3D Interest Points Detection Using Symmetric Surround-Based Surface Saliency

Yitian Zhao and Yonghuai Liu

Department of Computer Science, Aberystwyth University
yyz10@aber.ac.uk

Abstract. The detection of interest points is an important pre-processing step for the analysis of mesh surfaces. This paper proposes a new method for the detection of the interest points for 3D surface. Our method incorporates the Relative Distance-based Laplacian and *Retinex* to generate the symmetric surround-based surface saliency. The effectiveness of this method is demonstrated by studying the repeatability of the detected interest points under different perceptual conditions, such as different viewpoints and noise corruption. The results show that the proposed method achieves better performance than competitors.

Keywords: salient, surface, bilateral, region, repeatability.

1 Introduction

Many applications have benefited from the wide diffusion of 3D models. In recent years, there has been a growing interest in the use of detailed models for better representation. Consequently, most of the latest 3D scanners can generate huge quantities of data points within a limited time. This creates a number of challenges for storage, editing and transmission. These challenges place a greater burden on feature detection tasks. 3D interest points detection is helpful in capturing the property of a point or region on a surface. Interest points, also referred to as feature points, salient points, or keypoints, are those points which are distinctive in their locality and stable at all instances of an object, or of its category of objects [1]. Therefore, interest points detection has a wide range of applications in the fields of computer vision and graphics, such as mesh simplification, view point selection, point cloud matching, and object recognition. Use of interest points has the advantage of providing local features that are semantically significant and also invariant to rotation, scaling, noise, deformation, and articulation.

Due to the efficiency of visual persuasion in traditional art and technical illustrations, visual saliency has now been widely used in computer vision applications. Recent years have witnessed the rapid development of methods for saliency and interest points detection on 3D surfaces. Most of the current 3D interest point detection methods have been developed over the last decade, as have defined functions summarizing the geometrical content of localities on a

A. Petrosino (Ed.): ICIAP 2013, Part I, LNCS 8156, pp. 632–641, 2013.

3D model in multiple scales, and selected local extrema of those functions as interest points. Gelfand et al. [2] proposed a method that is related to the surface curvature, and which is invariant to rotation and translation. Lee et al. [3] produced an approach that used a center-surround operator on the local curvature as the discriminative feature and generated the mesh saliency. Castellani et al. [4] applied Difference-of-Gaussian (DoG) on various scales, and vertices that are highly displaced after the filtering are marked as interest points. Sun et al.al [5] proposed a Heat Kernel Signature (HKS) feature point detection and the maxima of the kernel were chosen as keypoints. Mian et al. [6] proposed a keypoint detection algorithm along with an automatic scale selection technique for subsequent feature extraction. The keypoints are detected by calculating the ratio of the covariance matrices. The 3D-Harris [7] approach calculated the derivatives by fitting a quadratic surface to a neighborhood of the vertex, then the interest points are located by the first order derivatives along two orthogonal directions on the 3D surface. A 3D-SIFT technique was proposed which constructed a scale space by applying 3D Gaussian filters with increasingly large scales to the voxelized model [8]. The extreme points are detected by searching the DoG space in both spatial and scale dimensions.

Some techniques were inspired by corresponding 2D approaches. However, such extensions are not always straightforward: for example, the keypoints represent interesting information at fine scales and thus, may be sensitive to noise and other transformations. Therefore, it is necessary to find larger and interesting structures to overcome the problems at fine scales. In this paper, we propose an algorithm to select interest points from salient regions. As we mentioned above, the sensitivity to noise affects the accuracy of the interest points detection. In order to overcome this problem, we employ the following three measures:

1. A **Relative Distance-Based Laplacian (RDL)** is used to characterize the local details of the geometry, since it uses relative distance rather than the absolute Euclidean distance. The local scale changes of features will not affect its value: as a result, it is able to capture both large and small scale features.
2. **Retinex** is a theory of color constancy that is usually applied for purposes of image enhancement. This method has been adapted for estimation of viewpoint invariant information in the saliency detection stage, so that the same salient regions will be independent of changes of viewpoint.

Our method proposes the RDL and *Retinex* as the relative importance values to define a new 'color' space. Based on the fused values from RDL and Retinex, the saliency is estimated for stability and continuity.

