Title: Cylindrical shape descriptors for unstructured point cloud registration

Authors:

Abstract

This paper presents a method for registration of unstructured point clouds. We firstly derive intrinsic shape context descriptors for 3D data organization. To replace the Fast-Marching method, a vertex-oriented triangle propagation method is applied to calculate the 'angle' and 'radius' in descriptor charting, so that the matching accuracy at the twisting and folding area is significantly improved. Then, a 3D cylindrical shape descriptor is proposed for registration of unstructured point clouds. The chosen points are projected into the cylindrical coordinate system to construct the descriptors. The projection parameters are respectively determined by the distances from the chosen points to the reference normal vector, and the distances from the chosen points to the reference tangent plane and the projection angle. Furthermore, Fourier transform is adopted to deal with orientation ambiguity in descriptor matching. Practical experiments demonstrate a satisfactory result in point cloud registration and notable improvement on standard benchmarks.

1. Introduction

Point cloud registration is a fundamental task in computer graphics and machine vision. Two kinds of three-dimensional (3D) point cloud data can be obtained directly by sensing devices. On one hand, there are structured-light based sensors such as the Microsoft Kinect which generate a structured point-cloud, sampled on a regular grid [1], and on the other hand, there are many time-of-flight (TOF) sensors such as the MESA 4500 TOF camera which yields an unstructured point-cloud [2]. These point clouds can be used in 3D reconstruction [3], object detection and recognition [4]. In these application cases, point clouds from different objects or views need to be aligned by a process which is often referred to registration. Registration algorithms are able to calculate the transformation that optimally maps two point-clouds each other.

In the wide range of registration algorithms, feature-based methods have gained popularity and good performance in many settings, including object recognition and 3D reconstruction. The choice for one of these algorithms generally depends on several important characteristics such as accuracy, computational complexity, and convergence rate, each of which depends on the application of interest. Moreover, the characteristics of most registration algorithms heavily depend on the data used, and most feature-based approaches are considered for structured point-cloud such as 3D meshes and curved surfaces. Classical works include the integral volume descriptors [5, 6], multiscale local features [7] and spin images [4], just to mention a few of many. Spin images as a general shape representation can be used for object recognition in complicated real scenes. However, the two distances used for generating spin images lack of angular distribution information, such shape representation is difficult to meet high accuracy registration.

Over the past years, features designed for rigid registration like spin images have been extended to non-rigid deformations. Some of the classical rigid descriptors are extended to the non-rigid case by replacing the Euclidean metric with its geodesic counterpart [8, 9]. Moreover, there are isometry invariant features. The heat kernel signature (HKS) [10] which is based on heat diffusion at a point. Based on the fundamental solutions of the heat equation,

a scale-invariant version of HKS was developed [11]. In recent years, variants of HKS have been proposed. The wave heat kernel signature (WHKS) is proposed as a solution to the excessive sensitivity of the HKS to low-frequency information in [12] which uses a quantum mechanical approach to capture multiscale details. By using volumetric distance, HKS was extended to volumetric data [13]. An another emerging method which use HKS as field function to generate an intrinsic shape context (ISC) descriptor approach to curved surfaces [14].

Early works about 3D shape context descriptors were explored in [15, 16], but the descriptors were affected by surface deformation because the descriptors were not intrinsic. Another work which explored the exploitation of intrinsic geometry was [17], but there the authors trivially deal with radial variation, by averaging over orientations. The descriptors proposed in [14] can retain and exploit the information contained in the radial variation around a point, however because of the huge curvature change at some particular positions on the surface, the information contained in the radial variation around a point will be ambiguous.

Feature based descriptors have so far achieved a more modest success in the analysis of surfaces, especially for surfaces represented by unstructured point clouds. Surfaces lack a global system of coordinates and can be associated with only a local vector structure. Such a system of coordinates holds only locally. Trying to use it globally like done in spin images makes the descriptor sensitive to shape deformations. Based on the experimental analysis of Spin-image method and ISC descriptor, we integrate Spin-image with shape context into cylindrical coordinate system to generate a 3D cylindrical shape descriptor.

