

Received November 5, 2020, accepted November 22, 2020, date of publication December 3, 2020, date of current version December 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3042299

Deep Neural Networks With Distance Distributions for Gender Recognition of 3D Human Shapes

HUI WANG¹, XIAOYANG LIN¹, NANNAN LI²,
FAN MENG¹, JUNJIE CAO³, (Member, IEEE), AND XIUPING LIU³

¹School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang 050043, China

²School of Information Science and Technology, Dalian Maritime University, Dalian 116026, China

³School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China

Corresponding author: Nannan Li (nannanli@dlmu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61972267, Grant 61802045, Grant 12004265, Grant 61976040, and Grant 61772104; in part by the Natural Science Foundation of Hebei Province under Grant F2018210100 and Grant A2020210001; in part by the Youth Talent Support Program of Universities of Hebei Province under Grant BJ2018003; in part by the Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, under Grant VRLAB2020A04; and in part by the National Science and Technology Major Project under Grant 2018ZX04041001-007 and Grant 2018ZX04016001-011.

ABSTRACT Automatic human gender recognition is an important and classical problem in artificial intelligence. Most of the previous gender recognition works are based on vision appearance and biometric characteristics. However, there are fewer gender recognition approaches for 3D human shapes. In this article, we propose a novel deep neural network learning method for gender recognition of 3D human shapes. Firstly, we introduce effective descriptors to distinguish male and female of 3D human shapes via probability distributions of biharmonic distances among points. Secondly, the above distances-based low-level descriptors are fed into a fully connected neural network for gender recognition. Furthermore, we construct a larger 3D human shape dataset for evaluation of the proposed gender recognition method by collecting and labeling human shape models. Compared with previous works, our method obtains higher recognition accuracy and has more advantages, such as posture invariant, robust to noises, and no need of landmarks or pre-alignment process.

INDEX TERMS Gender recognition, 3D human shape, neural network.

I. INTRODUCTION

Gender is an important physiological and demographic attribute of people. Many real-life applications utilize gender information such as human-computer interaction, surveillance system, commercial development, video game, and social security. For example, automatic gender recognition of 3D human shape would make the body virtual fitting system more user-friendly. Thus, human gender recognition has received wide attention in the past decades. Comprehensive reviews are provided in [1], [2]. Based on types of features that differentiate between male and female, previous gender recognition works can be mainly divided into vision appearances and biometric characteristics approaches [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Lei Wei¹.

With the rapid development of laser scanning, technologies of gender recognition from 3D human shapes are also investigated [3]–[6]. Geodesic distances between some manually placed anthropometric landmarks for each of the scanned subject from the 3D human dataset CAESAR (Civilian American and European Surface Anthropometry Resource) [7] are used for gender classification [3]. However, this method is lack of generality. Since the other 3D human datasets usually don't have manual given landmarks, and computation of automatic landmarks is difficult. Positions of the landmarks are also treated as feature vectors in the K-Nearest Neighbor (KNN) or Support Vector Machine (SVM) classifier for gender recognition [5]. However, this kind of approach is posture variants, which only can classify genders of human shapes in the same posture. The proposed gender recognition methods [4], [6] are also posture variant, and use the

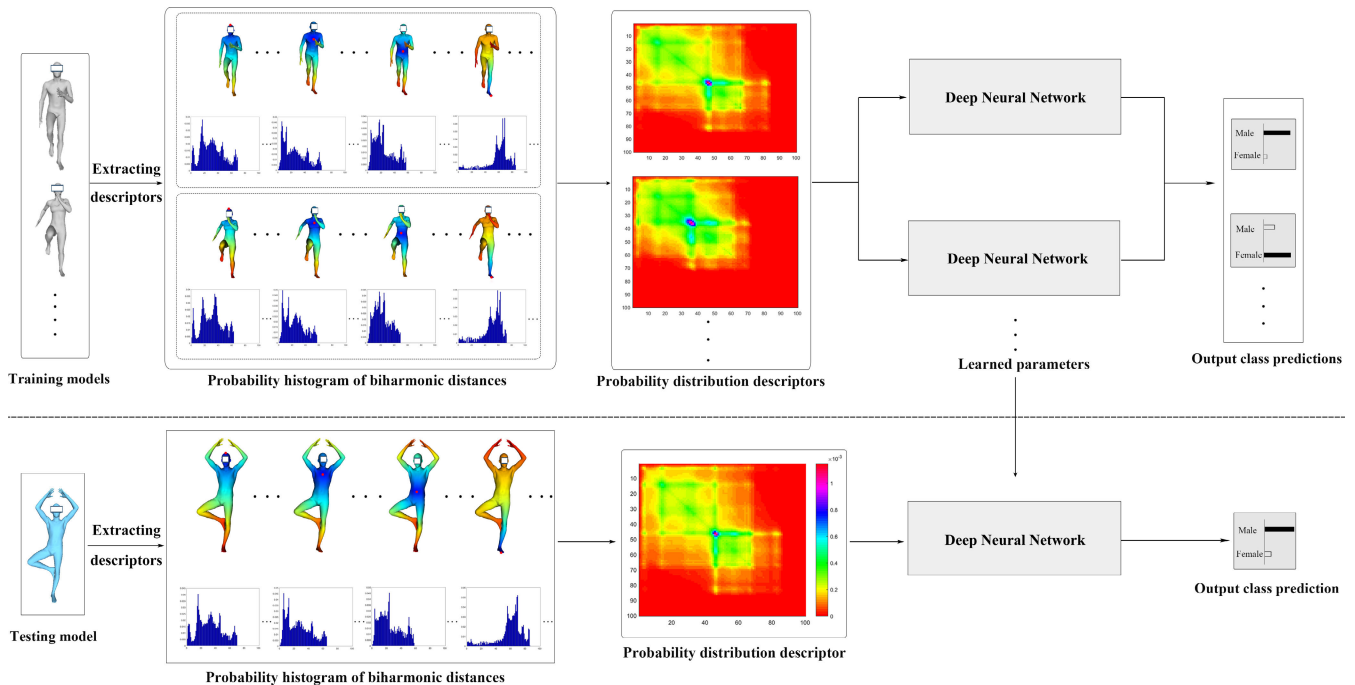


FIGURE 1. The framework of our proposed method mainly consists of two steps: extracting probability distribution descriptors based on biharmonic distances and deep neural network for gender recognition. The top and bottom parts are the training and testing process respectively.

traditional classifiers such as SVM, KNN, and Extra Trees Classification (ETC).

In view of the limitations of previous works, such as posture variant and need of landmarks or pre-alignment process, we propose a novel deep neural network method for gender recognition of 3D human shapes shown in Figure 1. Firstly, a new geometric descriptor, which is posture invariant, robust to noises, no need of landmarks or pre-alignment process, and can process shapes with different numbers of vertices, via probability distributions of biharmonic distances is proposed for distinguishing gender. Secondly, considering the success of geometric deep learning [8], the above distances based descriptors are fed into a fully connected neural network for gender classification. The proposed method obtains higher accuracy than previous works. Our main technical contributions can be summarized as follows:

- Propose an effective descriptor using probability distributions of biharmonic distances for distinguishing male and female 3D human shapes, which is posture invariant and no need of landmarks.
- Introduce a novel deep neural network from the above distance-based low-level descriptors for gender recognition of 3D human shapes.
- Construct a larger 3D human shape dataset with more than one hundred thousand of shapes for the evaluation of gender classification and future research.