2 Relative Distance-Based Laplacian

In this paper we propose to use the RDL, which is learned from the data, and better reflects the relative geometry. For two sets of points with similar neighboring relationships but different densities, the absolute distances between corresponding points differ dramatically from each other, but the relative distances are in

general similar. This is an advantage of the relative distance over the Euclidean distance in reflecting the relative density. The relative distance is used in case the distribution of the data is not uniform: the distance metric mainly focuses on the representation of the neighboring relationships between points [11]. We use the RDL for mesh processing. It represents each point of a mesh as the difference between the point and its neighborhood. The structure of a shape, either in terms of topology or geometry, can be modelled by matrices. We would then expect its set of eigenvalues to provide an adequate characterization of the shape [9]. The eigenvalues serve as compact global shape descriptors. They are sorted by their magnitudes so as to establish a correspondence for computing the similarity distance between two shapes. Given a point set M containing m points $x_1, x_2, ...x_m$: we begin by defining two kinds of relative distances between a pair of points x_i, $x_j \in M$ as follows, according to [11]:

1: Relative maximum distance: $rd\_max(x_i, x_j) = \frac{\|x_i - x_j\|^2}{max_{x_k \subset M}(\|x_i - x_k\|^2)}$

2: Relative average distance: $rd\_ave(x_i, x_j) = \frac{\|x_i - x_j\|^2}{ave_{x_k \subset M}(\|x_i - x_k\|^2)}$

The $max_{x_k \subset M}(\|x_i - x_k\|^2)$ and $ave_{x_k \subset M}(\|x_i - x_k\|^2)$ are the maximum and average Euclidean distance between x_i and other points belonging to M respectively. We note that the algorithm for computing the surface Laplacian is related to the set of algorithms for computing the Laplacian of point clouds in data analysis and machine learning by using the heat equation. We construct a weighted graph with n nodes, each corresponding to a point in the observed space. If x_i is one of the K nearest neighbors of x_j, or x_j is one of the K nearest neighbors of x_i, we construct an edge between points i and j. The heat kernel is used to weight the adjacency graph. Corresponding to the definition of the relative distance, there are two ways to calculate the heat kernels W:

Definition 1. *If node j is one of the K nearest neighbors of i, or vice versa, rd_max distance-based W_{ij} is defined as:*

$$W_{ij} = \frac{1}{\sqrt{4\pi t}} exp\{-\frac{(rd\_max(x_i, x_j))^2}{t}\} \tag{1}$$

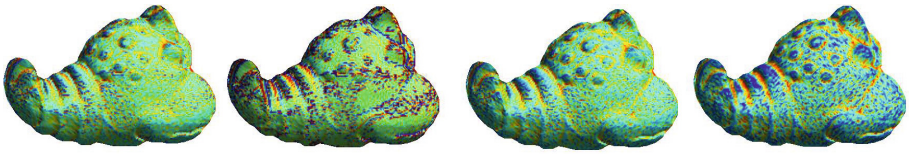
and otherwise, $W_{ij} = 0$. t is the parameter of the heat kernel.

Definition 2. *If node j is one of the K nearest neighbors of i, or vice versa, rd_ave distance-based W_{ij} is defined as:*

$$W_{ij} = \frac{1}{\sqrt{4\pi t}} exp\{-\frac{(rd\_ave(x_i, x_j))^2}{t}\} \tag{2}$$

and otherwise, $W_{ij} = 0$. Both of the relative distances are defined on the K nearest neighbors of point x_i.

Eigenmaps. We compute the eigenvectors and eigenvalues for the generalized eigenvector problem by: $Ly = \lambda Dy$, where D is a diagonal weight matrix, and its entries are column or row sums of W, since W is symmetric. $D_{ii} = \sum_j W_{ji}$. This leads to the geometric Laplacian, which we use in the remainder of the paper:



(a) (b) (c) (d)

Fig. 1. Example of two kinds of RDL on model *lobster*. (a) Maximum distance based RDL, t=1. (b) Maximum distance based RDL, t=3.(a) Average distance based RDL, t=1.(a) Average distance based RDL, t=3.

$L = D - W$. Let $y_0, y_1 ..., y_d$ be the solutions of $Ly = \lambda Dy$, ordered according to their eigenvalues, $0 = \lambda_0 \leq \lambda_1 \leq, ... \leq \lambda_d$. We leave out the eigenvector y_0 corresponding to the eigenvalue 0 and use the remaining eigenvectors for embedding in d dimensional Euclidean space,

$$x_i \rightarrow (y_1(i), y_2(i), ...y_d(i))^T \tag{3}$$

Fig. 1 shows RDL surfaces with two different kinds of relative distances. Both of them have the ability to represent the shape.