The contribution of this paper has three aspects. First, a VTP-ISC descriptor is proposed to represent the shape of mesh more accurate especially at the twisting and folding positions on the mesh, the matching accuracy is obviously improved. Second, we propose 3D cylindrical shape descriptors for unstructured point cloud registration. These 3D descriptors are not only based on the distance distribution but also the angle distribution of the points which are allowed to contribute to the descriptors, compare to 2D descriptors, richer surface information is included. Finally, we experimentally demonstrate that the proposed VTP-ISC descriptor and 3D cylindrical shape descriptor show notable performance in point cloud matching.

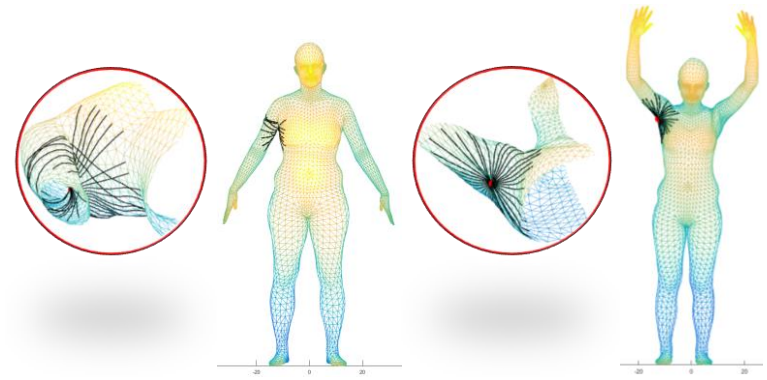
2. VTP-ISC descriptor

ISC descriptor is an extension of original shape context to surfaces [14]. The authors chart the surface by shooting geodesic outwards from the point being analysed, where ‘angle’ is treated as tantamount to geodesic shooting direction, and ‘radius’ as geodesic distance.

In the process of outward ray shooting, the propagation of the directions outwards from the 1-ring is done using the standard unfolding procedure. However, the propagation of direction will intersect at the twisting and folding positions on the mesh. On the premise that all the parameters of the descriptor are determined, the propagation direction of the vertex at right armpit is intersected obviously as shown in Fig 1a. It makes the vertices which contribute to the descriptors are ambiguously assigned to the angular bins.

Commented [HY1]: Geodesic lines from the same source should not intersect each other.

Check it with reference again!



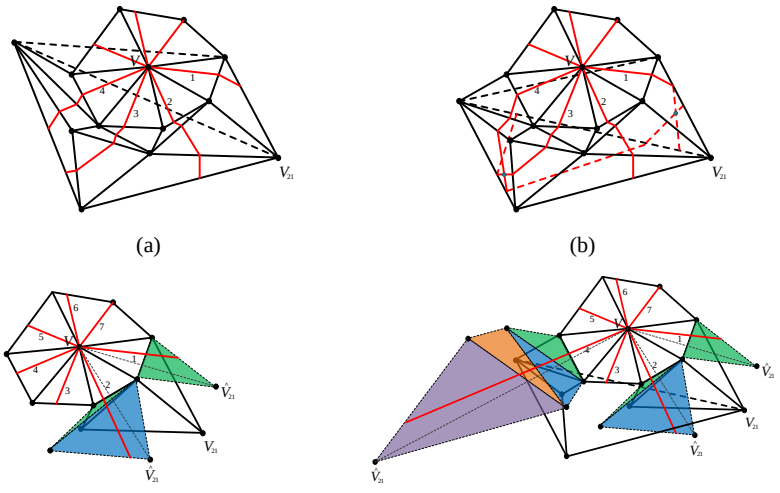
(a) (b)

Figure 1 propagation direction intersects at the twisting and folding areas on the mesh

Ambiguous assignments cause errors in constructing descriptors, then, defective descriptors lead mesh mismatch. As is demonstrated in Fig 2, according to the outward ray shooting method [14], V is a vertex on a mesh. The red lines are the central propagation directions of the angular bins. Numbers mark different angular bins. The vertices will be assigned to angular bins based on the angle value between the central line of the angular bins and the path to V on the propagation triangles. For example, V_{21} will be assigned to bin 1 or bin 2 (Fig. 2c). However, with the deformation of the mesh or the increasement of the radius of the descriptors, the propagation directions intersect. V_{21} will be assigned to bin 1 or bin 2 or bin 4 (Fig. 2d). The unreasonable ambiguous assignment will cause vertices mismatching.

Commented [HY2]: ISC description is not clear here.

Use formula and figures to illustrate ISC instead of text description.