The rest of this article is organized as follows. Related works are summarized in Section II. In Section III, our deep neural network learning method for gender recognition of 3D shapes is introduced. Experimental results and performance evaluation of our approach are described in

Section IV. At last, we give a conclusion and point out some possible future works.

II. RELATED WORK

A. GENDER RECOGNITION FROM 3D HUMAN SHAPES

Compared with the previous works of gender recognition based on vision appearances and biometric characteristics [1], [2], there are fewer approaches of gender recognition from 3D human shapes in the past [3]–[6].

Pairwise geodesic distances between the 73 manually placed anthropometric landmarks on shapes from the 3D human dataset CAESAR are used as features for gender recognition, where the classifiers are the traditional linear discriminant function, bayesian decision boundary, and SVM [3]. The anthropometric landmarks limit the generalization of this method for the other shape datasets, which usually are lack of consistent landmarks.

The other hand-crafted features are also used for gender recognition of 3D human shapes, such as distributions of normals on the chest regions [4], coordinates of the 73 landmarks of shapes from CAESAR dataset [5], and positions of all vertices [6]. Different classifiers are conducted on the above hand-crafted features, such as SVM, KNN, ETC, random forest, and so on. However, these three methods are posture variant, i.e., only can recognize genders of shapes in the same posture.

In addition to the disadvantages of posture variant and requiring landmarks or pre-alignment, another limitation of the above previous works is that the training datasets with about one or two thousand shapes are a bit smaller, which restricts the generalization ability of the learning. Our method

is both posture invariant and no need of landmarks or pre-alignment process and is trained on a larger dataset with more than one hundred thousands of 3D human shapes.

B. DESCRIPTORS VIA GEOMETRIC DISTANCES

Being invariant to isometric deformations makes the intrinsic distances, for example, geodesic [9] and Laplacian spectral distances [10], natural and useful for analysis of non-rigid 3D shapes.

The geodesic distance between two points is the length of the shortest path constrained on the shapes. Geodesic distances between the manually placed anthropometric landmarks have been used for gender recognition of 3D human shapes [3]. A deep learning method for 3D shapes retrieval with geodesic moments is proposed in [9]. However, the geodesic distances have several drawbacks such as being sensitive to topological noises and not globally shape-aware.

Spectral distances, e.g., biharmonic [11], diffusion [12], and commute-time [13] distances, can be defined through a filtering of the Laplacian eigenfunctions and eigenvalues, which are widely used in several applications [10]. The distributions of all pairwise spectral distances are used for unsupervised 3D shape recognition [14], [15], where dissimilarity criteria of the discrete histograms are the well defined χ^2 or earth mover distances. Descriptors of the biharmonic and commute-time distances are also used for shape retrieval [13], [16]. In this article, we propose a novel descriptor for classifying genders of 3D human shapes using the probability distributions of biharmonic distances from each point.

C. GEOMETRIC DEEP LEARNING

In the past years, deep learning, and in particular Convolutional Neural Network (CNN), has shown the ability to learn powerful image features from large collections of examples. Different from images whose grid-based representation is simple and regular, 3D geometric shapes usually have a variety of complex representations. Thus, there is no an universal framework for 3D geometric deep learning, and some classical works are proposed, such as deep learning from low-level hand-crafted features [17], multi-view CNN [18], voxel-based 3D CNN [19], point-based CNNs [20], graph or manifold based CNN [21] and so on. Please refer to the survey [8] for details. Each of the above 3D geometric deep learning approaches has advantages and disadvantages based on the applications. Another factor limiting the usability of deep learning on geometric data is that acquisition of 3D data is much more difficult than the 2D images and videos. In this article, the deep neural network is operated on the low-level biharmonic distance-based features for gender recognition of 3D human shapes.

III. METHODS

Because the 3D human shapes usually have different connectivity, it is difficult to generalize the CNN on the geometric shapes directly. The descriptors via probability density

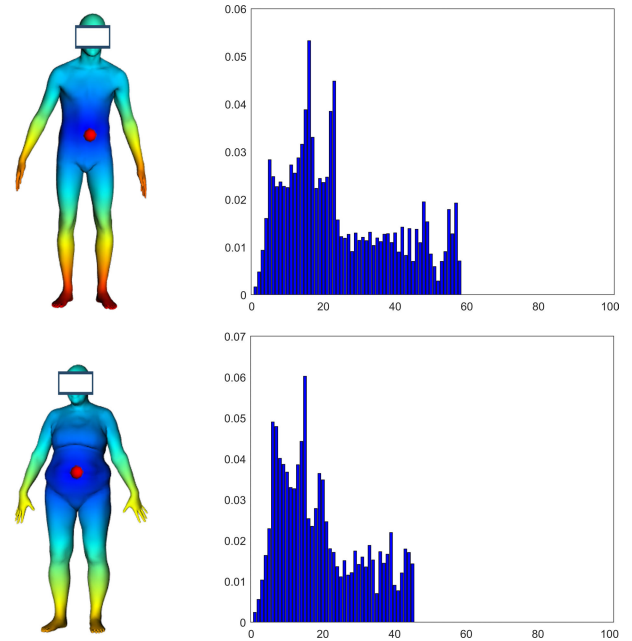


FIGURE 2. Left: the biharmonic distances from the red source vertices, where red color indicates a higher value and blue represents a lower value. Right: the corresponded probability density functions, which is different between on a male shape at the top row and a female shape at the bottom one.

functions of biharmonic distances proposed in Section III-A are fed into a deep neural network for gender classification described in Section III-B.

A. DESCRIPTORS VIA DISTRIBUTIONS OF BIHARMONIC DISTANCES

The biharmonic distances proposed in [11] are posture invariant, locally isotropic, globally shape-aware, and robust to noises, which are widely used for shape analysis. Distributions of pair wise biharmonic distances are used for shape recognition [14], [15]. The functional biharmonic distance map is defined as signatures for non-rigid shape retrieval [16]. In this section, we define novel descriptors via probability density functions of biharmonic distances for distinguishing male and female 3D shapes.

1) CONTINUOUS SETTING

The biharmonic distance $d(x, y)$ between two points x and y on a 3D human shape \mathbb{S} is defined as follows:

$$d(x, y) = \sqrt{\sum_{k=1}^{\infty} \frac{(\phi_k(x) - \phi_k(y))^2}{\lambda_k^2}}, \quad (1)$$

where $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \dots$ are eigenvalues of the positive defined Laplace-Beltrami operator and ϕ_k is the corresponding eigenfunction of the eigenvalue λ_k .