3 Symmetric Surround Saliency

In this section we propose an algorithm that aims to detect salient region and select interest points from a 3D surface. [14] treats the entire 2D image as the common surround for any given pixel, and then the saliency map is obtained by computing the Euclidean distance of the average CIELAB vector of all pixels of an input image with each pixel of a Gaussian blurred version of the same image. LAB color space is designed to approximate human vision. In this paper, we extend this scheme to generate 3D surface saliency. Initially, we employ the *Retinex* [10] for estimation of viewpoint invariant properties of the given surfaces. Normally, human perception and objective information with respect to vision are not in agreement, especially with respect to the regions that have been captured by camera or scanner but to which human eyes are relatively insensitive. The human brain interprets an image of a 3D shape differently from how photosensors or scanners may sense it by consciously correcting brightness, removing noise, shadows, glare, or reflections. After the application of the *Retinex*, the 3D surfaces are represented more faithfully by comparison with the original, simulating the human visual system. The *Retinex* algorithm for 3D surfaces has been proposed in our previous papers [12] [13] for surface details enhancement. Then Relative average distance-based Laplacian (RDL) and *Retinex* are used instead as the dimensions of LAB to estimate the saliency of the surfaces. For a given vertex p_i, where $p_i = (x_i, y_i, z_i)$, we estimate the attention value $\mathscr{E}$ in both RDL (L) and *Retinex* (R) frame:

$$\mathscr{E}(G_i, G_j) = \frac{\|G_i - G_j\|}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}} \tag{4}$$

where G is either of the geometric measures of a point: L and R, $\|\cdot\|$ calculates the Euclidean distance of vertices p_i and p_j in the mesh. The relative importance value $\mathscr{C}$ for each vertex is estimated as:

$$\mathscr{C}(p_i) = \sqrt{\sum_{i,j \in p} \mathscr{E}(R_i, R_j)^2 + \sum_{i,j \in p} \mathscr{E}(L_i, L_j)^2} \tag{5}$$

where $\mathscr{E}(\cdot,\cdot)$ indicates the visual attention values of the p_i due to p_j in terms of R and L, respectively.

We voxelize the surface that has been mapped with $\mathscr{C}$ by the bounding box, which it is divided into sub-boxes. The scale r of the voxelization is specified by the user. Let A and B be the faces that are located in two voxels. $\mathscr{C}'_A$ and ρ_A denote the mean relative importance value and mean position of the patch A. We define dissimilarity between faces A and B as:

$$diss(\mathscr{C}_A, \mathscr{C}_B) = \frac{\|\mathscr{C}'_A - \mathscr{C}'_B\|}{1 + c \cdot d(\rho_A, \rho_B)} \tag{6}$$

where $d(\rho_A, \rho_B)$ is the Euclidean distance between the mean positions of patches A and B. We set $c = 3$ in this paper as it achieved best performance. The dissimilarity measure is proportional to the difference in shape index and inversely proportional to the positional distance. We then initialize all elements of the dissimilarity map (DM) to 0, and start updating dissimilarity values as follows:

$$DM_A = DM_A + diss(\mathscr{C}_A, \mathscr{C}_B) \tag{7}$$

In practice, if the most similar faces (low dissimilarity faces) are significantly different from face A, then clearly all the other patches are also highly different from face A. Accordingly, considering the Q most similar faces is sufficient. Faces $\{q_k\}_{k=1}^{Q}$ can be identified from Eq. 6. The multi-scale saliency value of face A at scale r is defined as:

$$SM_A^r = 1 - exp\{-\frac{1}{Q}\sum_{k=1}^{Q} DM(A^r, q_k^r)\} \tag{8}$$

where $r \in R$. In this paper, the voxelisation scales $R = 1, 2, 3mm \cdots$, (if size $r = 3mm$, the size of each sub-voxel is 3*3*3). The scanner has a resolution of $0.7mm$ for the dataset we used in this paper. If $r = 1mm$, which is larger than $0.7mm$, one voxel could include more than one point. The smaller the scale r, the more accurate the saliency of the faces.

