Comparison of 2D/3D Features and Their Adaptive Score Level Fusion for 3D Face Recognition

Wael Ben Soltana, Di Huang, Mohsen Ardabilian, Liming Chen
LIRIS (Laboratoire d'InfoRmatique en Image et Systèmes d'information), Ecole Centrale Lyon
36 Avenue Guy de Collongue, 69134, Ecully, France
{wael.ben-soltana, di.huang, mohsen.ardabilian, liming.chen}@ec-lyon.fr

Chokri Ben Amar
REGIM (REsearch Group on Intelligent Machines), University of Sfax
National Engineering School of Sfax, 3038, Sfax, Tunisia
chokri.benamar@ieee.org

Abstract

3D face has been introduced in the literature to deal with the unsolved issues of 2D face recognition, namely lighting and pose variations. In this paper, we study and compare the distinctiveness of features extracted from both the registered 2D face images and 3D face models. Sparse Representation Classifier (SRC) is exploited to calculate all similarity measures which are compared with the ones by a baseline of Nearest Neighbor (NN). As individual features of 2D and 3D are far from distinctive for discriminating human faces, we further present an adaptive score level fusion strategy for multimodal 2D-3D face recognition. The novel fusion strategy consists of an offline and an online weight learning process, both of which automatically select the most relevant weights of all the scores for each probe face in each modality. The weights calculated offline are based on the EER value of each type of features, while the online ones are dynamically obtained according to matching scores. Both types of weights are then fused to generate a final weight. Tested on the complete FRGC v2.0 dataset, the best rank-one recognition rate using only 3D or 2D features is 79.72% and 77.89%, respectively; while the new proposed adaptive fusion strategy achieves 95.48% with a 97.03% verification rate at 0.001 FAR, highlighting the benefit of exploring both 3D and 2D clues as well as the effectiveness of our adaptive fusion strategy.

1. Introduction

Face is potentially the best biometrics for people identification and verification applications, since it is socially well accepted, non-intrusive and contactless. Unfortunately, all human faces are similar and hence offer the low distinctiveness as compared with the other biometrics such as fingerprint and iris [17]. Furthermore, dealing with 2D facial texture images, intra-class variations, due to factors as diverse as lighting, pose, etc., are often much greater than inter-class variations, making 2D face recognition techniques far from reliable in real conditions [41].

In recent years, 3D face recognition has been intensively investigated by the research community to handle the two major unsolved issues in 2D face recognition domain, i.e., illumination and pose variations [31] [4]. However, even if 3D face models are theoretically insensitive to illumination changes, they still need to be properly registered before the matching step. Moreover, the issue of facial expression influences is even more difficult than that in 2D modality, because 3D face models provide the exact shape information of facial surfaces.

In this paper, we study the distinctiveness of some popular 3D shape related features, such as curvatures and tangent vectors, compared with several well known 2D texture-based ones, including Gabor and LBP, for face recognition. Furthermore, we also investigate the potential benefit of fusing 2D and 3D features as suggested in several previous works [19] [37] [28] [3] [36] [6] [16] [20] [23] stating that multimodal 2D-3D face recognition is more accurate and robust than either of the single modality [4]. For instance, Lu et al. [21] applied a weighted sum rule to fuse a facial surface matching score with a 2D appearance based one. In [26], Mian et al. proposed a feature-based method extracting pose invariant features from 3D models which were further fused with SIFT features from 2D faces, and achieved a 96.1% rank-one recognition rate on the FRGC v2.0 database. The same authors [25] also designed another system which performs hybrid (feature-based and holistic) matching for multimodal face recogni-

tion and displayed a 97.37% rank-one recognition rate on the same dataset. An in-depth study of fusion strategies for 3D face recognition was carried out by Gokberk et al. [13] who discussed and compared various techniques for classifier combination such as fixed rules, voting- and rank-based fusion schemes. By fusing several experts based on off-the-shelf 3D and 2D features, they reported a 95.45% accuracy on the FRGC v2.0 dataset.

The fusion strategies utilized for face recognition can be classified essentially into three categories: complementary fusion, competitive fusion, and serial fusion. Complementary fusion [2] combines features extracted from multiple data sources to create a new feature set to represent the face. Competitive fusion generates a single score by fusing different scores provided by multiple classifiers. In this case, there are three fusion levels: score level [7], rank level [14] and decision level [9]. At the first level, similarity scores are combined by various techniques [11], for example, Sum Rule, Product Rule, Min Rule, Max Rule, etc. At the second level, sorted lists computed by classifiers are merged based on different approaches such as Borda Count and Logistic Regression [27]. At the third level, all the candidates of classifiers are fused by adopting several approaches [12], i.e., Majority Vote or Majority Vote with maximum confidence. Finally, serial fusion [20] filters out a number of the most similar classes provided by a simple classifier and then feeds these classes into a more complex and powerful second classifier.

