramid representation. Pattern Recognit. **44**, 2502–2515 (2011)

36. Chakraborty, S., Singh, S.K., Chakraborty, P.: Local directional gradient pattern: a local descriptor for face recognition. Multimed. Tools Appl. **76**, 1201–1216 (2017)

37. Yang, S., Bhanu, B.: Facial expression recognition using emotion avatar image. In: Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition, Santa Barbara, USA, pp. 866–871 (2011)

38. Rivera, A.R., Castillo, R., Chae, O.: Local directional number pattern for face analysis: face and expression recognition. IEEE Trans. Image Process. **22**(5), 1740–1752 (2013)

39. Dahmane, M., Meunier, J.: Emotion recognition using dynamic gridbased HoG features. In: Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition, Santa Barbara, USA, pp. 884–888 (2011)

40. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans. Image Process. **19**(6), 1635–1650 (2010)

41. Yin, Q.B., Kim, J.N.: Rotation-invariant texture classification using circular Gabor wavelets based local and global features. Chin. J. Electron. **17**(4), 646–648 (2008)

42. Melendez, J., Garcia, M.A., Puig, D.: Efficient distance-based per-pixel texture classification with Gabor wavelet filters. Pattern Anal. Appl. **11**(3), 365–372 (2008)

43. Jafari-Khouzani, K., Soltanian-Zadeh, H.: Radon trans-form orientation estimation for rotation invariant texture analysis. IEEE Trans. Pattern Anal. Mach. Intell. **27**(6), 1004–1008 (2005)

44. Varma, M., Zisserman, A.: A statistical approach to material classification using image patch exemplars. IEEE Trans. Pattern Anal. Mach. Intell. **31**(11), 2032–2047 (2009)

45. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using local affine regions. IEEE Trans. Pattern Anal. Mach. Intell. **27**(8), 1265–1278 (2005)

46. Zhang, B., Shan, S., Chen, X., Gao, W.: Histogram of gabor phase patterns (hgpp): a novel object representation approach for face recognition. IEEE Trans. Image Process. **16**(1), 57–68 (2007)

47. Doshi, N., Schaefer, G.: A comprehensive bench-mark of local binary pattern algorithms for texture retrieval. In: Proceedings of the International Conference on Pattern Recognition (ICPR), pp. 2760–2763 (2012)

48. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recognit. **29**(1), 51–59 (1996)

49. Pietikäinen, M., Hadid, A., Zhao, G., Ahonen, T.: Computer Vision Using Local Binary Patterns. Springer, Berlin (2011)

50. Murphy-Chutorian, E., Trivedi, M.: Head pose estimation in computer vision: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **31**(4), 607–626 (2009)

51. Wang, K., Wu, Y., Ji, Q.: Head pose estimation on low-quality images. In: Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition, Xi'an, China, 15–19 May 2018

52. Pantic, M., Rothkrantz, L.J.M.: Automatic analysis of facial expressions: the state of the art. IEEE Trans. Pattern Anal. Mach. Intell. **22**(12), 1424–1445 (2000)

53. Munasinghe, M.I.N.P.: Facial expression recognition using facial landmarks and random forest classifier. In: Proceedings of the IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), Singapore, Singapore, 6–8 June 2018

54. Hansen, D.W., Ji, Q.: In the eye of the beholder: a survey of models for eyes and gaze. IEEE Trans. Pattern Anal. Mach. Intell. **32**(3), 478–500 (2010)

55. Wu, Y., Ji, Q.: Facial landmark detection: a literature survey. Int. J. Comput. Vis. **127**(2), 115–142 (2018)

56. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Trans. Pattern Anal. Mach. Intell. **23**(6), 681–685 (2001)

57. Tzimiropoulos, G., Pantic, M.: Optimization problems for fast AAM fitting in-the-wild. In: Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 Dec. 2013, pp. 593–600

58. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012, pp. 2879–2886

59. Tzimiropoulos, G., Pantic, M.: Gauss–Newton deformable part models for face alignment in-the-wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 23–28 June 2014, pp. 1851–1858

60. Cao, X., Wei, Y., Wen, F., Sun, J.: Face alignment by explicit shape regression. Int. J. Comput. Vis. **107**(2), 177–190 (2014)

61. Zhang, H., Li, Q., Sun, Z., Liu, Y.: Combining data-driven and model-driven methods for robust facial landmark detection. IEEE Trans. Inf. Forensics Secur. **13**(10), 2409–2422 (2018)

62. Cootes, T.F., Taylor, C.J.: Active shape models—'Smart snakes'. In: Proceedings of the British Machine Vision Conference, Leeds, UK, 22–24 September 1992, pp. 266–275

63. Cristinacce, D., Cootes, T.F.: Feature detection and tracking with constrained local models. In: Proceedings of the 17th British Machine Vision Conference, Edinburgh, UK, 4–7 September 2006

64. Saragih, J.M., Lucey, S., Cohn, J.F.: Deformable model fitting by regularized landmark mean-shift. Int. J. Comput. Vis. **91**(2), 200–215 (2011)

65. Yu, X., Huang, J., Zhang, S., Yan, W., Metaxas, D.N.: Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model. In: Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013, pp. 1944–1951

66. Asthana, A., Zafeiriou, S., Cheng, S., Pantic, M.: Robust discriminative response map fitting with constrained local models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013, pp. 3444–3451

67. Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., Zhao, D.: The CAS-PEAL large-scale chinese face database and baseline evaluations. IEEE Trans. Syst. Man Cybern. (Part A) **38**(1), 149–161 (2008)

68. Murala, S., Maheshwari, R.P., Balasubramanian, R.: Local tetra patterns: a new feature descriptor for content-based image retrieval. IEEE Trans. Image Process. **21**(5), 2874–2886 (2012)

69. Mohammad, T., Ali, M.L.: Robust facial expression recognition based on local monotonic pattern (LMP). In: 14th International

Conference on Computer and Information Technology (ICCIT), IEEE, Dhaka, Bangladesh, pp. 572–576 (2011)

70. Liu, S., Zhang, Y., Liu, K.: Facial expression recognition under partial occlusion based on Weber Local Descriptor histogram and decision fusion. In: Proceedings of the 33rd Chinese Control Conference (CCC), Nanjing, China, pp. 4664–4668 (2014)

71. Islam, M.S.: Local gradient pattern—a novel feature representation for facial expression recognition. J. AI Data Min. **2**, 33–38 (2014)

72. Hafiane, A., Seetharaman, G., Zavidovique, B.: Median binary pattern for textures classification. In: Proceedings of the International Conference on Image Analysis and Recognition, Lecture Notes in Computer Science, vol. 4633, pp. 387–398. Springer, Berlin (2007)

73. Islam, M.S., Auwatanamongkol, S.: Facial expression recognition using local arc pattern. Trends Appl. Sci. Res. **9**, 113–120 (2014)

74. Yang, M., Zhang, L., Shiu, S.C.K., Zhang, D.: Monogenic binary coding: an efficient local feature extraction approach to face recognition. IEEE Trans. Inf. Forensics Secur. **7**(6), 1738–1751 (2012)

75. Hosseini, H., Marvasti, F.: Fast restoration of natural images corrupted by high-density impulse noise. EURASIP J. Image Video Process. **15**, 1–7 (2013)

76. Ng, P.E., Ma, K.K.: A switching median filter with boundary discriminative noise detection for extremely corrupted images. IEEE Trans. Image Process. **15**(6), 1506–1516 (2006)

77. Luisier, F.: Image denoising in mixed Poisson–Gaussian noise. IEEE Trans. Image Process. **20**(3), 696–708 (2011)

78. Barbu, T.: Variational image denoising approach with diffusion porous media flow. In: Abstract and Applied Analysis, pp. 1–8 (2013). https://doi.org/10.1155/2013/856876

79. Zhou, S.K., Chellappa, R., Zhao, W.: Unconstrained Face Recognition. Springer, New York (2005)

80. Choi, Sang-Il: Face recognition under illumination variation using shadow compensation and pixel selection. Int. J. Adv. Robot. Syst. **9**(130), 1–7 (2012)

81. Jaya Mohan, C., Deepak, M.S., Alphin Ezhil Manuel, M.L., Joy Winnie Wise, D.C.: Face recognition under expressions and lighting variations using artificial intelligence and image synthesizing. Int. J. Comput. Sci. Netw. Secur. (IJCSNS) **15**(9), 59–63 (2015)

82. Min, R., Hadid, A., Dugelay, J.L.: Improving the recognition of faces occluded by facial accessories. In: Proceedings of the International Conference on Automatic Face and Gesture Recognition, Santa Barbara, CA, USA, pp. 442–447 (2011)

83. Penev, P., Atick, J.: Local feature analysis: a general statistical theory for object representation. Netw. Comput. Neural Syst. **7**(3), 477–500 (1996)

84. Martínez, A.M.: Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. IEEE Trans. Pattern Anal. Mach. Intell. **24**(6), 748–763 (2002)

85. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: illumination cone models for face recognition under variable lighting and pose. In: PAMI (2001)

86. https://cswww.essex.ac.uk/mv/allfaces/faces94.html

87. Lyons, M.J., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with gabor wavelets. In: Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200–205, Nara, Japan (1998)

88. Lyons, M.J., Budynek, J., Akamatsu, S.: Automatic classification of single facial images. IEEE Trans. Pattern Anal. Mach. Intell. **21**(12), 1357–1362 (1999)

89. Dailey, M.N., Joyce, C., Lyons, M.J., Kamachi, M., Ishi, H., Gyoba, J., Cottrell, G.W.: Evidence and a computational explanation of cultural differences in facial expression recognition. Emotion **10**(6), 874–893 (2010)

Taner Cevik received the B.Sc. degree in computer engineering from Istanbul Technical University, Istanbul, in 2001 and completed Ph.D. at Istanbul University in 2012. He joined the Software Engineering Department at Aydin University in 2017 and continues to work as an Associate Professor. His research interests are image processing and wireless multimedia sensor networks. Mailing Address: Ataturk Mahallesi, Gazi Mustafa Kemal Caddesi Deran Sokak No: 6/5 Büyükçekmece Istanbul Turkey.

Fatih Sahin received the B.Sc. degree in Electronic Communication Engineering from Military Academy, Ankara, in 2000 and obtained his Master of Science degree in Computer Engineering from Sakarya University, Sakarya, in 2014. He pursues his Ph.D. at Istanbul Aydin University in Istanbul. His research interests are cryptography and image processing. Mailing Address: Nisbetiye Mh. Binbaşı Nazim Kaptan Sok. Jandarma Lojmanlari E-Blok no:15 Besiktas, Istanbul.