



ELSEVIER

Contents lists available at ScienceDirect

Computer Aided Geometric Design

www.elsevier.com/locate/cagd



Automatic extraction of generic focal features on 3D shapes via random forest regression analysis of geodesics-in-heat [☆]

Qing Xia ^a, Shuai Li ^{a,*}, Hong Qin ^b, Aimin Hao ^{a,*}^a State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, China^b Department of Computer Science, Stony Brook University, USA

ARTICLE INFO

Article history:

Received 9 May 2016

Received in revised form 12 August 2016

Accepted 26 October 2016

Available online 3 November 2016

Keywords:

Generic focal features

Automatic extraction

Random forest regression

3D shapes

Isometric deformation invariance

ABSTRACT

Flexible definition and automatic extraction of generic features on 3D shapes is important for feature-centric geometric analysis, however, existing techniques fall short in measuring and locating semantic features from users' psychological standpoints. This paper makes an initial attempt to propose a learning based generic modeling approach for user-central definition and automatic extraction of features on arbitrary shapes. Instead of purely resorting to certain local geometric extremes to simply formulate feature metrics, it enables the users to arbitrarily specify application-specific features on training shapes, so that similar features can be automatically extracted from other same-category shapes with isometric or near-isometric deformations. Our key originality is built upon an observation: the geodesic distance from one point to desired feature point on testing shape should be similar to the points on training shapes. To this end, we propose a novel regression model to bridge the massive random-sampled local properties and the desired feature points via incorporating their corresponding geodesic distances into the powerful random forest framework. On that basis, an effective voting strategy is proposed to estimate the locations of the user-specified features on new shapes. Our extensive experiments and comprehensive evaluations have demonstrated many attractive advantages of our method, including being fully-automatic, robust to noise and partial holes, invariant under isometric and near-isometric deformation, and also scale-invariant, which can well facilitate to many downstream geometry-processing applications such as semantic mesh segmentation, mesh skeleton extraction, etc.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction and motivation

Automatic feature extraction on 3D shapes is an important and fundamental technique in computer graphics, which is significant for many downstream feature-centric geometric applications, such as registration among different shapes (Huang et al., 2008; Lui et al., 2010), coarse or dense correspondence (Zhang et al., 2008; Lipman and Funkhouser, 2009), shape retrieval (Liu et al., 2006; Ovsjanikov et al., 2009), etc. Conventional feature extraction methods mainly exploit certain differential attributes to characterize prominent features (Zhang et al., 2005; Ohtake et al., 2004; Wang et al., 2015) based on local geometric measurements. However, these features are inevitably biased towards high-saliency shape extremities/pro-

[☆] This paper has been recommended for acceptance by Rida T. Farouki.

* Corresponding author.

E-mail addresses: lishuai@buaa.edu.cn (S. Li), ham@buaa.edu.cn (A. Hao).

trusions in some senses, while tending to ignore non-salient but semantically meaningful regions. Meanwhile, because these approaches only rely on objective properties relevant to shape geometry, the user-central semantic definition is extremely hard to be accommodated.

Motivated by the urgent demands of high-level shape understanding applications, more and more attention has been paid to the flexible definition and automatic extraction of user-specified features. Chen et al. firstly proposed a *Schelling Points* method (Chen et al., 2012) by employing decision trees (Witten and Frank, 2005) to combine multiple geometry properties into an analytical model, which can be used to predict the points specified by users (*Schelling Points*) on new shapes. However, although this method provides us a feasible way to map some kinds of user-specified features to certain geometric signals, it requires tedious labeling in a crowd-sourcing way, and the landmark detecting process heavily relies on curvature like local geometric properties, which is hard to be flexibly used for universal focal feature definition and extraction.

In this paper, we advocate a totally different vision for the application-specific flexible definition and automatic extraction of user-specified features on 3D shapes, which we name as focal features. The underlying observation supporting our novel idea is that, for a random-sampled point on the testing shape, its geodesic distance to the desired focal features should be similar to the points with similar properties on the training shapes of same category. Therefore, with massively random-sampled points on training shapes, we can use their local properties and their corresponding geodesic distances to the focal features to construct a random forest. Given testing shape, we randomly sample many points again, and use the random forest to estimate the corresponding geodesic distances of each point. Finally, we design an effective voting scheme to locate the focal feature points on testing shape with the estimated geodesic distances of the random-sampled points serving as vote sources. In particular, the primary contributions of this paper can be summarized as follows:

- We propose a novel generic modeling approach based on regression learning to flexibly define and automatically extract focal features on arbitrary 3D surface, which facilitates to bridge the large gap between objective geometric measurements and subjective application-specific feature recognition.
- We make the first attempt to incorporate the random-sampled local properties and their geodesic distances to the desired focal features into the powerful random forest regression framework, which enables semantic focal features for high-level shape understanding applications.
- We propose a novel consensus voting scheme by fully leveraging the mutual interactions of the spatially-different properties in a global-to-local sense, which gives rise to more precise and robust extraction of focal features even for noise-stained shapes with scattered votes' distribution.
- We deeply exploit the scale-invariant heat kernel signatures (Bronstein and Kokkinos, 2010) and geodesics-in-heat (Crane et al., 2013) to respectively serve as the replaceable descriptor and distance metric of sampled points, which facilitates the construction of intrinsic links between testing shape and training samples.

2. Related work

Closely relevant to the central theme, we now briefly review previous approaches and their related applications in two categories: feature extraction and random forest.

2.1. Feature extraction

Feature detection/extraction on 3D shapes is a classical problem in geometric modeling and computer graphics, which is also a fundamental problem for many downstream geometry processing tasks, such as shape correspondence (Zhang et al., 2008; Lipman and Funkhouser, 2009), shape registration (Huang et al., 2008; Lui et al., 2010), shape retrieval (Liu et al., 2006; Ovsjanikov et al., 2009), etc. The main goal of feature detection is to locate stable points or regions on a shape that carry discriminative significant/salient/semantic information, which maybe repeatedly occur on other similar shapes. Since such significance/saliency/semantics on shape is very subjective in definition, many approaches have been proposed with different application requirements. Researchers early tended to focus on the extremities or protrusions of the 3D shape. And these regions are commonly characterized by some local geometric properties or functions defined over the shape, including the local Gaussian curvature (Lipman and Funkhouser, 2009; Sahillioğlu and Yemez, 2011), differences of Gaussians (DoG) at multiple scales (Castellani et al., 2008; Zou et al., 2008), differences of diffusions (DoD) (Hua et al., 2008), ridge and valley lines (Ohtake et al., 2004), prominent tip points (Zhang et al., 2005), the Heat Kernel Signatures (Sun et al., 2009), etc. In addition, many researchers also resorted to extracting feature points in some special embedding domains, such as the multidimensional scaling (MDS) embedding (Katz et al., 2005) and the spectral embedding of the Laplace–Beltrami Operator (Dey et al., 2015; Ruggeri et al., 2010). Moreover, inspired by the success of 2D image feature detection, some techniques were also extended to handle 3D shapes. For example, 3D-Harris method (Pratikakis et al., 2010) was extended from 2D corner detection, and mesh-SIFT method (Maes et al., 2010) was derived from the famous scale invariant feature transform (SIFT) algorithm (Lowe, 2004). However, most of the aforementioned methods simply rely on the local/global geometric properties or the structures of the shape. It means the significance/saliency is fully defined by the shape itself while lacking of users' intention, and thus is not suitable for semantic tasks. Recently, *Schelling points* method (Chen et al., 2012) successfully integrated the geometric properties with users' intention to predict landmark points on new shapes.

