# Textured 3D Face Recognition using Biological Vision-based Facial Representation and Optimized Weighted Sum Fusion

Di Huang[1], Wael Ben Soltana[1], Mohsen Ardabilian[1], Yunhong Wang[2], and Liming Chen[1]
[1]Université de Lyon, CNRS, Ecole Centrale Lyon, LIRIS UMR 5205, 69134, Ecully, France
{di.huang, wael.ben-soltana, mohsen.ardabilian, liming.chen}@ec-lyon.fr
[2]School of Computer Science and Engineering, Beihang University, 100191, Beijing, China
yhwang@buaa.edu.cn

## Abstract

*This paper proposes a novel biological vision-based facial description, namely Perceived Facial Images (PFIs), aiming to highlight intra-class and inter-class variations of both facial range and texture images for textured 3D face recognition. These generated PFIs simulate the response of complex neurons to gradient information within a certain neighborhood and possess the properties of being highly distinctive and robust to affine illumination and geometric transformation. Based on such an intermediate facial representation, SIFT-based matching is further carried out to calculate similarity scores between a given probe face and the gallery ones. Because the facial description generates a PFI for each quantized gradient orientation of range and texture faces, we then propose a score level fusion strategy which optimizes the weights using a genetic algorithm in a learning step. Evaluated on the entire FRGC v2.0 database, the rank-one recognition rate using only 3D or 2D modality is 95.5% and 95.9%, respectively; while fusing both modalities, i.e. range and texture-based PFIs, the final accuracy is 98.0%, demonstrating the effectiveness of the proposed biological vision-based facial description and the optimized weighted sum fusion.*

## 1. Introduction

Face is one of the best biometrics for person identification and verification related applications, since it is socially well accepted, non-intrusive and contactless. Unfortunately, all human faces are similar and thereby offer low distinctiveness as compared with other biometrics, e.g., fingerprint and iris [1]. Moreover, when utilizing facial texture images, intra-class variations, due to factors as diverse as illumination and pose changes are often greater than inter-class ones, making 2D face recognition far from reliable in real conditions [2].

In recent years, 3D face recognition methods have been extensively investigated by the research community to deal the unsolved issues in 2D face recognition, i.e., illumination and pose changes [3] [4]. However, even if 3D face data are theoretically insensitive to illumination variations, they still need to be registered before matching step. Furthermore, the challenge of facial expression changes is even more difficult than in 2D modality, as 3D face models provide exact shape information of facial surfaces.

Since most of the current 3D imaging systems deliver 3D face models along with their aligned texture counterpart, a major trend in the literature of face recognition is to adopt both the 3D shape and 2D texture based modalities, arguing that the joint use of these two clues can generally achieves more accurate and robust accuracy than using only either of the single modality [5]-[13].

Most techniques in the literature for face recognition do not operate directly on original input facial images because faces are all similar and undergo the changes of illumination and pose. Instead, they try to look for an intermediate facial representation, for instance, eigenface [14], fisher face [15], LBP face [16] etc., aiming to highlight intra-class similarity and inter-class dissimilarity.

In this paper, we propose a novel biological vision-based facial description, namely Perceived Facial Images (PFIs), applied to both facial range and texture images for textured 3D face recognition. These PFIs simulate the response of complex neurons to gradient information in a neighborhood and own properties of being highly distinctive and robust to affine lighting and geometric transformations. Based on this intermediate facial representation, SIFT-based local feature matching is then used to calculate similarity scores between probe and gallery faces. Because the facial representation generates a PFI for each of quantized orientations of facial range and texture images, we further propose a score level fusion scheme that optimizes weights by a genetic algorithm in a learning step. Evaluated on the complete FRGC v2.0 dataset, the rank-one recognition rate using only 3D or 2D modality is 95.5% and 95.9% respectively; while combining these two modalities, i.e. range- and texture-based PFIs, the final accuracy is 98.0%, demonstrating the effectiveness

of the proposed biological vision-inspired facial description and the optimized weighted sum fusion.

The remainder of this paper is organized as: Section 2 introduces the biological vision-based facial representation that is applied on both facial range and texture images. The following SIFT-based local feature matching step is shown in section 3. Section 4 presents a weighted sum rule based score fusion method optimized by a genetic algorithm from a learning database. Section 5 analyzes and discusses these experimental accuracies achieved on the FRGC 2.0 dataset. Section 6 concludes the paper.

## 2. Biological Vision-based Facial Description

In order to improve the distinctiveness of the human faces and offering certain tolerance to lighting and pose changes, in this section, we introduce a novel biological vision-based facial description which can be applied to both facial range and texture images.

