Think locally, fit globally: Robust and fast 3D shape matching via adaptive algebraic fitting

Shaodi You [a,b], Diming Zhang [c,*]

[a] Data61 - CSIRO, Canberra, ACT, Australia
[b] Australian National University, Canberra, ACT, Australia
[c] Jiangsu University of Science and Technology, Zhenjiang, China

ABSTRACT

In this paper, we propose a novel 3D free form surface matching method based on a novel key-point detector and a novel feature descriptor. The proposed detector is based on algebraic surface fitting. By global smooth fitting, our detector achieved high computational efficiency and robustness against non-rigid deformations. For the feature descriptor, we provide algorithms to compute 3D critical net which generates a meaningful structure on standalone local key-points. The scale invariant and deformation robust Dual Spin Image descriptor is provided based on the 3D critical net. Our method is proved by solid mathematics. Intensive quantitative experiments demonstrate the robustness, efficiency and accuracy of the proposed method.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Free form 3D surface matching is the technique to find the correspondences between surfaces. There is a increasing demand for efficient and accurate 3D surface matching techniques because it serves as a fundamental method for a number of 3D based computer vision techniques, *e.g.*, tracking [35], recognition [3], classification [7,16,17], retrieval [1,40], registration [15,27,36], modeling [9,13,37], morphing [31,38] and BRDF estimation [21–23].

Matching surfaces is not a trivial task when the surface is deformable. Deformation occurs in many situations, *i.e.*, when the surface itself is non-rigid and when the sensor is not well calibrated. Non-rigid object is quite common: human, animals, liquid are all highly non-rigid. Sensor deformation is also common for structure from motion or a moving range sensor.

Mathematically, in global aspect, deformation can be modeled using a smooth warping function, however, this is not easy because the deformation could be highly non-linear and topological inconsistent. In local aspect, deformation can be modeled using linear transforms. However, these modeling could be highly sensitive to noise.

In general state-of-art sparse matching technique is desired to be:

Robust: It is basically desired a shape detector should be robust against noise, sampling resolution, data missing, and rigid deformations as well as non-rigid deformations.

Topology independent: 3D data are frequently used in many computer vision and graphic tasks, *e.g.* CAD model (complete and noiseless surface), range data (incomplete and noisy surface), structure from motion (sparse and noisy point cloud). And thus, a surface matching technique is desired to be generally applicable in various scenarios.

Efficient: Computational efficiency is always desired in computer vision tasks.

In this paper, we model the shape deformation by using algebraic surface fitting (ASF) [30]. The proposed local key-point detector - The 3D implicit polynomial detector (IP detector) - detects key-point on the original model according to the energy field provided by ASF. Our detector is fast by adopting an ASF technique named Adaptive Fitting [32]. The efficiency is further improved by our speedup strategy. The overall detecting on 100,000 points can be finished within 10s. Our detector is range data compatible, for ASF is insensitive to topology, boundary and data missing. By modeling the deformation, our detector achieves invariance against scaling and rotation; and robustness against noise, data missing and free form deformation. The invariance and robustness are demonstrated by mathematical proofs and experiments.

For feature description, we first generate a graphic structure on stand alone keypoints, call the 3D critical net. Based on it, we extent the local shape descriptor Spin Image [14] to Dual Spin Image (DSI) which achieves invariance against scaling and robustness

* Corresponding author.
  E-mail address: diming.zhang@gmail.com (D. Zhang).
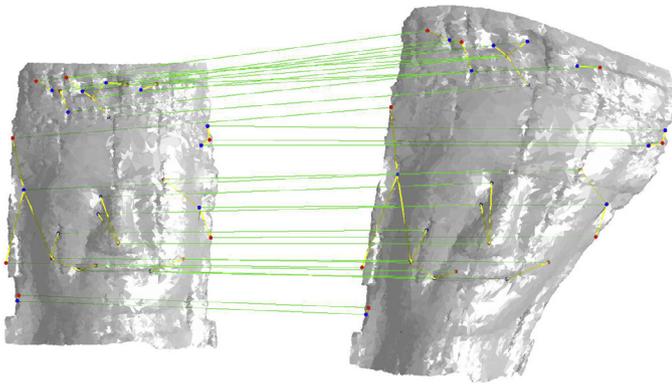
Fig. 1. Robust matching of 3D surfaces with free form deformation. Blue and red makers denote key-points found by algebraic surface fitting. Yellow lines are 3D critical net connecting key-points. Green lines show the match between surfaces. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

against free form deformation. The performance is demonstrated by quantitative matching experiments.

The contribution of this paper is threefold: First, we introduce the theoretical insight on tackling the non-rigid registration problem as an multi-scope algebraic geometry problem. Second, we propose a novel sparse feature detector by exploited the residual field of algebraic surface fitting. And third, we propose a context-aware feature descriptor based on critical net of the sparse feature points.

Fig. 2 illustrate the idea of the proposed method and Fig. 1 shows and example of the surface matching.

The rest of the paper is organized as follows: Section 2 discusses the related work in surface matching, non-rigid shape modeling. Section 3, the ASF, which is used to statistically model the shape deformation, is introduced. 3-layers method and adaptive fitting method is introduced which improves the fitting robustness and computational efficiency correspondingly. Base on it, in Section 4, the method to utilize the modeled deformation for point detection is provided along with mathematical proof. After that, in Section 5, 3D critical net is introduced and the local shape descriptor: Dual Spin Image (DSI) is introduced. Section 6 is the quantita-

tive experiments which demonstrate the robustness and efficiency of our detector and descriptor. Section 7 is the discussion and conclusion.

## 2. Related work

2.1. Matching rigid surfaces

Early research works are more interested in exploiting matching rigid surfaces, say, surface with no distortion. These methods are usually used in retrieval and alignment. These methods can be categorized as global feature based methods and local feature based methods.

Global feature based methods has a long history in computer vision. It characterize the global distribution of a 3D surface: Sederberg and Anderson [30] represent surface by algebraic surface fitting (ASF) and extract feature from polynomial coefficients. Kang and Ikeuchi [18] represent 3D models by mapping the normal orientation and normal distance of the surface point to a unit sphere; Osada et al. [25] introduce shape distributions on distance, angle, area and volume of random surface points. However, these global matching techniques do not work with partial data that come from range sensor. And they are not developed to be deformation-robust.

Local feature based surface matching has shown advantages in handling detailed or partial data. Local feature works from the neighborhood of points on the surface: Chua and Jarvis [5] compute point signatures that accumulate surface information along a 3D curve in the neighborhood of a point. Mian et al. extract feature by evaluating the 3D basis and principal curvature of local neighborhood [24]. Although these techniques work with incomplete data but they are not developed to be deformation-robust either. Some local shape information, *i.e.* local curvature, are sensitive to both noise and deformation.