For a fixed source point x , the cumulative distribution function $F_x(\delta)$ of biharmonic distances from x to the other points on \mathbb{S} is denoted by

$$F_x(\delta) = \int \chi(d(x, y) \leq \delta) d\mu(y), \quad (2)$$

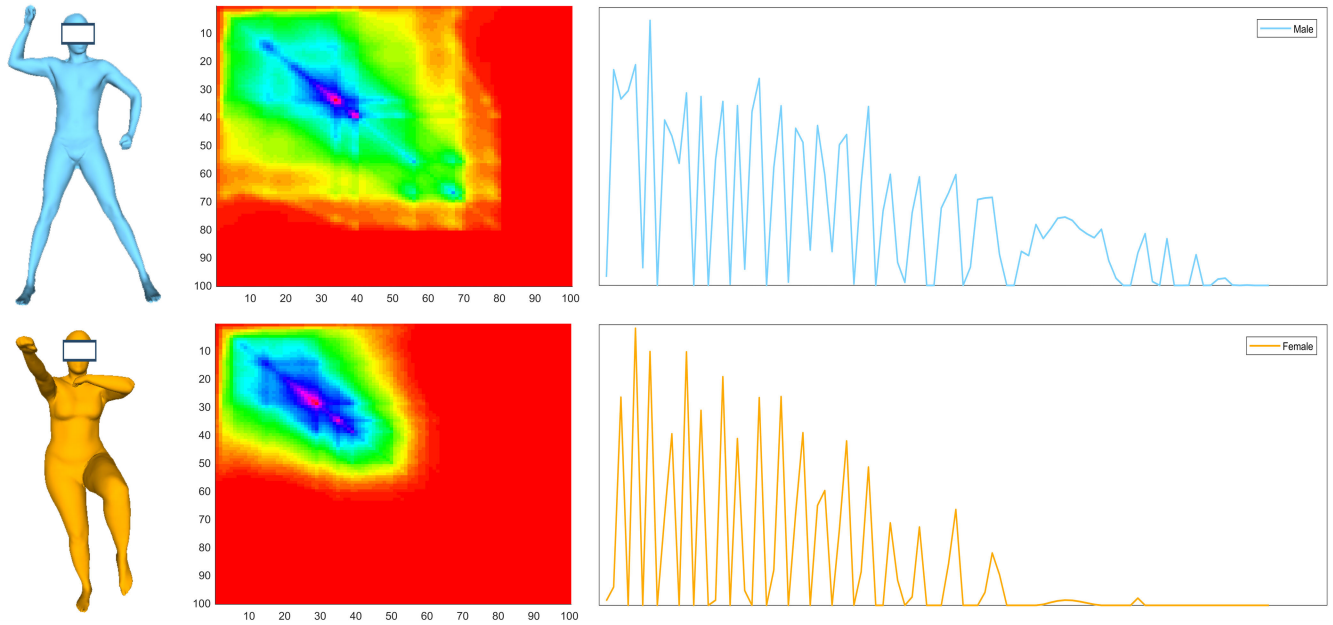


FIGURE 3. The probability distribution matrix \mathbf{D} (middle) and its flattened vector \mathbf{d} (right) of a male (top) and female (bottom) shape.

where χ is the indicator function and μ is the area measure. In this way, $F_x(\delta)$ is the area of points whose biharmonic distances to x are no larger than δ . $F_x(\infty) = \mu(\mathbb{S})$ is the area of the 3D human shape \mathbb{S} . We use the normalized cumulative distribution $\hat{F}_x(\delta) = F_x(\delta)/F_x(\infty)$ for convenience. The corresponding probability density function is defined as the derivative $f_x(\delta) = \frac{d}{d\delta}\hat{F}(\delta)$.

The probability density functions reflect many differences between male and female 3D human shapes shown in Figure 2. In this figure, the male shape is taller and thinner, and the biharmonic distances on the feet from the source point are larger than that of the female shape. The probability density function of the male shape has more spans on the horizontal axis compared with the one of the female shape. Thus, in this article, we use the probability density function of biharmonic distances from every point for classifying gender.

2) DISCRETE SETTING

The 3D human shape \mathbb{S} is represented by a triangle mesh \mathbb{M} with n vertices $\mathbf{v}_i, i = 1, 2, \dots, n$. We use the classical cotangent weight scheme with mixed area normalization for the discretization of the Laplace-Beltrami operator [22]. The biharmonic distance is approximated by the first K summands in Equation (1) as

$$d_D(\mathbf{v}_i, \mathbf{v}_j) = \sqrt{\sum_{k=1}^{K-1} \frac{(\phi_k(\mathbf{v}_i) - \phi_k(\mathbf{v}_j))^2}{\lambda_k^2}}, \quad (3)$$

using the first K smallest eigenvalues (except the zero eigenvalue) and their eigenvectors of the discrete Laplace-Beltrami matrix [11]. The probability density function $f_i(\delta)$ for vertex \mathbf{v}_i is discretized as a p dimensional histogram $\mathbf{B}_i = (b_{i1}, b_{i2}, \dots, b_{ip})$, where b_{ij} is the probability of vertices

in the j -th bin

$$b_{ij} = \sum_{k=1}^n \chi((j-1) * L \leq d_D(\mathbf{v}_i, \mathbf{v}_k) < j * L) \frac{\mu(\mathbf{v}_k)}{\mu(\mathbb{S})}, \quad (4)$$

L is the length of bin interval and $\mu(\mathbf{v}_k)$ is the mixed area of vertex \mathbf{v}_k defined in [22].

Since the number of mesh vertices differs from one shape to another one, the matrix $\mathbf{B} = (b_{ij}), i = 1, 2, \dots, n, j = 1, 2, \dots, p$ of the probability distributions of biharmonic distances cannot be consistently used. To avoid this limitation, we use the probability distribution matrix $\mathbf{D} = \mathbf{B}^T \mathbf{B}$. Since the matrix \mathbf{D} is symmetric, we flatten its upper triangular part to get a $m = p * (p + 1)/2$ dimensional vector \mathbf{d} as the final Probability Distribution Descriptor (PDD for short) for gender recognition of 3D human shapes. The probability distribution matrix and its flattened vector are demonstrated in Figure 3, which have many differences on male and female 3D human shapes.

B. DEEP NEURAL NETWORK

We use the Deep Neural Network (DNN for short) for the gender recognition of 3D human shapes, where the input layer is the m dimensional vector of the descriptor proposed in Section III-A and the output layer is a two dimensional vector for the probability scores of males and females demonstrated in Figure 1. From the input layer to the output layer, there are two hidden layers with h_1 and h_2 hidden units respectively. Thus, the architecture of our four layers fully connected neural network is $m-h_1-h_2-2$.

If the vector in l -th layer is denoted as $\mathbf{x}^l, l = 1, 2, 3, 4$. The feed-forward maps are denoted by

$$\mathbf{x}^{l+1} = \sigma^l(\mathbf{W}^l \mathbf{x}^l + \mathbf{b}^l), \quad (5)$$

where \mathbf{W}^l and \mathbf{b}^l are the weight matrix and bias vector, and σ^l is a nonlinear element-wise activation function. In this article, we use the ReLU activation function except for the last one, which incorporated the well-known softmax function for gender classification.