The proposed facial description is inspired by the study of Edelman et al. [17], who proposed a representation of complex neurons in primary visual cortex. These complex neurons respond to a gradient at a particular orientation and spatial frequency, but the location of gradient is allowed to shift over a small receptive field rather than being precisely localized. Our facial representation implements this idea.

### 2.1. Description of the complex neuron response

The proposed description aims at simulating the response of complex neurons, based on a convolution of gradients in specific orientations in a pre-defined circular neighborhood. The radius value can be varied experimentally for different applications.

Specifically, given an input image $I$, a certain number of gradient maps $L_1, L_2, \ldots, L_o$, one for each quantized direction $o$, are first computed. They are defined as:

$$L_o = \left( \frac{\partial I}{\partial o} \right)^+ \tag{1}$$

The '+' means that only positive values are kept to preserve the polarity of the intensity changes.

Each gradient map describes gradient norms of the input original image in an orientation $o$ at every pixel. We then simulate the response of complex neurons by convolving its gradient maps with a Gaussian kernel $G$, and the standard deviation of $G$ is proportional to the value of radius of the given neighborhood area, $R$, as in (2).

$$\rho_o^R = G_R * L_o \tag{2}$$

The purpose of the convolution with Gaussian kernels is to allow the gradients to shift within a neighborhood without abrupt changes.

At a given pixel location $(x, y)$, we collect all the values of the convolved gradient maps at that location and form the vector $\rho^R(x, y)$, and it thus has a response value of complex neurons for each orientation $o$.

$$\rho^R(x,y) = \left[ \rho_1^R(x,y), \cdots, \rho_O^R(x,y) \right]^t \tag{3}$$

This vector, $\rho^R(x, y)$, is further normalized to unit norm vector, which is called response vector and denoted by $\underline{\rho}^R$ in the following parts of this paper.

### 2.2. Facial description by response vectors

Now facial range and texture images can be represented by their perceived values of complex neurons according to their response vectors. Specifically, given a facial range or texture image $I$, we generate a new Perceived Facial Image (PFI) $J_o$ using complex neurons for each orientation $o$ defined as:

$$J_o(x,y) = \underline{\rho}_o^R(x,y) \tag{4}$$

Fig. 1 shows such a process. In our work, we generate 8 PFIs for 8 quantized directions respectively. Instead of the original facial range and texture images, the PFIs are further fed to SIFT-based local feature matching for face recognition.

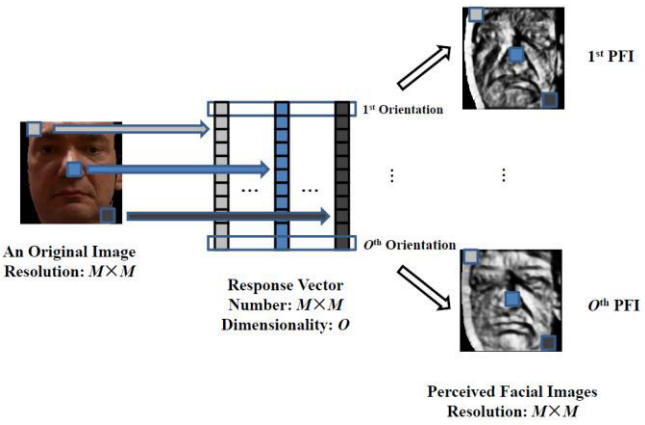



Figure 1: An illustration of the perceived facial images for each of the quantized orientations $o$.

### 2.3. The properties of distinctness and invariance

The generated PFIs potentially offer high distinctiveness because they highlight the details of local shape and texture variations. Meanwhile, they also have the property of being robust to affine lighting and geometric transformations.

As applied to 2D facial texture images, the PFIs offer the property of being robust to affine lighting transformations. Indeed, a PFI, $J_o$, is simply the normalized convolved facial gradient map at orientation $o$ according to (4), while affine

lighting variations usually add a constant intensity value on images, so it does not affect the computation of gradients. On the other hand, a change of image contrast in which the intensities of all the pixels are multiplied by a constant will result in the multiplication of gradient calculation; however, this contrast change will be cancelled by the normalization of the response vector.

Similarly, the PFIs of facial range images which contain 3D shape information are also invariant to affine geometric transformation leading to certain tolerance to pose changes.

The proposed PFIs can be made even rotation invariant if we choose to quantize directions starting from the principal direction of all gradients in the neighborhood. Nevertheless, we do not perform such a rotation normalization step to save computational cost, since 3D face models are generally in an upright frontal position in user cooperative applications.

## 3. SIFT based Local Feature Matching

Once PFIs of all quantized orientations are achieved from both range and texture face images, a local feature matching step is carried out on these widely-used SIFT based features [18] extracted from PFIs for similarity score calculation. It is well known that local feature based matching scheme is generally more robust to occlusion and pose changes.

