SIViP (2016) 10:129–137
DOI 10.1007/s11760-014-0712-x

ORIGINAL PAPER

Robust multimodal 2D and 3D face authentication using local feature fusion

A. Ouamane · M. Belahcene · A. Benakcha · S. Bourennane · A. Taleb-Ahmed

Received: 8 October 2013 / Revised: 17 October 2014 / Accepted: 18 October 2014 / Published online: 16 November 2014

Abstract In this work, we present a robust face authentication approach merging multiple descriptors and exploiting both 3D and 2D information. First, we correct the heads rotation in 3D by iterative closest point algorithm, followed by an efficient preprocessing phase. Then, we extract different features namely: multi-scale local binary patterns (MSLBP), novel statistical local features (SLF), Gabor wavelets, and scale invariant feature transform (SIFT). The principal component analysis followed by enhanced fisher linear discriminant model is used for dimensionality reduction and classification. Finally, fusion at the score level is carried out using two-class support vector machines. Extensive experiments are conducted on the CASIA 3D faces database. The evaluation of individual descriptors clearly showed the superiority of the proposed SLF features. In addition, applying the (3D + 2D) multimodal score level fusion, the best result is obtained by combining the SLF with the MSLBP + SIFT descriptor yielding in an equal error rate of 0.98 % and a recognition rate of RR = 97.22 %.

A. Ouamane (✉) · M. Belahcene
LMSE, Universite Mohamed Khider Biskra, Biskra, Algeria
e-mail: ouamaneabdealmalik@yahoo.fr

M. Belahcene
e-mail: belahcene_mebnef@yahoo.fr

A. Benakcha
LGEB, Universite Mohamed Khider Biskra, Biskra, Algeria
e-mail: ab.benakcha@gmail.com

S. Bourennane
Institut Fresnel, Universite de Marseille, Marseille, France
e-mail: salah.bourennane@fresnel.fr

A. Taleb-Ahmed
LAMIH UMR CNRS 8201, Valenciennes, France
e-mail: abdelmalik.taleb-ahmed@univ-valenciennes.fr

Keywords Iterative closest point · Multi-scale local binary patterns (MSLBP) · Statistical local features (SLF) · Scale invariant feature transform (SIFT) · Fusion

1 Introduction

A significant number of feature extraction approaches are proposed in literature to represent face images. The existing face recognition approaches can broadly be classified into two categories, global and local [1]. Global face recognition methods are usually based on statistical approaches in which features are extracted from the entire face image. Among global methods, principal component analysis (PCA) [2], fisher linear discriminant (FLD) [2], independent component analysis (ICA) [3], the space–frequency techniques such as Fourier transform [4]. Although the global face recognition techniques are the most common in face recognition, recently, lots of work is being done on local feature extraction methods as these are considered as more robust against variations in facial expressions, noise, and occlusion. These structure-based approaches deal with local information related to some interior parts of face images. Among the sparse descriptors, the scale invariant feature transform (SIFT) [5], Gabor wavelet [6], and local binary patterns (LBP) [7]. Early face recognition research was based on 2D appearance images [8]. However, an increasing number of 3D-shape-based face recognition algorithms have recently emerged with the advent of 3D scanners. Although the appearance of a face in a 2D image encodes the shape information of the face, aside from the face albedo, 2D face recognition alone has not been able to achieve the desired accuracy because of its sensitivity to illumination, pose, and expression variations. On the other hand, 3D face recognition can better handle pose variations (particularly in depth

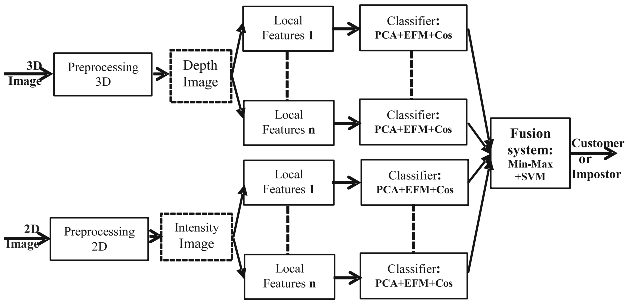

Fig. 1 Overview of the proposed framework

rotations) and the 3D data can be used to correct the pose of the corresponding 2D image (texture) as well [9].

1.1 Overview of our proposed approach

In this paper, we investigate how local features of 3D and 2D information contribute to face recognition when illumination, expressions and combined changes in expression under illumination are taken into account. All processes included in our training and test steps are fully automated. Our system, as illustrated in Fig. 1, includes four important steps which consist in:

1. Preprocessing: By translating and rotating one input 3D image to align one reference 3D image, face poses, and changed positions between the face and the equipment are normalized.
2. Feature extraction: Robust feature representation is very important to the whole system. It is expected that these features are invariant to rotation, scale, expression, and illumination. The existing work usually uses raw depth and intensity features. In our system, we combine one global feature and four local features.
3. Classification: PCA combined with enhanced fisher linear discriminant model (EFM) are used for reducing the dimensional space. Classification is performed using the normalized correlation metric.
4. Fusion system: It consists in the fusion of the classification results by support vector machines (SVM) method [10], and the score normalization is performed using Min_Max method [11] upstream is chosen for its simplicity.