During the training process, the input data include descriptors and labels $\chi_{train} = \{\mathbf{d}_i, y_i\}$, $i = 1, 2, \dots, N_{train}$, where \mathbf{d}_i is the PDD and y_i is the corresponding gender label of i -th 3D human shape, i.e., 0 for male and 1 for female. The minimized objective function is the cross entropy error, and $\{\mathbf{W}^l, \mathbf{b}^l\}$, $l = 1, 2, 3$ are the parameters to be learned. Pseudocodes of the training process are listed in Algorithm 1. After the training, given a new 3D human shape \mathbb{S} , we first compute its PDD \mathbf{d} in Section III-A and input it to the above neural network for predicting the gender label \hat{y} using the learned parameters.

Algorithm 1: The Training Process

Input: 3D human shapes and their gender labels:

$$\chi_{train} = \{\mathbb{S}_i, y_i\}, i = 1, 2, \dots, N_{train}.$$

Output: Learned parameters: $\{\mathbf{W}^l, \mathbf{b}^l\}$, $l = 1, 2, 3$.

- 1 **for** $i = 1 \rightarrow N_{train}$ **do**
 - 2 Calculating pairwise biharmonic distances of \mathbb{S}_i based on Equation (3);
 - 3 Computing the PDD \mathbf{d}_i of shape \mathbb{S}_i via Equation (4);
 - 4 **end**
 - 5 Initializing parameters $\{\mathbf{W}^l, \mathbf{b}^l\}$, $l = 1, 2, 3$;
 - 6 **for** $epoch = 1 \rightarrow N_{epoch}$ **do**
 - 7 Updating $\{\mathbf{W}^l, \mathbf{b}^l\}$, $l = 1, 2, 3$ by back propagation of the proposed DNN with softmax loss;
 - 8 **end**
-

IV. EXPERIMENTS

In this section, we conduct extensive experiments to demonstrate the effectiveness of the proposed deep neural networks using the probability distribution for gender classification of 3D human shapes. We first present the used datasets and then describe the implementation details.

Datasets. The previous works for gender recognition of 3D human shapes are mainly based on the commercial CAESAR dataset [7], where only one or two thousand 3D shapes are used for training and testing. The relatively smaller training dataset usually limits the learning generalization ability.

We collect 3D human shapes from some public datasets and manually label the genders, such as SCAPE [23], SHREC 2007 [24], TOSCA [25], AdobeData [26], Statistical Model [27], SHREC 2014 [28], SHREC 2015 [28]–[30], Breuckmann BodyScan [31], MoSh [32], MPI FAUST [33], SPRING [34], Dyna [35], MPII CAESAR [36], MPI D-FAUST [37], and DUTH [38], which are mainly aimed at registration, correspondence and other applications. The constructed 3D Human Shapes Gender Recognition Dataset (3D-HSRD for short) consists of more than one hundred thousand shapes with various appearances and postures for evaluation of the proposed method.

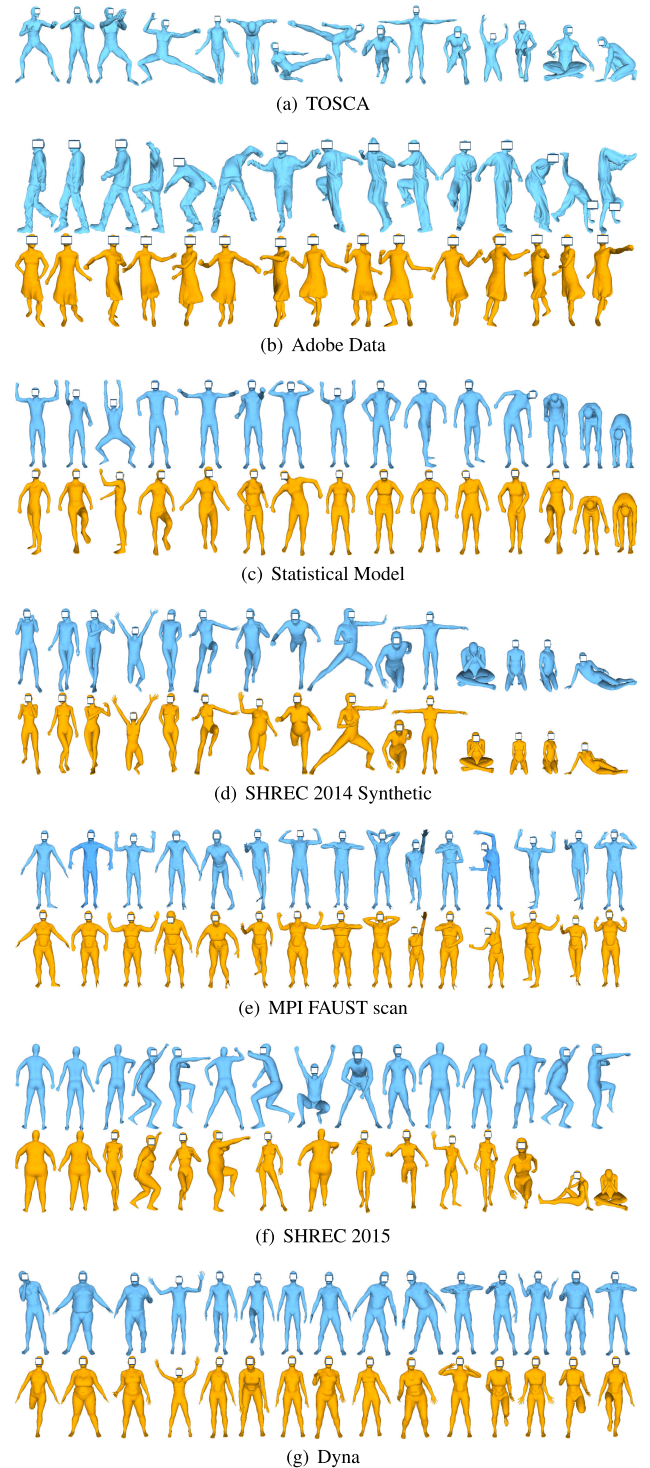


FIGURE 4. Sample shapes and gender labels of the 3D-HSRD, where male and female shapes are in blue and orange colors respectively.

Randomly selected shapes of the 3D-HSRD are illustrated in Figure 4. For the ethical issues and privacy protection of gender classification, we use opaque rectangle regions to cover the head regions of shapes in all figures for privacy protection. The numbers of male and female in the 3D-HSRD are listed in Table 1, where the 3D human shapes are classified into three types, i.e., real scans, registration scans,

TABLE 1. The Constructed 3D-HSRD has 100,202 3D Human Shapes With 50,929 Males and 49,273 Females.