2.2. Non-rigidity modeling

By limiting the deformation to isometric invariant, Elad et al. [8] use the Isomap embedding that embed the deformed shape to an isometric invariant canonical form. Then shape matching is performed between the canonical forms. Raviv et al. [26] make an
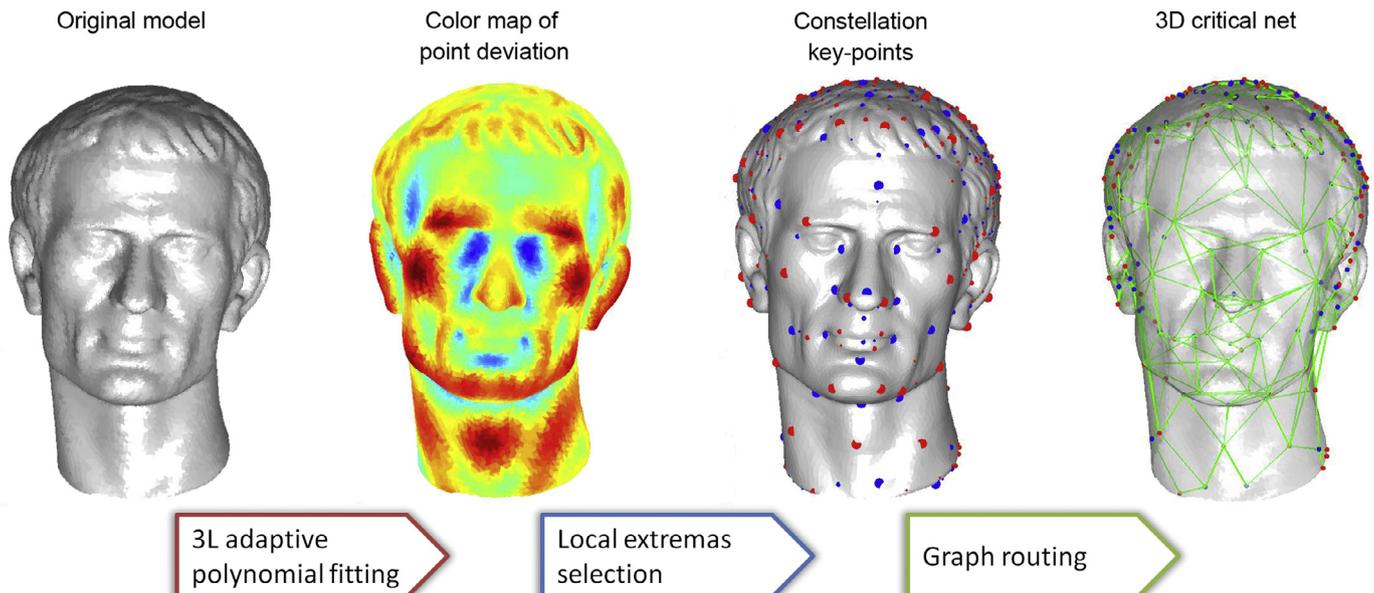
Fig. 2. Workflow of the proposed 3D shape key-point detector.
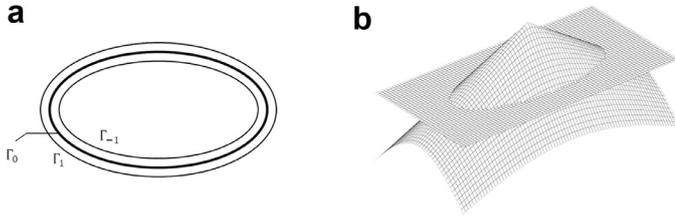
Fig. 3. Illustration of 3L fitting in 2D case. (a)Illustration of the 3 layers: Γ_0 the original shape, Γ_1 its dilation and Γ_{-1} its corrosion. (b) Viewing the polynomial as a 3D surface, the fitted shape $\widetilde{\Gamma}_0$ is the intersection of the polynomial with $z = 0$ plane.

extension that the canonical form is also affine-invariant. Lipman et al. [20] relax the restriction from isometric to conformal and embed the deformed shape to the conformal invariant canonical form.

Manifold learning, *e.g.* ISOMAP [33] and LLE [28] enable the access to deformable 3d shape matching. By limiting the deformation to isometric invariant, Elad et al. [8] use the ISOMAP embedding that embed the deformed shape to an isometric invariant canonical form. Then shape matching is performed between the canonical forms.

There are several limitations of the existing methods: 1. The isometric or conformal invariance is too strong which is scarcely satisfied in real practices. [19,39] 2. Furthermore, these techniques do not work with incomplete data which have arbitrary boundary because they are very sensitive to data integrity and boundary. 3. Embedding based methods are computational high-cost because they need to do eigen-decomposition on dense matrices. For example, the ISOMAP embedding is $O(n^3) - O(n^6)$ complex, where n is the number of data points. It needs about one minute to solve only 2000 points. Shape with more than 10,000 points, which is very common for both CAD and range data, cannot be solve in a reasonable time. Several speed up techniques are proposed. A Sahillioglu et al. reported their performance on finishing 20,000 points within three minutes. [29] However, the performance is still on satisfying for real practices.

2.3. Local and sparse matching

To avoid explicitly modeling the non-linear deformation, sparse correspondence take in the idea that local deformation can be well approximated by linear transforms. Sparse correspondences between deformed shape could be applied to non-rigid object retrieval, recognition, registration; it also enables dense matching of non-rigid shapes. Basing on sparse correspondences, Sharma et al. [31] and Tung et al. [35] explore the dense correspondence on deformed shapes and applied their methods to rigid shape morphing and motion tracking.

It is worthy to notice that, all the existing deformation compatible techniques are within the scope of Riemannian geometry, say, manifold learning techniques which aim to embed the deformed shape to a deformation invariant canonical form. In this paper, we try to deal with the deformation in an novel way: we implicitly model the deformation, rather than remove the deformation, we represent the deformation as an energy field. 1. we do not need to make very strong assumption on the deformation; 2. Modeling is statistical so that it is not topological sensitive or boundary sensitive, thus both CAD and range data are compatible; 3. The method is computational efficient.

3. Algebraic surface fitting

Algebraic surface fitting (ASF) has a long history in computer vision as being using in modeling 3D objects [4,30,32]. As a pre-

requisite of the proposed method, in this section, we introduce the basic idea as well as the algorithm.

In 3D case, implicit algebraic surface is defined by an n degree implicit polynomial (IP) as:

$$f_n(\boldsymbol{x}) = \sum_{0 \leq i,j,k, i+j+k \leq n} a_{ijk} x^i y^j z^k = 0,$$

$$\boldsymbol{x} = (x, y, z) \in \mathbb{R}^3. \tag{1}$$

Given 3D shape Γ_0 represented by a set of M points:

$$\{\boldsymbol{x}_m\} = \{(x_m, y_m, z_m)\}, m = 1, 2, \ldots, M, \tag{2}$$

given the degree of IP n, ASF is aiming to find the set of coefficients $\{a_{ijk} | 0 \leq i, j, k, i+j+k \leq n\}$ which minimizes the ℓ_2 error:

$$E = \sum_{m=1}^{M} (f_n(\boldsymbol{x}_m))^2. \tag{3}$$

The found IP is thus denoted as $\widetilde{f}_n(\cdot)$:

$$\sum_m \widetilde{f}_n(\boldsymbol{x}_m)^2 = \min_{f_n(\cdot)} E = \min_{f_n(\cdot)} \sum_m f_n(\boldsymbol{x}_m)^2. \tag{4}$$

The zero surface $\widetilde{\Gamma}_0$:

$$\widetilde{\Gamma}_0 = \{\boldsymbol{x} = (x, y, z) | f(\boldsymbol{x}) = 0\}, \tag{5}$$

is desired to be a smooth approximation of the input surface. Fig. (4) shows an example.

There are two challenges of ASF. The first is the over-fitting. And the second is the efficiency. Fortunately, both are solve as introduced in the following subsections. Section 3.1 introduces the 3L method which significantly robustifies the fitting. And Section 3.2 introduces the adaptive fitting which improves the performance.

3.1. 3L fitting

It is well known that the original polynomial fitting is very likely to fall into a over-fitting solution. 3L (3 layers) fitting, proposed by Blane et al. [4], has significantly improved the stability of IP fitting. For clarity, 2D shapes are used as examples, all the methods also work for 3D surfaces.

For a given 2D shape $\widetilde{\Gamma}_0$, which consists of M points: $\{\boldsymbol{x}_m = (x_m, y_m, z_m), m = 1, 2, \ldots M\}$, and its corresponding polynomial $\widetilde{f}(x, y)$, The polynomial $\widetilde{f}(x, y)$ could be considered as a surface in 3D space:

$$(x, y, z) = (x, y, \widetilde{f}(x, y)) \in \mathbb{R}^3 \tag{6}$$

Here, the 2D shape represented by the zero set:

$$\widetilde{\Gamma}_0 = \{\boldsymbol{x} = (x, y) | \widetilde{f}(\boldsymbol{x}) = 0\} \tag{7}$$

is the intersection of the polynomial surface with the horizontal plane passing through the origin at $z = 0$. An illustration is shown in Fig. 3(a).

Now, let's consider the gradient of polynomial $\nabla \widetilde{f}(x, y)$ in the vicinity of the 2D shape $\widetilde{\Gamma}_0$. If $\|\nabla \widetilde{f}(x, y)\|$ is small in the area, or even if there are singularities (place where $\nabla \widetilde{f}(x, y) = 0$), change in the shape $\triangle (x, y)$ would not result in a big change of the polynomial value $\triangle \widetilde{f}(x, y)$ according to the relation:

$$\triangle \widetilde{f}(x, y) \approx \nabla \widetilde{f}(x, y) \triangle (x, y) \tag{8}$$

According to (4), IP fitting is found by minimizing overall polynomial norm. The fitting is not robust if the polynomial surface is too flat in the vicinity of $\widetilde{\Gamma}_0$, because a large change of the shape would only result in a small change of the polynomial value.