1.2 Contributions of this paper

In this paper, we propose a new scheme to combine several methods of local feature extraction from depth and intensity images to overcome the problems due to illumination, expressions, and combined changes in expression under illumination. The main contributions of this paper are as follows:

1. A novel feature extraction method (statistical local features) is proposed. It is based on the calculation of statistical parameters in a neighborhood of the pixel such as the average, standard deviation, variance.
2. Study the fusion of two multimodal systems (multi-algorithms: built by the fusion of several local characteristics and multi-sensor: built by the fusion of 2D and 3D information).
3. Studying several feature extraction methods to gain insights into their complementarity.

The remainder of this paper is organized as follows. Section 2 introduces the preprocessing procedure, which is very important to robust recognition. Section 3 describes the features for face representation. Section 4 reports the experimental results. Finally, Section. 5 summarizes this paper.

2 Preprocessing

It is assumed in this paper that one face is described by one 3D point cloud captured by one 3D laser scanner. Each point cloud consists of thousands of points in the 3D space. These discrete points approximately describe the face surface. We use CASIA 3D face database. Each point is described with 3D spatial coordinates and corresponding RGB color coordinates. In this section, we describe how the original 3D data are preprocessed. That is, we exactly register the data and then obtain the depth and intensity images. This part prepares for the feature extraction in the next section. This preprocessing includes two main steps, registration of 3D face surfaces and acquisition of depth and intensity images.

2.1 Registration

We use iterative closest point (ICP) [12]. ICP has two attributions to this method. Firstly, it aligns all the faces with the first 3D face (neutral expression). Secondly, it examines whether the detected nose tip is correct.

2.2 Depth and intensity images

Depth and intensity images are obtained from registered 3D data. The data are converted into a 3D image depth (see Fig. 2a) and a color image (see Fig. 3a). In most images, the nose is the closest part of the face in 3D scanner; it has the highest value in depth between all points of the face. For each pixel, the average is calculated using the neighboring window of size 9×9 around it. Using a window of size 3×3 which calculates the sum of the depth of its corresponding pixels, the nose is detected as the coordinates of the center pixel of the window that returns the maximum value (see Fig. 2b). After detecting the nose, a sub-image centered on the center

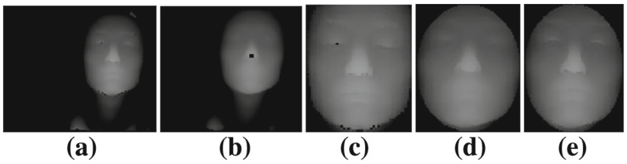

Fig. 2 Preprocessing of the depth image: **a** the depth image, **b** detecting the nose tip, **c** extracted sub-image, **d** mean image 5 × 5, **e** depth image after removing noise and filling holes

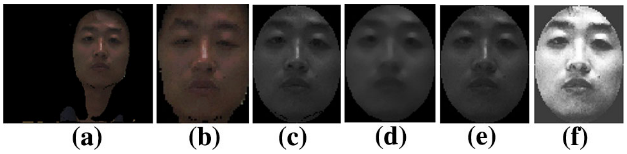

Fig. 3 Preprocessing of the intensity image: **a** the color image, **b** extracted sub-image, **c** intensity image, **d** mean image 5 × 5, **e** intensity image after removing noise and patching holes, **f** intensity image after histogram equalization

of the nose, with size 57 × 47, is extracted (see Figs. 2c, 3b). For RGB color images, we used the intensity images (see Fig. 3c). However, due to the quality of original 3D data, the depth and intensity images usually contain much noise, such as holes and outliers. We can obtain enhanced images by the following processes. The preprocessing of depth images includes noise removal and hole filling. We use the following scheme to remove the outliers. For each pixel, the mean is computed for the 5 × 5 neighboring window (see Figs. 2d, 3d). If the pixel intensity is less than a given threshold, this pixel is replaced by the mean pixel. The result is shown in Figs. 2e, 3e. The variation in the lighting strongly influences the presentation of the intensity images. To cope with this problem, histogram equalization is used to reduce the influence of the illumination variations (see Fig. 3f).