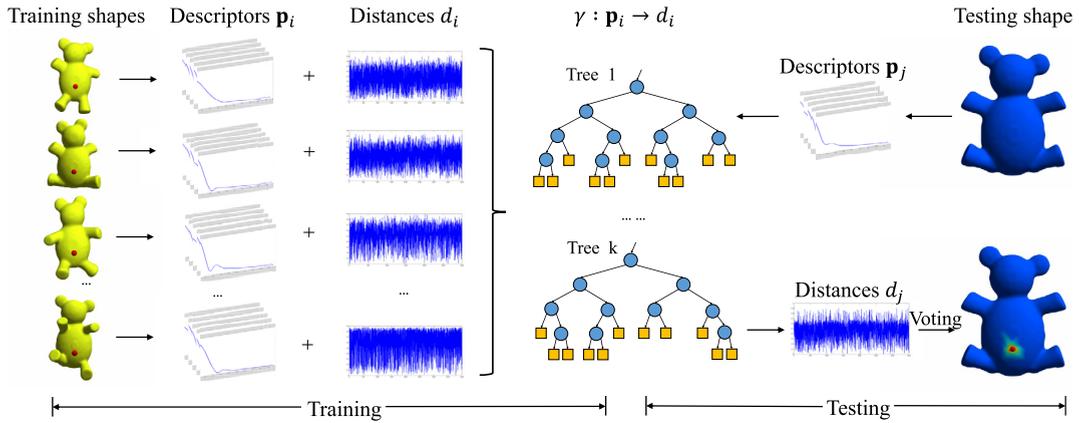


Fig. 1. The full pipeline of our method. The training shapes are listed in the left most column with yellow color and one focal feature point is indicated as red sphere on each teddy's belly. We firstly compute many randomly sampled points' descriptors and geodesic distances to the feature point, and then construct a randomized regression forest shown in the middle. When given a new shape colored blue, we also compute many descriptors as before and use the forest to predict their geodesic distances to the feature point. Finally, we use a voting scheme to compute the probability of each point to be the focal feature, where red indicates high probability. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Nevertheless, it heavily relies on curvature like local geometric properties, and their feature points are usually at the tips of protrusions or conspicuous saddles, which is hard to be flexibly used for universal focal feature definition and extraction.

2.2. Random forest

Random forest belongs to ensemble learning method, which constructs a multitude of decision trees at training time and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. Breiman (2001) firstly introduced the term “random forest” and further polarized its practical applications. In recent years, an explosion of random forest based methods have been proposed in various machine learning and computer vision applications, such as image classification (Bosch et al., 2007), anatomy detection and localization in CT slices (Criminisi et al., 2011), head pose estimation (Fanelli et al., 2011), object detection (Gall and Lempitsky, 2013), semantic segmentation of CT images (Montillo et al., 2011), etc. Specially, Criminisi et al. (2011) used regression forest to predict the offsets between the target anatomy and sampled points in the tomography scans for the locating of the anatomy's position, and in Chu et al. (2014), they used the same idea to automatically detect landmarks in cephalometric X-Ray images. In this paper, we will extend this strategy to user-specified focal feature detection on 3D shapes. However, for 3D shape surfaces, it does not allow us to naively leverage the axis-aligned offsets to locate a point due to the potential non-rigid transformations and spatial scales across shapes. Meanwhile, such principle will become much more complicated than that used in images, because few feature descriptors on 3D shapes can keep invariant under different kinds of shape variations.

3. Method overview

As shown in Fig. 1, the pipeline of our ensemble learning based method consists of two main stages: one is for training, and the other is for testing.

In the training stage, a few shapes with focal feature points marked by users serve as training shapes, of which, the involved steps are briefly summarized as follows:

- **Random samples description.** For each training shapes, we randomly sample many points and compute shape descriptors, denoted as $\{\mathbf{p}_1, \mathbf{p}_2, \dots\}$, to distinguish them against other points by depicting their local neighborhood.
- **Geodesic distance metric.** For each sampled point, we compute its geodesic distance to the focal feature point, denoted as $\{d_1, d_2, \dots\}$. This distance is used to vote the feature point's location in the final step.
- **Random forest construction.** With these descriptors and geodesic distances, we construct a random forest, which acts as a regression function γ that maps the point descriptor to its geodesic distance to desired feature point as $\gamma: \mathbf{p}_i \rightarrow d_i$. Here the random forest consists of an ensemble of decision trees, whose leaf nodes contain a few points sharing similar properties. It gives rise to meaningful geodesic distance prediction for the point on new testing shape.

In the testing stage, we try to extract the focal feature point on a new shape, and the involved steps are briefly outlined as follows:

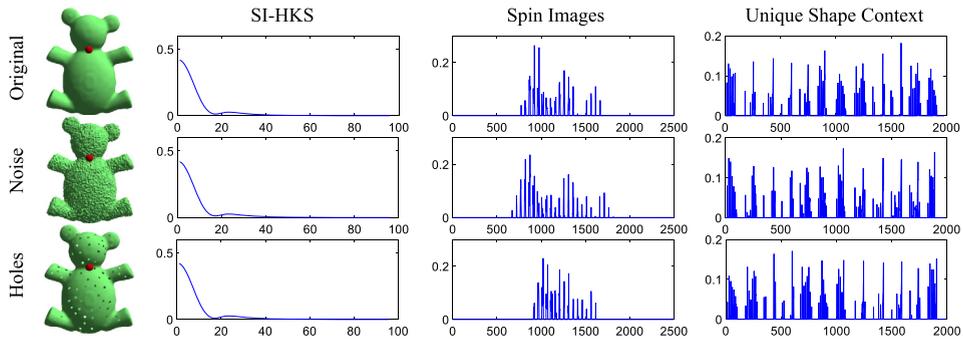


Fig. 2. Comparison of SI-HKS (Bronstein and Kokkinos, 2010), Spin Images (Johnson, 1997) and Unique Shape Context (USC) (Tombari et al., 2010). The descriptors are all rearranged into 1D vectors. The top row shows the 3 kinds of descriptors for the red point on the 3D shape, the middle row shows the descriptors when the shape is perturbed by Gaussian noise, and the bottom row shows the descriptors when the shape has many small holes. It is obvious that, SI-HKS is much more robust than the other two descriptors, both of which have large changes under perturbations while SI-HKS keeps nearly-invariant. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- **Random samples description.** For the given testing shape, we randomly sample many points and compute their descriptors $\{P_j\}$ by the same way as that in the training stage.
- **Geodesic distance prediction.** With these new descriptors, we feed them to the regression forest constructed in the training stage to predict their corresponding geodesic distances $\{d_j\}$ to the desired feature point.
- **Feature point localization.** With these predicted geodesic distances, we can estimate the probability of each location to be the desired feature point via a consensus voting strategy, and finally locate the feature point on the new shape.

In summary, according to the whole technological processes of our method, there are 4 main challenges we should address in our framework: 1) how to intrinsically describe the local properties of each random-sampled point, which will be detailed in Section 4.1; 2) how to efficiently and robustly calculate the geodesic distances (whose robustness has direct influence on the subsequent steps), which will be detailed in Section 4.2; 3) how to predict the geodesic distances to the unknown focal feature points based on the local properties of the sampled points on testing shapes, which is the core of our method and will be detailed in Section 4.3; 4) how to jointly estimate the location of desired feature points on testing shape with the obtained distances, which will be detailed in Section 4.4.

4. Automatic extraction of focal feature

As aforementioned, the main purpose of our method is to define and extract focal features on 3D shapes by bridging the random-sampled points' local properties with their geodesic distances to the desired feature points, and the testing shape's focal feature to be extracted is collectively voted by the predicted geodesic distances of many random-sampled points. In this section, we will show the details of each step in our framework addressing the problems mentioned in last section.

User-specified focal feature definition. The focal features in this paper are intuitively and flexibly defined via simple labeling of users. Similar to the way used in Chen et al. (2012), given a few shapes belonging to the same category, user can mark the interested feature points on each shape, which are expected to be automatically extracted on new shapes. For example, as shown in the left most column of Fig. 1, if the user wants to extract all the similar points on the belly of a series of teddy shapes, he should mark one exemplar point on the belly of each training shape. These marked points serve as defined focal features, which naturally convey the user-central interests/significance/saliency/semantics and can be located anywhere on the shape. Thus, they are very useful for application-specific tasks in shape understanding.