The 3L method (3 Layers) will search for polynomial surface with steep slides in the vicinity of 2D shape $\widetilde{\Gamma}_0$. For given shape
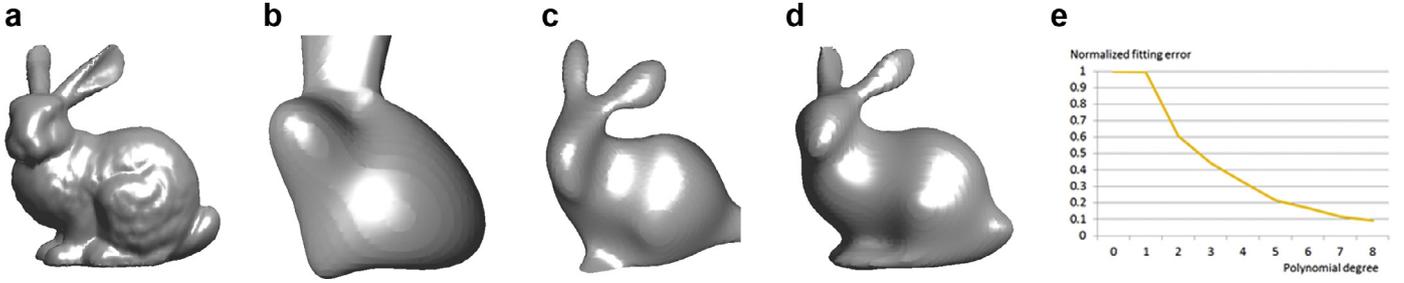
Fig. 4. Illustration of the incremental fitting. (a) The origin shape [41], (b) the approximate surface find by 6° ASF, (c) 8 degree, (d) 10°, (e) the normalized fitting error decreases while the polynomial degree increases.

Γ_0, 3L method first searches its dilation Γ_1 and its corrosion Γ_{-1}. Then the ASF is to find the optimal IP that satisfying:

$$f(\Gamma_0) = 0$$
$$f(\Gamma_1) = 1$$
$$f(\Gamma_{-1}) = -1 \qquad (9)$$

Fig. 3(a) is an illustration.

It is proved by Blane et al. that by conducting 3L method, singularities and small gradient polynomial $\triangledown \tilde{f}$ could be ruled out the in the vicinity of $\widetilde{\Gamma}_0$. Readers are encourage to refer to Appendix A of [4] for the proof.

3.2. Adaptive fitting

ASF is aiming to find the accurate smooth approximation of shapes. The accuracy could be calculated according to the fitting error (3). When the model is under-fitted. According to (1), the higher the polynomial degree is, the more the polynomial coefficients are, the more the accuracy will be[1]. Fig. 4 shows an example that the fitting error drops along with the increasing polynomial degree.

Notice that, according to (3), the fitting accuracy is not known until a given-degree-fitting is finished. There is a trade-off between fitting accuracy and computational efficiency: a low degree fitting is fast but might not meet the desired accuracy; a high degree fitting might guarantee the fitting accuracy but some unnecessary coefficients are calculated. Therefore, an appropriate polynomial degree that balances the accuracy and time cost is desired.

Fortunately, benefited from the adaptive solution of QR decomposition, the quadratic form can be solved in a similar adaptive strategy. Which means the internal optimization value calculated for n-degree fitting can be directly used in the $(n + 1)$-degree fitting. Readers are encouraged to refer to the detail in [32].

By far, we have introduced the original ASF, its robust modification 3L-ASF, its fast modification Adaptive-ASF. In the following, we will used the combined method and call it 3L-adaptive-ASF.

4. IP-based key-point detection

Our local key-point detection method is based on the polynomial value of ASF $\tilde{f}_n(\mathbf{x})$. We detect local extrema according to $\tilde{f}_n(\mathbf{x})$ and these extrema are considered to be invariant against scaling rotation, robust against deformations and noise.

The mathematic support of our method is consisted of two lemmas and one theorem, which will be introduced in Section 4.2.1 to Section 4.2.3. Strategy for key-point selection based on $\tilde{f}_n(\mathbf{x})$ are introduced in Section 4.3.1. After that, a speed up strategy is introduced in Section 4.3.2. The overall methodology is summarized in

---

[1] Here we assume the overfitting is well-solved by 3L fitting.
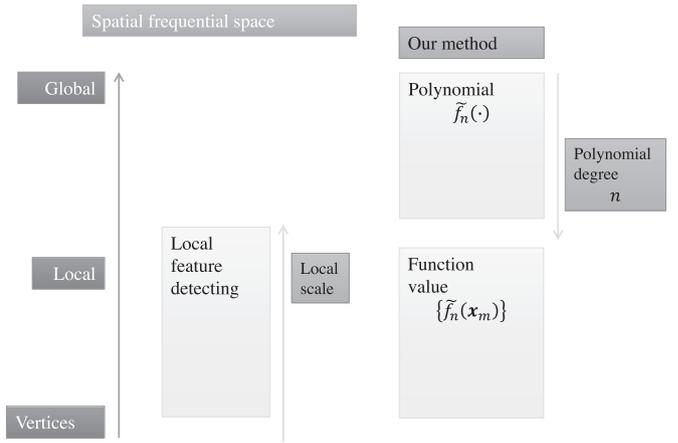
Fig. 5. The Spatial frequency space aspects of polynomial fitting. While most local detectors work from small scale to large scale. The polynomial fitting works in the opposite direction. The global information is extracted in the polynomial $\tilde{f}_n(\cdot)$; and the local information is distilled in function value $\{f_n(\mathbf{x})\}$.

Section 4.3.3. In Section 4.4, we will discuss the robustness of our detector.

4.1. Motivation

As introduced in Section 3.2, the 3L and incremental fitting strategy makes the state-of-art ASF robust and fast. ASF has a long history in computer vision as being using in modeling 3D shapes for recognition, retrieving, registration and *etc.* [4,30,32].

According to (1), Given a 3D shape Γ_0 consists of M points $\{\mathbf{x}_m\}$, the yielded polynomial $\tilde{f}(\cdot)$, the polynomial coefficients, is a good representation of the surface. Thus previous researches are focusing on the corresponding polynomial coefficients: $\tilde{f}(\cdot)$. Rotation, scaling and affine invariants are extracted from the polynomial coefficients $\{a_{ijk}\}$.

Mention that, feature extracted from $\tilde{f}(\cdot)$ are all global features. As discussed in the introduction, local feature are more information rich and versatile. In this research, we are aiming to develop a local key-point detector based on ASF. Thus, we consider ASF in a different view:

From $\tilde{f}(\cdot)$, we can get zero set $\widetilde{\Gamma}_0$ (see (5)) which is a smooth approximation of original shape Γ_0. Actually, for local key-point detection, the smooth part is not information rich. The polynomial values (residue) $\{\tilde{f}(\mathbf{x})\}$ of points $\{\mathbf{x}_m\}$ once were only considered as the error of ASF, contains all the local information. Unlike the original data $\{\mathbf{x}_m\}$, $\{\tilde{f}(\mathbf{x})\}$ are local information filtered by a global filter: $\tilde{f}(\cdot)$. This means, $\{\tilde{f}(\mathbf{x})\}$ contains both global and local information. Fig. 5 illustrates our idea. One explanation of $\{\tilde{f}(\mathbf{x})\}$ is introduced as follows:
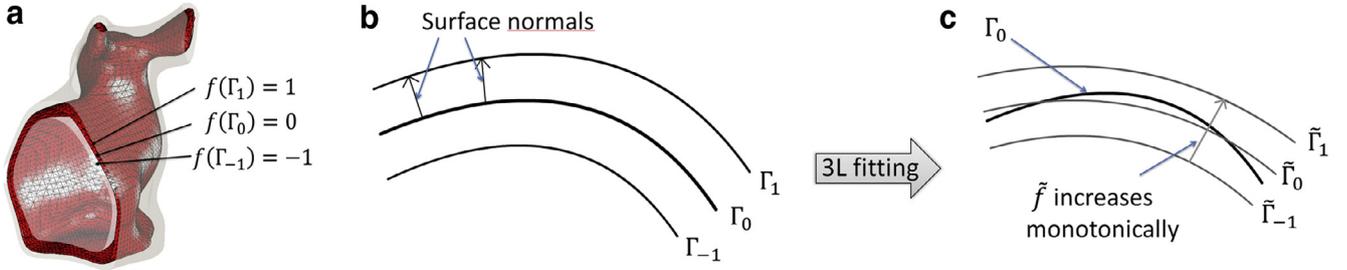
Fig. 6. Illustration of the point deviation represented by algebraic distance $\widetilde{f}$. (a) 3 layers to be fitted. (b) A local look of the 3 layers. (c) After the fitting, a monotonic field is generated around the original surface. Polynomial value $\widetilde{f}$ is the algebraic distance represents the deviation from $\widetilde{\Gamma}_0$.

4.2. Point deviation based on algebraic distance

4.2.1. Theorem
We claim the theorem:

Theorem. Given 3D shape Γ_0 consists of point set $\{x\}$ and its n-degree polynomial from 3L-adaptive-ASF $\widetilde{f}(\cdot)$ along with the induced smooth approximation $\widetilde{\Gamma}_0$. For each point x, its polynomial value $\widetilde{f}(x)$ is a locally signed algebraic distance [32] of how far it is deviated from $\widetilde{\Gamma}_0$.

