Combining Face Averageness and Symmetry for 3D-based Gender Classification

Baiqiang Xia, Boulbaba Ben Amor, Hassen Drira, Mohamed Daoudi, Lahoucine Ballihi

Combining Face Averageness and Symmetry for 3D-based Gender Classification

Baiqiang Xia, Boulbaba Ben Amor, Hassen Drira, Mohamed Daoudi, and Lahoucine Ballihi.

Abstract

Although human face averageness and symmetry are valuable clues in social perception (such as attractiveness, masculinity/femininity, healthy/sick, etc.), in the literature of facial attribute recognition, little consideration has been given to them. In this work, we propose to study the morphological differences between male and female faces by analyzing the averageness and symmetry of their 3D shapes. In particular, we address the following questions: (i) is there any relationship between gender and face averageness/symmetry? and (ii) if this relationship exists, which specific areas on the face are involved? To this end, we propose first to capture densely both the face shape averageness (AVE) and symmetry (SYM) using our Dense Scalar Field (DSF), which denotes the shooting directions of geodesics between facial shapes. Then, we explore such representations by using classical machine learning techniques, the Feature Selection (FS) methods and Random Forest (RF) classification algorithm. Experiments conducted on the FRGCv2 dataset show a significant relationship exists between gender and facial averageness/symmetry when achieving a classification rate of 93.7% on the 466 earliest scans of subjects (mainly neutral) and 92.4% on the whole

FRGCv2 dataset (including facial expressions).
Keywords:
3D Face, Gender Classification, Face averageness, Face symmetry, Dense Scalar Field, Feature selection, Random Forest.

1. Introduction

Human gender perception is an extremely reliable and fast cognitive process since the face presents a clear sexual dimorphism [1]. In human face analysis using machines [3], automatic gender classification is an active research area. Developed solutions could be useful in human computer interaction (intelligent user interface, video games, etc.), visual surveillance, collecting demographic statistics for marketing (audience or consumer proportion analysis, etc.), and security industry (access control, etc.). Research on automatic gender classification using facial images goes back to the beginning of the 1990s. Since then, significant progress has been reported in the literature [4, 5, 6, 7, 8]. Fundamentally, proposed techniques differ in (i) the format of facial data (2D still images, 2D videos or 3D scans); (ii) the choice of facial representation, ranging from simple raw 2D pixels or 3D cloud of points to more complex features, such as Haar-like, LBP and AAM in 2D, and shape index, wavelets and facial curves in 3D; and (iii) the classifiers, for instance Neural Networks, SVM, and Boosting methods [4].

1.1. Related work on 3D-based gender classification

Statistically, the male and the female faces present different morphological characteristics in geometrical features, such as in the hairline, the forehead, the eyebrows, the eyes, the cheeks, the nose, the mouth, the chin, the jaw, the neck, the skin and the beard regions [13]. Usually, the female brow tends to be more arched than that of the male (which is more horizontal), the noses and chins in male faces are more prominent than those in female faces [27], and men have a more acute nasolabial angle than women [26]. The 3D face scans, which capture the spatial structure of the facial surfaces, allow to capture these differences between male and female faces more easily compared to 2D texture images. Thus, the goal of 3D-based gender classification is to develop a fast and automatic approach which yields high classification performance compared to the 2D-based approaches.

In [9], *Liu et al.* analyze the relationship between facial asymmetry and gender. They impose a 2D grid on each 3D face to represent the face with 3D grid points. With the selected symmetry plane, which equally separates the face into right and left halves, the distance difference between each point and its corresponding reflected point is calculated as height differences (HD). In addition, the angle difference between their normal vectors is calculated as orientation differences (OD). The approach based on HD-face achieves 91.16% and the approach based on OD-face achieves 96.22%. However, these performances are reported on a private dataset of 111 full 3D neutral face models of 111 subjects, and 3D face manual landmarks are needed.

In [12], *Lu et al.* use Support Vector Machine (SVM) to classify ethnicity (Asian and Non-Asian) and gender (Male and Female). A merging of two frontal 3D face databases (UND and MSU databases) is used for the experiments. The best gender classification results using 10-fold cross-validation reported is 91%. However, this approach is based on six landmarks (inside and outside corners of the eyes, the nose tip, and the chin point) manually labeled. Moreover, the results are obtained only on neutral faces.

In [15], *Wu et al.* use 2.5D facial surface normals recovered with Shape From Shading (SFS) from intensity images for gender classification. The best average gender recognition rate reported is 93.6% with both shape and texture considered. However, seven manual landmarks are needed and a small dataset of neutral scans has been used to perform the experiments.

In [16], *Hu et al.* propose a fusion-based gender classification method from 3D frontal faces. Each 3D face shape is separated into four face regions using face landmarks. With the extracted features from each region, the classification is done using SVM on a subset of the UND dataset and another database captured by themselves. Results show that the upper region of the face contains the highest amount of discriminating gender information. Fusion is applied to the results of four face regions and the best result reported is 94.3%. Their experiments only involve neutral faces. In this study, no attention is given to facial expressions.

In [3], *Toderici et al.* employ MDS (Multi-Dimensional Scaling) and wavelets on 3D face meshes for gender classification. They use the 4007 3D scans of the 466 subjects from the FRGCv2 dataset for gender classification. Experiments are carried out subject-independently with no common subject used in the testing stage of 10-fold cross validation. With polynomial kernel SVM, they achieve 93% gender classification rate with the unsupervised MDS approach, and 94% classification rate with the wavelets-based approach. Both approaches significantly outperform the kNN and kernel-kNN approaches.

In [17], *Ballihi et al.* extract facial curves (26 level curves and 40 radial curves) from 3D faces for gender classification. The features are extracted from lengths of geodesics between facial curves from a given face to the Male and Female templates computed using the Karcher Mean Algorithm. The Adaboost algorithm is then used to select salient facial curves. They obtained a classification rate of 84.12% with the nearest neighbor classifier when using the 466 earliest scans of the FRGCv2 dataset as the testing set. They also performed a standard 10-fold cross-validation for the 466 earliest scans of FRGCv2, and obtain 86.05% with Adaboost.

Compared to [17], in the current paper, we represent mathematically facial bilateral symmetry and averageness for gender classification using Dense Scalar Fields. The DSFs denoting the shooting directions for geodesics between facial shapes, are both novel and interesting. We view this representation for gender classification as the main contribution of this paper. The set of facial deformations is a nonlinear space while the set of Dense Scalar Field (DSF) is a vector space. The only remaining challenge is the large dimensionality of DSF, which is handled using a feature-selection-based dimension reduction, followed by a Random Forest classifier. In terms of experimental performances, the present approach have achieved higher classification rates compared to [17]. In summary, the novelty of this paper is in representing bilateral symmetry and face averageness using DSF and its successful application to the gender classification problem.

1.2. Methodology and contributions

From the above analysis, existing works on 3D-based gender classification are based on local or global *low-level* feature extraction (see table 2 for a complete summary) followed by classical classification methods. To the best of our knowledge, no work has been done considering *high-level* cues, such as face averageness and bilateral face symmetry, except the study in [9] which investigates the relationship between facial symmetry and gender. Using sparse measures of height differences (HD), and orientation differences (OD) on a defined grid imposed on full 3D face models, their process requires manual landmarks on the face and the experiments are performed on a small dataset. The main contributions of this work are as follows :

- ☞ We introduce two *high-level* features, face averageness (AVE) and bilateral face symmetry (SYM), for 3D-based gender classification. These primary facial perception features are rarely considered in the literature of facial attribute recognition.

- ☞ We provide an interesting mathematical tool, named *Dense Scalar Field* (DSF) [18], to capture densely and quantitatively the averageness/symmetry differences on the face surface. The DSFs grounding on Riemanniann shape analysis are capable to densely capture the shape differences in 3D faces (such as averageness/symmetry differences).

- ☞ We propose a fully-automatic gender classification without any human interaction. We achieve competitive results compared to the approaches in the state-of-the-art on a challenging dataset, FRGCv2. Also, we provide a comprehensive study of the robustness of the proposed approach against age, ethnicity and expression variations.