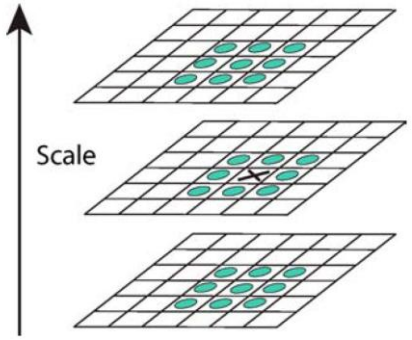



Figure 2: Extrema (maxima or minima) of difference-of-Gaussian images are detected by comparing a pixel (marked with X) to its 26 neighbors in 3×3 regions at the current and adjacent scales [18].

SIFT uses scale-space Difference-of-Gaussian (DOG) to detect keypoints in 2D images. The raw image is repeatedly convolved with Gaussians of different scales separated by a constant factor $k$ to produce an octave in scale space. As for an input image, $I(x, y)$, its scale space is defined as a function, $L(x, y, \sigma)$, produced by convolution of a variable scale Gaussian $G(x, y, \sigma)$ with $I$, and the DOG function $D(x, y, \sigma)$ thus can be computed from the difference of two nearby scales:

$$\begin{aligned} D(x,y,\sigma) &= (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \\ &= L(x,y,k\sigma) - L(x,y,\sigma) \end{aligned} \quad (5)$$

The extrema of $D(x, y, \sigma)$ can be detected by comparing each pixel value with those of its 26 neighbors within a 3×3 area at current and adjacent scales, and an illustration can be found in Figure 2. At each scale, gradient magnitude, $m(x, y)$, and orientation, $\theta(x, y)$, is computed by using pixel differences in (6) and (7).

$$\begin{aligned} m^2(x,y) = &(L(x+1,y) - L(x-1,y))^2 \\ &+ (L(x,y+1) - L(x,y-1))^2 \end{aligned} \quad (6)$$

$$\theta(x,y) = \tan^{-1} \frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)} \quad (7)$$

For each detected keypoints, a feature vector is extracted as a descriptor from the gradients of sampling points within its neighborhood. See Figure 3 for more details. To achieve the orientation invariance, coordinates and gradient orientations of sampling points in the neighborhood are rotated relative to the keypoint orientation. Then a Gaussian function is used to assign a weight to the gradient magnitude of each point. The points close to the keypoint are given more emphasis than the ones far from it (see [18] for more details of SIFT parameters). Orientation histograms of 4×4 sampling areas are computed, each with 8 bins. Therefore a final feature vector with a dimension of 128 (4×4×8) is produced.

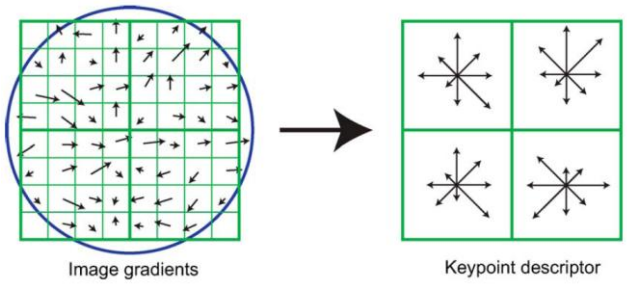



Figure 3: Computation of the keypoint descriptor [18].

SIFT operates on each PFI separately. As PFIs highlight local texture and shape changes of generally smooth facial images, many more keypoints can be detected than the ones if we directly apply SIFT on the original facial range and texture images. On the FRGC v2.0 database, it shows that the average number of keypoint detected on original facial range and texture images are only 41 and 67 respectively; while using PFIs for keypoint detection, the average numbers rise up to 116 and 304. Figure 4 shows an example of SIFT-based keypoint detection.

Given local features extracted from each corresponding PFI pair in the gallery and probe set respectively, two sets of facial keypoint sets can be matched. Matching one point to another is accepted only if the matching distance between them is less than a pre-defined threshold $t$ times the distance to the second closest match. In this work, $t$ is empirically set to 0.6 as used in [18]. Here, $N_{R(R,\,o)}$ and $N_{T(R,\,o)}$ denote the

number of matched keypoints of perceived range and texture image pair generated by the former biological vision inspired representation from corresponding facial range and texture images at orientation $o$ within the neighborhood area radius $R$, respectively.