4.2.2. Lemma1

Lemma 1. By 3L ASF, a locally monotonic space is generated around the 3D surface $\widetilde{\Gamma}_0$.

According to the idea of 3L fitting (9), other than the zero set $\widetilde{\Gamma}_0$, which is the smooth approximation of original shape Γ_0, we could also find the "1"-level-set $\widetilde{\Gamma}_1 = \{x|\widetilde{f}_n(x) = 1\}$ which is the smooth approximation of the dilation Γ_1. So does the "-1"-level-set: $\widetilde{\Gamma}_{-1} = \{x|\widetilde{f}_n(x) = -1\}$.

Because polynomials are everywhere smooth (infinite differential), actually, we could get any level set $\widetilde{\Gamma}_\theta$ for $-1 \le \theta \le 1$:

$$\widetilde{\Gamma}_\theta = \{x|\widetilde{f}_n(x) = \theta\} \tag{10}$$

This means, in the area bounded by $\widetilde{\Gamma}_1$ and $\widetilde{\Gamma}_{-1}$, denoted as Ω, the polynomial value is bounded by -1 and 1, and changes continuously. Fig. (6).

To prove Lemma 1, the left thing is to prove that: $\widetilde{f}(\cdot)$ is strict monotonic inside Ω. Or, equivalently, $\widetilde{f}(\cdot)$ has no singularity inside Ω:

$$\exists c > 0, \forall x \in \Omega, \|\nabla \widetilde{f}(\cdot)\| > c \tag{11}$$

As mentioned in Section 3.1, this part has already been proved by Blane et al. [4].

According to Lemma 1, given any point x on $\widetilde{\Gamma}_0$, by moving towards $\widetilde{\Gamma}_1$ its polynomial value $\widetilde{f}_n(x)$ increases from 0 to 1 continuous and strict monotonically and vice versa. Therefore $\widetilde{f}_n(x)$ is a signed algebraic distance to from point x to $\widetilde{\Gamma}_1$.

Mention that, Lemma 1 does not guarantee the polynomial value outside Ω. Make $\widetilde{\Gamma}_1$ and $\widetilde{\Gamma}_{-1}$ away from $\widetilde{\Gamma}_0$ will extend the area. However, $\widetilde{\Gamma}_1$ and $\widetilde{\Gamma}_{-1}$ are dilation and corrosion of $\widetilde{\Gamma}_0$, if they are too distanced, there shape resemblance will decrease. In practice, the distance is set to 5% in comparing with global shape size.

4.2.3. Lemma 2

Lemma 2. The points on 3D surface $\widetilde{\Gamma}_0$ are guaranteed to be located inside the locally monotonic space Ω with a very high probability.

According to Lemma 1, Ω is bounded by $\widetilde{\Gamma}_1$ and $\widetilde{\Gamma}_{-1}$ which are the smooth approximation of the original shape's corrosion Γ_1 and corrosion Γ_{-1} correspondingly. This means, it is not 100% guaranteed the original shape Γ_0 will locate inside Ω.

Actually, we have not specified the original shape Γ_0. The point distribution of Γ_0 can be very different. Noise and data missing could
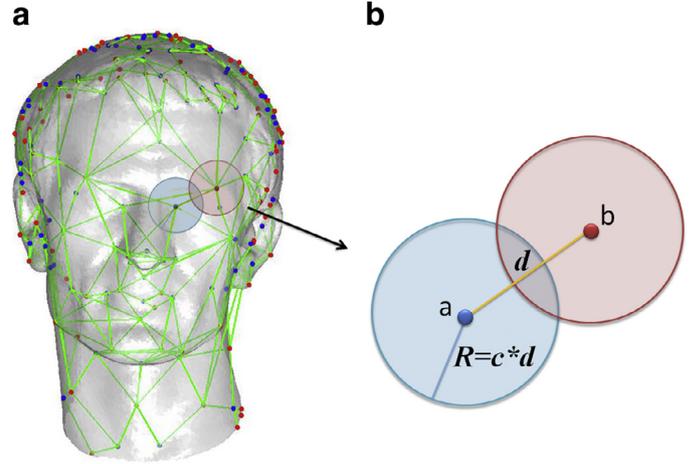
Fig. 7. Illustration of dual spin image (DSI). (a) Sampling area of DSI based on 3D critical net. (b) Structure of DSI.

also be including. Therefore, we approach this problem in the view of probability.

The points $\{x\}$ on Γ_0 are considered to be set of random points. After fitting, there corresponding polynomial values $\{\widetilde{f}_n(x)\}$ are also a set of random number. For any given point x, it is located inside Ω if its polynomial value satisfies:

$$-1 \le \widetilde{f}_n(x) \le 1 \tag{12}$$

If we could model the distribution of $\{\widetilde{f}_n(x)\}$, then we could calculate the probability.

Refer to (3), $\widetilde{f}_n(x)$ is the error of the ASF. And the algorithm presented in this paper is to minimize the over square error. A popular model for this situation is the "Normal distribution" centers at '0':

$$p_y(\widetilde{f}_n(x) = y) \approx N(0, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2\sigma^2}y^2\right\} \tag{13}$$

In practice, the exact distribution differs from model to model. In Fig. 9, we have shown the distribution of $\widetilde{f}_n(x)$ of several different models. Normal distribution centers at '0' is very plausible in these cases.
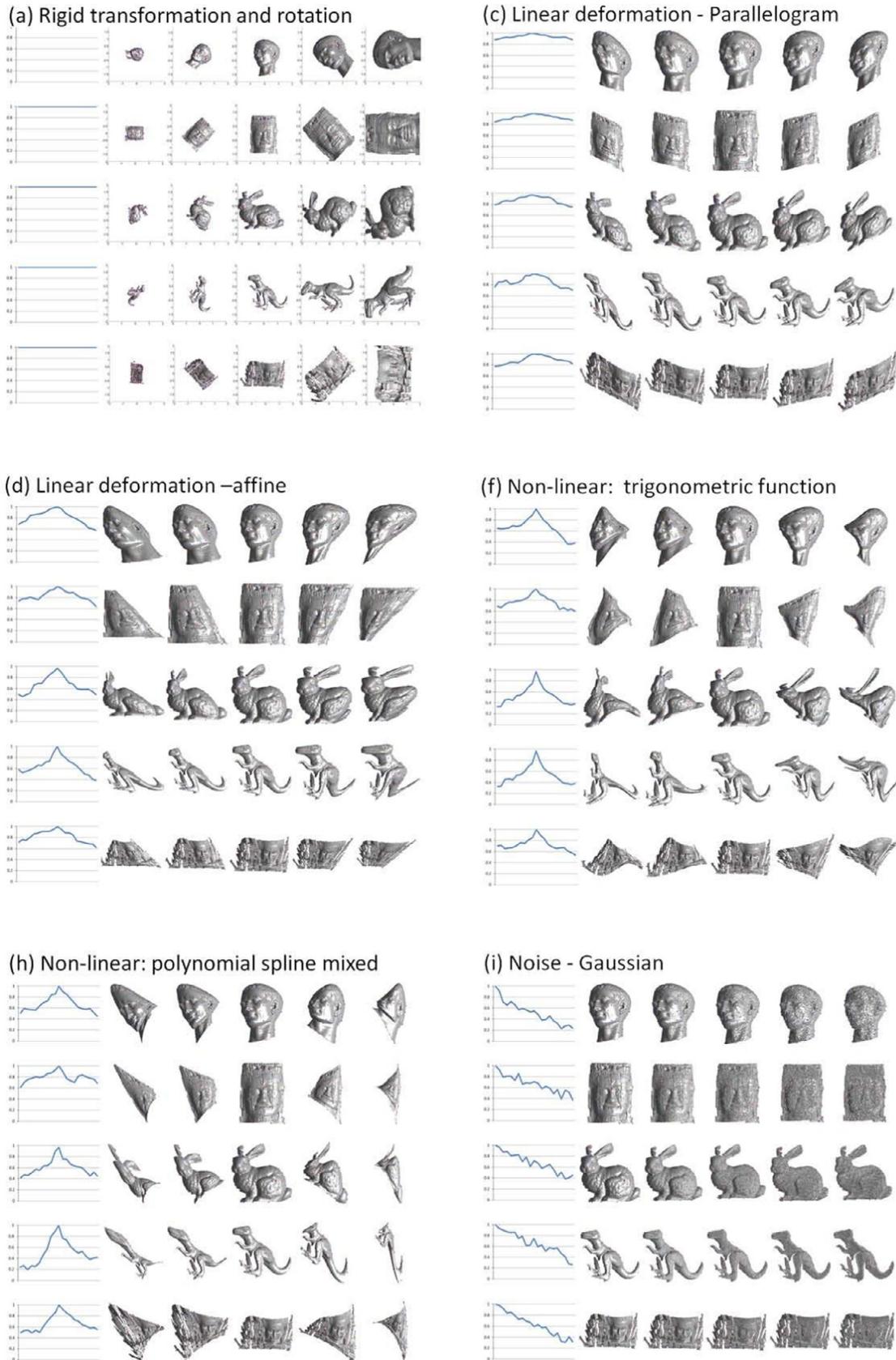
After modeling the distribution of $\widetilde{f}_n(x)$ we could calculate the probability that $-1 \le \widetilde{f}_n(x) \le 1$:

$$p(-1 \le \widetilde{f}_n(x) \le 1) = erf\left(\frac{1}{\sqrt{a}\sigma}\right) \tag{14}$$