4.1. Symmetry-aware intrinsic description for random-sampled points

Traditional point descriptors are usually defined as statistics of certain geometric properties within a small neighborhood around the point, such as Spin Images (Johnson, 1997) and 3D Shape Context (Frome et al., 2004; Tombari et al., 2010). However, even though these descriptors are widely-used, they still have drawbacks when facing perturbations as noise, as shown in Fig. 2. In this paper, we resort to the scale-invariant heat kernel signature (SI-HKS) (Bronstein and Kokkinos, 2010) to character the random-sampled points, which is an enhanced version of HKS (Sun et al., 2009), resolving the limitations of HKS by rewriting the signature in the Fourier domain. These built-in properties, such as isometry-invariance, robust to topological changes and noise (Yu et al., 2012), facilitate adaptability of our focal feature extraction on diverse shape perturbations.

Symmetry invariance. However, SI-HKS cannot well distinguish the symmetry, that points on different symmetric parts of the shape will have the same descriptors. As shown in Fig. 3(a), the descriptors of the points on the left and right arms/legs are very similar to each other, which is also suffered by almost all the existing isometric feature descriptors. Nevertheless, most

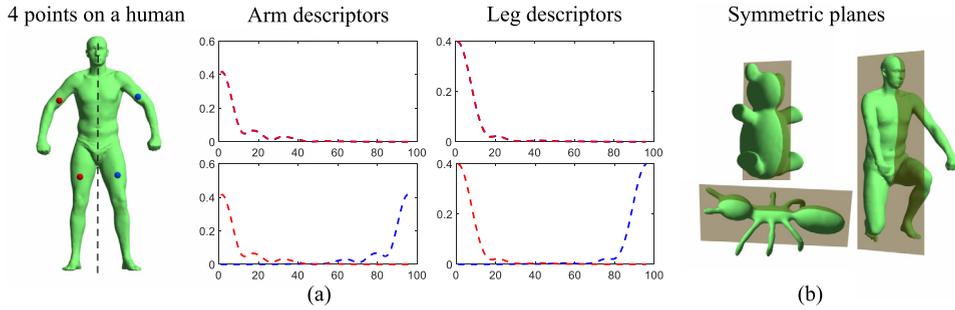


Fig. 3. Illustration of the symmetry-aware local descriptors. (a) From left to right, 4 points on a human shape; the descriptors of the points on symmetric arms; the descriptors of the points on symmetric legs. The descriptors in the first row are original SI-HKS, and the ones in the bottom are our symmetry-aware descriptors, wherein the SI-HKS representations are inverted for the points on the right side (colored in blue, likewise, red corresponds to left side). (b) The estimated symmetric planes of different kinds of shapes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

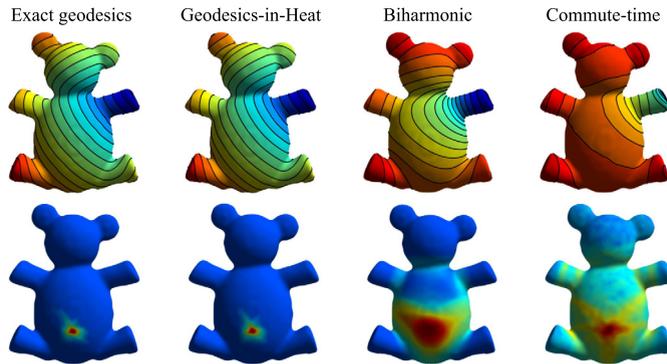


Fig. 4. Analysis on 3D shape distance metrics under our framework. The exact geodesics are computed using the method (Surazhsky et al., 2005), which also serves as the ground truth. With a location at teddy’s left hand as anchor point, all the four shapes in the top row are color-coded according to the distances, and blue indicates small distance and red indicates long one. The bottom row shows the extraction results using these distances as geodesic metric in our framework. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of the shapes in real world are symmetric (especially plane-symmetric), which cannot be avoided in practical applications. Therefore, in this paper, we propose a simple but effective strategy to solve such plane-symmetric problems. As shown in Fig. 3(b), we firstly employ the method in Kakarala et al. (2013) to estimate the approximated symmetry plane of each shape. Then, when computing the descriptor for each point, we firstly judge the relative location of this point with respect to the symmetry plane. As shown in the bottom row of Fig. 3(a), if this point lies on the side in which the plane’s normal points to, we directly employ original SI-HKS as the descriptor. Otherwise, all the components of the descriptor should be transformed in an inverse order. This strategy makes the symmetric descriptors on different sides become very different from each other.

4.2. Analysis on 3D shape distance metrics under our framework

In this paper we employ the *Geodesics in Heat* proposed by Crane et al. (2013) to calculate the geodesic distances. Comparing to existing distance metrics on 3D shapes, such as Biharmonic Distance (Lipman et al., 2010) and Commute-time Distance (Qiu and Hancock, 2007), it is better for approximation of true geodesic distances. As shown in Fig. 4, *Geodesics in Heat* is almost the same with exact geodesics, while the other two distances’ isolines are uneven spaced. And this uneven spacing of isolines means a small difference in distance may correspond to a large space on the shape surface, which will result in a much more scattered and mussy votes distribution in the final localization stage, as shown in the bottom row of Fig. 4. In fact, since *Geodesics in Heat* is a type of approximated geodesic distances based on the heat equation, it owns both advantages of geodesic and spectral distances, such as being robust, insensitive to noise and partial holes, and invariant to isometric deformations. Considering the geodesics may be heavily influenced by the size of the shape, to make the distances invariant to the spatial scales, we need further to normalize the distance to $[0, 1]$ for each point.

4.3. Learning based distance regressor construction

Here, we leverage regression forest to define a function that maps the point descriptor to its geodesic distance to desired focal feature. Regression forest is an ensemble of T binary decision trees. In each decision tree, the descriptor set is itera-

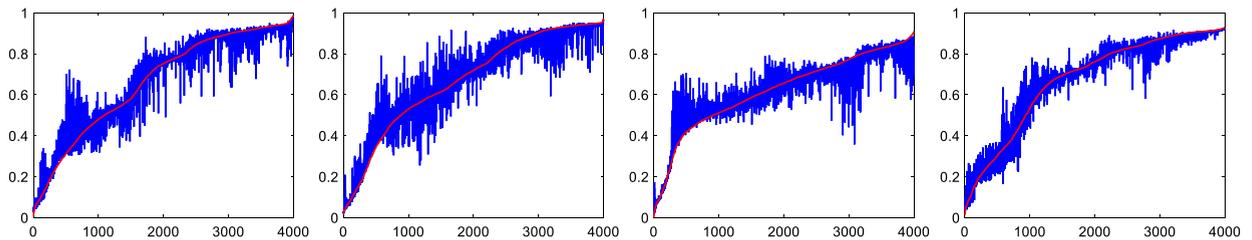


Fig. 5. Four regressed geodesic distances from random-sample points to the known desired feature points in Fig. 1. The red line indicates the ground truth distances to the target points, and the blue line indicates the predicted distances. Y axis stands for the normalized distances, and X axis stands for the order of the sorted random-sampled points. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

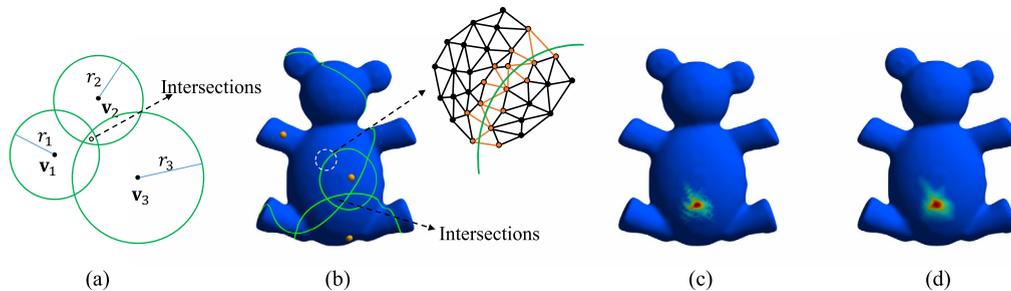


Fig. 6. Illustration of our joint voting strategy for locating the focal feature points. (a) An example of trilateration in 2D. (b) Our voting strategy on a 3D shape. (c) The voting result. (d) The average votes on each vertex, equal to its probability to be a feature point, red corresponds to high probability and blue corresponds to low probability. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tively split into many small clusters containing similar descriptors, and the predictions are obtained via regression within these clusters.