In the literature of 2D-3D multimodal face recognition, score level fusion, such as Sum Rule, Weighted Sum Rule and Product Rule, was extensively exploited [26] [25] [11] [24] [18] and proved effective. This paper also focuses on score level fusion and proposes an adaptive score level fusion strategy to combine features from complementary 2D and 3D modalities. The proposed fusion consists of an offline and an online weight learning process, both of which automatically select the most relevant weights of all scores for each probe from each modality. The weights calculated offline are based on the EER (Equal Error Rate) value of each feature type, while the online ones are dynamically obtained according to matching scores. Both types of weights are further fused to generate a final weight. Sparse Representation Classifier (SRC) [39] is applied to compute all the similarity measures compared with the ones produced by the baseline of Nearest Neighbor (NN). This scheme allows discarding wrong matches of individual feature type in test or by assigning a lower weight to a corresponding similarity of a gallery face. For instance, if 2D features are not discriminative enough due to illumination variations, 3D features will be assigned a proper weight in classifiers. The FRGC v1.0 database is used for training; the rank-one recognition rate on the complete FRGC v2.0 database is 95.48%; while verification rate is 97.03% at 0.001 FAR. As we will see later, individual features of 2D and 3D prove far from discriminative for reliable face recognition; however, the performance of final fusion is comparable to the state-of-the-art [26] [25] [13], highlighting the effectiveness of the proposed fusion strategy.

The remainder of this paper is organized as follows. The system overview is shown in section 2. Preprocessing and alignment in this work is introduced in section 3. Section 4 describes both the 2D and 3D feature set as well as Sparse Representation Classifier. The proposed fusion approach is presented detailedly in section 5. Section 6 analyzes and discusses experimental results on the FRGC database. Section 7 concludes the paper.

2. The System Overview

The framework of our system is shown in Figure 1 . The training and test stages contain the same steps: data preprocessing, feature extraction, SRC classification and score normalization. The fusion is carried out at similarity score level using a weighted sum rule. The weight associated to each feature type is determined not only offline according to its discriminative ability but also online according to similarity measures. At the training stage, the weights are calculated offline based on the EER value of individual features, while the weights at the test stage are dynamically achieved based on matching scores. Both types of weights are further combined to generate a final weight to improve fusion performance for face identification or verification.

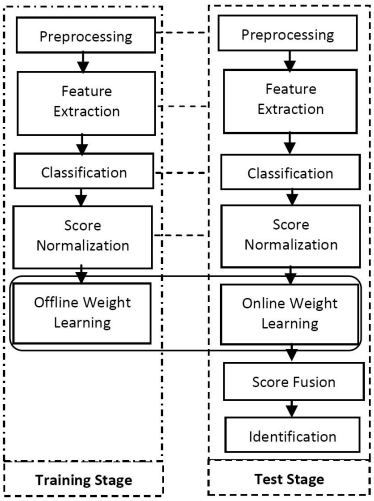

Figure 1. The framework of the face recognition system.

3. Preprocessing and Alignment

Each 3D face model has its concomitant 2D color (RGB) information, and 2D texture is densely registered to its corresponding 3D shape. In this work, since we focuse on the distinctiveness study of several popular 2D/3D features and

the benefit of their fusion for face recognition, we thus try to reduce at its maximum the impact by erroneous registration, and make use of 15 manually landmarked points to register 3D face models as accurately as possible (see Figure 2).

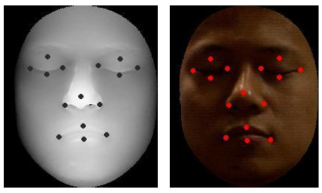

Figure 2. The preprocessed face sample with manual landmarks.

3.1. 3D Face Preprocessing

3D face models delivered so far by current 3D imaging systems are usually corrupted by spikes and holes. In this paper, we adopt the preprocessing technique in [33]. 3D face models are first cropped with a sphere of radius of 80 mm centered at nose tip. Basically, instead of using the median filter to remove spikes, we apply a decision-based median filtering technique by using the median filter only on vertexes predefined as potential spikes after a thresholding operation. This method can efficiently remove all the spikes without touching the properly scanned points. Once all the spikes are removed, holes are filled by fitting a mean square surface to the hole border located by searching vertexes possessing less than eight neighbors.