(a) (b)

(c) (d)

Figure 2 Propagation of direction outwards from 1-ring using the unfolding procedure. The triangles unfolded along the propagation directions (thick red lines, bottom) are in colour (bottom). The vertices assignment is determined by the angle between the propagation direction and the path of the

vertices on the propagation triangles (dashed black lines, bottom). (c) V_{21} will be assigned to bin 1 or bin 2; (d) V_{21} will be assigned to bin 1 or bin 2 or bin 4.

There are still two drawbacks exposed in this surface charting approach. First, the geodesic distance calculated by fast marching method (FMM) will be incorrect, when there are skinny triangles on the calculation path. Second, the propagation of direction will intersect at some twisting and folding areas on the mesh. These two disadvantages affect the two parameters that generate the shape context descriptors respectively.

Commented [HY3]: See the above question!

Based on the above two defects, we derive the surface charting method in descriptor generation process. For exact geodesic distance calculation, we use vertex-oriented triangle propagation (VTP) [18] to calculate the exact shortest paths from the chosen vertices to the central vertex being analysed. The shortest geodesic distance and angle of the chosen vertices can be uniquely determined. Hence, the shape context descriptors can be charted. About the other procedure, we follow the description in Bronstein's work [14].

Commented [HY4]: It is not clear why you use VTP here, only for exact geodesics?
In terms of ISC definition, you need to justify/prove why approximate geodesics do not work well.

You need to evidence the importance of using VTP here!

3. 3D cylindrical descriptor

3.1 Cylindrical shape descriptor

In our representation, each point in the 3D point cloud is associated with a 3D cylindrical descriptor which encodes global properties of the surface by using the object center coordinate. The correspondences between two point-clouds can be established by matching the cylindrical descriptors. Figure 3 shows the cylinder descriptor and 3D histograms of descriptors statistics for three points on a human model.

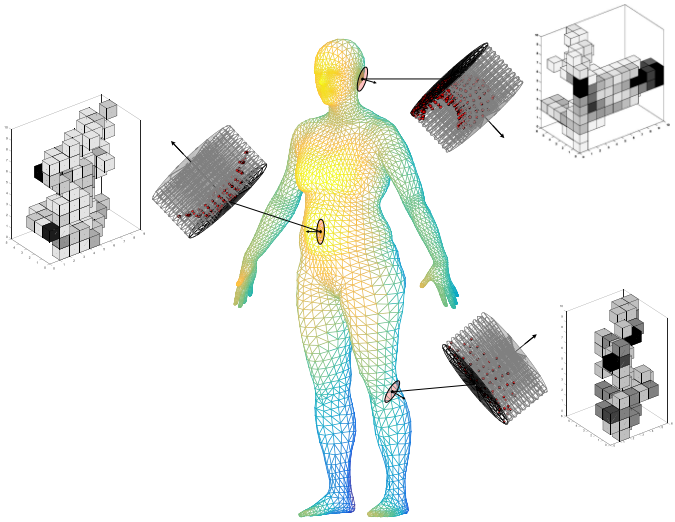


Figure 3. 3D cylinder descriptors of three points on a mesh of the human model.

A 3D cylinder descriptor is created for a point in the point-cloud as follows. A 3D accumulator indexed by α, β and θ is created. The coordinates (α, β, θ) are computed for a point in the point-cloud that is within the support of the descriptor. The box indexed by (α, β, θ) in the accumulator is then incremented. The resulting accumulator can be thought of as a

descriptor. Dark areas in the descriptor correspond to boxes that contain many projected points. 3D cylinder descriptors are constructed for every point for point cloud registration. To compare the 3D descriptors, the linear correlation coefficient can be used. The descriptor similarity can be measured by using the correlation coefficient between two 3D descriptors.

Commented [HY5]: What is correlation coefficient?