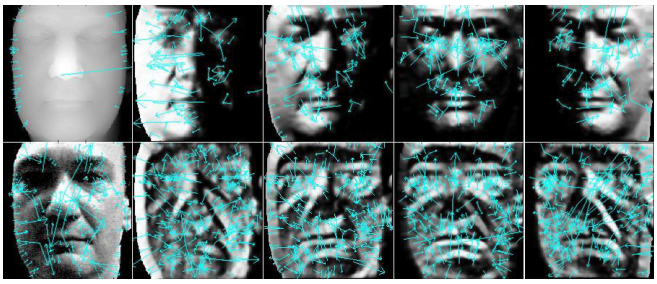


Figure 4: SIFT-based keypoint detection. The upper row lists an original range face and its perceived facial range images for the first four orientations; while the bottom one displays an original texture face and its first four perceived facial texture images.

The similarity measure, $N_{R(R,\,o)}$ or $N_{T(R,\,o)}$, is with a positive polarity (a bigger value means a better matching).

## 4. Optimized Weighted Score Sum Fusion

Given a textured 3D face model as a probe, the previous SIFT-based matching produces a similarity measure in each orientation of both facial texture and range images in the gallery. These similarity scores need to be further fused to deliver a final similarity score. In this paper, we develop a fusion scheme by using a weighted sum rule, as score level fusion has been extensively used in the literature of 2D/3D face recognition and has proved its efficiency in a number of non-trivial pattern recognition problems [19].

Formally, a weighted sum rule is defined as:

$$S = \sum_{i=1}^{N} w_i * S_i \qquad (8)$$

where $S_i$ is a similarity score; $w_i$ is its corresponding weight; $N$ is the number of modalities used for generating similarity scores. A bigger weight value indicates a higher importance; and a smaller one indicates a lower importance.

Our weighting scheme is learning-based, using a genetic algorithm [20] seeking an optimal set of weights through applications of selection, mutation, and recombination of a population.

Figure 5 shows the process of learning optimal weighting. First, a population is created by randomly generating individual 'chromosomes'. The chromosome length is the same with the number of variables (weights) corresponding to the number of similarity matrices. Given $N$ similarity measure matrices generated by the SIFT-based matching on different PFIs, each of chromosomes thus possesses $N$ gene positions representing $N$ different weights. At each iteration, a normalization process is first carried out to keep the sum of all the weights as one.

The chromosomes are used to encode trial solutions in a genetic algorithm. Iterative selection, crossover, and mutation are then exploited to make evolution of the population. At each generation, a new set of chromosomes is generated based on the fittest genes of previous generation to achieve a better solution. This fitness is calculated according to the produced similarity measure matrix in terms of recognition accuracy. Stochastic Universal Sampling [21] is applied to select chromosomes and to generate offspring. The operation of crossover leads to generate better offspring by exchanging characteristics of their parents. It enables the most efficient characteristics to be concentrated in an individual. The mutation randomly varies the genetic representation of an individual by adding a random value and tends to inhibit the possibility of converging to a local optimum, rather than the global one. The evolution is carried out until a desired solution is achieved, or a pre-specified number of iterations are ended. The final solution with a higher fitness represents the best vector of weights.

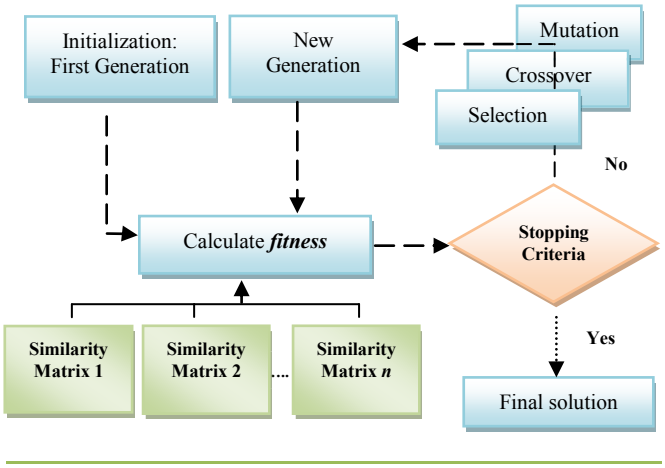



Figure 5: The process of the proposed learning strategy for optimized weighting

## 5. Experimental Results

The experiments were evaluated on FRGC v2.0 [22], one of the most comprehensive and popular databases. It contains 4007 textured 3D face models of 466 different subjects. One facial range image and its 2D texture counterpart were extracted from each of 3D face models. For 3D face data, a preprocessing step was introduced to remove spikes with a median filter and fill holes by using cubic interpolation. For 2D facial texture images, histogram equalization was used to reduce influence caused by non-affine lighting variations. Due to the properties of the PFIs and the use of local fea-

ture-based matching, we did not perform any registration on 3D face models. It is different from the work in [23] which also used both facial range and texture data. The faces were cropped by using a basic bounding box based on the mask provided by a 3D scanner indicating if the point is valid or not in that position. Cropped faces thus have pose, lighting and expression changes as well as occlusions caused by hair. The FRGC v1.0 database consisting of 943 neutral expression face models was adopted by our adaptive fusion approach for training optimized weights which were then used for test stage.