3.2. 3D Face Registration

3D face registration is a critical step in 3D face recognition. In order to avoid the impact of registration errors in our distinctiveness analysis of 2D/3D features and their fusion for face recognition, we employed a manual registration method, named Region-based Iterative Closet Point (R-ICP) [1] for 3D face models. A frontal face model with neutral expression is chosen from the FRGC v2.0 database as the reference model. For each model, the rigid facial region composed of nose and forehead, insensitive to facial expression variations, is cropped for alignment (see Figure 3). The alignment step exploits a coarse-to-fine strategy. The coarse step utilizes eleven facial landmarks located on the upper part of the face model and applies SVD to recover 3D rotation and translation in a rigid transformation. At the fine step, ICP is used for surface matching and to improve the estimates of translation and rotation parameters.

3.3. 2D Face Preprocessing

As 2D color information is densely registered to its corresponding 3D face data, the previously cropped and registered 3D face model also has its 2D texture counterpart. The positions of the eye inner corners are further used for rotation normalization. Finally, all the 2D color faces are converted to gray level, and resized to 80×92 pixels.

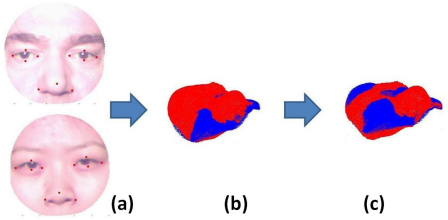

Figure 3. R-ICP alignment: (a) rigid regions of two textured 3D samples; (b) coarse alignment step (c) fine alignment step.

4. Feature Extraction and Classification

We explore the distinctiveness of features extracted from both 2D facial texture images and 3D face model data as well as their possible complementary contribution in a fusion process for face recognition. In classification, Sparse Representation Classifier (SRC) [39] is applied to all these features for its discriminative power.

4.1. 3D Features

A 3D face model captures exact geometry information of a facial surface. Various shape related features can thus be extracted from 3D face models, ranging from several basic ones such as point-clouds, range images, to more elaborated ones: normals, binormals, tangent vectors, curvatures etc., all of which describe shape variations over local patches inspired by differential geometry of 3D surfaces [42].

Most of these 3D features are quite popular and can be found in many works in the literature. For instance, Gokberk et al. [13] carried out an extensive study on several shape based features, including point-clouds, surface normals, depth images, curvatures and voxels, and found out that the principal curvatures along with the shape index and surface normal were three best features on the FRGC v1.0 dataset for face recognition.

In this work, we also consider the raw point-cloud representation of 3D face model as well as geometric features containing normal, binormal, tangent vector and curvature related features that have the potential for a higher accuracy to describe surface based events, and are well suited to represent the properties of facial regions, such as two cheeks, forehead, and chin. Four categories of curvature-based features are extracted. The first two types rely on the principal directions corresponding to maximum and minimum curvatures [35]. The last two are their derivatives, i.e., mean (H) and Gaussian (K) curvatures [13].

We further investigate another type of 3D features based on the anthropometric approach which advocates extracting a signature from some anthropometric points considered the most relevant; these points should be stable and discriminative. Faltemier et al. [9] [10] have shown that the portion of the face region surrounding nose was very stable. In [2], Arca et al. used segments of 2D face around the eyes and nose to extract 3D profiles. Inspired by these work and the

one by Perrot [29] on facial anthropometric measurements, a feature vector is generated including distance representing the edge length of every two anthropometric points; the angle formed by two edges with one same endpoint; and the ratio between horizontal nose width and vertical nose height. It is important to note that the measurements are on an absolute scale (mm) instead of a relative scale (pixels).

4.2. 2D Features

Three different approaches are introduced to extract features from 2D texture images. The first one is the simple pixel-based method that encodes grayscale intensity values into a single vector.

The second one is a non-parametric method, namely Local Binary Patterns (LBP) [34]. Its most important properties are the tolerance to monotonic illumination variations as well as the computational simplicity. LBP summarizes local structures of 2D images efficiently by comparing each pixel with its neighboring pixels. The neighborhood is defined as a set of sampling points evenly spaced around a circle centered at the pixel to be labelled, and the sampling points that do not fall in pixels are expressed using bilinear interpolation, thus allowing any radius value and any sampling point number. The general idea of LBP based facial representation is that a face can be seen as a composition of micro-patterns described by LBP. A facial image is thus first divided into several local regions from which LBP histograms are extracted. These LBP histograms are then concatenated, thereby describing both local and global information of the face.

The third feature is extracted by Gabor filters [38] which are spatially localized and selective to spatial orientations and scales. They are reputed to simulate receptive fields of simple cells in the mammalian visual cortex [22]. Since Gabor filters detect amplitude-invariant spatial frequencies of grayvalues of pixels, they are known to be robust to variations of illumination. Five different frequencies and eight equally spaced orientations are utilized to generate Gabor kernels. Hence, the dimension of Gabor filter based facial feature is 40 (5×8) times as much as that of original images.