Where $erf(\cdot)$ is the error function:

$$erf(a) = \frac{2}{\sqrt{\pi}} \int_0^a \exp(-y^2)dy \tag{15}$$

Some typical value of (15) is as follows:

Fig. 8. Tests on key-point repeatability. (a) Invariance against transformation and rotation. (b)–(h) Robustness against deformations. (h) Robustness against noise.
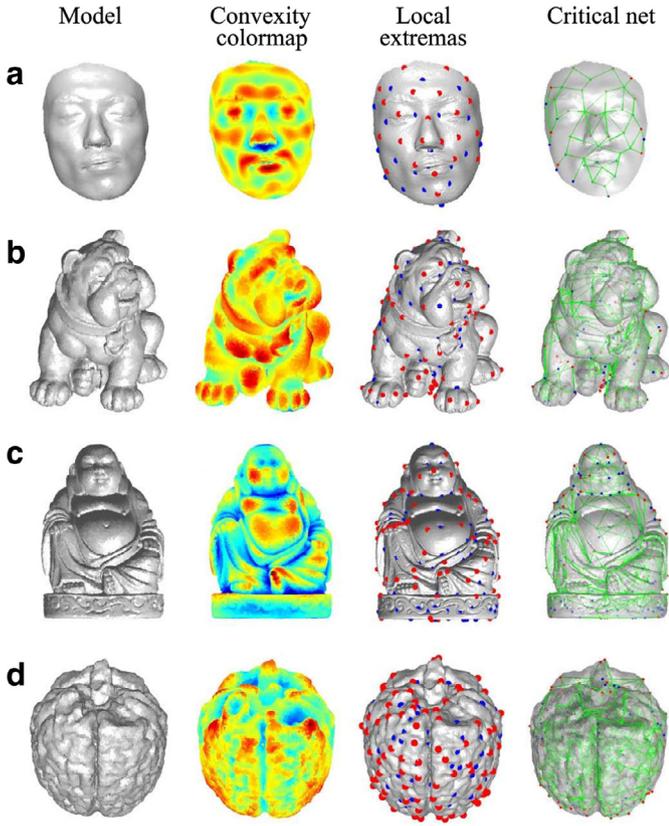
Model	Convexity colormap	Local extremas	Critical net

Fig. 9. Key-point detection on different 3D shapes.

If $\sigma < 0.39$, ($\sigma^2 < 0.16$), then $p(-1 \leq \widetilde{f}_n(\boldsymbol{x}) \leq 1) > 0.99$
If $\sigma < 0.32$, ($\sigma^2 < 0.10$), then $p(-1 \leq \widetilde{f}_n(\boldsymbol{x}) \leq 1) > 0.999$

By far σ is an internal parameter of normal distribution, now we interpret this value with the parameter from ASF. For the Normal distribution, σ^2 is the variance:

$$\sigma^2 = \frac{1}{M} \sum_{m=1}^{M} |\widetilde{f}_n(\boldsymbol{x}_m)|^2 - \left\{ \frac{1}{M} \sum_{m=1}^{M} \widetilde{f}_n(\boldsymbol{x}_m) \right\}^2 \tag{16}$$

For Normal distribution centers at '0',

$$\frac{1}{M} \sum_{m=1}^{M} \widetilde{f}_n(\boldsymbol{x}_m) = 0 \tag{17}$$

Substitute (17), into (16):

$$\sigma^2 = \frac{1}{M} \sum_{m=1}^{M} |\widetilde{f}_n(\boldsymbol{x}_m)|^2 \tag{18}$$

Compare (18) with (3), it is found σ^2 is exactly the average square fitting error $\bar{E}$:

$$\sigma^2 = \bar{E} = \frac{1}{M} \sum_{m=1}^{M} |\widetilde{f}_n(\boldsymbol{x}_m)|^2 \tag{19}$$

By far, we have proved that if the ASF is accurate enough, then the points on 3D surface Γ_0 are guaranteed to be located inside the locally monotonic space Ω with a very high probability. Typically, we fit the surface until the average square error is smaller than 1. In this situation, points on Γ_0 are located inside Ω with the probability greater than 0.999. It is enough for practices.

By far, we have proved Lemma 2. Together with Lemma 1, we have proved the theorem.

Fig. 9 shows examples of polynomial value of point on different models.

4.3. Methodology

4.3.1. Key-point selection

In Section 4.2, we have proposed the method to generate the reference and evaluation the deviation from reference by using ASF. For a given 3D shape Γ_0 consists of point set $\{\boldsymbol{x}\}$ and their polynomial value $\widetilde{f}_n(\boldsymbol{x})$, we select local extremas as key-points according to $\widetilde{f}_n(\boldsymbol{x})$.

The key-point selection is computational efficient. It only needs to compare their corresponding polynomial value (residue). The comparison could start from very small neighborhood, where very few comparisons are needed. To find local extremas in larger neighborhood, fewer comparisons are needed, because only those smaller-neighborhood-extremas need to be compared.

The radius of the neighborhood could also be used as an tag of the detected extrema. Fig. 2(c) is an example of the extremas found in neighborhood with different radius.

4.3.2. Speedup

Our method inherits the computational efficiency from the adaptive fitting. In this subsection, we introduce a down-sampling strategy which further speedup the detection without losing any local information.

The time complexity of adaptive ASF is $O(Mk^3)$ where M is the number of points of the model and k is the number of coefficients of the polynomial. Number of polynomial coefficients is determined by the degree of polynomial. Some typical numbers are listed as Table 1.

Reduce the polynomial degree will improve the computational efficiency, however, the tradeoff is the fitting accuracy. To support Lemma 2, an upper bound of the fitting accuracy is provided, which indicates the lower bound of the number of polynomial coefficients.

An alternative way of improve the computation efficiency is by reducing the number of points of the model: down-sampling. We mention that down-sampling is generally not applicable for local key-point detection, because down-sampling significantly reduce the detailed information from the model. However, in our method, down sampling could be introduced to ASF which is the most time consuming procedure. The original points $\{\boldsymbol{x}\}$ are used to train the polynomial $\widetilde{f}_n(\cdot)$. Then, $\widetilde{f}_n(\cdot)$ is used to calculate the polynomial value $\widetilde{f}_n(\boldsymbol{x})$. Referring to Table 1, to train a polynomial lower than 10°, at most 286 coefficients need to be trained. Typically, a 3D free form shape acquired by range sensor is consisted by 5000–500,000 points. It is very redundant to for the training.

To train k coefficients, at least k points are need. For the sake of training robustness, we give more redundancy that $10k$ points are used. In Section 6, we show the effectiveness of the down-sampling. As discussed above, the appropriate polynomial degree is not known until the training is finished. Typically, the training will terminate at 8°. Thus 3000 sampling points is sufficient for polynomial up to 10°. Some examples of performance improvement are shown in Table 2.

4.3.3. Algorithm

The overall key-point detection algorithm is summarized as follows:

1. For 3D shape Γ_0 which is consisted of M points: $\{\boldsymbol{x}\}$, ($M \gg 300$), down-sample the point-set and get the subset $\{\boldsymbol{x}\}_{sub}$.
2. Conduct 3L-adaptive-ASF of $\{\boldsymbol{x}\}_{sub}$ until the polynomial $\widetilde{f}_n(\cdot)$ satisfies the desired fitting accuracy. Typically, fit until the average square fitting error (see formula (3)) is smaller than 0.1.
3. Compute the polynomial value of all points $\{\boldsymbol{x}\}$ and get the value set $\{\widetilde{f}_n(\boldsymbol{x})\}$.
4. For each point $\boldsymbol{x}$ in $\{\boldsymbol{x}\}$, find its starting local neighborhood $\{nei(x)\}_1$ which contains about 6 neighbors. This could be done

Table 1
Number of polynomial coefficients of various degree.