In the experiments, we adopted the same protocol of the state of the art for the purpose of comparison. The first 3D face model with a neutral expression of each subject makes up a gallery set of 466 subjects. The remaining face models (4007-466=3541) were treated as probes, as in [23]-[25].

We designed three experiments: the first is to discuss the impact of the neighborhood area radius $R$ on the final performance; the second is to test the proposed method on face recognition and verification compared with the state of the art; the last one is to evaluate the robustness of the proposed approach to facial expression variations.

### 5.1. Radius analysis of neighborhood area

Recall that complex neurons respond to gradient information within a neighborhood which is defined as a circular region in our implementation. In our experiments, we tested different values of radius $R$ and study its impact on final performance of both facial texture and range images. Table 1 summarizes the results with different radii applied to PFIs in eight orientations: (a) on texture faces; (b) on range faces. Fusion I lists the results with the weighted sum rule as in [24]; while Fusion II gives performance using the proposed optimized weighted sum fusion strategy.

As we can see in Table 1, the neighborhood area with a smaller radius achieves a better result for range face images; while for texture faces, the best result is obtained when $R$ = 1.5. Compared with the approach in Fusion I, the optimized fusion (Fusion II) method always performs slightly better.

Table 1: Results when using different neighborhood area radius R on: (a) texture faces; (b) range faces.

(a)

| Texture | $R$=1.0 | $R$=1.5 | $R$=2.0 | $R$=2.5 | $R$=3.0 | $R$=3.5 |
|---|---|---|---|---|---|---|
| PFI1 | 0.7800 | 0.8142 | 0.8260 | 0.8323 | 0.8356 | 0.8320 |
| PFI2 | 0.8308 | 0.8469 | 0.8619 | 0.8531 | 0.8461 | 0.8351 |
| PFI3 | 0.8517 | 0.8718 | 0.8749 | 0.8797 | 0.8763 | 0.8709 |
| PFI4 | 0.8622 | 0.8749 | 0.8862 | 0.8777 | 0.8749 | 0.8633 |
| PFI5 | 0.8006 | 0.8190 | 0.8365 | 0.8243 | 0.8215 | 0.8066 |
| PFI6 | 0.8074 | 0.8297 | 0.8489 | 0.8602 | 0.8577 | 0.8503 |
| PFI7 | 0.8444 | 0.8602 | 0.8526 | 0.8563 | 0.8467 | 0.8201 |
| PFI8 | 0.8419 | 0.8563 | 0.8639 | 0.8746 | 0.8664 | 0.8599 |
| Fusion I | 0.9531 | 0.9574 | 0.9551 | 0.9551 | 0.9554 | 0.9438 |
| Fusion II | 0.9545 | **0.9585** | 0.9554 | 0.9571 | 0.9576 | 0.9478 |

(b)

| Range | $R$=1.0 | $R$=1.5 | $R$=2.0 | $R$=2.5 | $R$=3.0 | $R$=3.5 |
|---|---|---|---|---|---|---|
| PFI1 | 0.7995 | 0.8054 | 0.8068 | 0.7896 | 0.7676 | 0.7410 |
| PFI2 | 0.8433 | 0.8359 | 0.8291 | 0.8130 | 0.7868 | 0.7667 |
| PFI3 | 0.8904 | 0.8884 | 0.8797 | 0.8701 | 0.8537 | 0.8416 |
| PFI4 | 0.8509 | 0.8543 | 0.8371 | 0.8210 | 0.7970 | 0.7817 |
| PFI5 | 0.8291 | 0.8339 | 0.8258 | 0.8105 | 0.7862 | 0.7628 |
| PFI6 | 0.8924 | 0.8893 | 0.8718 | 0.8543 | 0.8396 | 0.8215 |
| PFI7 | 0.8475 | 0.8424 | 0.8201 | 0.7953 | 0.7673 | 0.7283 |
| PFI8 | 0.8820 | 0.8868 | 0.8690 | 0.8565 | 0.8407 | 0.8173 |
| Fusion I | 0.9514 | 0.9455 | 0.9367 | 0.9226 | 0.9110 | 0.9006 |
| Fusion II | **0.9548** | 0.9494 | 0.9407 | 0.9254 | 0.9167 | 0.9043 |

### 5.2. Comparison with the state of the art

In the literature, many tasks addressed the problem of 2D or 3D face recognition and used the FRGC v2.0 dataset for evaluation. Table 2 (a) lists performance comparisons between the proposed approach and several existing features only utilizing 2D texture images for face recognition; while Table 2 (b) lists a comparison between our method and the state of the art results only using facial range images or 3D face models for the same task.