Forest training. The training process constructs each regression tree and decides how to reasonably split the incoming descriptors at each node. For each sample descriptor \mathbf{p} , starting at the root of each tree, it is sent down along the tree structure. The j -th split node applies a binary test $h(\mathbf{p}, \theta_j)$ to determine whether the current sample point should be assigned to the left or right child node. Randomness is embedded via randomly choosing parameter θ_j at each node to maximize the information gain G (Criminisi and Shotton, 2013), which encourages decreasing the uncertainty of the prediction in each node. And this will result in many small clusters of descriptors at the leaf nodes, wherein they are similar to each other and have similar corresponding geodesic distances.

Forest testing. During testing, a new sample point's descriptor \mathbf{q} traverses the tree until it reaches to a leaf, and such testing descriptor is likely to end up in a leaf that is closely related to the most similar training descriptors. Thus, it is reasonable to use the statistics gathered in that leaf to predict the distance associated with the testing descriptor.

4.4. Joint voting based focal feature localization

In the next, we use these predicted distances to locate the feature point on the new shape. Locating an object is a common problem in our daily life, and the Global Positioning System (GPS) is proposed to solve this problem. A GPS receiver uses trilateration to determine its position on the surface of the earth by timing signals from three satellites in the Global Positioning System. As shown in Fig. 6(a), three satellites locate at $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. From each satellite, we can draw circles with radius r_1, r_2, r_3 respectively (the radius is measured by the time of signal transmission), and the receiver lies in the intersecting region of these three circles. This is similar to our purpose, here our random-sampled points correspond to the satellites and geodesic distances correspond to the timings of signals.

In the positioning system, if the timings of signals are precise enough, commonly 3 satellites are able to locate the receiver. However, in our case, the geodesic distances predicted by the forest are not accurate. As shown in Fig. 5, most predicted distances are not exactly equal to the ground-truth, whose values are oscillating around the ground truth, it means the predicted positions may fall into a neighborhood of the target feature point. Thus, we design a joint voting strategy to determine the location of the feature points, which is similar to that in Xu et al. (2009), where they use many pairs of points with similar descriptors to vote their Voronoi boundary as partial symmetry axis. As shown in Fig. 6(b), for each randomly sampled point, together with its corresponding predicted geodesic distance, we analogously compute a geodesic circle with the sample point as center as that in the GPS example. On the circle, all points' geodesic distance to the center are the same, equaling to the predicted geodesic distance. And then, the sampled point will give one vote to the two endpoints of the edges that this geodesic circle goes through (colored in orange). Besides, we weight the vote

Table 1
Performance statistics (in seconds).

Model	#V	#S	Training				Testing				
			sr_r	t_{rd}	t_{rg}	t_{rr}	sr_e	t_{ed}	t_{eg}	t_{er}	t_{ev}
Human	12.5K	10	0.5	9.56	8.31	66.12	0.1	3.82	3.09	0.03	1.33
Lion	5K	10	0.5	3.70	0.85	13.45	0.1	1.51	0.34	0.03	0.48
Teddy	12.5K	10	0.5	9.49	10.63	68.10	0.1	3.81	4.34	0.03	1.19
Ant	7.5K	10	0.5	6.42	3.32	26.63	0.1	2.52	1.27	0.02	0.74
Hand	8.7K	8	0.6	6.71	3.69	21.59	0.1	2.64	1.49	0.02	0.86

From left to right, #V: number of shape vertices, #S: number of shapes, sr_r : random-sampling rate on each training shape, t_{rd} : time cost for the descriptors' computation on each training shape, t_{rg} : time cost for geodesics computation on each training shape, t_{rr} : time cost for forest training, sr_e : random-sampling rate on testing shape, t_{ed} : time cost for descriptors computation on testing shape, t_{eg} : time cost for geodesics computation on testing shape, t_{er} : time cost for forest testing, t_{ev} : time cost for joint voting.

Table 2
Errors.

Model	①		②		③		Average
	#F1	Error1	#F2	Error2	#F3	Error3	
Human	5	2.46	6	2.57	2	2.14	2.46
Lion	6	2.85	1	3.09	1	4.16	3.04
Teddy	4	3.79	2	4.67	2	4.21	4.12
Ant	10	4.01	0	N/A	4	4.92	4.27
Hand	12	3.24	8	4.82	0	N/A	3.87

From left to right, #F1: number of points marked with ①, Error1: error of points marked with ①, #F2: number of points marked with ②, Error2: error of points marked with ②, #F3: number of points marked with ③, Error3: error of points marked with ③, average error of all feature points.

of this sampled point according to its geodesic distance as $w = e^{-\mu * d * d}$, because we think points closer to feature points should have more influences on the final localization. Here μ controls the influence range. Finally, we count the votes on each points, as shown in Fig. 6(c), the average votes in a small neighborhood can be described as the probability to be the feature point. As shown in Fig. 6(d), the desired focal feature point is the point with the highest average votes. Since the voting process of each sampled point is independent of each other, it can be implemented efficiently in CUDA.

5. Experimental results and evaluations

We have implemented our automatic focal feature extraction framework using C/C++ and CUDA on a PC with Intel Core i7-3370 CPU @ 3.40 GHz and Nvidia GeForce GTX 780 GPU. To verify the effectiveness and versatility, we have designed different kinds of experiments, including the focal feature extraction over isometric shapes and near-isometric shapes together with their comparisons to *Schelling Points*, verification tests on the insensitivities to noise/partial holes and spatial scales, and other exemplar downstream applications. And the time statistics of all experiments are detailed in Table 1, the errors are shown in Table 2.

Parameters setting. In all our experiments, the parameters involved in our method keep fixed unless otherwise stated. The HKS is computed with the same way as that in Bronstein and Kokkinos (2010), all the shapes are normalized to approximately have the same scale as the Shape Google database used in *Schelling Points*. We use the cotangent weight approximation for the Laplace–Beltrami operator, and set $k = 200$ as the number of the selected leading eigenfunctions. The time t is sampled logarithmically with base 2, whose exponent ranges from 1 to 25 with increments of 0.1. To obtain the SI-HKS, we use the first 20 lowest discrete frequencies, and then extend the dimension to 96 using cubic spline interpolation. The symmetric plane and geodesics are computed using the source code provided by the authors with default parameter settings. We implement the random forest regressor based on the Sherwood Library provided in Criminisi and Shotton (2013), 30 decision trees are used for each regression, the tree construction stops when the leaf node contains less than 10 descriptors or the information gain is no longer increased, and only 5 components of each descriptor are randomly sampled during the split node optimization. In the final voting stage, the weight factor is set to be $\mu = 5$, and the neighborhood of each point is defined by the points whose geodesic distances are less than $d = 0.02$.