An overview of the proposed approach is shown in Figure 1. Firstly, during the first step an algorithm commonly used for facial scans preprocessing is applied. Its includes hole filling, facial part cropping and 3D mesh smoothing

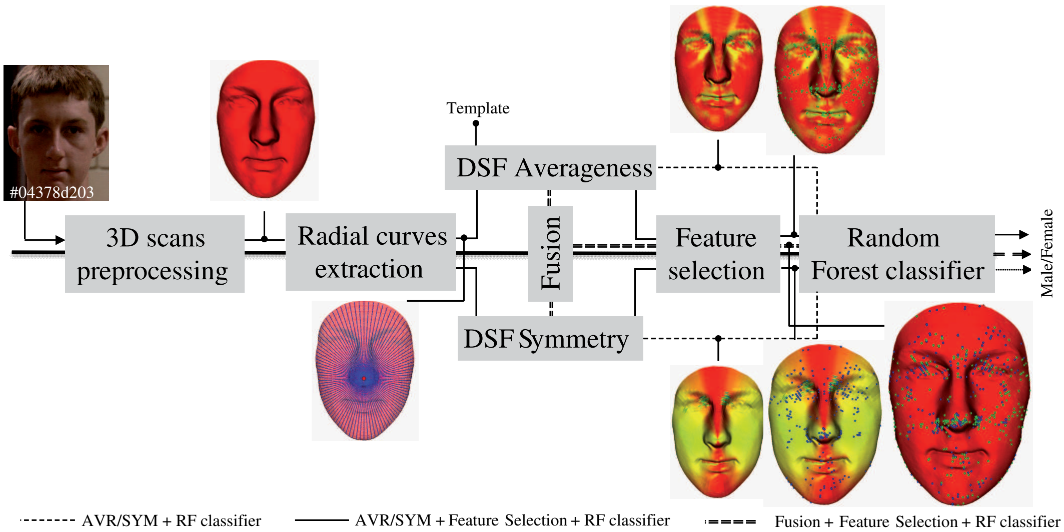

Figure 1: Flow chart of the proposed gender classification approach. There are various pipelines for gender classification. Namely, the pipelines are, (1) the symmetry DSF features *(SYM-Original)*, (2) the selected features of symmetry DSF features *(SYM-Selection)*, (3) the averageness DSF features *(AVE-Original)*, (4) the selected features of averageness DSF features *(AVE-Selection)*, (5) the fusion of symmetry and averageness DSF features by concatenation *(FUS-Original)*, and (6) the selected features of the fusion of symmetry and averageness DSF features *(FUS-Selection)*.

applied to each scan, together with nose tip detection and pose normalization, as proposed in [17] or [12]. We denote the preprocessed face as $\boldsymbol{S}$. The plane which equally separates the preprocessed face $\boldsymbol{S}$ into right and left halves is picked up as the middle plane. This plane $P(t, \overrightarrow{n_h})$ passes through the detected nose tip t and has a horizontal normal $\overrightarrow{n_h}$ from the frontal view. Secondly, a DSF extraction step goes after the preprocessing. Here, the preprocessed face $\boldsymbol{S}$ is approximated by a collection of radial curves defined over the facial region and stemming from the nose tip. Then, the *Dense Scalar Field* (DSF) features are computed, pair-wisely, to capture the shape differences (averagenesss/symmetry differences) between corresponding radial curves on each indexed point. Thus, we obtain two DSFs for each scan, an averageness DSF and a symmetry DSF. A fusion descriptor is then obtained for each scan by concatenating its averageness DSF and symmetry DSF. Thirdly, after DSF extraction, we investigate the two following classification pipelines. In the first pipeline, Random Forest classifier is applied directly on the obtained feature vectors - averageness DSFs, symmetry DSFs and fusion DSFs. In the second pipeline, we first apply a supervised feature selection (FS) algorithm on the averageness, symmetry and their fusion DSFs, then the Random Forest (RF) classifier is applied on the selected features for gender classification.

This work relates closely to the work previously published in [17], in terms of face representation by an indexed collection of radial curves, which is one of the first steps of our approach's pipeline. However, while this face parameterization is in common, the feature extraction step is completely different. Indeed, in [17], the features are extracted from ***lengths of geodesics*** between facial curves from a given face to the Male and Female templates. In contrast, this work considers the ***shooting vectors on the geodesics*** between facial curves to capture shape differences. The DSFs are computed to describe densely the *Symmetry* and *Averageness* of a given face. This allows to compute densely and and locally the facial features on each point of the face.

The rest of the paper is organized as follows: in section **2**, we highlight our methodology for extracting features that contain 3D facial averageness/symmetry difference; in section **3**, we detail the classifier, the feature selection method, and the fusion method for gender classification; experimental results and discussions are presented in section **4** while section **5** concludes the work.

2. Feature Extraction Methodology

As mentioned earlier, after the preprocessing, the next step of our approach is to extract densely the averageness and symmetry features from faces. Both of them are based on a Riemannian shape analysis of 3D face.

2.1. Background on Dense Scalar Field Computation

The idea to capture locally and densely face asymmetry and its averageness is to represent facial surface $\boldsymbol{S}$ by a set of parameterized radial curves emanating from the nose tip $\boldsymbol{t}$. Such an approximation can be seen as a solution to facial surface parameterization which approximates the local shape information. Then, a Dense Scalar Field (DSF), based on pairwise shape comparison of corresponding curves, is computed along these radial curves on each point. A similar framework has been used in [18] for 4D face expression recognition by quantifying deformations across 3D face sequences followed by a classification technique. More formally, a parametrized curve on the face, $\beta : I \to \mathbb{R}^3$, where $I = [0, 1]$, is represented mathematically using the *square-root velocity function* [19], denoted by $q(t)$, according to: $q(t) = \frac{\dot{\beta}(t)}{\sqrt{\|\dot{\beta}(t)\|}}$. This specific parameterization has the advantage of capturing the shape of the curve and providing simple calculus [19].

Let us define the space of such functions: $\mathcal{C} = \{q : I \to \mathbb{R}^3, \|q\| = 1\} \subset \mathbb{L}^2(I, \mathbb{R}^3)$, where $\|\cdot\|$ implies the $\mathbb{L}^2$ norm. With the $\mathbb{L}^2$ metric on its tangent spaces, $\mathcal{C}$ becomes a Riemannian manifold. Given two curves q_1 and q_2, let ψ

denote a path on the manifold $\mathcal{C}$ between q_1 and q_2, $\dot{\psi} \in T_{\psi}(\mathcal{C})$ is a tangent vector field along the path $\psi \in \mathcal{C}$. In our case, as the elements of $\mathcal{C}$ have a unit $\mathbb{L}^2$ norm, $\mathcal{C}$ is a hypersphere of the Hilbert space $\mathbb{L}^2(I, \mathbb{R}^3)$. The geodesic path ψ^* between any two points $q_1, q_2 \in \mathcal{C}$ is simply given by the minor arc of great circle connecting them on this hypersphere, $\psi^* : [0, 1] \to \mathcal{C}$, given by:

$$\psi^*(\tau) = \frac{1}{\sin(\theta)} \left(\sin((1-\tau)\theta)q_1 + \sin(\theta\tau)q_2\right) \tag{1}$$

and $\theta = d_{\mathcal{C}}(q_1, q_2) = cos^{-1}(\langle q_1, q_2 \rangle)$. We point out that $sin(\theta) = 0$ if the distance between the two curves is null, in other words $q_1 = q_2$. In this case, for each τ, $\psi^*(\tau) = q_1 = q_2$. The tangent vector field along this geodesic $\dot{\psi}^* : [0, 1] \to T_{\psi}(\mathcal{C})$ is given by (2):

$$\dot{\psi}^* = \frac{d\psi^*}{d\tau} = \frac{-\theta}{\sin(\theta)} \left(\cos((1-\tau)\theta)q_1 - \cos(\theta\tau)q_2\right) \tag{2}$$

Knowing that on a geodesic, the covariant derivative of its tangent vector field is equal to 0, $\dot{\psi}^*$ is parallel along the geodesic ψ^* and we shall represent it with $\dot{\psi}^*|_{\tau=0}$. This vector $\dot{\psi}^*|_{\tau=0}$ represents the initial velocity of the geodesic path connecting q_1 to q_2 and called also the shooting vector for this geodesic. Accordingly, (2) becomes:

$$\dot{\psi}^*|_{\tau=0} = \frac{\theta}{\sin(\theta)} \left(q_2 - \cos(\theta)q_1\right) \tag{3}$$

with $\theta \neq 0$. Thus, $\dot{\psi}^*|_{\tau=0}$ is sufficient to represent this vector field; the remaining vectors can be obtained by parallel transport of $\dot{\psi}^*|_{\tau=0}$ along the geodesic ψ^*. with the magnitude of $\dot{\psi_\alpha}^*$ at each point, located in curve β_α^S

with index k, we build a *Dense Scalar Field* (DSF) on the facial surface $\boldsymbol{S}$, $V_\alpha^k = |\dot{\psi}_\alpha^*|_{(\tau=0)}(k)|$. This *Dense Scalar Field* quantifies the shape difference between corresponding curves on each indexed point.