Finally, the symmetric surround saliency value at the point (v) is obtained as:

$$S(v) = \|SM'(v) - \mathscr{B}(v)\| \tag{9}$$

where $\mathscr{B}(v)$ is the corresponding vertex vector in the bilateral filtered version. Lee et al. [3] introduced the mesh saliency by using a center-surround operator on Gaussian-weighted mean curvature. The disadvantage of taking a Gaussian-weighted average is that it might make two opposite and symmetric vertices have

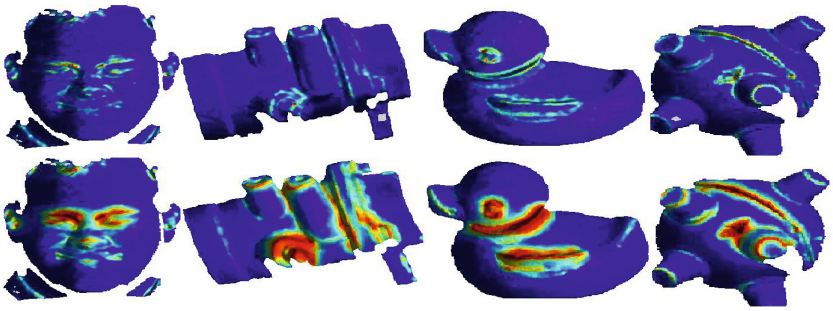


Fig. 2. Example of detected saliency on the model *par-face*, *valve*, *duck* and *frame*, respectively. Top row: Mesh saliency [3]. Bottom row: Our method.

the same saliency. To avoid this, we employ the bilateral filter operation instead of the Gaussian filter. The bilateral filtering for real function $f(v)$ on each vertex v as follows:

$$\mathscr{B}(f(v),\sigma) = \frac{\sum\limits_{u\in N(v,2\sigma)} f(u)W_c(\|u-v\|)W_s(|f(u)-f(v)|)}{\sum\limits_{u\in N(v,2\sigma)} W_c(\|u-v\|)W_s(|f(u)-f(v)|)} \tag{10}$$

where $N(v,\sigma) = \{u \| |u-v| < \sigma, u \; is \; a \; vertex\}$, denotes a neighborhood for a vertex v, the closeness spatial smoothing is the standard Gaussian filter with parameter σ_c: $W_c = exp[-u^2/(2\sigma_c^2)]$, the feature preserving weight function is $W_s(u) = exp[-u^2/(2\sigma_s^2)]$, where σ_s penalizes large variation in the scalar function $|f(u) - f(v)|$. Compared with the Gaussian operation, the output of the bilateral operation on a vertex v is also a weighted average of the surrounding vertices. However, the weight depends not only on the spatial distance, but also on the scalar function difference. For the purposes of this paper, we chose $\sigma = 8.0\varepsilon$, where ε is 0.3% of the length of the diagonal of the bounding box of the model [3]. Based on a number of experiments, we chose $\sigma_c = \sigma_s = 1.5\sigma$ in our implementation. Fig. 2 shows the results of saliency detection by mesh saliency [3] and the proposed method. Both of the methods capture the visually important features and locate the large curvature regions. However, our method detects the salient region more completely and faithfully than [3].

Interest Points Selection. The first step is the voxelization of the saliency mapped model. Voxels without saliency content are removed. The entropy of saliency values of points in a voxel is calculated based on the saliency of the histogram map as $H(X) = \sum_{i=1}^{n} - P(S_i) * log_2(P(S_i))$, where $P(S_i)$ is the probability of saliency value S_i in a local voxel and n denotes the total number of the saliency values. The entropy measures how the saliency of vertices in a voxel varies. The larger the variation, the larger the entropy, the more detail the voxel contains. Thus the entropy can be used to guide the points sampling. Normally, the maximum entropy will be chosen by default. The minimum distances between the samples are used to address the possibility that the generated samples