Table 2: Comparisons with the state-of-the-art using only 2D or 3D face data.

(a)

| 2D Approaches | Rank-one RR |
|---|---|
| Eigenface [28] | 0.498 |
| LBP Histogram [28] | 0.718 |
| Gabor [28] | 0.779 |
| Original Texture + SIFT | 0.793 |
| Texture LBP Face + SIFT | 0.448 |
| Texture PFIs + SIFT | **0.959** |

(b)

| 3D Approaches | Rank-one RR |
|---|---|
| Chang et al. [29] | 0.919 |
| Mian et al. [24] | 0.935 |
| Mian et al. [23] | 0.962 |
| Huang et al [30] | 0.938 |
| Kakadiaris et al. | 0.970 |
| Huang et al [31] | 0.972 |
| Original Range + SIFT | NA |
| Range LBP Face + SIFT | 0.801 |
| Range PFIs + SIFT | **0.955** |

It should be noted that in Table 2 (a), results based on eigenface, LBP histogram, and gabor are directly cited from [28], because the same experimental protocol was adopted, and these results were achieved based on optimized parameters and a Sparse Representation Classifier (SRC). When SIFT-based matching was directly applied on original texture images, the accuracy is 79.3%, while if we exploited the proposed PFI instead, the accuracy was improved to 95.9%,

which highlights its effectiveness to describe local texture changes. We also investigated LBP face (8 sampling points and the same radius value as PFIs) in the same framework, but it only achieved a recognition rate of 44.8%.

On the other hand, Table 2 (b) compares our method with several existing systems for 3D face recognition. Similarly, we also operated SIFT-based matching on raw range faces, but it did not achieve a reasonable performance, because the detected keypoints on original range images are too limited, and usually located on the face border as shown in Figure 4. While, in this case, LBP face enhances the distinctiveness of facial range image, and improves the accuracy to 80.1%. As we can see in Table 2 (b), our performance is comparable to the state of the arts on 3D face recognition.

Both the sub-tables of Table 2 illustrate that the proposed PFIs enhanced local texture and shape variations leading to satisfying recognition results.

Table 3 compares the proposed approach with the state of the art on both face recognition and verification tasks using textured 3D face models. In recognition, the rank-one recognition rate of the proposed approach outperforms all the others, while in verification, our verification rate at FAR = 0.1% is only slightly weaker than that in [23]. However, in this work, we did not perform costly 3D face registration in the preprocessing step while an ICP based fine registration was applied in [23].

Table 3: Comparison with the state-of-the-arts using textured 3D face data.

| Systems | Rank-one RR | VR@FAR=0.1% |
|---|---|---|
| Mian et al. [24] | 0.961 | 0.986 |
| Mian et al. [23] | 0.974 | 0.993 |
| Gokberk et al [25] | 0.955 | NA |
| Xu et al. [26] | NA | 0.975 |
| Maurer et al. [13] | NA | 0.958 |
| Husken et al. [11] | NA | 0.973 |
| Ben Soltana et al. [28] | 0.955 | 0.970 |
| Texture PFI + SIFT | 0.959 | 0.973 |
| Range PFI + SIFT | 0.955 | 0.971 |
| Multi-Modal PFI +SIFT | **0.980** | **0.989** |

### 5.3. Robustness to facial expression variations

In this experiment, the probe face scans were divided into two subsets according to their expression labels to evaluate its insensitiveness to facial expression variations. The first subset contains face scans with the neutral expression; while the other with face scans possessing non-neutral expressions. As a result, besides the experiment of Neutral vs. All, two additional experiments of Neutral vs. Neutral and Neutral vs. Non-Neutral were also carried out. In the Neutral vs. Neutral and Neutral vs. Non-Neutral experiment, only the neutral and non-neutral probe subsets were used, respectively.

Using the same experimental protocol, we also compared the performance of the proposed method in face recognition with the one by Mian et al. [24] and Ben Soltana et al. [28] for robustness analysis on facial expression variations (see Table 4). The results of our approach are 99.6% and 96.0% for Neutral vs. Neutral and Neutral vs. Non-Neutral experiment, respectively. The recognition rate on the first subset is comparable to the state-of-the-art, while we make great progress on the second subset, displaying a 96.0% rank-one recognition rate. The degradation when non-neutral facial expression faces are included drops by 3.6%, which is much lower than 7.3% in [24] and 7.9% in [28] as both shape and texture clues are combined. These results suggest that our method tends to be insensitive to facial expression changes. Table 5 lists the robustness comparison in face verification task, similar conclusions can be drawn.

Meanwhile, the results of 2D modality are always slightly better than those of 3D based one. Figure 6 indicates verification rates by the ROC curves in these three experiments in Table 5.