2.2. Face symmetry description

The idea of the face symmetry description is to capture the bilateral symmetry difference in the face by DSF. Symmetry difference is defined as the deformation from a face point to its corresponding symmetrical point on the other side of face. In practice, symmetry DSF is calculated on each indexed point of the corresponding symmetrical curves in the preprocessed face $\boldsymbol{S}$. Let β_α denote the radial curve that makes an angle α with the middle plane $P_{\boldsymbol{S}}(t, \overrightarrow{n_h})$ from the frontal view of $\boldsymbol{S}$, and $\beta_{2\pi-\alpha}$ denotes the corresponding symmetrical curve that makes an angle $(2\pi-\alpha)$ with $P_{\boldsymbol{S}}(t, \overrightarrow{n_h})$. The tangent vector field $\dot{\psi_\alpha}^*$ that captures the deformation from β_α to $\beta_{2\pi-\alpha}$ is then calculated. With the magnitude of $\dot{\psi_\alpha}^*$ at each point, located in the curve β_α with index k, we build a *symmetry Dense Scalar Field* (symmetry DSF) on the facial surface.

This *Dense Scalar Field* quantifies the shape difference between corresponding symmetrical curves on each point of the preprocessed face $\boldsymbol{S}$. Some examples illustrating this symmetry descriptor are shown in Figure 2. For each subject, face in column (a) shows the 2D intensity image; column (b) illustrates the preprocessed 3D face surface $\boldsymbol{S}$; column (c) illustrates the the 3D face $\boldsymbol{S}$ with extracted curves; column (d) shows the symmetry degree as a color-map of the DSF mapped on $\boldsymbol{S}$. The color bar is shown in the up-right corner. The hot colors mean the minimum difference (i.e. maximum symmetry) and cold colors signify the maximum difference (i.e. minimum

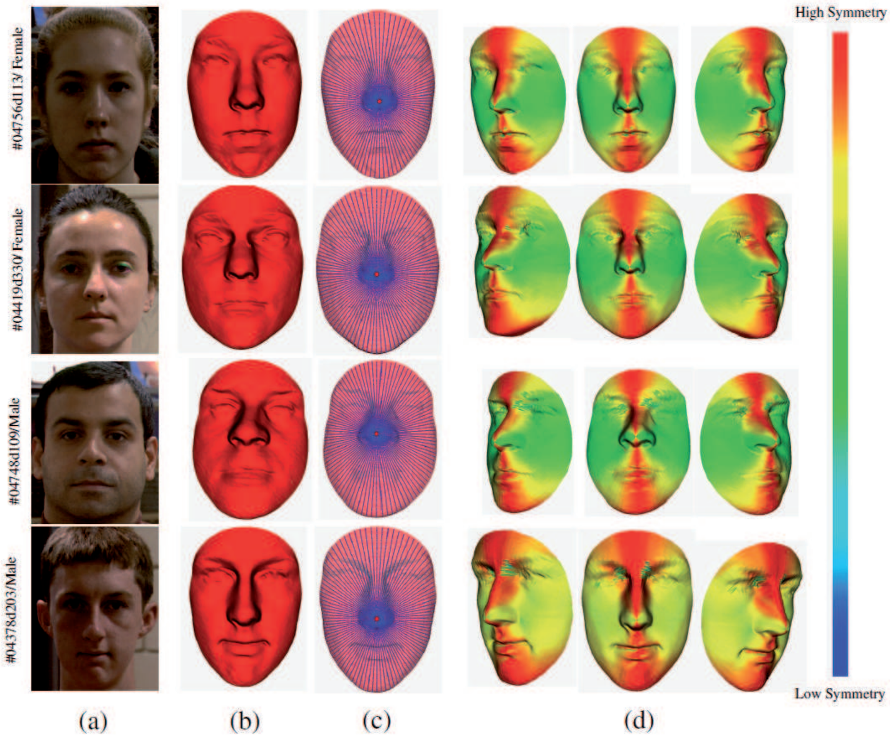

Figure 2: Illustrations of the symmetry DSFs on faces. (a) 2D intensity image; (b) preprocessed 3D face $\boldsymbol{S}$; (c) 3D face $\boldsymbol{S}$ with extracted curves; (d) color-map of symmetry DSF mapped on $\boldsymbol{S}$ with three poses. While the cold colors reflect lower symmetrical regions, the warm colors represent higher symmetrical parts of the face.

symmetry). The hotter the color, the higher is magnitude of the bilateral symmetry. In this work, the symmetry DSFs are generated with 200 radial curves extracted from each face and 100 indexed points on each curve. Thus, the size of each DSF is 20000. The average time consumed for extracting all 200 curves for each face is 1.048 seconds, and for generating the bilateral symmetry descriptor (symmetry DSF) on all the 200×100 points of each face is 0.058 seconds. The average preprocessing time consumed for each

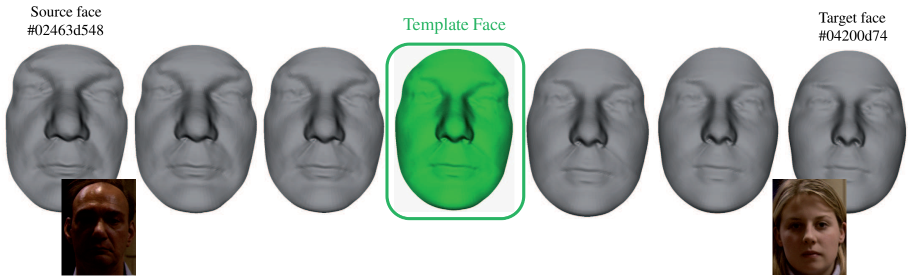

Figure 3: The averageness face template is defined as the middle point of the geodesic path between two representative faces randomly taken from the male and female classes in the FRGCv2 dataset.

scan is 0.116 seconds. The total computation time (including preprocessing) for each scan is less than 1.25 seconds. All our programs are developed in C++ and executed on Intel Core i5 CPU 2.53 GHZ with 4Go of RAM.

2.3. Face averageness description

As mentioned earlier, generally, male faces have more prominent features (forehead, eyebrows, nose, mouth, etc.) in comparison with female faces. Here, our aim is to capture the morphologcial sexual differences between male and female faces by comparing their shape differences to a defined face template. We assume that such differences change with the face gender. Thanks to DSF, presented in subsection 2.1, we are able to capture densely such shape differences as long as a face template is defined.

As shown in Figure 3, the face template is defined as the middle point of the geodesic path which connects a male face *(ID: 02463d548; Age: 48; White)* to a female face *(ID: 04200d74; Age: 21; White)* taken from the FRGCv2 dataset. With the two faces represented by collections of radial curves, we compute pair-wisely the geodesic path between corresponding curves using equation (1). By interpolation, we have the middle point of the geodesic which we take as the face template $\boldsymbol{T}$.

For a preprocessed face $\boldsymbol{S}$, let $\beta_\alpha^{\boldsymbol{S}}$ denote the radial curve that makes an angle α with the middle plane $P_{\boldsymbol{S}}(t, \overrightarrow{n_h})$ from the frontal view of $\boldsymbol{S}$, and β_α^T denotes the curve that makes the same angle α with $P_T(t, \overrightarrow{n_h})$ in the averageness face template $\boldsymbol{T}$. The tangent vector field $\dot{\psi_\alpha}^*$ that represents the projection of the deformation between the given face and the template face, in the tangent space associated with the template face, is then calculated on each point. Similar to the symmetry descriptor, with the magnitude of $\dot{\psi_\alpha}^*$ at each point, located in curve β_α^S with index k, we build an *averageness Dense Scalar Field* (averageness DSF) on the facial surface, $V_\alpha^k = |\dot{\psi}_\alpha^*|_{(\tau=0)}(k)|$. This *Dense Scalar Field* quantifies the shape difference between corresponding curves of $\boldsymbol{S}$ and $\boldsymbol{T}$ on each indexed point.

Figure 4 shows this averageness descriptor. For each subject, the face in column (a) shows the 2D intensity image; column (b) illustrates the preprocessed 3D face surface $\boldsymbol{S}$; column (c) shows the 3D face $\boldsymbol{S}$ with extracted curves; column (d) shows color-map of the Averageness DSF mapped on $\boldsymbol{S}$ with three poses. The hot colors mean the minimum difference (i.e. maximum averageness) and cold colors signify the maximum difference (i.e. minimum averageness). The hotter the color, the higher is the magnitude of the averageness.