Type	Dataset	#Male	#Female	#Shapes
Real scans	SHREC 2014 Real [28]	200	200	400
	MPI FAUST scan [33]	150	150	300
	Breuckmann BodyScan [31]	54	46	100
	SHREC 2007 [24]	5	11	16
Registration scans	SCAPE [23]	71	0	71
	TOSCA [25]	27	12	39
	Statistical Model [27]	90	251	341
	SHREC 2015 [28]–[30]	50	30	80
	MPI FAUST reg [33]	50	50	100
	MoSh [32]	10	10	20
	SPRING [34]	1,517	1,531	3,048
	MPII CAESAR [36]	2,013	2,288	4,301
	Dyna [35]	21,376	21,862	43,238
Synthetic shapes	MPI D-FAUST reg [37]	20,756	20,464	41,220
	Adobe Data [26]	1,675	325	2,000
	SHREC 2014 Synthetic [28]	100	100	200
	DUTH [38]	2,785	1,943	4,728

and synthetic shapes. The real scans are directly acquired in the real world with the specific multi-stereo 3D scanners. The registration scan is obtained by a non-rigid deformation of a given template shape to approximate the real scan. The synthetic shapes are generated by specific softwares or parametric modeling methods. We will make the shapes and gender labels of the 3D-HSRD public online for future research.

Performance evaluation measure. For a test 3D human shape dataset $\chi_{test} = \{\mathbb{S}_i\}$ and true gender label $y(\mathbb{S}_i)$ of each human \mathbb{S}_i , $i = 1, 2, \dots, N_{test}$, the classification accuracy of a gender recognition method on this test dataset is computed as follows

$$Accuracy = \frac{|\mathbb{S} : \mathbb{S} \in \chi_{test} \wedge \hat{y}(\mathbb{S}_i) = y(\mathbb{S}_i)|}{N_{test}}, \quad (6)$$

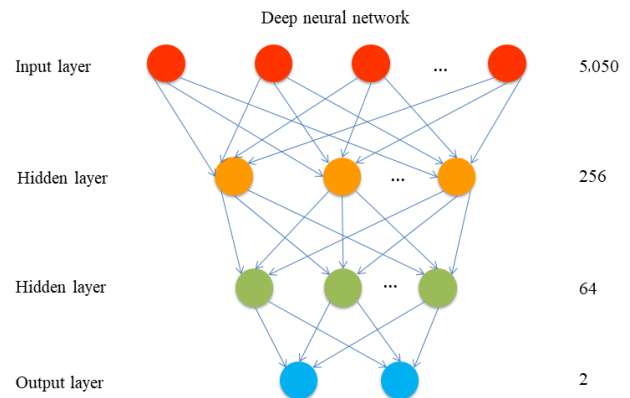
where $\hat{y}(\mathbb{S}_i)$ is the predicted gender label of the gender recognition method and the molecule in the above fraction is the number of humans that have correct classification labels.

Implementation details. All experiments are conducted on a laptop with an Intel Core i5-6200U CPU with 2.4 GHz and 12GB RAM. The algorithms for computing the PDDs in Section III-A are implemented in MATLAB, and the source codes of the DNN in Section III-B are written in python using the Tensorflow.

Because the biharmonic distances are not scaled invariant, we normalize each 3D human shape \mathbb{S} to have a unit area, i.e., $\mu(\mathbb{S}) = 1.0$. The default value of the parameter K is 100, which is the number of the first smallest eigenvalues of the Laplace-Beltrami operator for computing the approximated biharmonic distances. The length of bin interval L is set to $d_{max}/nBins$, where d_{max} is the maximum biharmonic distance for all the 3D human shapes in the 3D-HSRD, and $nBins = 99$ is the number of bins. We use the *hstc* function in MATLAB for computing the histogram in Equation (4) and obtain a 100 dimensional vector. Thus, the dimension of the final PDD d for each shape is 5,050.

The sizes of the two hidden layers in Section III-B are set as $h_1 = 256$ and $h_2 = 64$ respectively. Thus, the architecture

of our four layers deep neural network is 5,050-256-64-2 shown in Figure 5. We also add two dropout layers with 50% rates after the two hidden layers to prevent overfitting and enhance the generalization ability of the network. During the experiments, the 3D-HSRD is randomly split into a 70% train set and a 30% test set.

**FIGURE 5.** The architecture of our deep neural network for gender classification.

Computational times and complexity. The mainly computational time of the proposed method is on the computation of the PDD descriptor. The computational complexity for computing the PDD is $O(n^2)$, where n is the number of vertices of 3D human shape. It will take about three days for computing all of the PDDs of shapes in the 3D-HSRD. Table 2 shows the average computational times in ten times to compute the PDDs of five representative 3D human models. The training and testing times of our DNN on the 3D-HSRD are 39.47 minutes and 4.08 seconds.

Parameter sensitivity. The length of a descriptor is usually domain-specific and selected through experimentation. We set the number of $nBins$ to 63, 99, and 127 respectively for SHREC 2014 Real and SHREC 2015 through ten times experimentations. Each dataset is randomly split into 70% for training and 30% for testing. While increasing the

TABLE 2. Times for Computing the PDDs of Five Representative 3D Human Shapes in the 3D-HSRD.

#Vertices	#Faces	Time (second)
4,709	9,408	5.67
6,449	12,894	11.93
6,890	13,776	14.14
12,500	25,000	27.53
15,000	29,996	47.90

number of bins, the accuracy rate will increase slightly, but the time to calculate the descriptors and the training will also elevate as recorded in Table 3. Therefore, $nBins = 99$ is chosen for the trade-off between the computational times and accuracy.

TABLE 3. Average Classification Accuracy and Training and Testing Times of SHREC 2014 Real [28] and SHREC 2015 [29] Using $nBins$ With Values of 63, 99, and 127 Respectively for the Bi-Harmonic Distance Descriptors.

Dataset	Accuracy (%)			Time (second)		
	63	99	127	63	99	127
SHREC 2014 Real	98.0	98.1	98.8	11.293	16.599	20.335
SHREC 2015	99.4	99.4	100	4.046	5.756	9.172

We also try to change the number of hidden units in the two hidden layers, the recognition accuracy is also above 99.4%. Furthermore, if the number of hidden layers is altered to be three or four, the deep neural network also obtains about 99.0% accuracy. In addition, when using 10% shapes of the 3D-HSRD for training, the gender classification accuracy on the remaining 90% shapes is 99.0%. It is verified that the proposed training method is not overfitting.

A. RESULTS

1) COMPARISONS WITH PREVIOUS WORKS

We run the proposed DNN and the SVM as a baseline algorithm on the PDDs of shapes in the 3D-HSRD ten times and compare their average accuracies listed in the last two rows in Table 4. It can be seen that the mean classification accuracy 99.5% of the proposed method PDD+DNN is higher than 82.1% of the PDD+SVM. The average training and testing times of our DNN and the SVM in ten times are 0.659 and 2.074 hours. During the training process, some times of the SVM are not convergent. So the averaged accuracy of the SVM is lower and the computational time is longer.

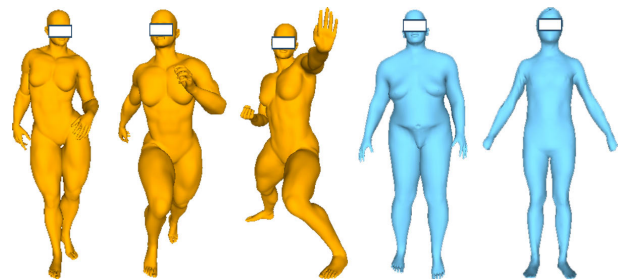
In Table 4, we also compare our method with previous works on gender recognition of 3D human shapes, where the classifiers are all of SVM, and the hand-crafted features are Geodesic Distances (GD) between landmarks [3], Normal Distributions (ND) on the chest region [4], Landmarks Positions (LP) [5], and Vertices Coordinates (VC) [6] respectively. Firstly, the previous works of gender recognition are either posture variant or need of landmarks and pre-alignment, while our method has the advantages of both posture invariant and no need of any pre-processing step. Secondly, the proposed geometric deep learning method obtains higher accuracy on a much larger dataset.