Table 4: Rank-one results using expression protocol on the FRGC v2.0 dataset.

| | **Subset I** | **Subset II** | **Degradation** |
|---|---|---|---|
| Mian et al. [24] | 99.4% | 92.1% | 7.3% |
| Ben Soltana et al. [28] | 98.6% | 90.7% | 7.9% |
| Texture PFI + SIFT | 98.8% | 92.1% | 6.7% |
| Range PFI + SIFT | 98.5% | 91.7% | 6.8% |
| **Multi-Modal PFI +SIFT** | **99.6%** | **96.0%** | **3.6%** |

Subset I: Neutral vs. Neutral
Subset II: Neutral vs. Non-Neutral

Table 5: Comparison of verification rates at 0.001 FAR using the expression protocol on the FRGC v2.0 dataset.

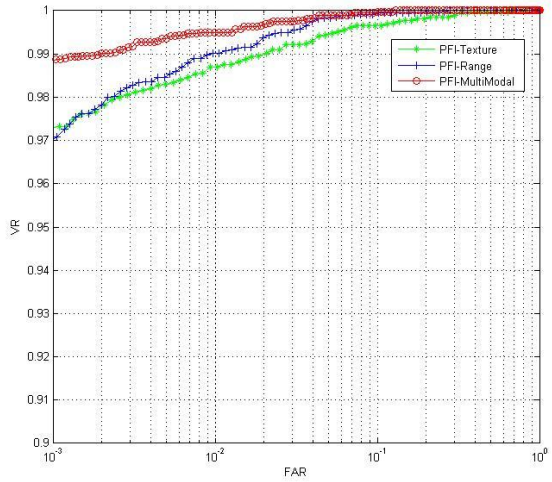
| | **VR I** | **VR II** | **VR III** |
|---|---|---|---|
| Mian et al. [24] | 97.4% | 99.9% | 92.7% |
| Texture PFI + SIFT | 97.3% | 99.7% | 93.7% |
| Range PFI + SIFT | 97.1% | 99.4% | 93.5% |
| **Multi-Modal PFI +SIFT** | **98.9%** | **99.9%** | **97.1%** |

VR I: Neutral vs. All
VR II: Neutral vs. Neutral
VR III: Neutral vs. Non-Neutral

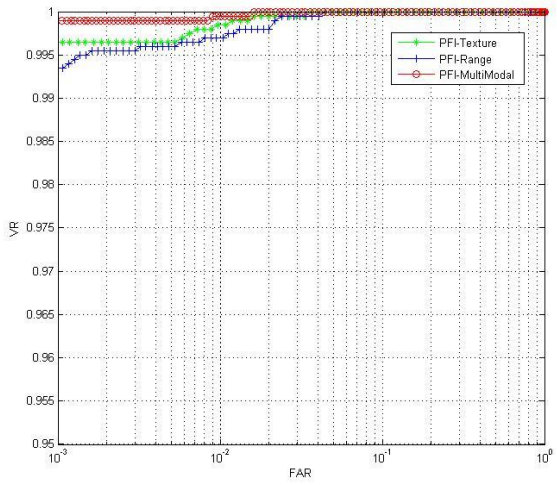
## 6. Conclusions

Biological vision is the basis of many local descriptors in the literature, e.g. SIFT [18], and probably the more recent DAISY [27]. This paper proposed a novel biological vision-inspired facial representation, namely Perceived Facial Images (PFIs), and we applied it to both facial texture and range images for the issue of textured 3D face recognition. As compared with other intermediate facial representations, e.g. Eigenface or Fischer face, these proposed PFIs simulate

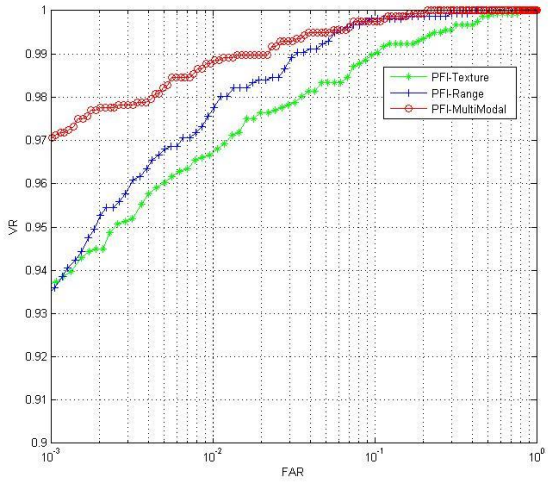
the response of complex neurons to gradient information in different orientations within a given neighborhood, thereby highlighting local details of range and texture face images and increasing their distinctiveness. As compared with LBP faces that also encode the difference between a central pixel and its neighbors, the PFIs are more informative as they take into account gradient information in several directions. PFIs are also likely less sensitive to noises than LBP faces because gradient information is summarized within a neighborhood convolved by a Gaussian kernel. Finally, PFIs also possess the properties of being robust to affine illumination and geometric transformations. Additionally, the designed score level fusion strategy further improved the final result when combining the results of PFIs of different orientations as well as fusing the accuracies of these two modalities. The experiments carried out on the FRGC v2.0 database showed the efficiency of the proposed method.