3. Gender classification

In this work, face averageness and symmetry are different types of information in the 3D facial shapes. Each of them provides a perspective (maybe

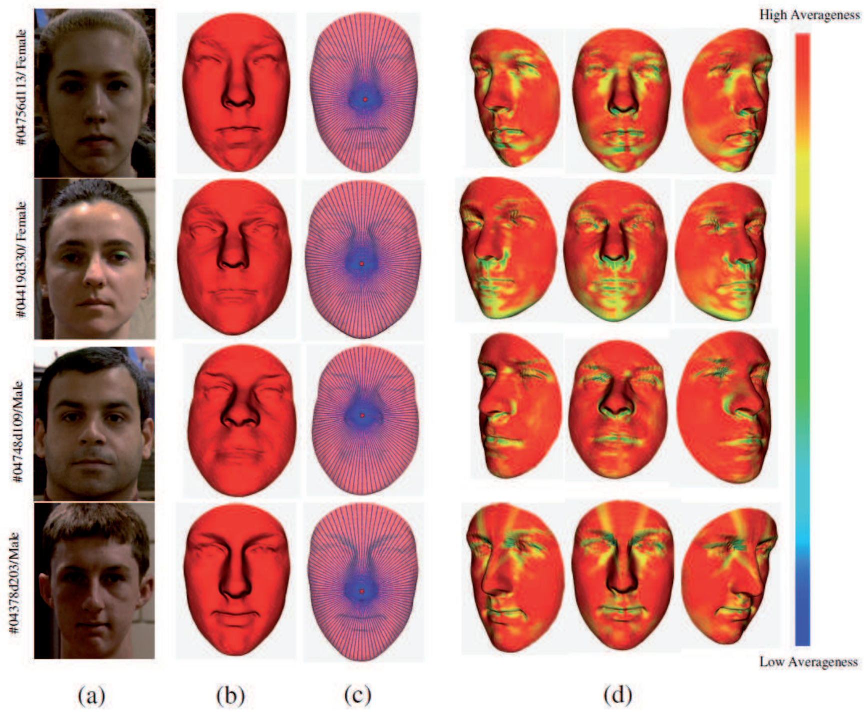

Figure 4: Illustrations of the averageness DSFs on faces. (a) 2D intensity image; (b) preprocessed 3D face surface $\boldsymbol{S}$; (c) the 3D face $\boldsymbol{S}$ with extracted curves; (d) color-map of the Averageness DSF mapped on $\boldsymbol{S}$ with three poses. While the cold colors reflect lower averageness, the warm colors represent higher averageness on the face.

correlated perspectives) in face perception. Thus, we first study individually their relationship with gender, then we combine them to find out if it enhances the gender classification results, which means that they contribute to gender classification in different ways. In practice, we use an *early fusion method* which consist in concatenating the *averageness DSF* and *symmetry DSF* features of each scan, to form the *fusion DSF* description. Then, we explore the performance of the Random Forest algorithm with the *avera-*

geness DSF, the *symmetry DSF* and the *fusion DSF* in different scenarios, in combination of Feature Selection methods. It has been demonstrated by *Perez et al.* in [29], that different types of information (such as gray scale intensity, range image and LBP texture) contributes to face based gender classification differently, and the fusion of multi-information yields a better classification performance.

3.1. Feature Selection

The size of the features is another important characteristic of the approach. As pointed out by *Bekios-Calfa et al.* in [28], in limited computational resource contexts, such as the mobiles, the development of resource-limited algorithms is important for applications of computer vision and pattern recognition. In their work, they make use of LDA techniques to reduce feature size. In our work, we use feature selection methods to select a much smaller set of the features to reduce the computational cost. Compared with LDA techniques, feature selection methods do not tranferm the meaning and values of feature, thus they allow to track back to the corresponding point on the face.

Feature subset selection is the process of identifying and removing as much irrelevant and redundant information as possible [22]. It is a central problem in machine learning. The earliest approaches for feature selection were ***the filter*** methods. These algorithms use heuristics based on general characteristics of the data to evaluate the merit of feature subsets. Another school of approaches argues that the bias of a particular induction algorithm should be taken into account when selecting features. This method, called ***the wrapper*** [23], uses an induction algorithm along with a statistical resampling technique such as cross-validation to estimate the final accuracy of feature subsets. The filter methods operate independently of any learning algorithm. The undesirable features are filtered out of the data before the learning begins. They are generally much faster than wrapper methods, especially on data of high dimensionality. Since the averageness, symmetry and fusion DSFs are really dense and possibly redundant after DSF extraction, we use a feature selection procedure on the DSFs to get rid of the irrelevant and redundant features. For the merits of filter methods, we chose a filter, named Correlation-based-Feature-Selection (CFS) [22]. It is an algorithm that couples the evaluation formula based on an appropriate correlation measure and a heuristic search strategy. The central hypothesis of CFS is that good feature sets should contain features that are highly correlated with the class, yet uncorrelated with each other. The feature evaluation formula (Pearsons correlation coefficient), based on ideas from test theory, provides an operational definition of this hypothesis. Within CFS, we try two heuristic search strategies, the Best-First search strategy and the Greedy-Step-Wise search strategy. The Best-First search strategy [24] is an AI search strategy that allows back-tracking along the search path. It moves through the search space by greedy hill-climbing augmented with a back-tracking facility. When the path being explored becomes non-improving, the Best-First search will back-track to a more promising previous subset and continue the search from there. The stopping criterion is the number of consecutive non-improving nodes (5 in our experiments) that result in no improvement. For Greedy-Step-Wise, it performs a greedy forward or backward search through the space of attribute subsets. It stops when the addition/deletion of any remaining attributes results in a decrease in evaluation.

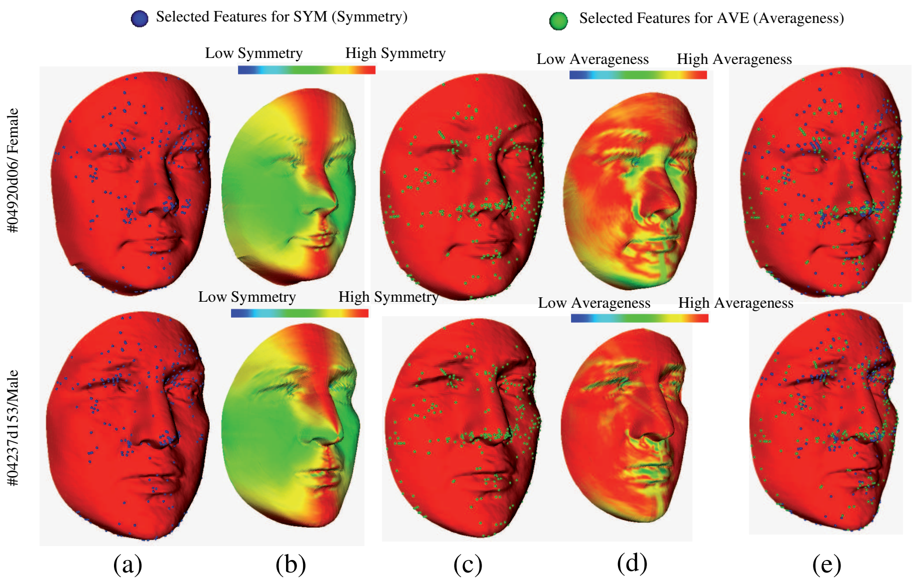

Figure 5: Feature selection. (a) selected points of symmetry DSF in the face; (b) color-map of original symmetry DSF; (c) selected points of averageness DSF in the face; (d) color-map of original averageness DSF; (e) selected points of both averageness DSF and symmetry DSF in face.

After Feature selection, we retain 301 salient points for averageness DSF, 271 salient points for symmetry DSF, and 365 salient points for the fusion. The feature selection procedure significantly reduces the size and complexity of original DSF description. Figure 5 shows the selected features of averageness DSF and symmetry DSF in faces. Column (a) maps the selected features of symmetry DSF in the face; Column (b) shows the color-map of original symmetry DSF on the face ; Column (c) maps the selected points of averageness DSF in the face ; Column (d) shows the original averageness DSF on the face; Column (e) maps the selected points of both averageness DSF and symmetry DSF in the face. For both averageness DSF and symmetry DSF, we observe dense distribution of salient points around the nose and eyes regions. More salient points exist in forehead regions in averageness DSF, and more salient points exist in cheek regions in symmetry DSF. These observations show that averageness DSF and symmetry DSF share both similarities and differences. In other words, they are complementary in face description.