3 Features for face representation

3.1 Multi-scale local binary patterns (MSLBP)

The original LBP operator was later generalized to deal with different neighborhoods [13]. A local neighborhood is defined as a set of sampling points evenly spaced on a circle which is centered at the pixel to be labeled, and the sampling points that do not fall within the pixels are interpolated using bilinear interpolation, thereby allowing for any radius and any number of sampling points in the neighborhood. Formally, given a pixel at (x_c, y_c), the resulting LBP can be expressed in decimal form as:

$$\mathrm{LBP}_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(i_p - i_c)2^p \tag{1}$$

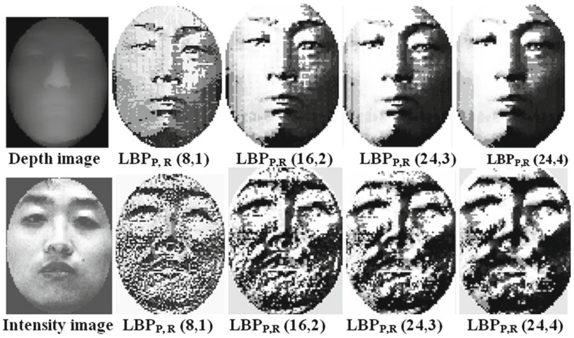

Fig. 4 The multi-scale LBP for facial depth and intensity images

where i_c and i_p are respectively gray-level values of the central pixel and P surrounding pixels in the circle neighborhood with a radius R, and function $s(x)$ is defined as:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x \prec 0 \end{cases} \tag{2}$$

Given a facial depth and intensity images, we generate a set of multi-scale LBP for facial representation. Some examples are illustrated in Fig. 4. In this figure, the number of sampling points varies from 8 to 24 points, and the radius value varies from 1 to 4 pixels.

3.2 Proposed statistical local features (SLF)

The main purpose of the proposed method is to compute some statistical parameters in the neighborhood of the pixel using different sizes and number of neighboring points. The calculated parameters are:

3.2.1 The mean

It is defined as:

$$\mathrm{mean}_{P,R}(x_c, y_c) = \frac{1}{P}\sum_{p=0}^{P-1} i_p \tag{3}$$

where i_c and i_p are respectively gray-level values of the central pixel and P surrounding pixels in the circle neighborhood with a radius R.

3.2.2 Standard deviation

Standard deviation shows how much variation exists from the average. It is defined as:

$$\mathrm{std}_{P,R}(x_c, y_c) = \sqrt{\frac{1}{P}\sum_{p=0}^{P-1}(i_p - \mathrm{mean}_{P,R}(x_c, y_c))^2} \tag{4}$$

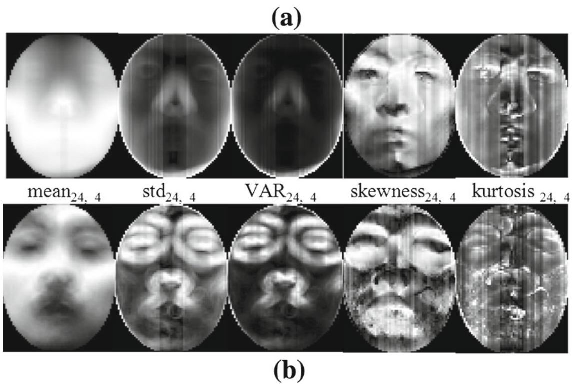

Fig. 5 The statistic local features (SLF), **a** facial depth image, **b** intensity image

3.2.3 Variance

The variance is a measure of how far a set of numbers is spread out. It is one of several descriptors of a probability distribution, describing how far the numbers lie from the mean (expected value). It is defined as:

$$\mathrm{VAR}_{P,R}(x_c, y_c) = \frac{1}{P}\sum_{p=0}^{P-1}(i_p - \mathrm{mean}_{P,R}(x_c, y_c))^2 \tag{5}$$

3.2.4 Skewness

Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the central point. It is defined as:

$$\mathrm{skew}_{P,R}(x_c, y_c) = \frac{\frac{1}{P}\sum_{p=0}^{P-1}(i_p - \mathrm{mean}_{P,R}(x_c, y_c))^3}{\left(\sqrt{\frac{1}{P}\sum_{p=0}^{P-1}(i_p - \mathrm{mean}_{P,R}(x_c, y_c))^2}\right)^{3/2}} \tag{6}$$

3.2.5 Kurtosis

Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. It is defined as:

$$\mathrm{kur}_{P,R}(x_c, y_c) = \frac{\frac{1}{P}\sum_{p=0}^{P-1}(i_p - \mathrm{mean}_{P,R}(x_c, y_c))^4}{\left(\sqrt{\frac{1}{P}\sum_{p=0}^{P-1}(i_p - \mathrm{mean}_{P,R}(x_c, y_c))^2}\right)^{2}} \tag{7}$$

Some examples for SLF features are illustrated in Fig. 5.