4.3. Sparse Representation Classifier

Sparse representation for signal classification (SRSC) is proposed in [15]. SRSC incorporates reconstruction properties, discriminative power and sparsity for robust classification. In [39], a general classification method for image-based object recognition is proposed based on a sparse representation computed by $L1$-minimization. The approach based on sparse representation can often achieve high performance using a data dictionary [40].

Let us assume that we have k distinct classes and n_i feature vectors (1-D vector), named $v_{i,j} \in R_m$ available for training from the i_{th} class, $i = 1, 2, ..., k$ and j is the index of the training sample, $j = 1, 2, ..., n_i$. All the training data from the i_{th} class are placed in a matrix A_i such that $A_i = [v_{i,1}, v_{i,2}, ..., v_{i,n_i}] \in R_m \times n_i$. We develop a dictionary matrix A for all k classes by concatenating A_i, $i = 1, 2, ..., k$ as follows:

$$A = [A_1, A_2, ..., A_k] \in R_{m \times n_i \cdot k} \tag{1}$$

A test pattern y can be represented as a linear combination of all n training samples ($n = n_i \times k$):

$$y = Ax \tag{2}$$

where x is an unknown coefficient vector. From equation 2 it is relatively straightforward to note that only those entries of x that are non-zero correspond to the class of y. This means that if we are able to solve equation 2 for x we can actually find the class of the test pattern y. Fortunately, in compressed sensing [5] [8], as long as the solution to equation 2 is known sufficiently sparse, an equivalent $L1$-norm minimization:

$$(L1) : x_1 = \arg\min \|x\|_1 ; \quad Ax = y \tag{3}$$

can be solved as a good approximation to equation 2. Favorably, equation 3 can be solved by standard linear programming techniques. In fact, conventional recognition methods relying on the Euclidean distance to find the nearest neighbor can be modeled as a minimization problem similar to equation 3, except that the objective to be minimized is $L2$-norm. With the solution x_1 to equation 3, we can compute the residual between a given probe face and each gallery face as:

$$r_i = \left\| y - \sum_{j=1}^{k} x_{1_{i,j}} v_{i,j} \right\|_2 \tag{4}$$

The identity of the given probe face is then determined as the one with the smallest residual.

5. Adaptive Score Level Fusion

The 3D face model and its registered 2D texture map belong to different modalities whose fusion may have complementary contribution for face recognition. In this work, we adopt a competitive fusion method and propose an adaptive score level fusion scheme using a weighted sum rule. The weight associated to each feature type is generated by an offline and an online weight learning step, both of which automatically select the most relevant weights of all scores for each probe from each modality. The weights calculated offline are based on the EER value of individual features, while the online ones are dynamically obtained according to matching scores. Both types of weights are further combined to produce a final weight. SRC is introduced to compute all the similarity measures.

5.1. Score Normalization

Before the fusion step, it is important to normalize scores achieved by different types of features. Normalization aims at mapping scores into a common scale and range before their fusion.

Here, we denote a raw matching score as s, from a set S of scores produced by different features, and its corresponding normalized score as n. Three normalization approaches are introduced, and their performance is also compared in our experiments.

Min-Max (MM) [11] [24] [32] maps matching scores to the range of [0, 1]. The function $max(s)$ and $min(s)$ specify the maximum and minimum of the score range respectively:

$$n = \frac{s - \min(s)}{\max(s) - \min(s)} \tag{5}$$

Z-Score (ZS) [11] [32] transforms all matching scores to a distribution with the mean of 0 and the standard deviation of 1. The operators $mean(s)$ and $std(s)$ denote the arithmetic mean and standard deviation respectively:

$$n = \frac{s - mean(s)}{std(s)} \tag{6}$$

Tanh (TH) [11] [32] provided by so-called robust statistical techniques, maps matching scores to the [0, 1] range:

$$n = \frac{1}{2}[\tanh(\frac{s - mean(s)}{std(s)} \times 0.01) + 1] \tag{7}$$

5.2. Fusion Strategy

At the offline stage, we employ the approach proposed in [32] to assign a weight to scores of a particular feature based on its corresponding value of Equal Error Rates (EER) from a learning dataset.

Denote the EER value of each feature m as e^m, $m = 1, 2, ..., M$. Then, the weight P^m associated to the score of feature m is calculated as:

$$P^m = \frac{\frac{1}{u}}{e^m},\ u = \sum_{m=1}^{M} \frac{1}{e^m},\ \sum\nolimits_{m=1}^{M} P^m = 1,\ 0 \leq P^m \leq 1 \tag{8}$$

It should be noted that the weight P^m is inversely proportional to its corresponding error, and the weights of more accurate scores are always bigger than those of less accurate ones.