Random sampling related analysis. Fig. 7(a) and Fig. 7(b) evaluate the accuracy and computation time cost of the regression when using different numbers of training shapes and different sampling rates on each training shape. The accuracy is computed by the mean square error of the predicted geodesic distances for each feature point. From Fig. 7(a), we can see that, the more points are sampled, the less error is produced. However, there is no significant decline in error when the sampling rate is larger than 50%, and the training time is increasing exponentially with the growth of sampling rate. Therefore, in all of our experiments, the sampling rate on training shape is setting around 50%, which may vary with the number of training shapes. Fig. 7(b) indicates the fact that, the more shapes used for training, the more accurate the regression is. Because more training shapes can cover more variances on shapes. When testing on a new shape, there is higher probability to find a more similar shape in the database, that is, the testing descriptor will end at a leaf node with more similar training descriptors, which gives rise to more accurate regression. The two figures on the bottom of (a) and (b) are both the plots of

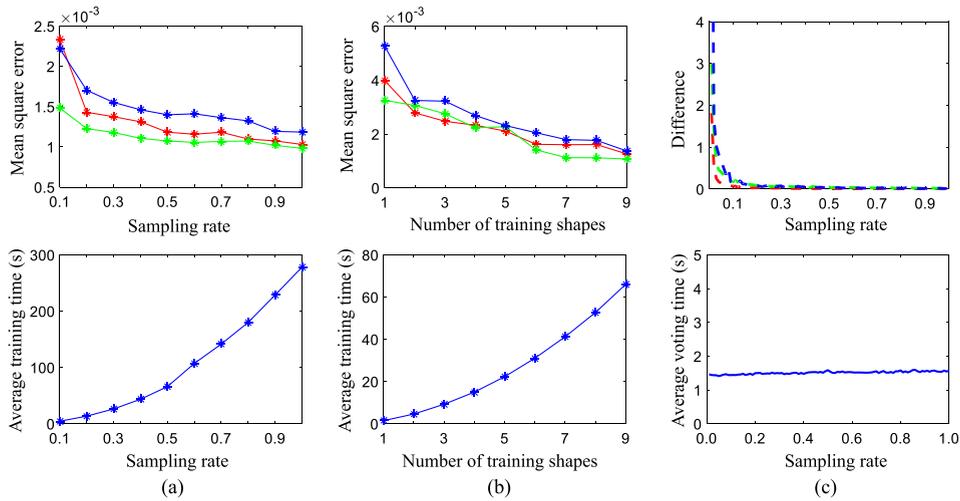


Fig. 7. Accuracy evaluations under different sampling rates on each training shape (a) and different numbers of training shapes (b), and the change of the votes' distribution under different sampling rates on testing shape (c). In (a), (b) and (c), the quantitative evaluations of 3 feature points located at chest, stomach and head (see Fig. 8) are shown on the top, which are colored in red, green and blue respectively. And the average time costs are correspondingly shown on the bottom. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the training time with respect to sample numbers. It should be noted that, an exact complexity analysis on random forest training is nontrivial. However, we can obtain an approximated time complexity. Actually, the training process is to build a random forest with a few decision trees. Assume we have n training samples, ideally each tree's depth is $\log(n)$. On each level of the tree, for all n samples, it should determine which child the sample will go to. Thus, the approximated time complexity of building a random forest is $O(n \log(n))$, and the plots also reveal the same complexity trend. Moreover, Fig. 7(c) evaluate the change of the normalized votes' distribution and the time cost of voting process when using different sampling rates on testing shape. The change is computed by the norm of the difference of two votes' distributions. We can see from the figure that, when the sampling rate is larger than 10%, the difference of votes' distributions under two successive rates is very close to 0, that means the votes' distribution are nearly unchanged when sampling rate is larger than 10%. The voting time only very slightly increases for larger sampling rate because of our parallel implementation. However, we should also consider the computation efficiency of SI-HKS and geodesic distances, so we only sample 10% points on the testing shape in all our experiments, which is enough to accurately extract the focal feature points. Besides, since the location of our focal feature is independent of the definition of feature type but depending on many other random-sampled points over the shape, the choices (like sampling rate) are universal for different shapes and different types of feature points, so long as the shapes belong to the same shape category.

Invariance to isometric and near-isometric shapes. As mentioned above, SI-HKS and geodesics are both invariant under isometric deformations. Fig. 8 shows two examples of our automatic focal feature extraction on isometric shapes. From the results, we can see that, our method can precisely predict the focal feature points on a new shape that undergoes isometric deformations. Fig. 9 shows three examples of our automatic focal feature extraction on near-isometric shapes. Even though these shapes seems similar to each other, there still exist small non-isometric deformations among them. However, our method can still produce satisfactory feature extraction results on these shapes. Comparing to the *Schelling Points* (Chen et al., 2012), our method requires much less manually-labeled points on each training shape than those in Chen et al. (2012). For example, our method only need 8 user labeled points to exact 8 focal feature points on the teddy shape (one labeled point for one feature point). In sharp contrast, as shown in the left and right most columns of Fig. 9, to produce result with similar number of feature points, Schelling Points method requires several tens or even hundreds of manually-labeled points. Besides, the method in Chen et al. (2012) highly relies on curvature-like local properties, and their feature points are usually at the tips of protrusions or conspicuous saddles. In sharp contrast, our method leverages random forest to predict the geodesic distances from many random-sampled points to the desired focal points, and these feature points are independent of the geometric properties of themselves, but related to the properties of other sampled points. Thus, the locations of our feature points can be arbitrary. As shown in the 3rd row of Fig. 9, our method successfully extracts the points at ②, however, Chen et al. (2012) ignore these points, even though there are points specified at this region in their training shapes (see the fingers in the right most column). Therefore, our method is more flexible for defining different kinds of focal features. It should be noted that, when we visualize the voting result of multiple feature points, we have squared the votes on each point and set the normalized votes less than 0.2 to be 0, and it makes the vote bumps smaller and not be interfered by each other for clearer illustration.

Robustness to noise, holes and scales. Fig. 10 shows the experiment results over the shapes with noise, holes and different spatial scales. The noise is produced by mesh perturbation using a Gaussian with 0.3 of the mesh's mean edge length as mean value. The holes are produced by randomly removing a few points and their adjacent triangles, and the

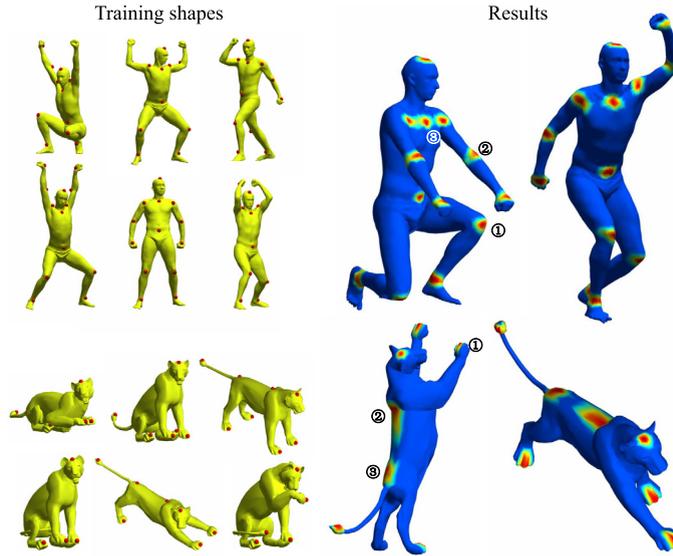


Fig. 8. Two examples of the automatic focal feature extraction on isometric shapes. The shapes on the left are training shapes with user-specified focal features, indicated by red sphere, and the shapes on the right are the automatic extraction results of our method. The space-varying colors represent the probability to be the desired focal feature. Here ① indicates salient points at shape extremities or tips, ③ indicates non-salient but semantically-meaningful points at flat regions, and ② indicates the points locating somewhere in between the above two kinds of regions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