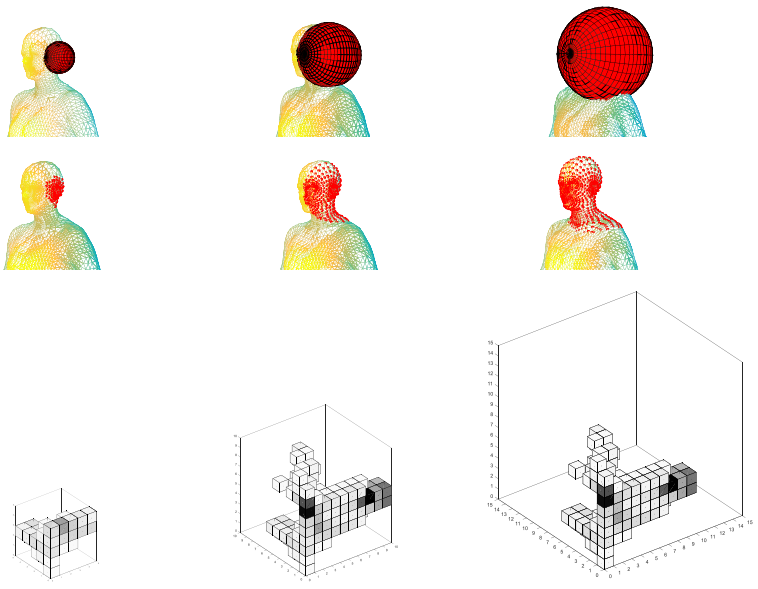
Your descriptor indeed takes into account FFT in section 3.3. Thus, you should give a complete description for your descriptor in case of any potential guess. And then introduce 3 parameters respectively…

Commented [HY6]: Which 3 parameters?

Use bullet point to describe each parameter……

3.2 Descriptor generation parameters

There are three parameters need to be determined based on the model. Support area size **R** is the radius of sphere area which is centered on the analysed point. Points within the **R** range around the analysed point are allowed to create descriptor. By setting the support area size, the amount of global information in a descriptor can be controlled. For a fixed descriptor resolution, decreasing support area size will decrease the descriptiveness of a cylinder descriptor because the amount of global shape included in the descriptor will be reduced. However, decreasing cylinder descriptor will also reduce the chances of clutter corrupting a descriptor. Figure 4 shows cylinder descriptor for a single point on the human model as the support area size is increased. This figure shows that as support area size decreases, the descriptiveness of the descriptors decreases.



(a) (b) (c)

Figure 4 The effect of support area size on 3D cylinder descriptor. As support area size increases, the descriptiveness of the descriptors decreases. By varying the support area size, cylinder descriptors can vary smoothly from global to local representations.

Descriptors for all points in the model are created for a particular support area size. Each descriptor is compared to all of the other descriptors from model, and the Euclidean distances between the point and the points corresponding to the best matching descriptors are computed. After repeating this matching for all descriptors on the model, the median Euclidean distance (match distance) is computed. By repeating this procedure for multiple support area size

using the human model, Figure 5 is created. Match distance is a single statistic that describes the correctness of constructing the cylinder descriptor, the lower the match distance, the more correct the matches.

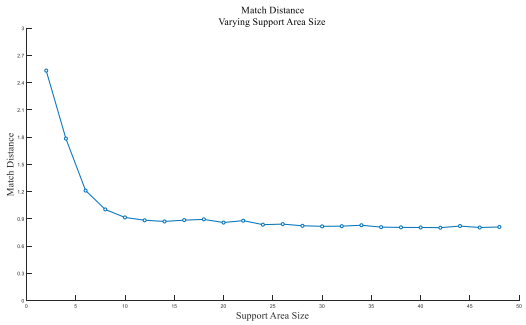


Figure 5 Effect of cylinder descriptor generation parameters support area size on match distance.

According to the geometry, the descriptor in this paper is divided into small boxes according to number on the cylindrical coordinate axis. n_α (the bin number of α), n_β (the bin number of β) and n_angle (the bin number of angle). Although the descriptor can have any number of α, β and angle, for simplicity, we generally make the number equals to τ. This results in the descriptor whose box size can be described by one parameter. We define τ which is the number of α, β and angle to be the *descriptor resolution*. Descriptors generated for the human model using different descriptor resolutions are shown in Figure 6. In this example, the descriptor, with resolution $\tau = 5$, generated for a descriptor is not very descriptive of the global shape of the model. The descriptor with resolution $\tau = 15$ does not have enough averaging to eliminate the effect of surface sampling. The descriptor with resolution $\tau = 10$ has the proper balance between encoding global shape and averaging of point positions.