At the online stage, for each gallery face g, we compute a score $S^{g,f}$ using feature type f with that of the probe face by SRC. All the similarities $S^{g,f}$ are then sorted in a descending order. We assign to each score $S^{g,f}$ a weight $w^{g,f}$ which is a function of its ordered position $p^{g,f}$.

Specifically, the weight $w^{g,f}$ is defined as:

$$w^{g,f} = f(p) \propto \ \ln(N_g / p^{g,f}) \tag{9}$$

where N_g is the number of subjects in the gallery set.

The online stage gives more importance to scores ranked at the first positions and aims at discarding wrong matching of each feature type in test by assigning a lower weight to its corresponding similarity with a gallery sample.

The final similarity score between the gallery face g and the probe face is:

$$S_{Final}(g) = \sum_{f \in Features} P^f \cdot w^{g,f} \cdot S^{g,f} \tag{10}$$

The probe face is recognized as the one in the gallery which obtained the highest final score according to equation 10.

6. Experimental Results

6.1. Database

The FRGC [30] dataset is selected for experiments. Each face data consists of one 3D face model and its registered 2D color image. The 3D face model is available in the form of four matrices, each of size 480×640. The first matrix is a binary mask indicating valid pixels or points in the other three matrices that contain x, y, and z-coordinates of vertices respectively. The associated 2D color images have one-to-one correspondences to their 3D face models, and they are correctly registered to 3D faces in most cases, but some examples of incorrect registration can be found. All the data are divided into three sets based on acquisition time, namely, Spring2003, Fall2003, and Spring2004.

The FRGC v1.0 dataset (Spring2003) consisting of about 900 face models is used for training while the FRGC v2.0 (Fall2003 and Spring2004) is used for validation, containing 4007 3D face models of 466 different subjects. Among these subjects, 57% are male and 43% are female, with age distribution: 65% 18-22 years old, 18% 23-27 and 17% 28 years or over. The dataset was collected during 2003-2004 academic year, and thus includes time lapse. Facial expression variations are labelled with neutral and non-neutral.

6.2. Experiment Settings

In the training step, 116 subjects that possess more than four face models are selected from the FRGC v1.0 to learn weights of all types of features and to train parameters of subspace based approaches such as LDA. Then, these parameters and basis images are saved and used for gallery images as well as probe images of the FRGC v2.0 dataset.

One 3D face scan with a neutral expression is selected from each subject to make a gallery of 466. The remaining 3D face scans (4007-466=3541) are treated as probes and divided into two subsets according to their expression labels. The first subset contains face scans with the neutral expression (1979 probes), and the other one with non-neutral face scans (1562 probes).

For evaluating the proposed approach, experiments are designed for both face recognition and verification. In each of the two tasks, three experiments are carried out: Neutral vs. Neutral, Neutral vs. Non-Neutral, and Neutral vs. All. In Neutral vs. Neutral and Neutral vs. Non-Neutral, only the neutral and non-neutral probe subset is utilized respectively. Such an experimental protocol enables detailed analysis of the proposed 2D/3D features and their fusion to facial expression variations.

6.3. Results and Analysis

First, Linear Discriminant Analysis (LDA) is applied to reduce the dimensionality of following features: grayscale textures; normal, binormal and tangent vectors; Gaussian

Table 1. Rank-one recognition rates of individual feature type on the FRGC v2.0 database - Neutral vs. All.

	SRC	NN
Gabor Filters	77.89%	57.92%
LBP	71.82%	47.64%
Intensity Image	49.82%	40.30%
Anthropometric Measurements	46.48%	42.60%
3D Points*	–	58.57%
Gaussian Curvature	59.02%	56.28%
Mean Curvature	71.62%	70.87%
Maximum Curvature	67.81%	66.82%
Minimum Curvature	66.73%	66.25%
Binormal Vectors	70.63%	69.78%
Normal Vectors	70.01%	69.26%
Tangent Vectors	79.72%	79.13%

Table 2. Rank-one recognition rates of individual feature type on the FRGC v2.0 database - Neutral vs. Neutral.

	SRC	NN
Gabor Filters	84.59%	62.81%
LBP	78.83%	54.53%
Intensity Image	56.85%	45.65%
Anthropometric Measurements	55.99%	51.58%
3D Points*	–	72.26%
Gaussian Curvature	73.67%	70.88%
Mean Curvature	85.25%	83.74%
Maximum Curvature	82.57%	81.44%
Minimum Curvature	81.05%	80.85%
Binormal Vectors	84.49%	83.29%
Normal Vectors	83.78%	82.57%
Tangent Vectors	89.64%	88.86%

(H), Mean (K), Minimum and Maximum curvatures; as well as anthropometric features. Then, two similarity measures of each feature are compared: a baseline of NN with the Euclidean distance and SRC. Table 1 lists rank-one recognition rate of each feature on the FRGC v2.0 dataset. SRC is not applied to 3D point-clouds which are directly matched using ICP.