2) COMPARISONS WITH OTHER GEOMETRIC DEEP LEARNING

As far as we know, the proposed method is the first one using geometric deep learning for gender recognition. Previous geometric deep learning are mainly used for shape classification [18], [39], filtering [19] and correspondence [21]. It is unable to use the deep learning frameworks of filtering [19] and correspondence [21] for gender recognition. We compare our method with the geometric deep learning for shape classification such as MVCNN [18] and PointNet [39] on the 3D-HSRD for gender recognition in Table 5. Due to the classic MVCNN and PointNet are mainly aimed at shape classification, they obtain lower accuracies on the new application of gender recognition.

3) PERFORMANCES

Figure 6 demonstrates that the proposed method can recognize the right genders of some 3D human with confused sex characteristics, which are difficult to distinguish their genders by human eyes. The tall and strong female shapes in orange color look like a male, but the proposed method can still recognize them as female. We also can rightly recognize female shapes in blue color with confused gender features, such as prominent chests, narrow shoulders, and slim figures. The above powerful recognition ability is mainly due to the 3D-HSRD has enough training samples with various human shapes and postures demonstrated in Figure 4.

**FIGURE 6. Our method can rightly recognize genders of shapes with confused sex characteristics, where female and male shapes are in orange and blue colors respectively.**

B. ROBUST TO NOISES

Due to the biharmonic distances are less sensitive to noises [11], the proposed gender recognition method (PDD+DNN) is robust both to gaussian and topological noises.

1) GAUSSIAN NOISES

We perform the proposed method on 3D human shape artificially corrupted by Gaussian noises on the AdobeData [26] dataset with 2,000 dressed human shapes, where the variance is P% of the mean edge length of each shape. Three group experiments are shown in Table 6, where shapes in the train and test sets are either clean or corrupted by Gaussian noises in different levels. In each time, the dataset is randomly split into a 70% train set and a 30% test set. It is demonstrated that the proposed gender recognition method is robust to the

TABLE 4. Comparisons With Previous Works. The Proposed Method (PDD+DNN) Achieves Higher Accuracy on the 3D-HSRD With Much More Shapes and Has the Advantages of Both Posture Invariant and No Need of Landmarks or Alignment.

Features	Classifier	Train	Test	Accuracy(%)	Posture invariant	No landmarks	No alignment
GD [3]	SVM	500	500	96.1	✓	×	✓
ND [4]	SVM	1,224	1,224	96.0	×	✓	×
LP [5]	SVM	1,224	1,224	98.9	✓	×	✓
VC [6]	SVM	450	149	75.0	×	✓	✓
PDD	SVM	70,141	30,061	82.1	✓	✓	✓
PDD	DNN	70,141	30,061	99.5	✓	✓	✓

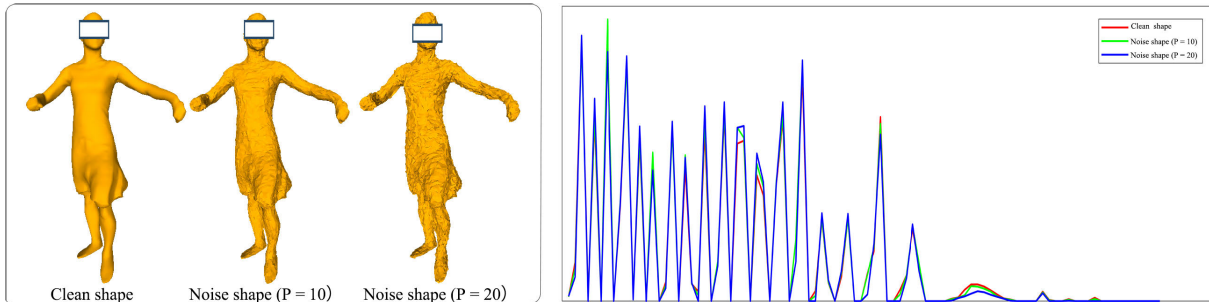


FIGURE 7. The proposed PDD is robust to Gaussian noises. Left: clean shape and shapes artificially corrupted with different levels of noises. Right: the corresponded three PDDs, which are very close.

TABLE 5. Comparison of Ours and Other Deep Learning Methods.

Method	Accuracy(%)
MVCNN [18]	53.4
PointNet [39]	91.9
Ours	99.5

TABLE 6. The Accuracy of the Proposed Method on the AdobeData [26] Dataset is Artificially Corrupted by Gaussian Noises in Different Levels, where P is with a Value of 10 or 20.

Train set	Test set	Accuracy(%)	
		P = 10	P = 20
Clean	Noisy	100	99.9
Noisy	Clean	99.6	96.5
Noisy	Noisy	100	100

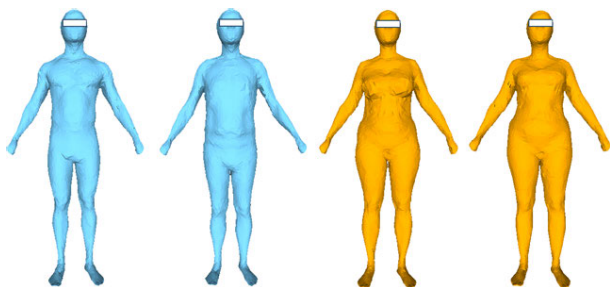


FIGURE 8. The proposed method can recognize genders of human shapes with real world noises.

scanned measurement noises, which is mainly due to the PDD based on the biharmonic distances that are insensitive to noises [11] shown in Figure 7. Figure 8 demonstrates that the proposed method can recognize genders of human shapes with real world noises, where male and female shapes are in blue and orange colors respectively.

2) TOPOLOGICAL NOISES

Due to inaccuracies in the scanning and merging process, the scanned 3D human shapes inevitably have some topological noises, such as holes, handles, and tunnels. It is desirable that the gender recognition method is robust to topological noises.

However, the previous works are usually based on geodesic distances which are sensitive to topological noises [3], [5]. Since the biharmonic distances are insensitive to topological noises [11], the proposed PDD and the gender recognition method are robust to topological noises shown in Figure 9.

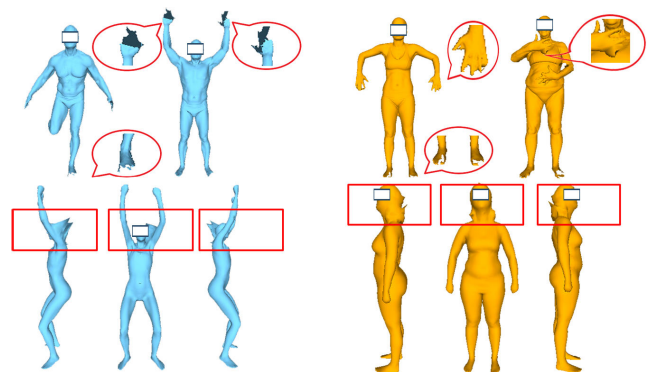


FIGURE 9. The proposed method can recognize genders of human shapes with topological noises shown in rectangles.