3.3 Overview of Gabor wavelet filters

In this paper, we use 2D Gabor filters of depth and intensity images to characterize a person. The Gabor wavelets

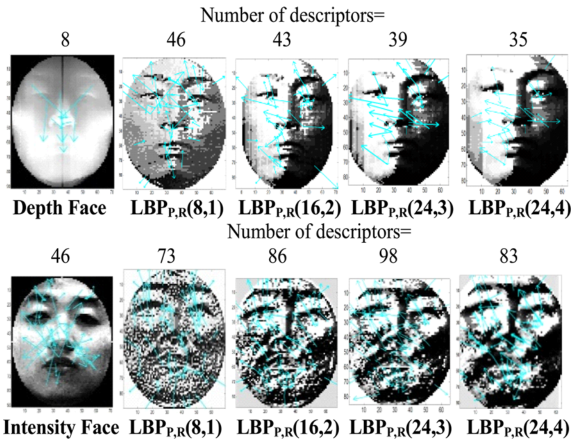

Fig. 6 The SIFT-based keypoints detected from an original depth, intensity facial image, and four associated $\mathrm{LBP}_{P,R}$

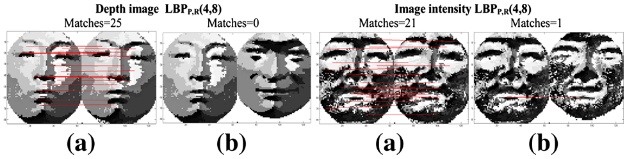

Fig. 7 SIFT matches between learning and evaluation faces belonging to **a** the same identity and **b** different identities

represent the properties of spatial localization, orientations, and spatial frequency selectivity. The representation of faces using Gabor wavelet has been successfully used in 2D and 3D face recognition [14]. This representation of an image describes the facial characteristics of both the spatial frequency and spatial relations.

3.4 Scale invariant feature transform (SIFT)

The SIFT [5] is a local 2D feature calculated at keypoint locations. The interested reader is referred to Lowes paper [5] for the details of the keypoint localization and the SIFT feature extraction.The SIFT operator works on each $\mathrm{LBP}_{P,R}$ separately. Because $\mathrm{LBP}_{P,R}$ highlight the local characteristics of smooth facial image depth and intensity. Many SIFT-based keypoints can be detected for the following step more than in the original images. Same statistical work was done along with the experiments on CASIA 3D faces database. The average number of descriptors extracted from each of $\mathrm{LBP}_{P,R}$ depth is 52 and $\mathrm{LBP}_{P,R}$ intensity is 162 while that of each original facial depth image is limited to 14, and intensity is limited to 63. The SIFT descriptors for these faces were then computed using Lowes code [15]. Figure 6 shows the SIFT-based keypoints extracted from one range and intensity face image and its four associated $\mathrm{LBP}_{P,R}$. To calculate the similarity between a learning and evaluation face, their SIFT descriptors were matched using the Euclidean distance (see Fig. 7).

4 Experimental results

4.1 The CASIA 3D database

We use the CASIA 3D face database [16] to test our proposed authentication system. The basis is constructed by a 3D scanner Minolta VIVID 910 non-contact working in the fast mode. This database contains 123 subjects, each subject having 37 or 38 images with individual variations of poses, expression, illumination, and combined changes in expression under illumination and pose as expressions. This database contains complex variations that are difficult to any algorithm. In this section, we studied the variations of illumination (images: 1, 2, 3, 4, 5), expressions (images: 6, 7, 8, 9, 10) and the combined changes in expression under illumination (images: 11, 12, 13, 14, 15). Therefore, we used 15 images for each subject. We used an assessment protocol of separating people into two classes, client and impostor. Customer group contains 100 subjects, while the impostor group is divided into 13 impostors for evaluation and 10 for testing. The repartition of images in different sets is given in Table 1.

Table 1 Distribution of photos in different sets

Together	Customer	Impostor
Learning	500 images (1, 4, 8, 9, 10)	0 images
Evaluation	500 images (2, 6, 7, 14, 15)	195 images (1:15)
Test	400 images (3, 5, 11, 12, 13)	150 images (1:15)

4.2 Global feature (PCA + EFM)

For this part, we use a holistic approach. The characteristic vector of 2D and 3D image is built by concatenation of rows of depth and intensity image. We use PCA + EFM method for reduction and separation of space and normalized correlation for similarity measure. Table 2 shows the error rate in the entire evaluation and testing for a comprehensive approach (PCA + EFM) feature extraction. (EER: equal error rate and RR: the recognition rate (RR $= 100 - \text{FRR} - \text{FAR}$). FRR: the false reject rate and FAR: the false accept rate, FN: feature number is the number of feature extracted by enhanced fisher linear discriminant model (EFM). It is generally computed experimentally [11]. We vary FN from 10, 20, ..., 200, and then, we select the one which gives the best result. P: number of points in the neighborhood pixel, R: radius). The table shows that PCA + EFM gives a poor performance for depth information (3D). Indeed, the RR criterion for instance is 89.36 %, whereas the RR is 93.14 % when relative intensity information(2D) and multimodal fusion (3D and 2D) are used.