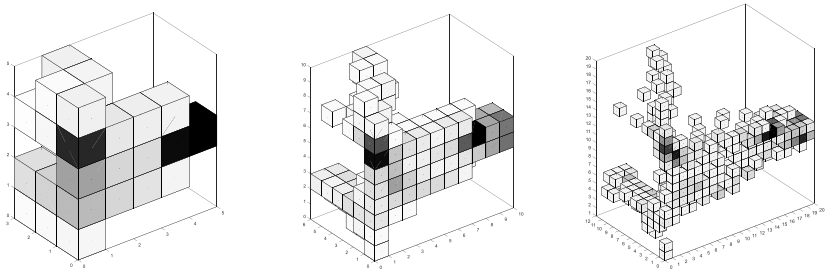


Figure 6 The effect of descriptor resolution on cylinder descriptor appearance. Three cylindrical descriptors of increasing descriptor resolution for a point on the human model are shown. Setting the descriptor resolution to 10 creates descriptive cylinder descriptors while averaging during point accumulation to eliminate the effect of individual vertex positions.

In cylinder descriptors generation for point clouds, an appropriate descriptor resolution needs to be determined. Figure 7 shows the effect of descriptor resolution on match distance.

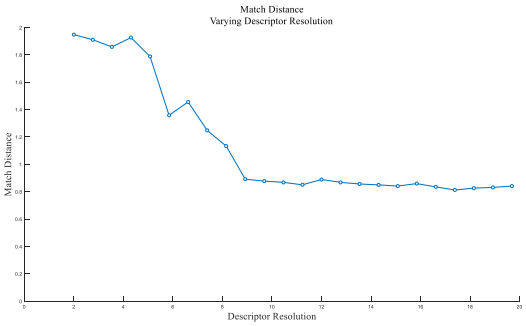


Figure 7 Effect of cylinder descriptor generation parameters descriptor resolution on match distance.

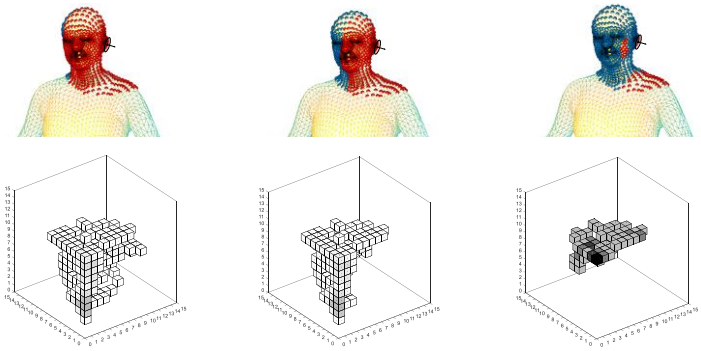
Unlike spin-image method, cylinder descriptors are used for unstructured point cloud. Each point in the cloud only contains the 3D position and an nondirectional normal vector. Support angle defined in the spin-image method is useless for point cloud without mesh structure. To limit the effect of self-occlusion and clutter during the descriptor matching. We define a division plane to segment the points in the support area (colour points in Figure 8). This division plane passes through the barycentre of the points in support area and normal to the normal vector of the point being analysed. The points in the support area are segmented into two parts (points in red which are allowed to contribute the descriptor and points in blue which are abandoned). In addition, we define the direction from the point being analysed to the barycentre as $\overrightarrow{n_b}$ and randomly pick a normal direction of the point being analysed $\overrightarrow{n_p}$. If

$$\cos\left(\overrightarrow{n_b} \cdot \overrightarrow{n_p}\right) \geq 0,$$

$\overrightarrow{n_p}$ is defined as positive direction, otherwise, $\overrightarrow{n_p}$ is defined as negative direction. Then, the distance from the division plane to the barycentre along the positive normal direction of the point being analysed is μ. As is shown in Figure 8, the descriptor generated for three different positions of the division plane.

Commented [HY7]: Illustrate your "spin-image" by graph/figure

Your description should be more intuitive through illustrating graph/figures...



$\mu = 4$ $\mu = 0$ $\mu = -4$

Figure 8 The effect of division plane on cylinder descriptor appearance. As the division plane moves in the negative direction, the number of points contributing to the descriptor decreases.

The division plane is used to reduce the number of the points on the opposite side of the model that contribute to the cylinder descriptor. This plane decreases the effect of occlusion on descriptor matching and also has the effect of decreasing the descriptiveness of descriptors. Figure 9 shows the effect of division plane on cylinder descriptor match distance. The graph shows that as the division plane moves in the negative direction, the match distance increases because the cylinder descriptors are becoming less descriptive. However, a small μ is necessary for robustness to clutter and occlusion.