From Table 1, we can see that SRC performs better than NN for each feature studied here and this improvement is significant in 2D modality. All the individual 3D and 2D features do not provide enough distinctiveness for representing faces, only displaying 79.72% with the best 3D feature by tangent vectors and 77.89% with the best 2D feature by Gabor filters. The following two tables show the performance of these features on facial expression variations.

Table 2 displays the performance of each feature in the case of Neutral vs. Neutral. As we can see from the table, all the features, either 3D or 2D, achieve better performance than the case of Neutral vs. All. Still, none of these 3D or 2D features reports enough distinctiveness for face recognition, and the best rank-one recognition rate is 89.64% by tangent vectors. With an average gain of about 10 percent, the most significant improvement is achieved by 3D features, confirming the intuition that 3D facial shape descriptors are much more sensitive to facial expression variations than 2D features. This intuition is further confirmed by figures in table 3 which evidences performance degradation is much greater for 3D features than the 2D ones in the case of Neutral vs. Non-Neutral.

Table 3. Rank-one recognition rates of individual feature type on the FRGC v2.0 database - Neutral vs. Non-Neutral.

	SRC	NN
Gabor Filters	69.40%	49.49%
LBP	63.06%	41.68%
Intensity Image	40.91%	30.65%
Anthropometric Measurements	34.44%	31.24%
3D Points*	–	41.23%
Gaussian Curvature	40.46%	37.85%
Mean Curvature	54.35%	53.63%
Maximum Curvature	49.10%	48.23%
Minimum Curvature	48.59%	47.96%
Binormal Vectors	53.07%	52.78%
Normal Vectors	52.56%	51.87%
Tangent Vectors	67.16%	66.73%

Table 4. Rank-one recognition rates of different score normalization methods with different fusion schemes.

	Max-Min	Z-Score	Tanh
Rank fusion	88.87%		
Product Rule	51.60%	67.64%	27.40%
Simple Sum Rule	87.49%	85.12%	86.44%
Weighted Sum offline	93.98%	91.89%	89.33%
Weighted Sum online	90.45%	85.23%	89.69%
Adaptive Method	95.48%	92.01%	94.69%

Table 5. Rank-one recognition rates of 2D modality, 3D modality and their fusion: (I) Neutral vs. All; (II) Neutral vs. Neutral; (III) Neutral vs. Non- Neutral.

	I	II	III
2D + 3D	95.48%	98.64%	90.65%
2D	85.31%	89.54%	76.63%
3D	84.07%	91.61%	67.99%

Table 4 compares the proposed adaptive score level fusion strategy with some other popular fusion approaches using three normalization schemes. As we can see from the table, all the fusion strategies, with the exception of product rule, when combined with the Max-Min normalization, ameliorate rank-one recognition accuracy as compared with the performance of individual features either in 3D or 2D. This result further confirms that 3D and 2D features are complementary modalities whose adequate fusion can improve the final decision. Moreover, the fusion strategies by the proposed two weighted sum rules, both the offline and online, perform better than the other fusion methods including rank fusion, product rule and simple sum rule. They display the best performance (95.48% recognition rate) when they are further combined according to the proposed adap-

Table 6. Some comparison results.

	EER	VR@0.1% FAR	Rank-one RR
Adaptive fusion	0.009	97.03%	95.48%
Gokberk et al. [13]	–	–	95.45%
Mian et al. [26]	–	98.60%	96.10%

tive score level fusion. For comparison purpose, we also evaluate the fusion scheme proposed in [26] to combine the previous 2D and 3D features and obtained a 85.31% rank-one recognition rate which is more than 10% drop compared to the one achieved by the proposed adaptive fusion scheme.

Table 5 shows performance of 2D modality, thus only fusing 2D features, 3D modality with all the 3D features and multimodal 2D-3D on the three predesigned experiments. These figures further evidence that 3D shape related features perform slightly better than 2D features in the presence of the neutral expression and unfortunately drop much more quickly in performance than 2D features in the presence of the non-neutral expression.

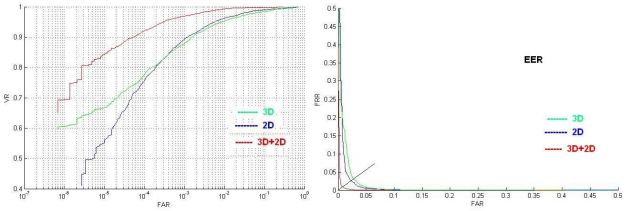

Figure 4. ROC and EER curves.