Polynomial degree	0	1	2	3	4	5	6	7	8	9	10
No.of coefficients	1	4	10	20	35	56	84	120	165	220	286

Table 2
Computing time of models used in the experiments.

Model	Caesar	Face-YO	Bulldog	Buddha	Brain	Bayon face1	Bunny	Bayon face 2	Bayon face 3
Place	Fig. 2	Fig. 9	Fig. 9	Fig. 9	Fig. 9	Fig. 11	Fig. 4	Fig. 11(b)	Fig. 11(c)
Vertices	21,627	13,746	25,378	25,003	73,249	7713	8171	10,304	10,347
Time (s)	3.12	1.16	10.27	10.26	15.22	1.16	11.86	2.91	2.87
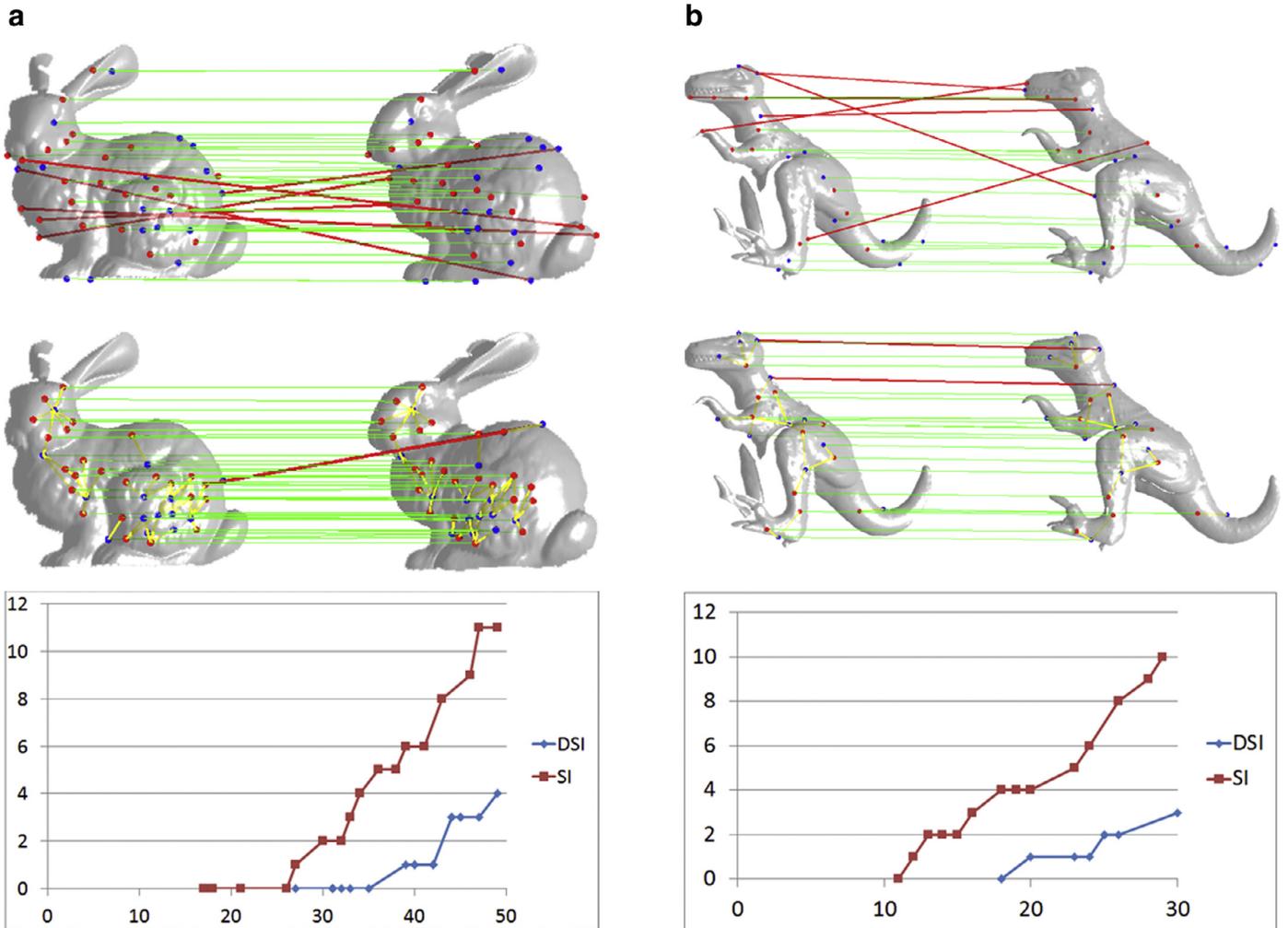
Fig. 10. Shape matching by using SI and DSI. First row: matching by SI, Second row: matching by DSI. Third row, result curve. (a)(b) Matching of range data on objects taken in 2 different directions. The x-axis the number of points detected while the y-axis is the accuracy.

by setting the radius of the neighborhood d_1 as small as the surface sampling resolution. If the surface is stored in mesh format, this could also be done by searching for first step neighbors.

5. Select extremas in each neighborhood $\{nei(x)\}_1$ according to their polynomial value $\tilde{f}_n(\pmb{x})$. Denote the selected extremas as $\{ext\}_1$ and tag the neighborhood radius d_1.

6. For extrema set $\{ext\}_i$ found in radius d_i, Set $d_{i+1} = 2d_i$. If two points are located in a d_{i+1}-neighborhood, compare their polynomial value and found new local extremas. The new local extrema set are denoted as $\{ext\}_{i+1}$.

7. Repeat Step 6 until d_i is 10–20% in comparing with the global shape size.

4.4. Robustness against deformations and noise

Our proposed key-point detector is considered to be robust against deformations. As mentioned above, the detector select points based on their deviation from the shape's smooth approximation $\tilde{\Gamma}_0$. Deformations, including affine deformation and nonlinear deformation, are considered to be smooth changes. This means these kind of changes can also be approximated by ASF so that it will not bring much change to the points deviation.

Local detectors would be sensitive to noise, *e.g.*, those points selected from local curvature or normal are not robust against noise. Fortunately, inherited from ASF, the proposed IP detector is robust against noise. ASF models the shape globally and noise is

considered to be local. This means the reference $\widetilde{\Gamma}_0$ is almost unchanged. To demonstrate the robustness, we conduct systematical repeatability experiments in Section 6.1.2.

5. Pairwise feature description

To achieve high accurate matching, it is expected that the detected key-point should be repeatable, and the feature descriptor should be invariant or robust against deformations. Such invariance and robustness can be achieved by introducing a structure to the standalone constellation key-points. In recently years, "structure of features" strategy is proposed and many research works claimed the improvement in matching accuracy [6,11].

Our detector finds local minimums and local maximums on 3D shape. We can generate structures between nearing minimums and maximums. A similar idea for 2D image is called Critical Net [12]. In Section 5.1 the 3D critical net is provided. Based on 3D critical net, the Dual Spin Image descriptor is provided in Section 5.2.

5.1. 3D critical net based key-point pairs

The critical net is one of the ideas to efficiently generate structures between local minimums and local maximums. However, the original algorithm only works for 2D images. We provide the algorithm for 3D mesh key points and the polynomial value in this subsection.

3D mesh can be considered as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is referring to the vertices which are denoted as $\{\boldsymbol{x}\}$ in this paper. $\mathcal{E}$ is referring to edges connecting the vertices, which can be easily obtained from 3D mesh. To construct path between key points, connection between two vertices is defined as:

Connection: for any $a, b \in \mathcal{V}$, there is a connection between a and b on the grid $\mathcal{G}$, denoted as $a \to b$, if there exists an ascending path from a to b.

For 3D mesh $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, all undirected edges $\mathcal{E}$ can be converted to directed edges $\mathcal{E}_f$ by comparing the polynomial values $\widetilde{f}_n(\boldsymbol{x})$ of the two points linked by the edge. Then, finding a path between two key points can be done by canonical graph theory methods. Specifically, we adopt a width-first search on directed graph in our method.

The Critical Net on 3D mesh is defined as:

Critical Net on 3D mesh: The critical net of a 3D mesh $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a directed acyclic graph: $\mathcal{G}_f = (\mathcal{V}_f, \mathcal{E}_f)$ where $\mathcal{V}_f$ is the local extremas detected by IP detector and $\mathcal{E}_f = \{(a, b) \in \mathcal{V}_f | a \to b\}$ denoted directed connection between local extremas.