3) PERFORMANCE ON REAL SCANS

We also perform the proposed method on the real scan dataset MPI D-FAUST scan [37], which consists of 20,756 male and 20,464 female real scans with larger real world topological noises shown in Figure 10. Due to the scans have about one hundred thousand vertices, we simplify each scan to ten thousand vertices for fast computation of PDD. This dataset

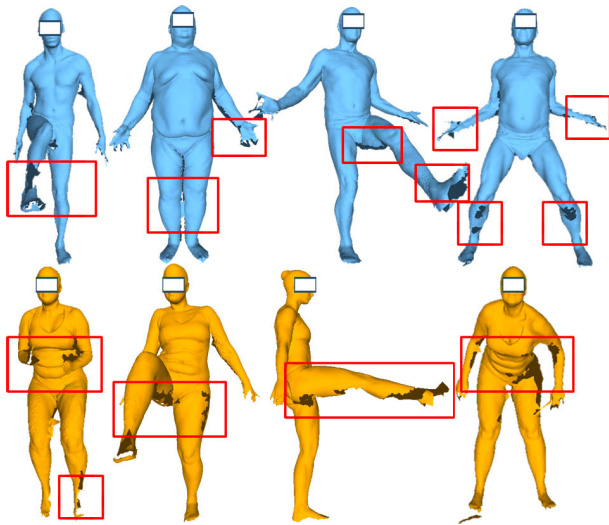


FIGURE 10. Sample shapes of MPI D-FAUST scan [37] with larger real world topological noises.

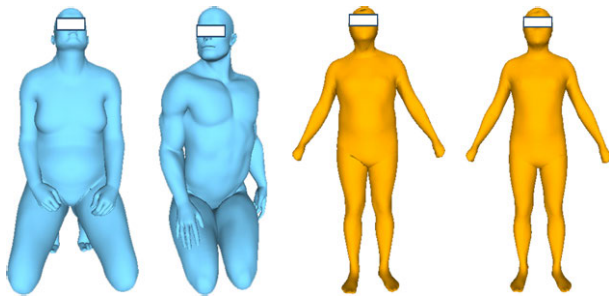


FIGURE 11. Failure cases: the male shapes in blue color are recognized as female, and the female shapes in orange color are classified as male.

is randomly split into a 70% train set and a 30% test set. The testing accuracy of the proposed method (PDD+DNN) is 96.7%, which is much better than the PDD+SVM with an accuracy of 73.5%. This is possible because the gender recognition is a non-linear problem, while the SVM is a linear classifier.

V. CONCLUSION, LIMITATIONS AND FUTURE WORKS

In this article, a novel deep learning method for gender recognition of 3D human shapes is introduced. Our gender recognition method mainly consists of two steps, i.e., extracting descriptors based on probability distributions of biharmonic distances and deep neural networks for gender recognition. Furthermore, we construct a larger 3D human shape dataset called 3D-HSRD with more than one hundred thousand of shapes for the evaluation of gender classification and further research of the other applications. Compared with previous works, the proposed approach obtains higher recognition accuracy and has more desirable properties, such as posture invariant, robust to noises, and no need of landmarks or pre-registration process.

However, the proposed method still has some limitations. Firstly, fewer shapes are still recognized incorrectly as shown in Figure 11, which are randomly selected from

the tested shapes with wrong predicted genders. Secondly, compared with the much larger ImageNet in computer vision, the 3D-HSRD is still relatively smaller, which limits the generalization ability of the learning method to some extent.

In the future, we would use parametric generation models, such as SCAPE [23], SMPL [40], CAPE [41], to synthesize more 3D human shapes, especially the dressed shapes, for enlarging the 3D-HSRD. Furthermore, we also investigate the deep learning methods for recognition of the other attributes of 3D human shapes, for example, height, weight, posture, and so on.

REFERENCES

- [1] C. B. Ng, Y. H. Tay, and B. Goi, "Recognizing human gender in computer vision: A survey," in *PRICAI 2012: Trends in Artificial Intelligence*, vol. 7458. Berlin, Germany: Springer, 2012, pp. 335–346.
- [2] Y. Wu, Y. Zhuang, F. Lin, X. Long, and W. Xu, "Human gender classification: A review," *Int. J. Biometrics*, vol. 8, nos. 3–4, pp. 275–300, 2016.
- [3] S. Wuhrer, C. Shu, and M. Rioux, "Posture invariant gender classification for 3D human models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2009, pp. 33–38.
- [4] J. Tang, X. Liu, H. Cheng, and K. M. Robinette, "Gender recognition using 3-D human body shapes," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 41, no. 6, pp. 898–908, Nov. 2011.
- [5] J. Tang, X. Liu, H. Cheng, and K. M. Robinette, "Gender recognition with limited feature points from 3-D human body shapes," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2012, pp. 2481–2484.
- [6] N. Pablo, P. Bruno, C. Celia, R. Virginia, G.-J. Rolando, and D. Claudio, "Gender recognition using 3D human body scans," in *Proc. IEEE Biennial Congr. Argentina (ARGENCON)*, Jun. 2018, pp. 1–6.
- [7] K. M. Robinette, S. Blackwell, H. Daanen, M. Boehmer, and S. Fleming, "Civilian American and European summary anthropometry resource (CAESAR), final report, volume I: Summary," Air Force Res. Lab., Wright-Patterson Air Force Base, OH, USA, Tech. Rep. 45433-7022, 2002.
- [8] J. Masci, E. Rodolà, D. Boscaini, M. M. Bronstein, and H. Li, "Geometric deep learning," in *Proc. SIGGRAPH ASIA Courses*, 2016, pp. 1–50.
- [9] L. Luciano and A. B. Hamza, "A global geometric framework for 3D shape retrieval using deep learning," *Comput. Graph.*, vol. 79, pp. 14–23, Apr. 2019.
- [10] G. Patané, "A unified definition and computation of Laplacian spectral distances," *Pattern Recognit.*, vol. 93, pp. 68–78, Sep. 2019.
- [11] Y. Lipman, R. M. Rustamov, and T. A. Funkhouser, "Biharmonic distance," *ACM Trans. Graph.*, vol. 29, no. 3, pp. 1–11, Jun. 2010.
- [12] K. Wang, Z. Wu, T. Jia, S. Ali, J. Zhao, G. Yang, and M. Zhou, "A scale-invariant diffusion distance for non-rigid shape analysis," in *Proc. Int. Conf. Virtual Reality Visualizat.*, Aug. 2014, pp. 290–295.
- [13] H. H. Mohamed, S. Belaid, W. Naanaa, and L. B. Romdhane, "Local commute-time guided MDS for 3D non-rigid object retrieval," *Int. J. Speech Technol.*, vol. 48, no. 9, pp. 2873–2883, Sep. 2018.
- [14] M. Mahmoudi and G. Sapiro, "Three-dimensional point cloud recognition via distributions of geometric distances," *Graph. Models*, vol. 71, no. 1, pp. 22–31, Jan. 2009.
- [15] M. M. Bronstein and A. M. Bronstein, "Shape recognition with spectral distances," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 1065–1071, May 2011.
- [16] J. Ye and Y. Yu, "A fast modal space transform for robust nonrigid shape retrieval," *Vis. Comput.*, vol. 32, no. 5, pp. 553–568, May 2016.
- [17] S. Bu, Z. Liu, J. Han, J. Wu, and R. Ji, "Learning high-level feature by deep belief networks for 3-D model retrieval and recognition," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2154–2167, Dec. 2014.
- [18] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3D shape recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 945–953.
- [19] P.-S. Wang, C.-Y. Sun, Y. Liu, and X. Tong, "Adaptive O-CNN: A patch-based deep representation of 3D shapes," *ACM Trans. Graph.*, vol. 37, no. 6, pp. 1–11, Jan. 2019.