4.3 Multi-scale LBP (MSLBP)

For this part, we use the MSLBP local method. Table 3 shows the error rate in all tests and evaluation by this method of feature extraction. The number of sampling points varies from 8 points to 24 points, and the radius value varies from 1 pixel to 4 pixels. LBP method gives better results for 3D information in the case of 2D information for the four radius values (R). The fusion of four radius values (MSLBP) improves per-

Table 2 Performance of the PCA + EFM throughout evaluation and test set

3D image					2D image					3D and 2D image			
Eval	Test				Eval	Test				Eval	Test		
EER	FAR	FRR	RR	FN	EER	FAR	FRR	RR	FN	EER	FAR	FRR	RR
7.24	4.24	6.4	89.36	10	3.36	2.94	4.8	92.26	30	**2.61**	**2.61**	**2.86**	**93.14**

Best performance is indicated in bold

Table 3 Performance of the multi-scale LBP method throughout evaluation and test set

(P,R)	3D image					2D image					3D and 2D image			
	Eval	Test				Eval	Test				Eval	Test		
	EER	FAR	FRR	RR	FN	EER	FAR	FRA	RR	FN	EER	FAR	FRA	RR
1-(1,8)	4.74	3.11	5.2	91.68	50	5.90	6.93	7.4	85.66	60	3.82	4.26	3.6	92.14
2-(16,2)	6	5.79	3.8	90.40	50	4.22	5.41	5.2	89.38	100	3.56	4.08	3	92.92
3-(24,3)	6.17	5.79	4	90.2	50	4.37	5.46	4.8	90.04	70	3.79	4.08	2.9	93.12
4-(24,4)	6.17	5.52	4.2	90.27	50	4.17	6.58	4.40	89.02	100	3.61	4.35	3.2	92.44
MSLBP	5.36	4.63	3.8	91.56	–	4.62	4.98	4.40	90.61	–	**0.95**	**1.25**	**4.2**	**94.54**

Best performance is indicated in bold

Table 4 Performance of the SLF throughout evaluation and test set

	(P,R)	3D image					2D image					3D and 2D image			
		Eval	Test				Eva	Test				Eval	Test		
		EER	FAR	FRR	RR	FN	EER	FAR	FRR	RR	FN	EER	FAR	FRR	RR
1-mean	(8,1)	6	3.89	5.4	90.7	10	3.36	2.73	4.4	92.86	50	2.17	1.74	3.6	94.66
	(16,2)	5.56	4.12	4.8	91.07	10	3.39	2.49	4.4	93.10	50	2.17	1.79	3	95.20
	(24,3)	5.20	4.78	4.8	90.41	10	3.23	1.99	5.2	92.8	30	**1.95**	**0.96**	**3.6**	**95.43**
	(24,4)	6	4.03	5.2	90.76	10	3.58	2.28	5.4	92.31	20	2.42	1.05	4.2	94.74
	MS	5.18	3.96	5	91.04	–	3.18	2.41	3.8	93.78	–	1.3	0.68	4.2	95.11
2-Std	(8,1)	11.59	14.39	10.4	75.2	30	3.76	3.3	4.8	91.89	60	3.57	3.36	4.2	92.16
	(16,2)	10.44	8.17	8.4	83.42	10	2.84	2.37	4.2	93.42	70	2.76	2.19	3.8	94.00
	(24,3)	9.24	7.79	7.6	84.6	20	3	2.13	3.2	94.66	80	1.96	2.4	2.8	94.80
	(24,4)	8.22	6.19	6.6	87.2	30	3.15	1.7	3.2	95.09	90	**3.36**	**1.69**	**3**	**95.3**
	MS	5.63	4.86	7.4	87.74	–	3.04	2.41	2.6	94.98	–	1.63	1.26	3.6	95.13
3-Var	(8,1)	11.8	8.62	11	80.38	20	5.61	4.76	7	88.23	70	4.62	4.22	4.6	91.17
	(16,2)	10.42	9.66	9.4	80.93	30	3.76	3.42	4.2	92.37	70	2.98	3.2	3.4	93.40
	(24,3)	11.19	11.34	10.2	78.46	40	2.97	2.48	3.4	94.11	70	**2.81**	**2.68**	**3.2**	**94.11**
	(24,4)	10.98	13.17	8.6	78.22	30	4.56	3.71	4	92.28	50	3.61	4.18	3.2	92.61
	MS	10.21	10.24	8.2	81.55	–	3.36	2.84	3.8	93.35	–	1.4	1.6	4.4	93.99
4-Skew	(8,1)	11.80	5.29	5.4	89.30	80	5.61	4.26	5.2	90.53	80	3.79	3.58	4.4	92.02
	(16,2)	5.39	4.54	4.4	91.05	50	3.78	3.88	5.6	90.51	60	3.81	3.45	3.6	92.94
	(24,3)	4.82	4.46	4.8	90.74	60	3.38	3.3	4	92.70	90	3.40	3.2	3.2	93.60
	(24,4)	4.61	3.03	4.2	92.76	50	3	3.23	3.6	93.16	50	**2.83**	**2.38**	**2.4**	**95.21**
	MS	5.16	4.28	4.4	91.31	–	2.96	3.06	4	92.94	–	1.61	1.72	3.8	94.48
5-Kur	(8,1)	14.84	14.08	11	74.91	20	18.78	20.33	18.6	61.06	20	12.98	13.51	14.20	72.28
	(16,2)	15.64	18.82	13.8	67.37	20	11.36	16.26	14.4	69.33	20	8.59	12.38	13.40	74.22
	(24,3)	20.44	23.30	20.6	56.10	20	10.18	15.46	12.4	72.13	40	9.16	14.74	13.40	71.86
	(24,4)	13.56	14.20	11.6	74.20	10	6.6	8.4	9	82.59	50	**5.79**	**7.16**	**9.40**	**83.43**
	MS	11.59	13.28	10.6	76.12	–	4.76	7.11	15	77.88	/	5.16	7.24	8.40	84.35
1+2+3	+4+5	3.6	2.06	3.80	94.13	–	4.98	2.14	2.6	95.25	/	0.99	0.80	4.00	95.20
1+2+	3+4	3.24	1.87	3.60	94.52	–	5.04	2.03	2.6	95.36	/	**1.2**	**0.88**	**2.80**	**96.32**