In short, to generate 3D CN on 3D mesh: for each given local minimum as the start point, if any local maximum can be arrived by a path along which polynomial value $\{\widetilde{f}_n(\boldsymbol{x})\}$ increase monotonically, this maximum is connected to the minimum.

Fig. 7(d) is an example of critical net on 3D human face.

5.2. Dual spin image

By using 3D critical net, local feature can be more robust and distinctive. We provide the dual Spin Image feature (DSI):

Dual Spin Image Feature: For 3D mesh $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with 3D critical net $\mathcal{G}_f = (\mathcal{V}_f, \mathcal{E}_f)$ DSI are defined as:

$$\mathcal{D}_{ab} = (SI(V_R(a)), SI(V_R(b))), \tag{20}$$

where a, b are key points detected by IP detector $a, b \in \mathcal{V}_f$, and they are connected in the critical net: $a \to b$. $V_R(a)$ denotes the R-neighborhood of a, where R is the radius of the neighborhood which is determine by the Euclidean distance between a, b: $R = c * d(a, b)$. c is a constant. $SI(\cdot)$ denotes the original Spin Image descriptor on single patch [14]. Fig. 7 is an illustration of DSI.

According to the definition, DSI is scale invariant for it adopts the scale information from 3D critical net: the distance d from the minimum a to the maximum b. DSI is more robust against deformations because the feature is anisotropic which means the area covered by DSI will move along with the deformations. Systematical matching experiments are arranged in Section 6.2.

5.3. Algorithm for DSI

Similar as Section 4.3.3, we summarize our algorithm for obtaining DSI:

a) For a meshed surface: obtain its residual field $\widetilde{f}(\cdot)$ and its local extrema (key points) $\{ext\}$ using the algorithm described in Section 4.3.3.

b) Divide the key points $\{ext\}$ as local minimum $\{L\}$ and local maximum $\{U\}$.

c) For each of the local minimum $l \in \{L\}$, search the ascending path on the meshed surface. Say, each hop between mesh vertices must have no decrease of the polynomial residual value. Specifically, such search is done by the canonical wide-first graph search algorithm. All the valid paths end at local maximums: $N(l) = \{u \in \{U\} and (l, u) are connected\}$.

d) For each pair (l, u) connected by the ascending path, obtain its dual spin image descriptor DSI(l,u) with spin image radius $R = cd$, as illustrated in Fig. 7.

6. Experiments

We demonstrate the effectiveness of the proposed method by intensive quantitative experiments. In Section 6.1 we first demonstrate the effectiveness of the key-point detector. After that, in Section 6.2 we demonstrate the effectiveness the feature descriptor and the overall surface matching method.

6.1. Key-point detection

We first quantitatively evaluate the key-point detector in three aspects: robustness against deformation, robustness against noise and efficiency. After that, we show the generally applicability on a variety of data.

6.1.1. Robustness against deformation

We have conducted a systematical repeatability test of our detector.

Tested deformations: Generally, models deformations resulted from moving camera are modeled by using sinuous function or polynomial splines [2,34]. According to that, we deformed the models by using sinuous function or polynomial splices. To start with, some linear deformations are also tested. Here we list the tested deformations:

Linear: scaling-and-rotation, affine.

Non-linear: trigonometric function, polynomial spline.

Testing strategy:

1. For a model selected from the dataset, called the original model, the corresponding deformed model is generated by a deformation function which is known a priori.
2. Detect key points on original model: $\{KP1\}$.
3. Detect key points on deformed model: $\{KP2\}$.
4. Remove the deformation of the deformed model by the deformation function known a priori, so that $\{KP2\}$ is converted to $\{KP2'\}$
5. Compare $\{KP1\}$ and $\{KP2'\}$: For each point in $\{KP1\}$, if there is a corresponding point in $\{KP2'\}$, this point is repeatable. Corresponding distance is determined by a threshold: about 1.5 sampling resolution.

Model and key-points	Matching
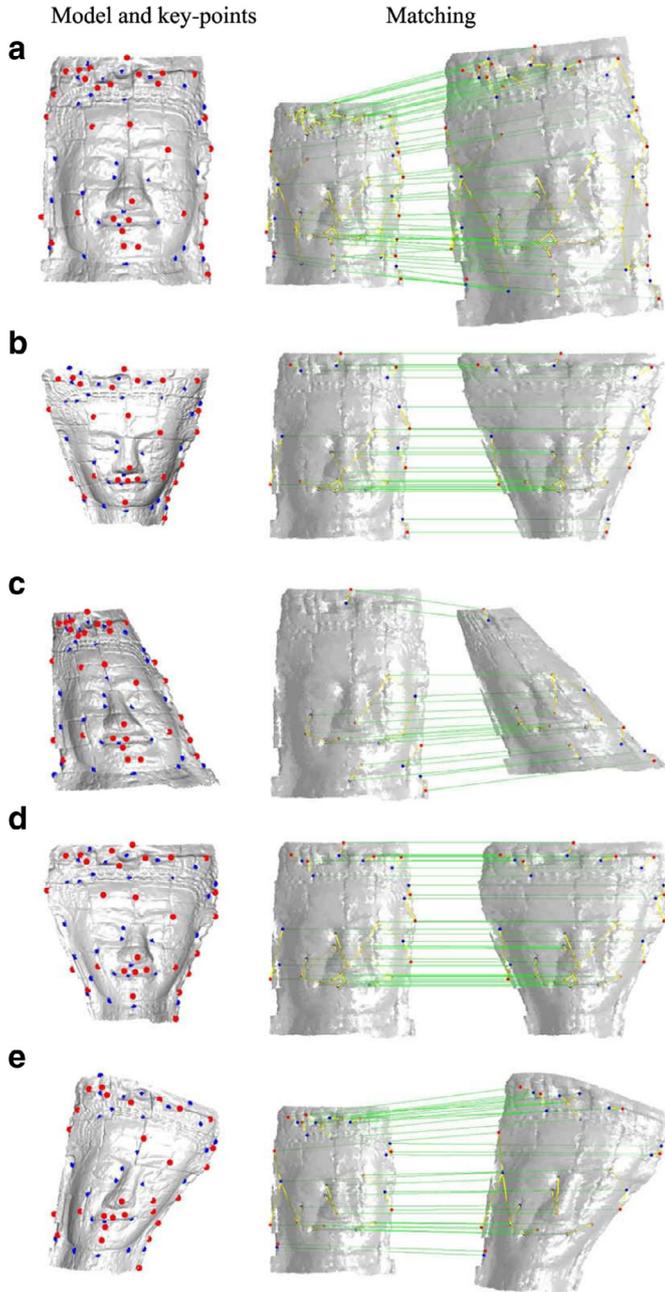
Fig. 11. Feature detecting and matching of deformed surfaces. (a) The original surface data from Bayon Digital Archival Project [10]. The matching is conducted between the original surface and surface with scaling and rotation. (b)(c) Surfaces with affine deformations. (d)(e) Surfaces with free form deformations.

6. Calculate the repeatability according to following formula:

$$repeatability = \frac{repeatable \quad points}{\min(points \quad in \quad \{KP1\}, \{KP2\})} \quad (21)$$

Fig. 8(a)–(e) is the test and result.

According to the test, we found the repeatability could be greater than 0.5 for a wide range of deformation.

6.1.2. Robustness against noise

We conduct repeatability test on noise. The testing strategy is the same as the one used in deformation tests with a difference in how we generate the models. The models with noise is generated by following method:

1. For each model, it is first centered to the original point. Then, it is normalized such that the variance of the vertices is 1. (The average squared-distance to the zero point of all the vertices on the model is 1.)
2. With the centered and normalized model as the original model, the corresponding noised model is generated by adding Gaussian noise to each vertex along its normal. The noise is quantized by varying σ from 0 to 0.01.

The result on the whole data set is shown Fig. 8(f).
Fig. 8(i) is the test and result.

6.1.3. Computational efficiency

The computational efficiency is evaluated with the following environment: CPU: Intel Core 2 Q9400 2.67 GHz; RAM: 4 GB; OS: Win7x64; Codes: Matlab 7.9; No parallelization. The computation time of the models used in this paper are listed in Table 2.

6.1.4. General applicability

Our detector is general applicable to different type of models. Experiments of models from Aim@Shape 3D shape repository [41] are shown in Fig. 9.

6.2. Descriptor and matching of range data

We conducted 3D surface matching by using the Dual Spin Image descriptor. We first quantitatively evaluate the precision and recall of the matching, and show the significant improvement for the single Spin Image. After that, we show the effectiveness on a variety of deformation and scenarios.

6.2.1. Quantitative measurement of matching accuracy.

We evaluate the matching precision and recall using the same evaluation metric proposed in previous subsection. As illustrated in Fig. 10, the proposed descriptor finds significant more matching than the single spin image while has a significant lower error rate.