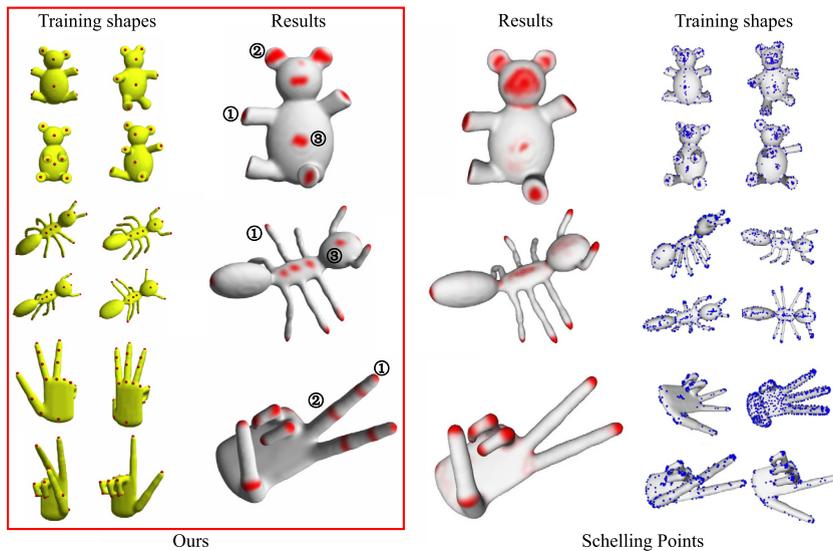


Fig. 9. Three examples of automatic focal feature extraction on near-isometric shapes with comparison to Schelling Points (Chen et al., 2012). The results of our method are listed on the left side marked by a red rectangle, and the results of Schelling Points are listed in the right side. The results of *Schelling Points* are obtained from their project website, and our results use a similar color style as they do for visual inspection. The intensity of the red color indicates the probability to be a feature point. Here ① indicates salient points at shape extremities or tips, ③ indicates non-salient but semantically-meaningful points at flat regions, and ② indicates the points locating somewhere in between the above two kinds of regions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

scales are indicate by the size of each shape visualized in Fig. 10. From the extraction results, we can observe that, even with excessive noise, many holes and different spatial scales, our approach can still precisely recognize the focal feature points. However, noise and holes indeed may influence the SI-HKS, and consequently reduce the precision of geodesic distances prediction, of which, the only difference between these results is that, the resulted votes' distribution on perturbed shape is more scattered than that on the original shape (see the second column of Fig. 10(a)). Because the affected predicted distances will be a little longer or shorter than the original ones. Thus, it means the predicted location will lie inside a larger neighborhood around the desired focal feature point. However, benefiting from our random sampling scheme and ensemble learning enabled random forest regressor, in addition that our geodesic distances distribution are hardly changed

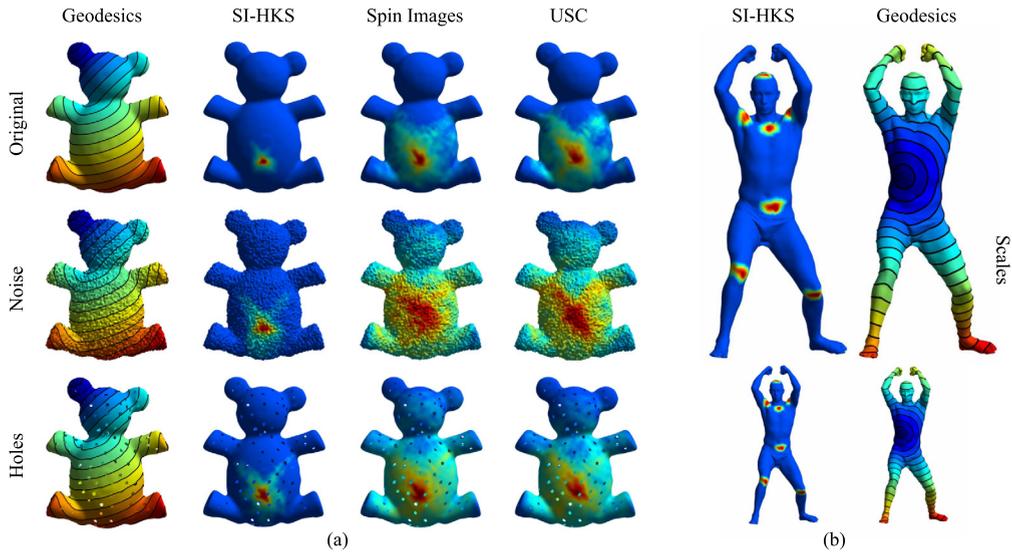


Fig. 10. The automatic feature extraction results of our method over different shape groups with noise, holes and spatial scales. (a) Extraction results using SI-HKS, Spin Images and Unique Shape Context as local point descriptors in each column. (b) Extraction results of our method under different scales. The isolines on each shape indicate the geodesic distance distribution with certain surface points as anchors. (For interpretation of the colors in this figure, the reader is referred to the web version of this article.)

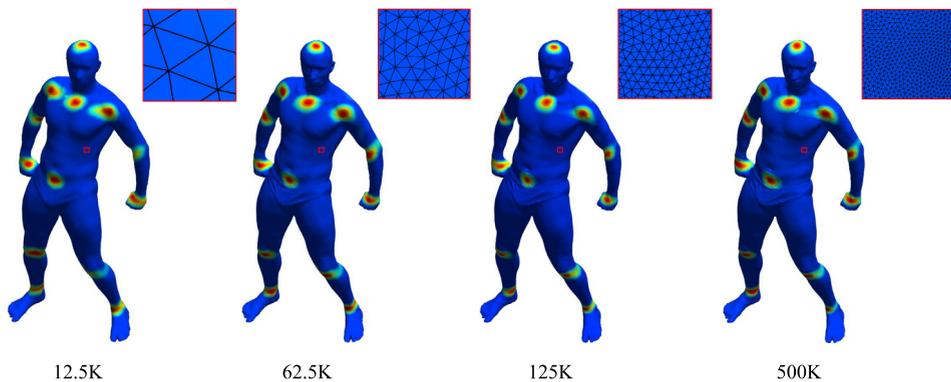


Fig. 11. The focal feature extraction results over different-resolution shapes. The shapes are produced by re-sampling the shape in Fig. 8 with 12.5K, 62.5K, 125K, 500K vertices. The detailed triangulation of the region marked by a red rectangle on each shape is shown in the upper right. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

under noise and holes, our method can still correctly locate the focal feature point even with a more scattered votes' distribution, which exactly proves the robustness of our method.

Influence of different descriptors and distance metrics. Fig. 10(a) shows the feature extraction results of our method when respectively using SI-HKS (Bronstein and Kokkinos, 2010), Spin Images (Johnson, 1997) and Unique Shape Context (USC) (Tombari et al., 2010) as local point descriptors. As we know, both Spin Images and Shape Context are sensitive to non-rigid deformations. When using them to compute descriptors in our framework, we only consider the points in certain neighborhood, that is, the descriptors are locally constructed. We can see that, the results based on SI-HKS is much better than those using the other two descriptors for all the three shape groups. On one hand, even though the Spin Image and Shape Context are locally constructed, many descriptors may still be influenced by the deformations, wherein some bad predictions would make the voting sources more scattered. On the other hand, these two descriptors are relatively sensitive to noise and holes, while SI-HKS is both nearly-invariant and robust under isometric deformations and noise or holes. As shown in Fig. 4, the result based on Geodesics-in-Heat is nearly the same with that based on Exact geodesics (which serves as ground truth). In contrast, the results based on other distance metrics are much more scattered because of their produced uneven distance iso-lines. Of which, a small difference in distance may give rise to large space on the surface, especially for the points that are far from the focal feature point, and thus it will make the final voting sources fall in a much more large region even with definitely-accurate distance prediction.