Best performance is indicated in bold

formance for 2D and 3D information and the multi-sensor fusion (3D + 2D) with an EER = 0.95 % overall evaluation and RR = 94.54 % in the test set.

4.4 Statistical local features (SLF)

For this part, we use the proposed local method (SLF). Table 4 shows the error rate in every test and evaluation for SLF method. The number of sampling points varies from 8 points to 24 points, and the radius value varies from 1 pixel to 4 pixels. First, the fusion of the four radii ($R = 1, 2, 3, 4$) for different neighborhoods does not improve performance for five statistical descriptors available. We also notice that for $R = 3$, $R = 4$ and the number of points $P = 24$ (maximum neighborhood size in our application), we obtain a better result for all statistical descriptors. Therefore, the increase of the number points in the vicinity improves the performance in the case of statistical descriptors. The four descriptors (mean, standard deviation, variance, skewness) give almost the same results. Kurtosis is the worst descriptor. It confirms the results of visual perception (image quality) obtained in images $kurtosis_{24,4}$ (see Fig. 6) The fusion of the five parameters of our local features improves face authentication. Performance without the kurtosis is better than the fusion of five statistical parameters with EER = 1.20 % overall evaluation and RR = 96.32 % in the test set.

4.5 Gabor wavelets

The family of Gabor filters is characterized by a number of resolutions or frequencies and orientations. In this work, we concatenated for each resolution the eight directions in the feature vector. Gabor filters have a complex shape that can be exploited. It is important to use the information provided by

Table 5 Performance of the LBP$_{P,R}$ + SIFT descriptor throughout evaluation and test set

(P,R)	3D image				2D image				3D and 2D image			
	Eval	Test			Eval	Test			Eval	Test		
	EER	FAR	FRR	RR	EER	FAR	FRR	RR	EER	FAR	FRR	RR
1- (8,1)	6.73	5.96	5.60	88.44	8.97	8.33	10	81.67	4.30	2.56	5	92.44
2-(16,2)	6.30	3.11	7.00	89.89	6.09	5.33	7.2	87.47	5.31	5.62	3	91.37
3-(24,3)	7.25	4.9	6.40	88.7	4.18	3.34	5	91.66	3.43	0.64	5	91.66
4-(24,4)	6.34	4.89	4.60	90.51	3.7	5.11	5.4	89.49	2.85	2.3	4.2	93.50
1+2+3+4	4.67	2.41	4.6	92.98	2.91	2.46	4.6	92.94	**2.48**	**2.26**	**3.00**	**94.73**

Best performance is indicated in bold

Table 6 Performance of five methods of feature extraction throughout evaluation and test set