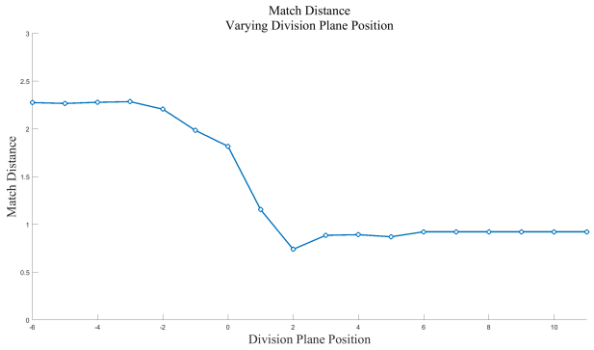


Figure 9 Effect of cylinder descriptor generation parameters division plane position on match distance.

3.3 Cylindrical descriptor matching

Since our descriptor is intrinsic, it cannot rely on a global system of coordinates and constructs a local cylindrical coordinate system around each point. This introduces rotation ambiguity: the zero-degree coordinate axis of cylindrical coordinate system picks randomly when generating the descriptor, resulting in our cylindrical descriptor to be defined up to some unknown phase $S_{\alpha,\beta,\theta+c \bmod 2\pi}$. Hence, we opt for using the *Fourier transform modulus* (FTM) technique to achieve rotation invariance. We observe that transforming the descriptor in the Fourier domain with regard to the angular coordinate θ,

$$F_\theta\{S_{\alpha,\beta,\theta}\}(\omega) = \sum_\theta S_{\alpha,\beta,\theta} exp(-l\theta\omega)$$

$$F_\theta\{S_{\alpha,\beta,\theta+c}\}(\omega) = F_\theta\{S_{\alpha,\beta,\theta}\}(\omega) exp(-l\theta\omega)$$

Taking the absolute value, we have $|F_\theta\{S_{\alpha,\beta,\theta+c}\}(\omega)| = |F_\theta\{S_{\alpha,\beta,\theta}\}(\omega)|$, eliminating the effects of orientation ambiguity without relying on orientation selection. After rotation disambiguity, cylinder descriptors from similar points of two different objects will be linearly related because the number of points that fall in corresponding boxes will be similar. Hence, we use the normalized linear correlation coefficient to compare linearly related descriptors. Given two cylindrical descriptors U and V with N boxes each, the linear correlation coefficient $R(U, V)$ is

$$\mathrm{R}(U,V) = \frac{N\sum u_i v_i - \sum u_i \sum v_i}{\sqrt{(N\sum u_i^2 - (\sum u_i)^2)(N\sum v_i^2 - (\sum v_i)^2)}}$$

Commented [HY8]: This matching strategy is not sufficient.

Do you know “shape context”?
There are lots of papers to address 3D shape context.

Compare your descriptor with shape context…

Additionally, Suggest to use ICP scheme or KNN for matching!

R is between -1 (anti-correlated) and 1 (completely correlated), and it measures the normalized error using the distance between the data and the best least squares fit line to the data. When R is high, the descriptors are similar; and when R is low the descriptors are not similar. The correlation coefficient imposes an ordering on point correspondences, so good and bad correspondences can be differentiated.

4. Experiment

This paper apply the derived VTP-ISC descriptor and 3D cylindrical descriptor, and evaluate the performance on FAUST dataset [19]. The dataset contains 300 scans of 10 people shapes are represented as triangular meshes with about 5000 vertices. Figure 10 compares the VTP-ISC descriptor with original ISC descriptor [14] which demonstrated better performance on this dataset especially at sampled points. As can be read from the graphs, the VTP-ISC descriptor (purple) compares favorably in terms of cumulative error distribution and average error. For the construction of the descriptors, radius is set to R = 10; 5 linearly spaced radial bins and 16 angular bins are used. Both kinds of descriptors are created based on HKS.