3.2. Gender classification based on Random Forest

Face-based gender classification is a binary classification problem which estimates the gender c of a given test face into Male or Female $\mathbf{c} \in \{Male, Female\}$. We carry out gender classification experiments with the well-known machine learning algorithm, Random Forest. Random Forest is an ensemble learning method that grows many classification trees $t \in \{t_1, .., t_T\}$ [25]. To classify a new face from an input vector (DSF-based feature vector $v = V_\alpha^k$), each tree gives a classification result and the forest chooses the classification having the most votes. In the growing of each tree, firstly, $\boldsymbol{N}$ instances are sampled randomly with replacement from the original data, to make the training set. Then, if each instance comprises of $\boldsymbol{M}$ input variables, a constant number $\boldsymbol{m}$ $\boldsymbol{(m<<M)}$ is specified. At each node of the tree, $\boldsymbol{m}$ variables are randomly selected out of the $\boldsymbol{M}$ and the best split on these $\boldsymbol{m}$ variables is used to split the node. The process goes on until the tree grows to the largest possible extent, without pruning.

The performance of the forest depends on the correlation between any two trees, and the strength of each individual tree. The forest error rate increases when the correlation decreases, or the strength increases. Reducing $\boldsymbol{m}$ reduces both the correlation and the strength. Increasing it increases both.

Thus, an optimal $\boldsymbol{m}$ is needed for the trade-off between the correlation and the strength. In Random Forest, the optimal value of $\boldsymbol{m}$ is found by using the oob-error rate (out-of-bag-error rate). It is reported that face classification by Random Forest achieves a lower error rate than some popular classifiers, including SVM [20]. As far as we know, there is no reported work in the literature of face-based gender classification using Random Forest.

4. Experiments

The FRGCv2 database was collected by researchers from the University of Notre Dame [21] and contains 4007 3D face scans of 466 subjects with differences in gender, ethnicity, age and expression. For gender, there are 1848 scans of 203 female subjects and 2159 scans of 265 male subjects. The ages of subjects range from 18 to 70, with 92.5% in the $18-30$ age group. When considering ethnicity, there are 2554 scans of 319 White subjects, 1121 scans of 99 Asian subjects, 78 scans of 12 Asian-southern subjects, 16 scans of 1 Asian and Middle-east subject, 28 scans of 6 Black-or-African American subjects, 113 scans of 13 Hispanic subjects, and 97 scans of 16 subjects subjects whose ethnicity are unknown. About 60% of the faces have a neutral expression, and the others show expressions of disgust, happiness, sadness and surprise. All the scans in FRGCv2 are near-frontal. With this dataset, we conducted two experiments. The first one is to examine the robustness of our approach to age and ethnicity variations. It uses the 466 earliest scan of each subject in FRGCv2, of which more than 93% are neutral-frontal. The second one extends to examine the robustness of our approach to variations of expression. It considers all the 4007 scans in FRGCv2, about

40% of which are expressive faces. For these experiments, the results are generated in a subject-independent fashion, using a 10-fold cross-validation setup.

4.1. Data preprocessing

The 3D face models present some imperfections, such as the holes (caused by the absorption of the laser in the dark areas like eyebrows and eyes and by the self-occlusions), the hair, and the spikes (caused by acquisition noise). Thus, a preprocessing step is needed to limit their influence. Firstly, through boundary detection, link-up and triangulation, holes are filled in each scan. Secondly, since the scans in FRGCv2 are all near-frontal, the nose tip is detected with a simple algorithm. The nose tip is detected by analyzing the peak point of the face scan in the depth direction. Then, the mesh is cropped with a sphere centered at the nose tip to discard the hair and the neck regions. Finally, a smoothing filter is used to distribute evenly the 3D vertices which capture the original 3D shape. We next perform the well-known Iterative Closest Point (ICP) algorithm to normalize the poses of the obtained meshes according to a reference mesh (frontal). The symmetry plane is then picked up as the plane that has as origin the nose tip and has an horizontal normal. In practice, the preprocessing step is performed automatically on the whole FRGCv2 dataset without any manual intervention. We obtained 4005 well preprocessed scans after preprocessing. The failed two scans (with scan id 04629d148 and 04815d208) were resulted from wrong nose tip detection. Considering the ratio of failure is rather tiny (2/4007<0.0005), we omit the influence of the two failed scans for the results generation.

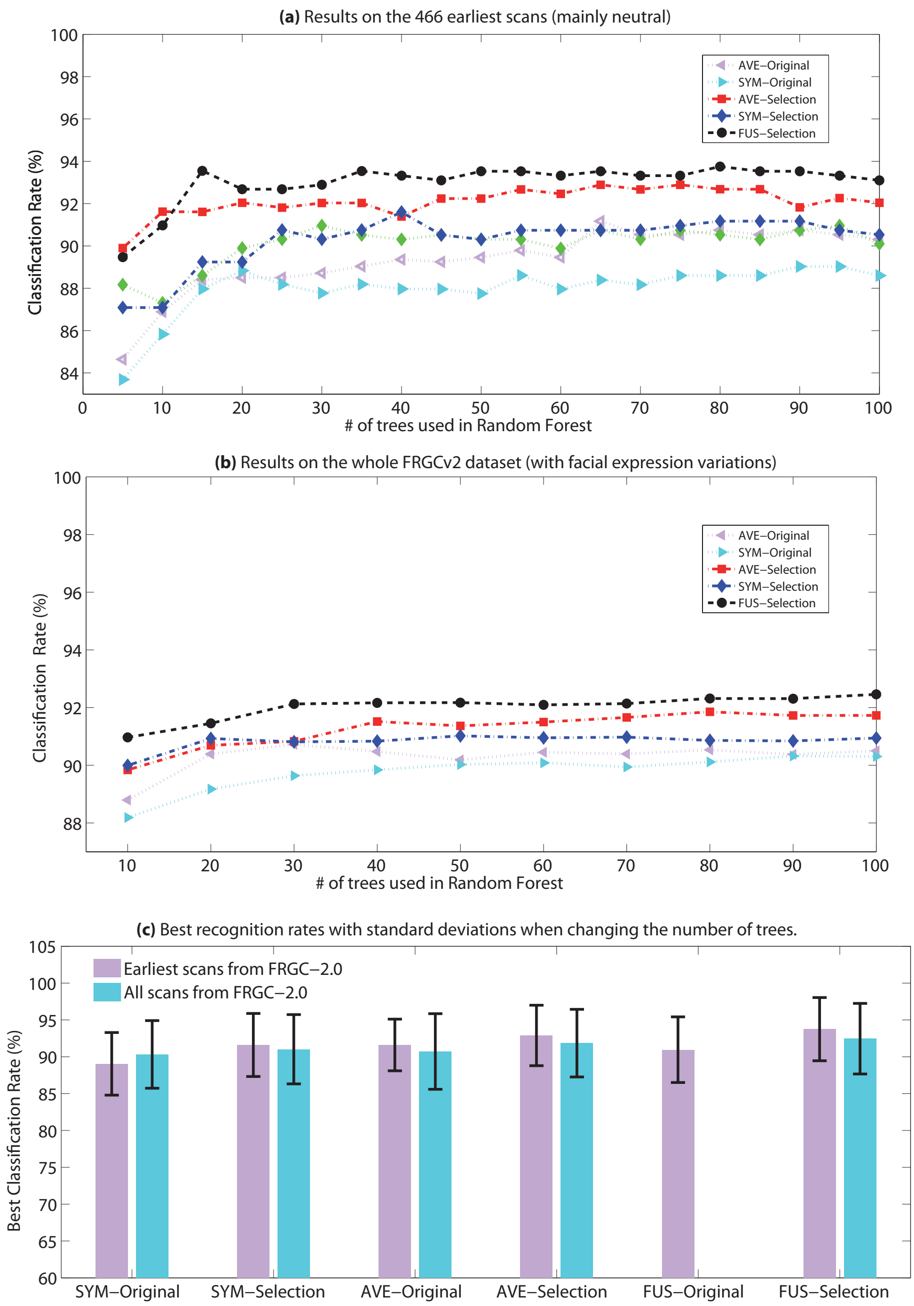

Figure 6: The reported results of the proposed methods[1] using Random Forest with different number of trees.