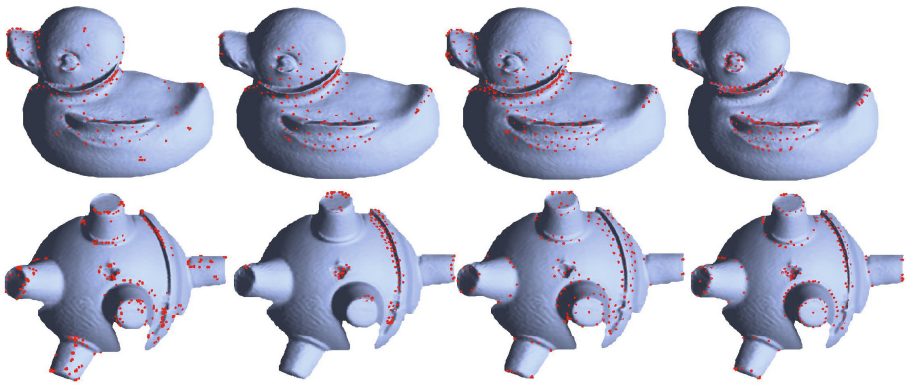


Fig. 3. Examples of the results of interest points detectors on *duck* and *frame*. (a) 3D-SIFT [8]. (b) 3D-Harris [7]. (c) MBO [6]. (d) Our method.

are close to the boundary between two or more adjacent voxels which might be too close to each other. Fig. 3(d) shows the detected interest points on the *duck* and *frame* by applying the proposed method.

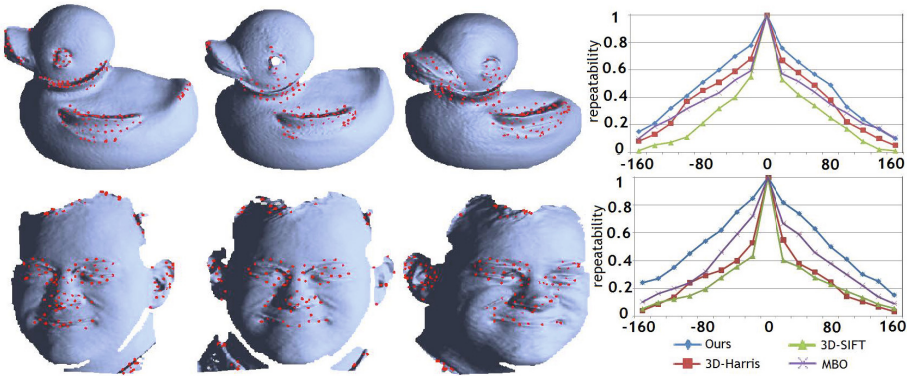
4 Experimental Evaluation and Discussion

In this section we perform a comprehensive evaluation of our interest point detector, investigating its performance under different variations of input data: viewpoints changes and noise corruption. The performance of the proposed method is also compared with existing state-of-the-art 3D surface point detectors: 3D-Harris [7], 3D-SIFT [8] , and MBO [6]. Fig. 3(a)-(c) show the outcomes of the different 3D interest point detectors on models *duck* and *pat-face*.

To evaluate the proposed and other interest point detectors, we study the repeatability of their detection of interest points. This characteristic measures the ability of the detector to find the same set of keypoints on different instances of a given model, where the differences may vary under conditions of changes viewpoint or noise corruption. Once the correspondence of the points under these different conditions has been obtained, we define the repeatability rate of interest points:

$$rep = \frac{|np_i \cap np_r|}{|np_r|} \tag{11}$$

where np_i is the number of interest points found under one of the instances and np_r is the number of interest points detected from the original given model (reference model). For a perfect detector, it detects the same interest points in the first and last frame, i.e. $rep = 1$. In this paper, we select only 1% of the total number of the vertices as interest points, as too few keypoints may not be



(a) θ=0° (b) θ=20° (c) θ=40° (d) Repeatability

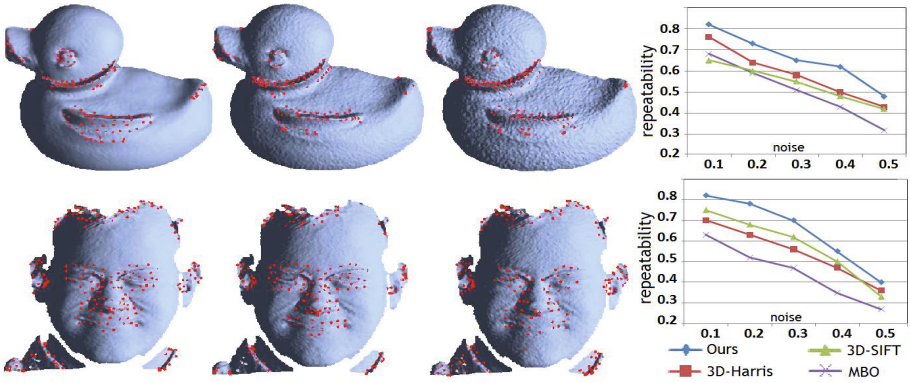
Fig. 4. (a)-(c) Model *duck* and *pat-face* subject to a rotation of θ at intervals of 20° around an unknown rotation axis are used to test the robustness of the proposed method. (d) Repeatability of interest points on the *duck* and *pat-face* with different detectors under different rotation of θ from 0° to ±160°.

enough to represent the global shape or supply further geometrical verification, while too many points leads to unnecessary waste of computational resources.