Robustness to shape resolution. Our focal feature extraction is also robust when the testing shape has different resolutions. We re-sampled the human shape in Fig. 8 with 12.5K, 62.5K, 125K, 500K vertices. As we know, $10 \times 12.5K = 1.25K$

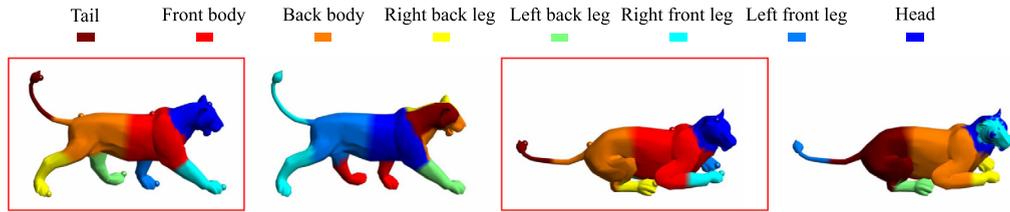


Fig. 12. Semantic segmentation based on our focal feature extraction. The first and third segmentations, marked with red rectangles, are the results using our feature points as initial centers for *k*-means clustering. And the second and fourth ones are the results using random initial centers. The color indicates the semantic labels of each segments, which is same with the label of the feature point. However, randomly initialized clustering centers give rise to bad segmentations without semantic meaning. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

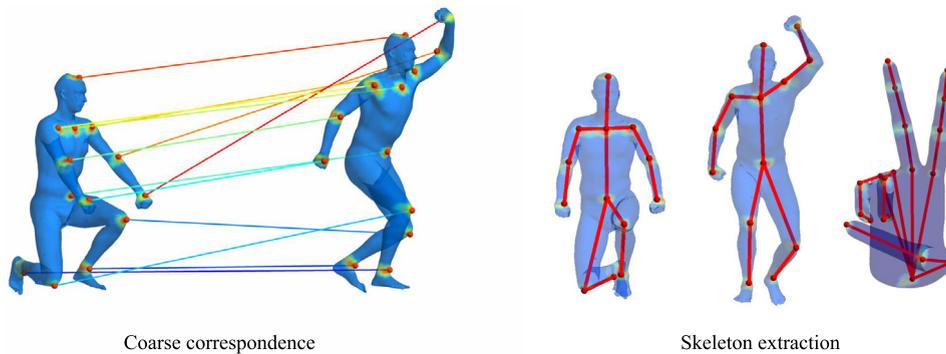


Fig. 13. Coarse correspondence and skeleton extraction applications based on our automatic focal feature extraction method. (For interpretation of the colors in this figure, the reader is referred to the web version of this article.)

points can cover the entirely human surface and produce precise voting result (in Fig. 8), so we randomly sample 1.25K points on all these 4 shapes and compute their descriptors, and then invoke the forest constructed for Fig. 8 to predict distances, the final voting results are shown in Fig. 11. We can see that, the focal feature extraction results on different-resolution shapes are almost the same with each other, wherein the feature points are all successfully extracted despite the point density on each shape. Because of the simplicity and effectiveness of our method, we are able to deal with 3D shape with up to 500K points within acceptable time and memory storage.

Applications of our method. Fig. 12 shows an example of semantic segmentation based on our feature extraction. Firstly, we extract 8 feature points just the same as those in Fig. 8, and then we conduct *k*-means clustering for each points on the shape by using the 8 feature points as the initial center of each cluster. Although the clustering only considers the coordinates and the SI-HKS properties of each point, it gives rise to high quality semantic segmentation. If we assign the feature points with different semantic labels during training, such as head, legs and tail, the segmented region on the testing shapes can share the same label with the corresponding feature point in this region. Besides, as shown in Fig. 13, our extracted feature points naturally have 1-to-1 correspondence among different shapes, and thus can be used to produce coarse correspondences among shapes, which can also serve as the input to the coarse-to-fine correspondence method (Sahillioglu and Yemez, 2011) to extended dense correspondences. Moreover, as shown in Fig. 13, our extracted feature points are also consistent across different shapes. Therefore, if we define a coarse skeleton on the training shape using such feature points, we can also directly extract a similar skeleton on the testing shapes, which can be used as the initialization of strictly skeleton extraction algorithm or directly used for simple animation generation.

Performance and accuracy analysis. Table 1 shows the performance statistics of all our experiments. We can see that, given a new shape, it only takes a few seconds to predict the location of the desired focal feature points, which makes our method very suitable for practice use. The errors are shown in Table 2. To better illustrate the capacity of our method, we divide the feature points into 3 kinds. The first kind is the feature point that lies on the shape extremities or tips (salient points, marked with ① in each figure). The second kind is the point that lies on a large smoothed or flat region (non-salient but semantically-meaningful points, marked with ③ in each figure). The third kind is the point that lies somewhere in between the above two regions (marked with ② in each figure). Each error is measured by the Euclidean distance between the predicted feature point and its desired location divided by the average mean edge length of the testing shape. We can see that the predicted feature points always lie in the 3-ring or 4-ring neighborhood of the ground truth points. And the errors of points at different regions have no obvious distinctions, because the final location of our focal feature is jointly decided by many other sampled points rather than itself, thus the error is irrelevant to where the desired point lies. Specially, of which, some errors are caused by the users' inaccuracy selection, because the strict correspondence cannot be well guaranteed among the feature points on training shapes.



Fig. 14. The extraction results for the user-specified feature point on stomach in Fig. 8. Here the testing shapes are respectively perturbed as: missing head, missing arms, and non-isometric deformations. Our method works for the first case but fails for the last two ones. (For interpretation of the colors in this figure, the reader is referred to the web version of this article.)

6. Conclusion

In this paper we have detailed a novel generic model to flexibly definite and automatically extract psychology involved focal features on 3D mesh surface. The key idea is to bridge the random-sampled local properties of the global shape with the location of the desired focal features via ensemble learning. Our feature extraction methodology and its system framework involve several novel technical components, including: scale-invariant and symmetry-sensitive random samples' representation, robust measurement of relative displacement on 3D shape, well-designed random forest regressor to map the random samples' descriptors to the corresponding geodesic distances, and a joint voting strategy for locating the focal feature points. Moreover, we have also designed diverse types of experiments over different kinds of shapes, which all demonstrate the advantages and great potentials of our method.

Limitations. Despite the attractive methodology properties of our method, it still has some limitations. The first one is that, the performance is still inefficient for interactive applications. Even though we have implemented the voting process in parallel based on GPU, the computations of geodesics and descriptors are still conducted on CPU, and they are a little expensive (especially when the shape has massive vertices). Moreover, as shown in Fig. 14, when we remove the head of the human model in Fig. 8, our method still works and successfully extracts the point on the stomach. Because many points are distant from this hole, their SI-HKS are not affected, and the geodesic distance distribution is not changed (the longest distance is still from left hand to right leg or right hand to left leg). However, if we remove the arms, it will change the geodesic distance distribution, and thus our method fails in this case. Furthermore, our method is mainly based upon an assumption that, the geodesic distances between points with similar properties should be similar across shapes. Meanwhile, we use geodesic distances as metric to locate target feature point, and use SI-HKS as point description (which is only invariant under isometry). Thus, currently our method can only accommodate the shapes with isometric or near-isometric deformations. As shown in the last sub-figure of Fig. 14, when testing a man with non-isometric deformation, our method fails.

Future works. In our upcoming work, we will endeavor our efforts to improve the efficiency by also computing descriptors and geodesic distances on GPU in a parallel way, so that our method can be used for interactive applications. Moreover, since our technical foci is to automatically extract user-specified feature points on the training shapes, it conceptually has some relations with the techniques like (Ovsjanikov et al., 2012; Solomon et al., 2012) in terms of constructing maps between shapes. Maybe we can extend our idea to conduct regional maps across shapes. A simple way to achieve this is to encode the region with a center point and a radius, so that the problem can be casted to extract the center. Besides, to make our method be applicable for shapes with non-isometric deformation, we should further develop a flexible descriptor that only depicts the relative locations of points, so that the invariance of descriptors can be achieved from the nature of similar shapes' intrinsic structures. Furthermore, if region based and non-isometric detection are achieved, we will also exploit the practical application potentials of our method in medical fields, such as automatic detection of different parts of a human body (head, hand and foot), or identification of the i -th column of a human spine, which are significant for computer-aid diagnosis, etc.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61190120, 61190121, 61190125, 61672149, 61602341, 61532002 and 61672077).

Appendix A. Supplementary material

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.cagd.2016.10.003>.¹

¹ This supplementary video introduces the basic idea of our method using animations and also shows the results provided in the manuscript.