(**a**)




(**b**)




(**c**)

Figure 6: ROC curves using texture PFIs, range PFIs, and Multi-Modal PFIs respectively in the experiments with neutral faces enrolled: (a) Neutral vs. All. (b) Neutral vs. Neutral. (c) Neutral vs. Non-neutral.

## References


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

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

[3] A. Scheenstra, A. Ruifrok, and R. C. Veltkamp. A survey of 3D face recognition methods. International Conference on Audio- and Video- based Biometric Person Authentication, 2005.

[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.

[5] 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. International Conference on Pattern Recognition, 2000.

[6] 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.

[7] T. Papatheodorou and D. Reuckert. Evaluation of automatic 4D face recognition using surface and texture registration," IEEE International Conference on Automated Face and Gesture Recognition, 321–326, 2004.

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

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

[10] K. Chang, K. W. Bowyer, and P. Flynn. Face verification from 3d and grey level cues. Multimodal User Authentication Workshop, 25-32, 2003.
[11] 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.
[12] X. Lu and A. K. Jain. Integrating range and texture information for 3d face recognition. IEEE Workshop on Applications of Computer Vision, 2005.
[13] T. Maurer et al. Performance of geometrix activeID$^{TM}$ 3d face recognition engine on the FRGC data. IEEE Workshop on FRGC Experiments, 2005.
[14] M. Turk, and A. Pentland. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 13(1): 71–86, 1991.
[15] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7): 711–720, 1997.
[16] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(12): 2037–2041, 2006.
[17] S. Edelman, N. Intrator, and T. Poggio. Complex cells and object recognition. Unpublished manuscript: http://kybele.psych.cornell.edu/~edelman/archive.html, 1997.
[18] D. G. Lowe. Distinctive image features from scale invariant keypoints. International Journal of Computer Vision, 60(4): 91-110, 2004.
[19] D. Ruta and B. Gabrys. Classifier selection for majority voting. Journal of Information Fusion, 6(1): 63–81, 2005.
[20] Y. H. Said. On genetic algorithms and their applications. Handbook of Statistics, 2005.
[21] J. E. Baker. Reducing bias and inefficiency in the Selection algorithm. International Conference on Genetic Algorithms and their Application, 1987.
[22] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, K. I. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. International Conference on Computer Vision and Pattern Recognition, 2005.
[23] A. S. Mian, M. Bennamoun, and R. Owens. An efficient multimodal 2d-3d hybrid approach to automatic face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(11): 1927–1943, 2007.
[24] A. S. Mian, M. Bennamoun, and R. Owens. Keypoint detection and local feature matching for textured 3d face recognition. International Journal of Computer Vision, 79(1): 1–12, 2008.
[25] B. Gokberk, H. Dutagaci, A. Ulas, L. Akarun, and B. Sankur. Representation plurality and fusion for 3d face recognition. IEEE Transactions on Systems, Man and Cybernetics-B: Cybernetics, 38(1):155–173, 2008.
[26] C. Xu, S. Li, T. Tan, and L. Quan. Automatic 3D face recognition from depth and intensity Gabor features. Pattern Recognition, 42(9): 1895-1905, 2009.
[27] E. Tola, V. Lepetit, and P. Fua. A fast local descriptor for dense matching. IEEE International Conference on Computer Vision and Pattern Recognition, 2008.
[28] W. Ben Soltana, D. Huang, M. Ardabilian, L. Chen, and C. Ben Amar. Comparison of 2D/3D Features and Their Adaptive Score Level Fusion for 3D Face Recognition. 3D Data Processing, Visualization and Transmission, 2010
[29] K. I. Chang, K. W. Bowyer, P. J. Flynn. Adaptive rigid multi-region selection for handling expression variation in 3D face recognition. IEEE Workshop on Face Recognition Grand Challenge, 2005.
[30] D. Huang, G. Zhang, M. Ardabilian, Y. Wang, and L. Chen. 3D face recognition using distinctiveness enhanced facial representations and local feature hybrid matching. IEEE International Conference on Biometrics: Theory, Applications and Systems, 2010.
[31] D. Huang, M. Ardabilian, Y. Wang, and L. Chen. A Novel Geometric Facial Representation based on Multi-Scale Extended Local Binary Patterns. IEEE International Conference on Automated Face and Gesture Recognition, 2011.