4.2. Robustness to variations of age and ethnicity

Among the 466 earliest scans, 431 scans are neutral-frontal and 35 are expressive-frontal. In our 10-fold cross validation setup, the 466 scans are randomly partitioned into 10 folds with each fold containing $46-47$ scans. In each round, 9 of the 10 folds are used for training while the remaining fold is used for testing. The average recognition rate and standard deviation for 10 rounds then give a statistically significant performance measure. The relationship between the gender classification result and the number of trees used in the Random Forest is depicted in Figure 6(a). It demonstrates that a significant relationship exists between gender and facial averageness and facial symmetry considered separately. We note also that both the fusion and the feature selection improve the gender classification results. In fact, the fusion descriptor outperforms individual averageness and symmetry descriptor. This implies that facial averageness and symmetry relate to gender in different ways. At the same time, results after the feature selection always override the results without feature selection. This means that the original averageness DSF and symmetry DSF contain redundant information. Gender-related features are distributed unequally in the facial regions. The best gender classification rate is 93.78%, achieved by 80-Tree Random Forest with the fusion descriptor after feature selection. This result is detailed in the confusion matrix in Table 1. The recognition rate for females (92.02%) is

[1]Methods as described in Figure 1 : (1) the symmetry DSF features *(SYM-Original)*, (2) the selected features of symmetry DSF features *(SYM-Selection)*, (3) the averageness DSF features *(AVE-Original)*, (4) the selected features of averageness DSF features *(AVE-Selection)*, (5) the fusion of symmetry and averageness DSF features by concatenation *(FUS-Original)*, and (6) the selected features of the fusion of symmetry and averageness DSF features *(FUS-Selection)*.

slightly lower than for male ones (95.44%). It is probably due to the fact that more male faces were used for training. We also performed a 10-fold 100-repetition experiment with Random Forest under the same setting, which resulted at an average classification rate of 92.84% with a standard deviation of 3.58%.

Table 1: Confusion matrix of RF-based classification.

%	Female	Male
Female	91.63	8.37
Male	4.56	95.44
Recognition Rate	$=93.78 \pm 4.29\%$	

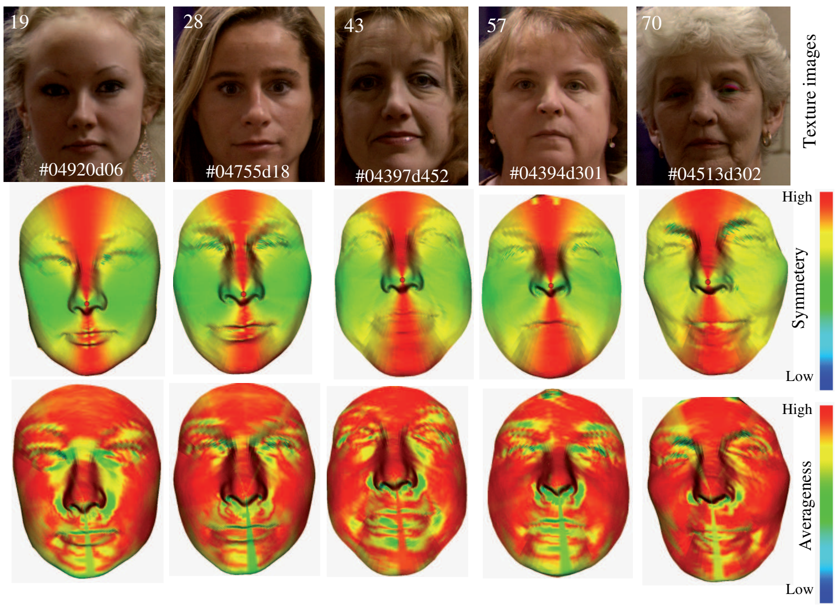

Figure 7: DSFs on faces with different Age.

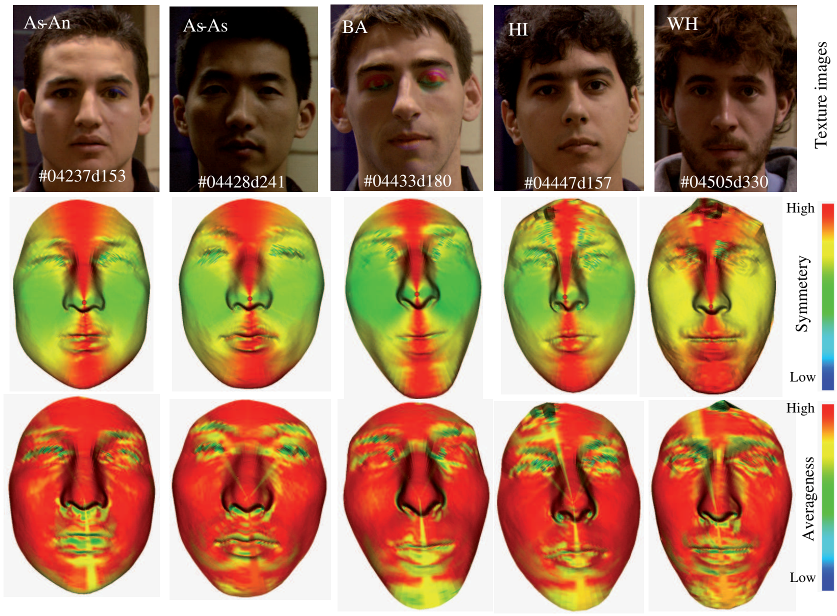

Figure 8: DSFs on faces with different Ethnicity.

Figure 7 illustrates the color-maps of symmetry DSF and averageness DSF on female faces with age differences and Figure 8 illustrates the color-maps of symmetry DSF and averageness DSF on male faces with differences in ethnicity. The information related to age, ethnicity and identity of scans are presented in the 2D images in the upper row of each figure. Based on the middle rows of Figure 7 and Figure 8, we can observe that the bilateral symmetry of both genders convey a visually symmetrical pattern, where the color-map of left-face is globally in symmetry with the right-face, although subtle local asymmetry exists. Low-level deformations (red color) are usually located near the middle plane and high-level deformations (yellow and green colors) happen more frequently in further areas. The asymmetry, in female faces, change obviously more smoothly than in male faces. On the other hand, with the lower rows of Figure 7 and Figure 8, we observe that female faces exhibit more deformations in mouth, nose and eye regions to deform from the averageness face template. More subtly, in cheek and forehead regions, the color is more consistent in male faces. All of these observations above stay relatively consistent with changes of age and ethnicity. We believe that these common patterns contribute to the robustness of our approach to variations of age and ethnicity to some extent.

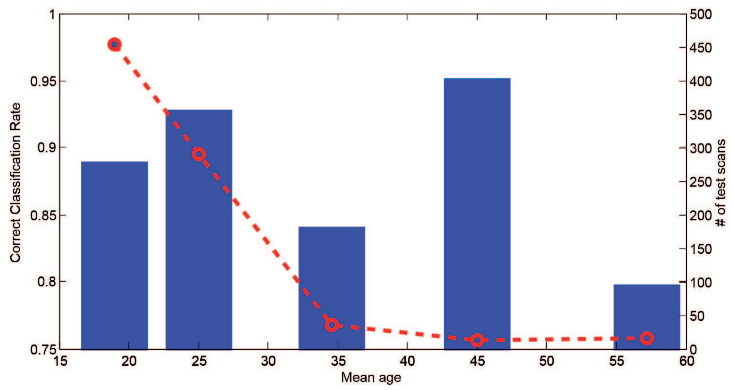

Figure 9: Gender classification results of different age group (the blue bars show the average recognition rate of each age group, and the red line shows the number of scans in this age group).

As it is well known that face perception is strongly affected by age [30], we provide Figure 9 to analyze gender classification performance for different age groups. In this figure, the blue bars show the average recognition rate for each age group, and the red line shows the number of scans in the same age group. We could confirm that gender classification is strongly influenced by the age. Generally, although the gender classification results decrease from above 90% to about 80% when increasing the age, all these results are near or above 80%. That is to say the performance of our approach stays relatively high with age variation. Moreover, due to unbalanced age distribution of scans in FRGCv2 dataset, we see the number of scans decreased significantly when the age is increased. We assume that this is also a reason for the decrease of the gender classification results.