References

- Bosch, A., Zisserman, A., Munoz, X., 2007. Image classification using random forests and ferns. In: Proceedings of the 11th IEEE International Conference on Computer Vision. IEEE, pp. 1–8.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Bronstein, M.M., Kokkinos, I., 2010. Scale-invariant heat kernel signatures for non-rigid shape recognition. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 1704–1711.
- Castellani, U., Cristani, M., Fantoni, S., Murino, V., 2008. Sparse points matching by combining 3D mesh saliency with statistical descriptors. *Comput. Graph. Forum* 27 (2), 643–652.
- Chen, X., Saparov, A., Pang, B., Funkhouser, T., 2012. Schelling points on 3D surface meshes. *ACM Trans. Graph.* 31 (4), 29.
- Chu, C., Chen, C., Nolte, L.P., Zheng, G., 2014. Fully automatic cephalometric x-ray landmark detection using random forest regression and sparse shape composition. In: Proceedings of ISBI International Symposium on Biomedical Imaging.
- Crane, K., Weischedel, C., Wardetzky, M., 2013. Geodesics in heat: a new approach to computing distance based on heat flow. *ACM Trans. Graph.* 32 (5), 152.
- Criminisi, A., Shotton, J., 2013. Decision Forests for Computer Vision and Medical Image Analysis. Springer Science & Business Media.
- Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E., 2011. Regression forests for efficient anatomy detection and localization in CT studies. In: Proceedings of the 2010 International MICCAI Conference on Medical Computer Vision: Recognition Techniques and Applications in Medical Imaging. MCV'10. Springer-Verlag, pp. 106–117.
- Dey, T.K., Fu, B., Wang, H., Wang, L., 2015. Automatic posing of a meshed human model using point clouds. *Comput. Graph.* 46, 14–24.
- Fanelli, G., Gall, J., Van Gool, L., 2011. Real time head pose estimation with random regression forests. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 617–624.
- Frome, A., Huber, D., Kolluri, R., Bülow, T., Malik, J., 2004. Recognizing objects in range data using regional point descriptors. In: Computer Vision-ECCV 2004. Springer, pp. 224–237.
- Gall, J., Lempitsky, V., 2013. Class-specific hough forests for object detection. In: Decision Forests for Computer Vision and Medical Image Analysis. Springer, pp. 143–157.
- Hua, J., Lai, Z., Dong, M., Gu, X., Qin, H., 2008. Geodesic distance-weighted shape vector image diffusion. *IEEE Trans. Vis. Comput. Graph.* 14 (6), 1643–1650.
- Huang, Q.-X., Adams, B., Wicke, M., Guibas, L.J., 2008. Non-rigid registration under isometric deformations. *Comput. Graph. Forum* 27 (5), 1449–1457.
- Johnson, A.E., 1997. Spin-images: a representation for 3-d surface matching. Ph.D. thesis. Microsoft Research.
- Kakarala, R., Kaliamoorthi, P., Premachandran, V., 2013. Three-dimensional bilateral symmetry plane estimation in the phase domain. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 249–256.
- Katz, S., Leifman, G., Tal, A., 2005. Mesh segmentation using feature point and core extraction. *Vis. Comput.* 21 (8–10), 649–658.
- Lipman, Y., Funkhouser, T., 2009. Möbius voting for surface correspondence. *ACM Trans. Graph.* 28 (3), 72.
- Lipman, Y., Rustamov, R.M., Funkhouser, T.A., 2010. Biharmonic distance. *ACM Trans. Graph.* 29 (3), 27.
- Liu, Y., Zha, H., Qin, H., 2006. Shape topics: a compact representation and new algorithms for 3D partial shape retrieval. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2. IEEE, pp. 2025–2032.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* 60 (2), 91–110.
- Lui, L.M., Thiruvenkadam, S.R., Wang, Y., Thompson, P.M., Chan, T.F., 2010. Optimized conformal surface registration with shape-based landmark matching. *SIAM J. Imaging Sci.* 3 (1), 52–78.
- Maes, C., Fabry, T., Keustermans, J., Smeets, D., Suetens, P., Vandermeulen, D., 2010. Feature detection on 3D face surfaces for pose normalisation and recognition. In: Proceedings of 2010 Fourth IEEE International Conference on Biometrics: Theory Applications and Systems. IEEE, pp. 1–6.
- Montillo, A., Shotton, J., Winn, J., Iglesias, J.E., Metaxas, D., Criminisi, A., 2011. Entangled decision forests and their application for semantic segmentation of ct images. In: Information Processing in Medical Imaging. Springer, pp. 184–196.
- Ohtake, Y., Belyaev, A., Seidel, H.-P., 2004. Ridge-valley lines on meshes via implicit surface fitting. *ACM Trans. Graph.* 23 (3), 609–612.
- Ovsjanikov, M., Ben-Chen, M., Solomon, J., Butscher, A., Guibas, L., 2012. Functional maps: a flexible representation of maps between shapes. *ACM Trans. Graph.* 31 (4), 30.
- Ovsjanikov, M., Bronstein, A.M., Bronstein, M.M., Guibas, L.J., 2009. Shape google: a computer vision approach to invariant shape retrieval. In: Proceedings of NORDIA, vol. 1.
- Pratikakis, I., Spagnuolo, M., Theoharis, T., Veltkamp, R., 2010. A robust 3d interest points detector based on harris operator. In: Proceedings of Eurographics Workshop on 3D Object Retrieval, vol. 1. Citeseer.
- Qiu, H.J., Hancock, E.R., 2007. Clustering and embedding using commute times. *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (11), 1873–1890.
- Ruggeri, M.R., Patanè, G., Spagnuolo, M., Saupe, D., 2010. Spectral-driven isometry-invariant matching of 3D shapes. *Int. J. Comput. Vis.* 89 (2–3), 248–265.
- Sahillioğlu, Y., Yemez, Y., 2011. Coarse-to-fine combinatorial matching for dense isometric shape correspondence. *Comput. Graph. Forum* 30 (5), 1461–1470.
- Solomon, J., Nguyen, A., Butscher, A., Ben-Chen, M., Guibas, L., 2012. Soft Maps Between Surfaces. *Comput. Graph. Forum*, vol. 31. Wiley Online Library, pp. 1617–1626.
- Sun, J., Ovsjanikov, M., Guibas, L., 2009. A concise and provably informative multi-scale signature based on heat diffusion. *Comput. Graph. Forum* 28 (5), 1383–1392.
- Surazhsky, V., Surazhsky, T., Kirsanov, D., Gortler, S.J., Hoppe, H., 2005. Fast exact and approximate geodesics on meshes. In: ACM Transactions on Graphics (TOG), vol. 24. ACM, pp. 553–560.
- Tombari, F., Salti, S., Di Stefano, L., 2010. Unique shape context for 3D data description. In: Proceedings of the ACM Workshop on 3D Object Retrieval. ACM, pp. 57–62.
- Wang, S., Li, N., Li, S., Luo, Z., Su, Z., Qin, H., 2015. Multi-scale mesh saliency based on low-rank and sparse analysis in shape feature space. *Comput. Aided Geom. Des.* 35, 206–214.
- Witten, I.H., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.
- Xu, K., Zhang, H., Tagliasacchi, A., Liu, L., Li, G., Meng, M., Xiong, Y., 2009. Partial intrinsic reflectional symmetry of 3D shapes. In: ACM Transactions on Graphics (TOG), vol. 28. ACM, p. 138.
- Yu, W., Li, M., Li, X., 2012. Fragmented skull modeling using heat kernels. *Graph. Models* 74 (4), 140–151.
- Zhang, E., Mischaikow, K., Turk, G., 2005. Feature-based surface parameterization and texture mapping. *ACM Trans. Graph.* 24 (1), 1–27.
- Zhang, H., Sheffer, A., Cohen-Or, D., Zhou, Q., Van Kaick, O., Tagliasacchi, A., 2008. Deformation-driven shape correspondence. *Comput. Graph. Forum* 27 (5), 1431–1439.
- Zou, G., Hua, J., Dong, M., Qin, H., 2008. Surface matching with salient keypoints in geodesic scale space. *Comput. Animat. Virtual Worlds* 19 (3–4), 399–410.