Feature	3D image				2D image				3D and 2D image				
Extraction	Eval	Test			Eval	Test			Eval	Test			
	EER	FAR	FRR	RR	EER	FAR	FRR	RR	EER	FAR	FRR	RR	Runtime (s)
1- PCA+EFM	7.24	4.24	6.4	89.36	3.36	2.94	4.80	92.26	2.61	2.86	4	93.14	0.371
2-MSLBP	5.36	4.63	3.80	91.56	4.62	4.98	4.40	90.61	0.95	1.25	4.20	94.54	0.374
3-SLF	3.24	1.87	3.60	94.52	5.04	2.03	2.60	95.36	**1.20**	**0.88**	**2.80**	**96.32**	**0.399**
4-Gabort	5.41	4.08	4.20	91.71	2.56	1.18	3.40	95.42	1.57	0.66	3.20	96.13	1.337
5-LBP+SIFT	4.67	2.41	4.60	92.98	2.91	2.46	4.60	92.94	2.48	2.26	3.00	94.73	1.084
3+2	2.55	2.02	3.00	94.97	2.18	1.03	3.40	95.56	1.16	0.76	2.80	96.43	0.472
3+1	3.55	2.22	3.40	94.38	2.61	0.86	3.80	95.33	1.35	1.08	2.60	96.31	0.469
3+4	3.18	2	3.00	95.09	4.64	2	2.40	95.68	1.20	0.76	2.60	96.64	1.435
3+5	2.39	1.74	2.00	96.26	3.56	1.89	2.40	95.70	**0.98**	**0.97**	**1.80**	**97.22**	**1.182**
3+5+4	2.35	1.61	1.60	96.78	2.37	1.78	2.80	95.41	1.00	0.76	2.20	97.03	2.22
3+5+2	1.81	0.94	4.20	94.86	1.63	0.74	3.20	96.06	0.96	0.69	2.20	97.10	1.255
3+5+1	2.38	1.60	2.60	95.79	2.04	0.96	3.60	95.44	0.84	0.75	2.80	96.44	1.252
1+2+3+4+5	1.64	0.87	4.00	95.12	1.58	0.68	3.40	95.92	0.77	0.56	3.00	96.43	2.358

Best performance is indicated in bold

the real part and the imaginary part of Gabor coefficients. We use the filtered phase responses of Gabor filters as in [11], we have shown that the filtered phase more relevant in this application. We use Gabor wavelets for each resolution and the fusion of five resolutions. The best results are obtained when the resolution is λ= 4, that is, EER = 1.57% and RR = 96.13%. The fusion of five resolutions does not improve.

4.6 MSLBP + SIFT

Table 5 shows the error rate in the entire evaluation and testing with the extraction of characteristics for LBP. The number of sampling points varies from 8 points to 24 points, and the radius value varies from 1 pixel to 4 pixels. Subsequently, SIFT is computed from the MSLBP 2D data. The table shows that the fusion of four $LBP(P, R)$ $(R = 1, 2, 3, 4$ and $P = 8, 16, 24)$ plus SIFT gives the best result with a EER = 2.48 % and RR = 94.73 %.

4.7 Fusion of feature representation

Table 6 compares the error rates, scores and computational load for the five considered descriptors and the fusion of these descriptors. Experiments were conducted on Matlab implementation on a Intel i5 2.50 GHz CPU processor with a 8 GB RAM. From this table, we can infer that: The SLF features give the best results with EER = 1.20 % overall evaluation and RR = 96.32 % in the test set and a low runtime equal 0.399 s. This shows the effectiveness of the proposed descriptor. Compared with all global and local descriptors studied, which justifies the effectiveness of our descriptor. The fusion of our proposed descriptor SLF with the descriptor MSLBP combined to SIFT gives the best results Indeed, we obtain an

Table 7 Comparison of recognition rate with state-of-art(CASIA databases)

Authors	image	RR
Xu et al. [14]	Gabor + a hierarchical selecting scheme embedded in LDA and AdaBoost learning	93.3
Wang et al. [17]	CPDM, Gabor, LBP, and PCA fusion	95.61
Ming et al. [18]	Robust sparse bounding sphere representation (RSBSR)	94
Ming [10]	Orthogonal spectral regression (ROSR)	96.13
Our method	SLF and MLBP +SIFT fusion	**97.22**

Best performance is indicated in bold

EER = 0.98 %, RR = 97.22 % and a runtime equal 1.182 s. The fusion of all considered descriptors does not improve the performance compared with the fusion of two descriptors SLF and MSLBP combined to SIFT. The performance of the proposed system is compared with the state-of-art in 3D and 2D face recognition for CASIA databases . The comparisons are based on the recognition rate as shown in Table 7. The results show that the proposed system achieves higher average recognition rate compared with the current systems in the literature tested using the same database.