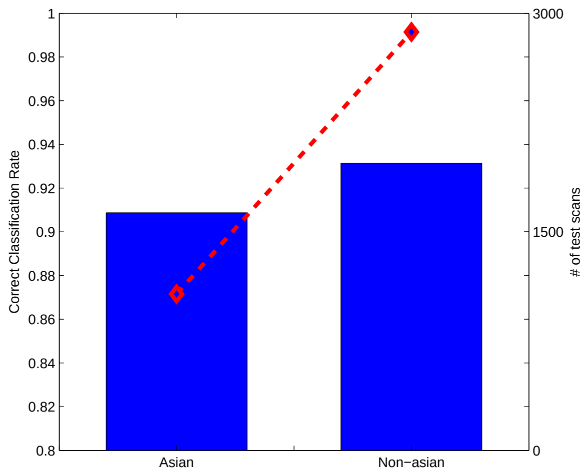

Figure 10: Gender classification results of different ethnicity group (the blue bars show the average recognition rate of each age group, and the red line shows the number of scans in this ethnic group).

Figure 10 analyzes the relationship between the obtained classification rate when varying the ethnicity. Here, the whole FRGCv2 dataset is separated into Asian and Non-Asian groups. We can see that the gender classification rates, shown by the blue bars, stay above 90% when varying the ethnicity. The classification rate of Non-Asian group is $3-4$ percent higher than that of the Asian group. This is probably due to a more sufficient training step has been involved with Non-Asian group, since it contains more than two times of the number of the scans of the Asian group, as shown in the figure by the red line.

4.3. Robustness to expression variations

In this experiment, with all the preprocessed scans of FRGCv2, we first performed the DSF extraction for averageness, symmetry and fusion descriptors, and then did the 10-fold subject-independent cross-validation with Random Forest. For each round, the scans of 46 subjects are randomly selected for testing, and the scans of the remaining subjects are dedicated to the training. For all the 10 rounds of experiments, no common subjects are used in training/testing. The relationship between the classification result and the number of trees used in Random Forest is shown in Figure 6(b). We note again that both fusion and feature selection improve the results. The best result achieved with the fusion and feature selection is 92.46% $\pm$ 4.79 with 100-Tree Random Forest. We argue this result by the fact that the majority of the selected features are located on the facial areas which are less affected by the expressions in particular the nose, the eyebrows, and the forehead as illustrated in Figure 5. Considering the FRGCv2 dataset is a challenging dataset which contains as many as 4007 scans with various changes in age, ethnicity and expression, we claim even more confident that a significant relationship exists between gender and 3D facial averageness/symmetry, and our method is effective and robust to ethnicity and expression variations.

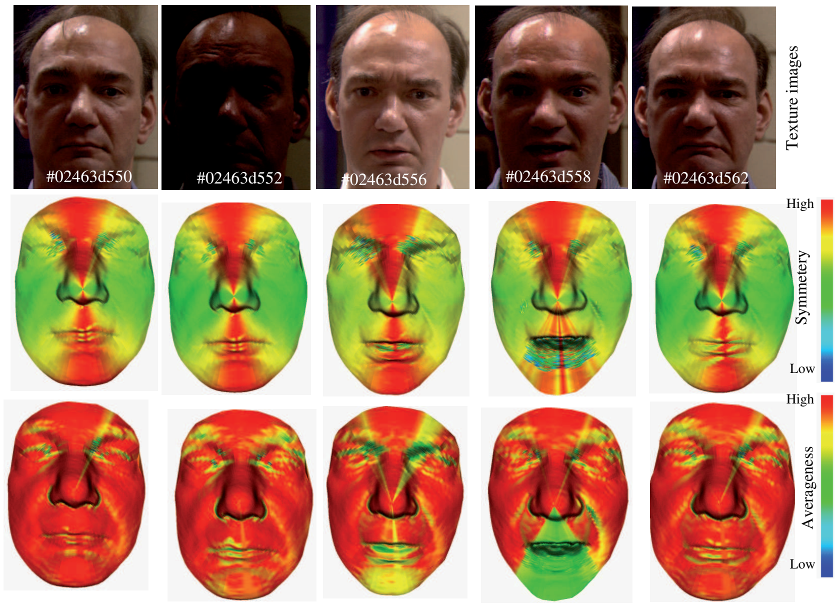

Figure 11: DSFs on faces with different expressions.

Figure 11 shows color-maps of DSFs generated for a subject with different expressions. Similar to the observations in Figure 7 and Figure 8, we perceive again in the middle row of Figure 11 that the symmetry deformations on both sides of the face are globally in symmetry, although tiny local asymmetry exists in areas like eye corners and lips. Low-level deformations (red) always locate near the middle plane and high-level deformations (yellow and green) occur more frequently in farther areas. With the lower rows of Figure 7 and Figure 11, we observe again that female faces require more deformation in mouth, nose and eye regions to deform from the averageness face template. In cheek and forehead regions, the color is more consistent in male faces. All these visible patterns do not change significantly with expression variations. We assume that these patterns contribute to the robustness of our approach to expression changes. Figure 6(c) shows the best gender recognition results (shown as bars) and their standard deviation (shown as black lines) in our experiment. It shows that the gender recognition rate increases with both fusion and feature selection, and the performances of all the approaches change little between the 466 earliest scans protocol and the whole FRGCv2 dataset protocol. It means our approach is even relatively robust to the size of the training set.

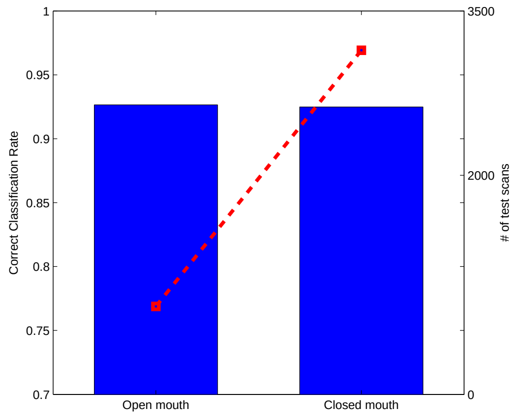

Figure 12: Gender classification results of different expression group (the blue bars show the average recognition rate of each age group, and the red line shows the number of scans in this expression group).

Again, in Figure 12, we illustrate the effects of expression variations on the proposed approach. We separated the FRGCv2 dataset into Open-mouth and Closed-mouth groups. Despite the fact of the unbalanced number of training scans in Open-mouth and Closed-mouth groups, as shown by the red line, the results shown by the blue bars in the figure are all above 90%, and the results between these two groups are comparable with each other.

4.4. Comparison with state of the art

Table 2 gives a comparison of this work with previous studies in 3D-based gender classification. With differences in the dataset, landmarking, experiment settings and so on, it is difficult to compare and rank these works simply according to the result values. Compared with our work, works in [9], [14], [15] are based on relatively smaller dataset which leave doubts about the statistical significance of their performances on larger and more challenging datasets. Works in [9], [12], [14], [15] require manual landmarking, thus they are not fully-automatic. Works in [9], [14], [15], [16] use different experimental settings other than the most prevailing 10-fold cross-validation. Our work addressed gender classification in a fully automatic way without manual landmarking. Experimented on a large dataset, FRGCv2, which contains challenging variations in expression, age and ethnicity, and reached competitive results with literature. The nearest works to ours are done by *Ballihi et al.* in [17] and *Toderici et al.* in [3]. With the 466 Earliest scans of FRGCv2 and standard 10-fold cross-validation, *Ballihi et al.* achieved 86.05% classification rate, while we achieved a much higher result of 93.78% by combining facial shape averageness and bilateral asymmetry. In [3], *Toderici et al.* also performed automatic 10-fold cross-validation on the FRGCv2 dataset in a subject-independent fashion. In general, we have achieved comparable results than them. They achieve about 1% higher gender classification rate than us. While we achieve a lower standard deviation which signifies better stability of the algorithm than theirs[2].

[2]During the work, we found 8 scans of a subject (id 04662, female indeed) had been mislabeled as male in the FRGCv2 metadata. We corrected them before the experiments.

Table 2: Comparison of our approach to earlier studies.