The proposed approach is also evaluated by face verification tasks. Figure 4 shows ROC curves of 2D modality, 3D modality and multimodal 2D-3D fusion at 0.001 FAR, achieving 89.52%, 88.08% and 97.03% verification rate respectively. It also reports their corresponding EER values: 0.0249, 0.0294, and 0.009.

Table 6 compares performance achieved by the proposed adaptive score level fusion technique with the state-of-the-art. The proposed adaptive score level fusion scheme displays comparable performance, and it is worth emphasizing that the focus of this work aims at studying the distinctiveness of some popular 3D and 2D features and the potential benefit of their fusion for the task of face recognition.

7. Conclusion

In this paper, we studied and compared the distinctiveness of popular features extracted from both the registered 2D face images and 3D face models and the benefit of their fusion for the task of 3D face recognition. Sparse Representation Classifier was explored to calculate all similarity measures compared with the ones by a baseline of NN. We also proposed an adaptive score level fusion based on weighted sum rule. The weight associated to each feature combines the offline weight computed using the EER value on a learning dataset with the online one from the matching scores. Tested on the FRGC v2.0 dataset, the best 3D and 2D features display a rank-one recognition rate of 79.72% and 77.89% respectively using SRC while the novel adaptive fusion scheme achieves 95.48% accuracy with 97.03% verification rate at 0.001 FAR.

This study shows that: 1) none of these 2D texture and 3D shape related features is distinctive enough for reliable face recognition; 2) 3D shape based features perform better in the case of Neutral vs. Neutral than its counterpart 2D texture related ones, while they prove much more sensitive to facial expression variations; 3) 2D and 3D features are complementary and their competitive fusion generally ameliorates overall performance; 4) the new adaptive fusion strategy shows its effectiveness when fusing these non-discriminative 2D and 3D features.

8. Acknowledgement

This work was partially carried out within the French FAR3D project supported by ANR under the grant ANR-07-SESU-004 FAR3D.

References

[1] B. B. Amor, M. Ardabilian, and L. Chen. New experiments on icp-based 3d face recognition and authentication. *Intl. Conf. on Pattern Recognition*, 2006. 3

[2] S. Arca, R. Lanzarotti, and G. Lipori. Face recognition based on 2d and 3d features. *Intl. Conf. on Knowledge-Based Intelligent Information and Engineering Systems*, pages 455–462, 2007. 2, 3

[3] C. Beumier and M. Acheroy. Face verification from 3d and grey level cues. *Pattern Recognition Letters*, 22(12):1321–1329, 2001. 1

[4] K. W. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition. *Computer Vision and Image Understanding*, 101(1):1–15, 2006. 1

[5] E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. on Information Theory*, 52(2):489–509, 2006. 4

[6] K. Chang, K. W. Bowyer, and P. Flynn. Face verification from 3d and grey level cues. *Multimodal User Authentication Workshop*, pages 25–32, 2003. 1

[7] J. Cook, M. Cox, V. Chandran, and S. Sridharan. Robust 3d face recognition from expression categorization. *Intl. Conf. on Biometrics*, pages 271–280, 2007. 2

[8] D. Donoho. Compressed sensing. *IEEE Trans. on Information Theory*, 52(4):1289–1306, 2006. 4

[9] T. Faltemier, K. W. Bowyer, and P. Flynn. 3d face recognition with region committee voting. *Intl. Symp. on 3D Data Processing, Visualization, and Transmission*, 2006. 2, 3

[10] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn. Using multi-instance enrollment to improve performance 3d face recognition. *Computer Vision and Image Understanding*, 112(2):114–125, 2008. 3

[11] A. Godil, S. Ressler, and P. Grother. Face recognition using 3d facial shape and color map information: Comparison and

combination. *Biometric Technology for Human Identification, SPIE*, 2005. 2, 5

[12] B. Gokberk and L. Akarun. Comparative analysis of decision level fusion algorithms for 3d face recognition. *Intl. Conf. on Pattern Recognition*, 2006. 2

[13] B. Gokberk, H. Dutagaci, A. Ulas, L. Akarun, and B. Sankur. Representation plurality and fusion for 3d face recognition. *IEEE Trans. on Systems Man and Cybernetics-Part B: Cybernetics*, 38(1):155–173, 2008. 2, 3, 7

[14] B. Gokberk, A. Salah, and L. Akarun. Rank-based decision fusion for 3d shape-based face recognition. *Intl. Conf. on Audio- and Video-Based Biometric Person Authentication*, pages 1019–1028, 2005. 2