5 Conclusion

In this work, we presented an automatic multimodal authentication algorithm based on 2D intensity and 3D depth face image. Firstly, we used a comprehensive approach based on the reduction of space PCA followed by EFM , secondly, by four local methods: MSLBP (based on the coding of a local neighborhood), SLF (based on the calculation of statistical parameters in a few neighborhood of the pixel), SIFT and Gabor wavelets. The application is carried out on the CASIA 3D database according to a protocol proposed for addressing major problems in the field of 3D facial recognition and multimodal, including: variations in illumination, expressions, variations combined in various expressions. From all the experiments carried out, we can say that:

- MSLBP is a good descriptor in the case of modality depth 3D versus 2D intensity modality, a significant improvement in performance is obtained by fusion of two modalities 3D and 2D.
- The best results are obtained by SLF descriptor if the neighborhood size becomes significant ($R = 3$, $R = 4$).
- The fusion of four LBP ($R = 1, 2, 3, 4$ and $P = 8, 16, 24$) + SIFT gives the best result with TEE = 2.85 % and RR = 93.50 %.
- For Gabor wavelets, the best result is obtained with the first resolution ($\lambda = 4$) with EER = 1.57 % and RR = 96.13 %. The fusion of the five resolutions will not improve performance.
- Local descriptor statistics SLF gives the best result compared with all global and local descriptors studied, which justifies the effectiveness of our proposed descriptor (we obtained TEE = 1.2 and RR = 96.32 %.
- The fusion of our proposed descriptor SLF with the MSLBP + SIFT descriptor gives the best result with EER = 2.39 % and RR = 96.26 % of the depth image (3D) and EER = 3.56 % and RR = 95.70 % for the intensity image (2D). Finally, the multi-fusion algorithms (3D + 2D) gives a TEE = 0.98 % and RR = 97.22 %.

Our approach is fully automatic and has been tested on different shifts, expressions and illuminations. The performances obtained are stable. For the future work we propose to:

- Study large rotations of the head,
- Improve the detection of the nose (since our algorithm uses only the most salient point),
- Study the fusion at features in the case of our method local statistics (SLF) for an adaptive selection of the best features.

References

1. Hjelmas, E., Low, B.K.: Face detection: a survey. Comput. Vis. Image Underst. **83**, 236–274 (2001)
2. Eskandari, M., Toygar, O.: Fusion of face and iris biometrics using local and global feature extraction methods. Signal Image Video Process. **8**(6), 995–1006 (2014)
3. Buciu, I., Kotropoulos, C., Pitas, I.: Comparison of ICA approaches for facial expression recognition. Signal Image Video Process. **3**(4), 345–361 (2009)
4. Sao, A.K., Yegnanarayana, B.: On the use of phase of the Fourier transform for face recognition under variations in illumination. Signal Image Video Process. **4**(3), 353–358 (2010)
5. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
6. Lajevardi, S.M., Hussain, Z.M.: Automatic facial expression recognition: feature extraction and selection. Signal Image Video Process. **6**(1), 159–169 (2012)
7. Hadid, Abdenour, Dugelay, Jean-Luc, Pietikinen, Matti: On the use of dynamic features in face biometrics: recent advances and challenges. Signal Image Video Process. **5**(4), 495–506 (2011)
8. Lu, J., Plataniotis, K., Venetsanopoulos, A.: Face recognition using LDA-based algorithms. IEEE Trans. Neural Netw. **14**(1), 195–200 (2003)

9. Yurtkan, K., Demirel, H.: Entropy-based feature selection for improved 3D facial expression recognition. Signal Image Video Process. **8**(2), 267–277 (2014)
10. Ming, Y.: Rigid-area orthogonal spectral regression for efficient 3D face recognition. Neurocomputing **129**(10), 445–457 (2014)
11. Ouamane, A., Belahcene, M., Benakcha, A., Boumehrez, M., Ahmed, A.T.: The classification of scores from multi-classifiers for face verification. Sens. Transducers J. **145**(10), 116–118 (2012)
12. Chen, Y., Medioni, G.: Object modeling by registration of multiple range images. In: Robotics and Automation, pp. 2724–2729
13. Huang, D., Ouji, K., Ardabilian, M., Wang, Y., Chen, L.: 3D Face recognition based on local shape patterns and sparse representation classifier. Lect. Notes Comput. Sci. **6523**, 206–216 (2011)
14. Xu, C., Li, S., Tan, T., Quan, L.: Automatic 3D face recognition from depth and intensity Gabor features. Pattern Recognit. **42**(9), 1895–1905 (2009)
15. Lowe, D.: Demo Software: SIFT Keypoint Detector. http://www.cs.ubc.edu.ca/lowe/, 2006
16. Xu, C., Wang, Y., Tan, T., Quan, L.: 3D Face recognition based on G-H shape variation. Lect. Notes Comput. Sci. **3338**, 233–243 (2005)
17. Wang, X., Ruan, Q. Ming, Y.: 3D Face recognition using corresponding point direction measure and depth local features. In: IEEE 10th International Conference on Signal Processing (ICSP), pp. 86–89 (2010)
18. Ming, Y., Ruan, Q.: Robust sparse bounding sphere for 3D face recognition. Image Vis. Comput. **30**(8), 524–534 (2012)