6.2.2. Applicability on various deformation and scenario.

We show the applicability on various free form deformations. As show in Fig. 11, the matching is robust on both linear and non-linear deformations.

7. Conclusion

In this paper, we have proposed an IP based detector. It provides a new aspect of algebraic fitting which is used to be considered only working globally. The IP detector is computational efficient and generally applicable to free form 3D surfaces. It is invariant against global scaling and rotations; it is also robust against free form deformation, change in topology, data missing and noise. The 3D critical net is efficiently generated and introduces global information to detected local key-points. The Dual Spin Image descriptor based on 3D net is more reliable than individual Spin Image for it is invariant against scaling and robust against free form deformation. The future work includes: 1) A more versatile single feature descriptor. Single spin image is restricted to circle neighborhood. In handling free form deformation, freedom in selecting feature neighborhood is expected. 2) The further discovering of 3D critical net. Affine deformation can be accurately calculated from 3 points. A tri-feature descriptor is expected to be invariant against affine deformation. 3) New matching strategy, based on critical net, tree matching is also applicable, which holds potential in shape retrieval.

References

[1] C.B. Akgul, B. Sankur, Y. Yemez, F. Schmitt, 3d model retrieval using probability density-based shape descriptors, IEEE Trans. Pattern Anal. Mach. Intell. 31 (6) (2009) 1117–1133.

[2] A. Banno, T. Masuda, T. Oishi, K. Ikeuchim, Flying laser range sensor for large-scale site-modeling and its applications in bayon digital archival project, Int. J. Comput. Vis. 78 (2–3) (2008) 207–222.

[3] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 24 (4) (2002) 509–522.

[4] M.M. Blane, Z. Lei, H. Civi, D.B. Cooper, The 3l algorithm for fitting implicit polynomial curves and surfaces to data, IEEE Trans. Pattern Anal. Mach. Intell. 22 (3) (2000) 298–313.

[5] S.J. Chua, R. Jarvis, Point signatures: a new representation for 3d object recognition, Int. J. Comput. Vis. 25 (1) (1997) 63–65.

[6] G. Csurka, C.R. Dance, L. Fan, J. Willamowski, C. Bray, Visual categorization with bags of keypoints, in: ECCV International Workshop on Statistical Learning in Computer Vision, 2004.

[7] F. Chen, R. Ji, L. Cao, Multimodal learning for view-based 3d object classification, Neurocomputing 195 (2016) 23–29.

[8] A. Elad, R. Kimmel, On bending invariant signatures for surfaces, IEEE Trans. Pattern Anal. Mach. Intell. 25 (10) (2003) 1285–1295.

[9] H. Gao, Q. Huang, X. Xu, X. Li, Wireless and sensorless 3d ultrasound imaging, Neurocomputing 195 (2016) 159–171.

[10] B. Falcidieno, AIM@SHAPE project presentation, in: Shape Modeling Applications, 2004.

[11] L. Fei-Fei, P. Perona, A bayesian hierarchical model for learning natural scene categories, in: IEEE Conference on Computer Vision and Pattern Recognition, 2005.

[12] S. Gu, Y. Zheng, C. Tomasi, Critical nets and beta-stable features for image matching, in: European Conference on computer vision, 2010.

[13] K. Ikeuchi, K. Hasegawa, A. Nakazawa, J. Takamatsu, T. Oishi, T. Masuda, Bayon digital archival project, in: Virtual Systems and Multimedia, 2004.

[14] A. Johnson, Spin-images: a representation for 3-d surface matching. doctoral dissertation, 1997, The Robotics Institute, Carnegie Mellon Univ.

[15] Y. Fu, W. Dong, 3d magnetic resonance image denoising using low-rank tensor approximation, Neurocomputing 195 (2016) 30–39.

[16] M. Liu, D. Zhang, S. Chen, H. Xue, Joint binary classifier learning for ECOC-based multi-class classification, in: IEEE Transactions on Pattern Analysis and Machine Intelli-gence, 2016. (In Press).

[17] M. Liu, D. Zhang, D. Shen, Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment, IEEE Trans. Med. Imaging 35 (6) (2016) 1463–1474.

[18] S. Kang, K. Ikeuchi, Determining 3-d object pose using the complex extended gaussian-image, in: IEEE Conference on Computer Vision and Pattern Recognition, 1991.

[19] O. van Kaick, H. Zhang, G. Hamarneh, D. Cohen-Or, A survey on shape correspondence, Comput. Graph. Forum 30 (6) (2011) 1681–1707.

[20] Y. Lipman, T. Funkhouser, Mäbius voting for surface correspondence, Proc. ACM SIGGRAPH 28 (3) (2009).

[21] F. Lu, Y. Matsushita, I. Sato, T. Okabe, Y. Sato, Uncalibrated photometric stereo for unknown isotropic reflectances, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 1490–1497. Portland, Oregon, USA, June 23–28

[22] F. Lu, I. Sato, Y. Sato, Uncalibrated photometric stereo based on elevation angle recovery from BRDF symmetry of isotropic materials, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 168–176. Boston, Massachusetts, USA, June 7–12,

[23] F. Lu, Y. Matsushita, I. Sato, T. Okabe, Y. Sato, From intensity profile to surface normal: photometric stereo for unknown light sources and isotropic reflectances, IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) 37 (10) (2015) 1999–2012.

[24] A. Mian, M. Bennamoun, R. Owens, On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes, Int. J. Comput. Vis. 89 (2–3) (2010) 348–361.

[25] R. Osada, T. Funkhouser, B. Chazelle, D. Dobkin, Shape distributions, ACM Trans. Graph. 21 (4) (2002) 807–832.

[26] D. Raviv, A.M. Bronstein, M.M. Bronstein, R. Kimmel, N. Sochen, Affine-invariant diffusion geometry for the analysis of deformable 3d shapes, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[27] M. Rouhani, A.S. Domingo, The richer representation the better registration, IEEE Trans. Image Process. 22 (2013) 5036–5049.

[28] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.

[29] Y. Sahillioglu, Y. Yemez, 3d shape correspondence by isometry-driven greedy optimization, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

[30] T.W. Sederberg, D.C. Anderson, Implicit representation of parametric curves and surfaces, Comput. Vis., Graph. Image Process. 28 (1) (1984) 72–84.

[31] A. Sharma, R. Horaud, J. Cech, E. Boyer, Topologically-robust 3d shape matching based on diffusion geometry and seed growing, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[32] G. Taubin, Estimation of planar curves, surfaces and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 13 (11) (1991) 1115–1138.

[33] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.

[34] S. Thrun, J.J. Leonard, Simultaneous localization and mapping, Springer handbook of robotics, Springer, 2008, pp. 871–889.

[35] T. Tung, T. Matsuyama, Dynamic surface matching by geodesic mapping for 3d animation transfer, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

[36] R. Xiao, J. Yang, J. Fan, D. Ai, G. Wang, Y. Wang, Shape context and projection geometry constrained vasculature matching for 3d reconstruction of coronary artery, Neurocomputing 195 (2016) 65–73.

[37] S. You, R.T. Tan, R. Kawakami, Y. Mukaigawa, K. Ikeuchi, Waterdrop stereo, 2016a, Arxiv preprint arXiv:1604.00730.

[38] S. You, Y. Matsushita, S. Sinha, Y. Bou, K. Ikeuchi, Multiview rectification of folded documents, 2016b, ArXiv preprint 1606.00166.

[39] W. Zeng, X. Gu, Registration for 3d surfaces with large deformations using quasi-conformal curvature flow, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[40] Y. Zhang, F. Jiang, S. Rho, S. Liu, D. Zhao, R. Ji, 3D object retrieval with multi-feature collaboration and bipartite graph matching, Neurocomputing 195 (2016) 40–49.

[41] The Stanford 3D Scanning Repository. http://graphics.stanford.edu/data/3Dscanrep/.

Shaodi You is a researcher at National ICT Australia and an adjunct lecturer at Australian National University. He serves as the Chair of Computer Society of IEEE ACT Section from 2016. He receives his Ph.D. and M.E. degrees from The University of Tokyo, Japan in 2015 and 2012 and his bachelor's degree from Tsinghua University, P. R. China in 2009. His research interests are physics based vision, image/video enhancement and machine learning.

Diming Zhang receives his B.E. degree in science from the Soochow University, Jiangsu, China, in 2009, and the Master degree in software engineering from the Southeast University, in 2011. In 2011, he joined the Department of Computer Engineering, Jiangsu University of Science and Technology, as a Lecturer. His current research interests are computer vision, operation system and parallel computing.