Reference	*Dataset*	*Auto*	*Features*	*Classifiers*	*Experiment settings*	*Results*	*Shape/ Texture*
Ballihi et al. [17]	466 *earliest scans of FRGCv2*	*Yes*	*facial curves*	*Adaboost*	*10-fold cross-validation*	86.05%	*Shape*
Toderici et al. [3]	*All scans of FRGCv2*	*Yes*	*Wavelets*	*Polynomial-SVM*	10*-fold cross-validation*	*Male :* $94 \pm 5\%$ *Female :* $93 \pm 4\%$	*Shape*
Hu et al. [16]	*729 UND scans and 216 private scans*	*Yes*	*Curvature based shape index*	*RBF-SVM*	5*-fold cross-validation*	94.03%	*Shape*
Han et al. [14]	*61 3D scans in GavabDB*	*No*	*Geometry Features*	*RBF-SVM*	5*-fold cross-validation*	$82.56 \pm 0.92\%$	*Shape*
Wu et al. [15]	*Needle maps of* 260 *subjects from UND*	*No*	*PGA features*	*Posterior Probability*	*200 train/60 test, 6 repetitions*	$93.6 \pm 4\%$	*Shape+ Texture*
Lu et al. [12]	1240 *scans from UND and MSU*	*No*	*Grid element values*	*Posterior Probability*	10*-fold cross-validation*	$91 \pm 3\%$	*Shape+ Texture*
Liu et al. [9]	*111 full 3D scans of 111 subjects*	*No*	*Variance Ratio in HD and OD faces*	*linear classifier*	*half train/ half test, 100 repetitions*	*HD:*$91.16 \pm 3.15\%$ *OD:*$96.22 \pm 2.30\%$	*Shape*
Our work[1]	*466 earliest scans of FRGCv2*	*Yes*	*AVE+SYM DSFs*	*Random Forest*	10*-fold cross-validation*	$93.78 \pm 4.29\%$	*Shape*
Our work[2]	***All scans of FRGCv2***	*Yes*	*AVE+SYM DSFs*	*Random Forest*	*10-fold cross-validation*	$92.46 \pm 3.58\%$	*Shape*

5. Conclusion

In this paper, we have proposed a fully automatic approach based on 3D facial averageness/symmetry differences for gender classification. We have proposed to use our *Dense Scalar Fields* grounding on Riemannian Geometry to capture densely facial averageness and its bilateral symmetry. The remaining challenge is the large dimensionality of the DSFs, which is handled using a feature-selection-based dimension reduction, followed by a Random Forest classifier. Despite the wide range of age, ethnicity and facial expressions, our method achieves a gender classification result of 93.78% $\pm$ 4.29% with 466 earliest scans of subjects, and 92.46% $\pm$ 3.58 on the whole

FRGCv2 dataset. We have also demonstrated that a significant relationship exists between the gender and these two high-level cues in face perception, the face averageness and symmetry. Our approach is competitive with state-of-the-art approaches. One of the limitations of the proposed approach is the dependence on near-frontal pose of faces to compute the symmetry and the averageness DSFs.

References

[1] A. Cellerino and D. Borghetti and F. Sartucci, "Sex differences in face gender recognition in humans", *Brain Research Bulletin*, vol. 63, 2004, pp. 443-449.

[2] V. Bruce and AM. Burton and E. Hanna and P. Healey and O. Mason and A. Coombes and R. Fright and A. Linney, "Sex discrimination: how do we tell the difference between male and female faces?", *Perception*, vol. 22, 1993, pp. 131152..

[3] G. Toderici and S. O'Malley and G. Passalis and T. Theoharis and I. Kakadiaris, "Ethnicity- and Gender-based Subject Retrieval Using 3-D Face-Recognition Techniques", *International Journal of Computer Vision*, vol. 89, 2010, pp. 382-391.

[4] J. Ylioinas and A. Hadid and M. Pietikinen, "Combining Contrast Information and Local Binary Patterns for Gender Classification", *Image Analysis*, vol. 6688, 2011, pp. 676-686.

[5] E. Makinen and R. Raisamo, "An experimental comparison of gender

classification methods", *Pattern Recognition Letters*, vol. 29, 2008, pp. 1544-1556.

[6] W. Yang and C. Chen and K. Ricanek and C. Sun, Changyin, "Gender Classification via Global-Local Features fusion", *Biometric Recognition*, vol. 7098, 2011, pp. 214-220.

[7] C. Shan, "Learning local binary patterns for gender classification on real-world face images", *Pattern Recognition Letters*, vol. 33, 2012, pp. 431-437.

[8] N. Kumar and A. Berg and P.N. Belhumeur and S. Nayar, "Describable Visual Attributes for Face Verification and Image Search", *Pattern Analysis and Machine Intelligence*, vol. 33, 2008, pp. 1962 -1977.

[9] Y. Liu and J. Palmer, "A quantified study of facial asymmetry in 3D faces", *Analysis and Modeling of Faces and Gestures*,2003, pp. 222-229.

[10] LG. Farkas and G. Cheung, "Facial asymmetry in healthy North American Caucasians. An anthropometrical study", *Angle Orthod*,vol. 51, 1981, pp. 70-77.

[11] A. Little and B. Jones and C. Waitt and B. Tiddeman and D. Feinberg and D. Perrett and C. Apicella and F. Marlowe, "symmetry Is Related to Sexual Dimorphism in Faces: Data Across Culture and Species", PLoS ONE, vol. 3, 2008, pp. e2106 .

[12] X. Lu and H. Chen and A. Jain, "Multimodal facial gender and ethnicity identification", *Proceedings of the 2006 international conference on Advances in Biometrics*, 2006, pp. 554-561.

[13] "The main differences between male and female faces", *www.virtualffs.co.uk.*

[14] X. Han and H. Ugail and I. Palmer, "Gender Classification Based on 3D Face Geometry Features Using SVM", *CyberWorlds*,2009, pp. 114-118.

[15] J. Wu and W. A. P. Smith and E. R. Hancock, "Gender Classification using Shape from Shading",*International Conference on Image Analysis and Recognition*, 2007, pp. 499-508.

[16] Y. Hu and J. Yan and P. Shi, "A fusion-based method for 3D facial gender classification", *Computer and Automation Engineering (ICCAE)*, vol. 5, 2010, pp. 369-372.

[17] L. Ballihi and B. Ben Amor and M. Daoudi and A. Srivastava and D. Aboutajdine, "Boosting 3D-Geometric Features for Efficient Face Recognition and Gender Classification", IEEE Transactions on Information Forensics & Security, vol. 7, 2012, pp. 1766-1779.

[18] H. Drira and B. Ben Amor and M. Daoudi and A. Srivastava and S. Berretti, "3D Dynamic Expression Recognition based on a Novel Deformation Vector Field and Random Forest", *21st International Conference on Pattern Recognition*, 2012.

[19] A. Srivastava and E. Klassen and S. H. Joshi and I. H. Jermyn, "Shape Analysis of Elastic Curves in Euclidean Spaces", *Pattern Analysis and Machine Intelligence*, vol. 33, 2011, pp. 1415 -1428.

[20] A. Z. Kouzani and S. Nahavandi and K. Khoshmanesh, "Face classification by a random forest", *TENCON 2007-2007 IEEE Region 10 Conference*, 2007, pp. 1-4.

[21] P. J. Phillips and P. J. Flynn and T. Scruggs and K. W. Bowyer and J. Chang and K. Hoffman and J. Marques and J. Min and W. Worek, "Overview of the face recognition grand challenge", *Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 947 - 954.

[22] Mark A. Hall, "Correlation-based Feature Subset Selection for Machine Learning", *PhD thesis, Department of Computer Science, University of Waikato*, 1999, chapter 3-4.

[23] R. Kohavi, "Wrappers for Performance Enhancement and Oblivious Decision Graphs". *PhD thesis, Stanford University*, 1995, chapter 4.

[24] E. Rich and K. Knight, "Artificial Intelligence", *McGraw-Hill College*, 1991.

[25] L. Breiman, "Random Forests", *Machine Learning*, vol. 45, 2001, pp 5-32.

[26] Lines PA, Lines RR, Lines CA., "Profilmetrics and facial esthetics". *Am J Orthod*, 1978, 73:640-57.

[27] Bradley N. Lemke, "Surgical Anatomy of the Face". *Arch Opthalmol*, 1995, 113(8):982.

[28] Bekios-Calfa J, Buenaposada JM, Baumela L., "Revisiting linear dis-

criminant techniques in gender recognition". *IEEE Trans Pattern Anal Mach Intell*, 2011 Apr, 33(4):858-64.

[29] C. Perez, J. Tapia, P. Estvez, C. Held., "Gender Classification From Face Images Using Mutual Information and Feature Fusion". *International Journal of Optomechatronics - INT J OPTOMECHATRONICS*, 2012, vol. 6, no. 1, pp. 92-119.

[30] Ferrario VF, Sforza C, Ciusa V, Dellavia C, Tartaglia GM, "The effect of sex and age on facial asymmetry in healthy subjects: a cross-sectional study from adolescence to mid-adulthood", *J Oral Maxillofac Surg*, 2001 Apr, 59(4):382-8.