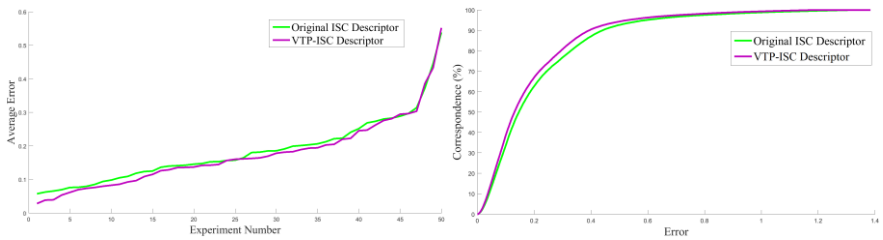
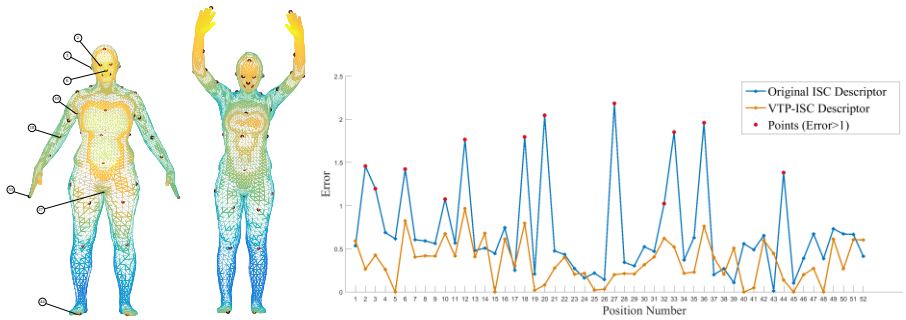


Figure 10 Cumulative and average errors achieved on the FAUST dataset [19] by the VTP-ISC descriptor and compared to the original ISC descriptor [14].

Figure 11a visualizes the sampled positions on first pair of models which are model No.10 and model No.59. These sampled points are almost at the twisting and folding positions on the mesh. The numerical symbols on the model correspond to the red points in Figure 11b which with errors more than 1centimeter.



(a) (b)

Figure 11 Matching errors contrast of the sampled points at the twisting and folding positions on No.10 and model No.59.

Figure 11b shows the matching errors contrast of 52 unsymmetrical sampled points. Compared to using original ISC descriptor (blue line), using the VTP-ISC descriptor (yellow line) leads smaller matching error especially at the severely distorted positions (red points). As is shown in Table 1, the VTP-ISC descriptor can effectively reduce the matching error at the twisting and folding positions on the mesh.

Table 1 Model No.10-Model No.59 Match Experiment

	Average error (5000 points)	Average error (52 points)
Original Descriptor	0.1808	0.7050
VTP-ISC Descriptor	0.1632	0.3532

Figure 12 compares our 3D cylinder descriptor with the methods of ISC descriptor [14] and Spin-image [4] which demonstrated state of the art results on FAUST dataset. For ISC descriptor and Spin-image, we use the models with triangular mesh and vertices to perform matching experiments. However, for the proposed descriptor, we only use the vertices. Following Bronstein's method [14], in the construction of ISC descriptors, radius is set to R = 10; 5 linearly spaced radial bins and 16 angular bins are used, the descriptors are also created based on HKS. In the construction of Spin-image, bin size is set to 1; image width is set to 10 and support angle is set to 60 degrees. According to Section 3.2, here descriptor parameters are set to these values: support area = 10; Descriptor resolution = 10; The division plane at the position +2. As can be found from the graphs, the result by this paper (blue) has better performance in terms of cumulative error distribution and average error.

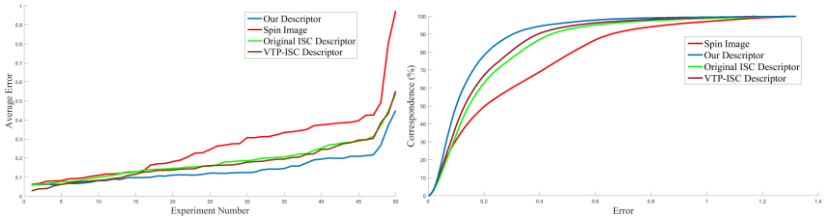


Figure 10 Results of the proposed descriptor, spin image, ISC descriptor and VTP-ISC descriptor on the FAUST shape matching benchmark.

Conclusion

Summary. We have developed a VTP-ISC descriptor which significantly reduces the matching error at the twisting and folding positions on the mesh. Moreover, the proposed 3D cylindrical descriptor demonstrates state of the art results for unstructured point-cloud registration.