[15] K. Huang and S. Aviyente. Sparse representation for signal classification. *Intl. Conf. on Neural Information Processing Systems*, 2006. 4

[16] M. Husken, M. Brauckmann, S. Gehlen, and C. v. d. Malsburg. Strategies and benefits of fusion of 2d and 3d face recognition. *IEEE Workshop on Face Recognition Grand Challenge Experiments*, 2005. 1

[17] A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Trans. on Circuits and Systems for Video Technology*, 14(1):4–20, 2004. 1

[18] I. A. Kakadiaris, G. Passalis, G. Toderici, N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(4):640–649, 2007. 2

[19] S. Lao, Y. Sumi, M. Kawade, and F. Tomita. 3d template matching for pose invariant face recognition using 3d facial model built with iso-luminance line based stereo vision. *Intl. Conf. on Pattern Recognition*, 2000. 1

[20] X. Lu and A. K. Jain. Integrating range and texture information for 3d face recognition. *IEEE Workshop on Applications of Computer Vision*, pages 155–163, 2005. 1, 2

[21] X. Lu, A. K. Jain, and D. Colbry. Matching 2.5d face scans to 3d models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(1):31–43, 2006. 1

[22] S. Marcelja. Mathematical description of the responses of simple cortical cells. *Journal of Optical Society of America A*, 70(11):1297–1300, 1980. 4

[23] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, and G. Medioni. Performance of geometrix activeidtm 3d face recognition engine on the frgc data. *IEEE Workshop on Face Recognition Grand Challenge Experiments*, 2005. 1

[24] A. Mian, M. Bennamoun, and R. Owens. Face recognition using 2d and 3d multimodal local features. *Intl. Symp. on Visual Computing*, 2006. 2, 5

[25] A. S. Mian, M. Bennamoun, and R. Owens. An efficient multimodal 2d-3d hybrid approach to automatic face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(11):1927–1943, 2007. 1, 2

[26] A. S. Mian, M. Bennamoun, and R. Owens. Keypoint detection and local feature matching for textured 3d face recognition. *Intl. Journal of Computer Vision*, 79:1–12, 2008. 1, 2, 7

[27] M. M. Monwar and M. Gavrilova. Fes: A system for combining face, ear, and signature biometrics using rank level fusion. *Intl. Conf. on Information Technology: New Generations*, pages 922–927, 2008. 2

[28] T. Papatheodorou and D. Reuckert. Evaluation of automatic 4d face recognition using surface and texture registration. *Intl. Conf. on Automated Face and Gesture Recognition*, pages 321–326, 2004. 1

[29] R. Perrot. Use of anthropological methods in the identification of unknown individuals: Human remains and armed robbers. *The Forensic Scientist Online Journal*, 1997. 4

[30] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. *IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, 2005. 5

[31] A. Scheenstra, A. Ruifrok, and R. C. Veltkamp. A survey of 3d face recognition methods. *Intl. Conf. on Audio- and Video- based Biometric Person Authentication*, 2005. 1

[32] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. Jain. Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(3):450–455, 2005. 5

[33] P. Szeptycki, M. Ardabilian, and L. Chen. A coarse-to-fine curvature analysis-based rotation invariant 3d face landmarking. *Intl. Conf. on Biometrics: Theory, Applications and Systems*, 2009. 3

[34] A. H. T. Ahonen and M. Pietikäinen. Face recognition with local binary patterns. *European Conf. on Computer Vision*, 2004. 4

[35] H. Tanaka, M. Ikeda, and H. Chiaki. Curvature-based face surface recognition using spherical correlation. *IEEE Intl. Conf. on Automatic Face and Gesture Recognition*, pages 372–377, 1998. 3

[36] F. Tsalakanidou, D. Tzocaras, and M. Strintzis. Use of depth and color eigenfaces for face recognition. *Pattern Recognition Letters*, 24(9):1427–1435, 2003. 1

[37] Y. Wang, C. Chua, and Y. Ho. Facial feature detection and face recognition from 2d and 3d images. *Pattern Recognition Letters*, 23(10):1191–1202, 2002. 1

[38] L. Wiskott, J. M. Fellous, N. Kruger, and C. v. d. Malsburg. Face recognition by elastic bunch graph matching. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997. 4

[39] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009. 2, 3, 4

[40] Z. Xie and J. Feng. Kfce: A dictionary generation algorithm for sparse representation. *Signal Processing*, 89(10):2072–2077, 2009. 4

[41] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Survey*, pages 399–458, 2003. 1

[42] S. W. Zucker. Differential geometry from the frenet point of view: Boundary detection, stereo, texture and color. *Handbook of Mathematical Models in Computer Vision*, 2005. 3