- [20] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, and B. Chen, "PointCNN: Convolution on x -transformed points," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 828–838.
- [21] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond Euclidean data," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 18–42, Jul. 2017.
- [22] M. Meyer, M. Desbrun, P. Schröder, and A. H. Barr, "Discrete differential-geometry operators for triangulated 2-manifolds," in *Visualization and Mathematics III*. Berlin, Germany: Springer, 2003, pp. 35–57.
- [23] D. Angelov, P. Srinivasan, D. Koller, S. Thrun, and J. Rodgers, "SCAPE: Shape completion and animation of people," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 408–416, Jul. 2005.
- [24] D. Giorgi, S. Biasotti, and L. Paraboschi, "Shape retrieval contest 2007: Watertight models track," *SHREC Competition*, vol. 8, no. 7, pp. 1–11, 2007.
- [25] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, *Numerical Geometry of Non-Rigid Shapes*. New York, NY, USA: Springer, 2008.
- [26] D. Vlastic, I. Baran, W. Matusik, and J. Popović, "Articulated mesh animation from multi-view silhouettes," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–9, Aug. 2008.
- [27] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H.-P. Seidel, "A statistical model of human pose and body shape," *Comput. Graph. Forum*, vol. 28, no. 2, pp. 337–346, Apr. 2009.
- [28] D. Pickup et al., "Shape retrieval of non-rigid 3D human models," in *Proc. Eurograph. Workshop 3D Object Retr.*, 2014, pp. 101–110.
- [29] D. Pickup, X. Sun, P. L. Rosin, R. R. Martin, and Z. Cheng. (2015). *SHREC'15-Canonical Forms for Non-Rigid 3D Shape Retrieval*. [Online]. Available: <http://www.cs.cf.ac.uk/shaperetrieval/shrec15/>
- [30] Z. Lian, A. Godil, B. Bustos, M. Daoudi, J. Hermans, S. Kawamura, Y. K. G. Lavoué, H. Nguyen, R. Ohbuchi, Y. Ohkita, Y. Ohishi, F. Porikli, M. Reuter, I. Sipiran, D. Smeets, P. Suetens, H. Tabia, and D. Vandermeulen, "SHREC'11 track: Shape retrieval on non-rigid 3D watertight meshes," in *Proc. Eurograph. Workshop 3D Object Retr.*, 2011, pp. 79–88.
- [31] A. Giachetti, E. Mazzi, F. Piscitelli, M. Aono, A. B. Hamza, T. Bonis, P. Claes, A. Godil, C. Li, M. Ovsjanikov, V. Patraucean, C. Shu, J. Snyders, P. Suetens, A. Tatsuma, D. Vandermeulen, S. Wuhler, and P. Xi, "Automatic location of landmarks used in manual anthropometry," in *Proc. Eurograph. Workshop 3D Object Retr.*, 2014, pp. 93–100.
- [32] M. Loper, N. Mahmood, and M. J. Black, "MoSh: Motion and shape capture from sparse markers," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 1–13, Nov. 2014.
- [33] F. Bogo, J. Romero, M. Loper, and M. J. Black, "FAUST: Dataset and evaluation for 3D mesh registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3794–3801.
- [34] Y. Yang, Y. Yu, Y. Zhou, S. Du, J. Davis, and R. Yang, "Semantic parametric reshaping of human body models," in *Proc. 2nd Int. Conf. 3D Vis.*, Dec. 2014, pp. 41–48.
- [35] G. Pons-Moll, J. Romero, N. Mahmood, and M. J. Black, "Dyna: A model of dynamic human shape in motion," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 1–14, Jul. 2015.
- [36] L. Pishchulin, S. Wuhler, T. Helten, C. Theobalt, and B. Schiele, "Building statistical shape spaces for 3D human modeling," *Pattern Recognit.*, vol. 67, pp. 276–286, Jul. 2017.
- [37] F. Bogo, J. Romero, G. Pons-Moll, and M. J. Black, "Dynamic FAUST: Registering human bodies in motion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5573–5582.
- [38] C. Veinidis, I. Pratikakis, and T. Theoharis, "Unsupervised human action retrieval using salient points in 3D mesh sequences," *Multimedia Tools Appl.*, vol. 78, no. 3, pp. 2789–2814, Feb. 2019.
- [39] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 652–660.
- [40] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, "SMPL: A skinned multi-person linear model," *ACM Trans. Graph.*, vol. 34, no. 6, pp. 1–16, Nov. 2015.
- [41] Q. Ma, J. Yang, A. Ranjan, S. Pujades, G. Pons-Moll, S. Tang, and M. J. Black, "Learning to dress 3D people in generative clothing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6468–6477.



HUI WANG received the Ph.D. degree in computational mathematics from the Dalian University of Technology. He is currently an Associate Professor with the School of Information Science and Technology, Shijiazhuang Tiedao University, China. His research interests include computer graphics, digital geometry processing, and image processing.



XIAOYANG LIN received the B.S. degree from the School of Mathematics and Computer, Hebei Normal University for Nationalities. She is currently pursuing the M.S. degree in computer technology with Shijiazhuang Tiedao University.



NANNAN LI received the B.S. and Ph.D. degrees in computational mathematics from the Dalian University of Technology, Dalian. She has been an Associate Professor and held a postdoctoral position with the School of Information Science and Technology, Dalian Maritime University. Her research interests include computer graphics, differential geometry analysis, and machine learning.



FAN MENG received the Ph.D. degree in electromagnetic field and microwave technology from the Beijing University of Posts and Telecommunications. He is currently a Lecturer with the School of Information Science and Technology, Shijiazhuang Tiedao University, China. His current research interests include integrated optics and image processing.



JUNJIE CAO (Member, IEEE) is currently an Associate Professor with the School of Mathematical Sciences, Dalian University of Technology, China. He is also working on computer graphics and machine learning. More details can be found in: <http://jjcao.github.io/>.



XIUPING LIU received the Ph.D. degree in computational mathematics from the Dalian University of Technology, China. She is currently a Professor with the School of Mathematical Sciences, Dalian University of Technology. Her research interests include shape modeling and analyzing, and computer vision.

...