Limitations. In the process of descriptor construction, the parameters calculated from one model are not optimal for all models. Therefore, using one set of parameters to create descriptors for all models cannot always achieve optimal matching results. Besides, the division plane is a rough segmentation for choosing points which are allowed to contribute to the descriptors.

Future work. We intend to pursue is applying the proposed descriptor for 3D vision tracking especially for unstructured point cloud which obtained by TOF cameras directly. Another work which we intend to pursue is adopting the descriptor as a feature of deep learning, e.g. for training the deep neural network [20] that combines with other features such as RGB and depth to perform 3D point cloud tracking.

[1] Z. Zhang, "Microsoft Kinect Sensor and Its Effect," *IEEE Multimedia,* vol. 19, pp. 4-10, 2012.

[2] I. T. Ringbeck, "A 3D TIME OF FLIGHT CAMERA FOR OBJECT DETECTION," 2007.

[3] W. Hu and Y. Qu, "3D reconstruction from multi-view point cloud based on particle system," in *Iita International Conference on Control, Automation and Systems Engineering*, 2009, pp. 500-503.

[4] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Transactions on Pattern Analysis & Machine Intelligence,* vol. 21, pp. 433-449, 2002.

[5] S. Manay, B. W. Hong, A. J. Yezzi, and S. Soatto, "Integral Invariant Signatures," in *Computer Vision - ECCV 2004, European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings*, 2004, pp. 87-99.

[6] Gelfand, Natasha, Mitra, J. Niloy, Guibas, J. Leonidas*, et al.*, "Robust global registration," *Proceedings of Eurographics Symposium on Geometry Processing,* p. 197, 2005.

[7] M. Pauly, R. Keiser, and M. Gross, "Multi-scale Feature Extraction on Point-Sampled Surfaces," *Computer Graphics Forum,* vol. 22, pp. 281-289, 2003.

[8] A. Elad and R. Kimmel, "On bending invariant signatures for surfaces," *Pattern Analysis & Machine Intelligence IEEE Transactions on,* vol. 25, pp. 1285-1295, 2003.

[9] A. B. Hamza and H. Krim, "Geodesic Object Representation and Recognition," in *International Conference on Discrete Geometry for Computer Imagery*, 2003, pp. 378-387.

[10] J. Sun, M. Ovsjanikov, and L. Guibas, "A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion," *Computer Graphics Forum,* vol. 28, pp. 1383-1392, 2010.

[11] M. M. Bronstein and I. Kokkinos, "Scale-invariant heat kernel signatures for non-rigid shape recognition," in *Computer Vision and Pattern Recognition*, 2010, pp. 1704-1711.

[12] M. Aubry, U. Schlickewei, and D. Cremers, "The wave kernel signature: A quantum mechanical approach to shape analysis," in *IEEE International Conference on Computer Vision Workshops*, 2011, pp. 1626-1633.

[13] D. Raviv, M. M. Bronstein, A. M. Bronstein, and R. Kimmel, "Volumetric heat kernel signatures," *3dor Proceedings of the Acm Workshop Ond Object Retrieval,* pp. 39-44, 2010.

[14] M. M. Bronstein, "Intrinsic shape context descriptors for deformable shapes," in *Computer Vision and Pattern Recognition*, 2013, pp. 159-166.

[15] M. Kortgen, "3D shape matching with 3D shape contexts," in *Central European Seminar on Computer Graphics*, 2003.

[16] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik, "Recognizing Objects in Range Data Using Regional Point Descriptors," in *European Conference on Computer Vision*, 2004, pp. 224-237.

[17] Y. Shi, P. M. Thompson, G. I. de Zubicaray, S. E. Rose, Z. Tu, I. Dinov*, et al.*, "Direct mapping of hippocampal surfaces with intrinsic shape context," *Neuroimage,* vol. 37, pp. 792-807, 2007.

[18] Y. Qin, X. Han, H. Yu, Y. Yu, and J. Zhang, "Fast and exact discrete geodesic computation based on triangle-oriented wavefront propagation," *Acm Transactions on Graphics,* vol. 35, p. 125, 2016.

[19] F. Bogo, J. Romero, M. Loper, and M. J. Black, "FAUST: Dataset and Evaluation for 3D Mesh Registration," in *Computer Vision and Pattern Recognition*, 2014, pp. 3794-3801.

[20] R. Q. Charles, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," pp. 77-85, 2016.