4.1 Changes of Viewpoint

This experiment evaluates the susceptibility of the proposed detector and its competitors to changing viewpoints. We set the models *buddha* and *lobster* at rotation angle 0° as the reference viewpoint and calculated the repeatability of interest points from 16 alternative viewpoints, of which 8 viewpoints were anti-clockwise rotated (20°,40°,...,160°) and 8 viewpoints were clockwise rotated (-20°,-40°,...,-160°). The effect of these rotations is shown in Fig. 4(a-c).

Fig. 4(d) shows the repeatability score of correspondences. The best results are obtained by our method for both *duck* and *pat-face* surfaces. It shows excellent tolerance with small rotations, such as ±20°. This is due to the high detection accuracy on the salient region, and selection of distinctive points from the salient region as the interest points. The repeatability score for a viewpoint change of 20 degrees on model *duck* achieves repeatability rates around 0.8 by our method, the 3D-Harris achieves a 0.68 repeatability rates and MBO gives a repeatability rate around 0.6, while 3D-SIFT only achieves 0.55. With increasing rotation angles, repeatability rates fall, since the larger the rotation angles the larger the difference between the points distributions. For instance, our detector is able to find approximately 85% of the overlapping points in the image with a rotation angle difference of 20° on model *pat-face*. However, this falls to 60% after the rotation angle increases to 60°. The second best performer is MBO, which achieves an accuracy of 50% at a rotation angle of 60°, as much as 10% lower than our method. The worst approach in this case is 3D-SIFT, which detected



(a) $\sigma = 0.1$ (b) $\sigma = 0.2$ (c) $\sigma = 0.3$ (d) Repeatability

Fig. 5. Interest points detected with different noise levels.(a) $\sigma = 0.1$. (b) $\sigma = 0.2$.(c) $\sigma = 0.3$. (d) Repeatability of interest points on the *duck* (top) and *pat-face* (bottom) with different detectors applied under different noise level.

only 30% of overlapping points; a reduction of almost 30% compared with our detector. In summary, the proposed method achieves the best performance.

4.2 Noise Corruption

To demonstrate the robustness of our method and others, we also added different levels of random Gaussian white noise with a standard deviation of σ to the original data. However, the distribution of detected interest points on the surface remains, as shown in Fig 5(a)-(c). These outcomes show the adaptive neighborhood estimation of the proposed method. After the surface was noise corrupted, the local tessellations around a vertex changed considerably. In this case, our proposal estimated good neighborhoods to mitigate the noise aspects.

The resulting interest points repeatability rates for models *duck* and *pat-face* before and after noise addition are shown in Fig. 5(d). The results were obtained by adding different levels of Gaussian noise. As Gaussian noise is generated randomly, we ran the experiments multiple times at each noise level and report the average repeatability rate. Fig. 5(d) illustrates the repeatability of interest points under different levels of noise. All detectors have nearly linear repeatability curves. However, it can be observed that MBO is clearly more sensitive to this type of transformation, since MBO is relies on the principal curvature to measure the keypoints and the curvature is known to be sensitive to noise. As we expected, the noise level affects the calculation of repeatability: as σ increases, repeatability rates decrease for all the detectors. 3D-Harris achieves a high repeatability rate in the *duck* case. 3D-SIFT presents a relatively stronger tolerance than MBO. The proposed method demonstrates the highest robustness. By contrast with the alternative methods, the proposed method has the most stable repeatability results under noise corruption.

5 Conclusions

In this paper, we propose and demonstrate a novel saliency detection method for mesh surfaces. We define a relative distance-based Laplacian, which can be used in the event that the distribution of the data is not uniform and the distance metric in the algorithm mainly focuses on the representation of neighboring relationship between points. The generated relative distance-based mesh Laplacian is fused with the *Retinex* algorithm in order to estimate a new importance value for the surface. Detection of salient regions or interest points of generic objects is one of those computer vision problems without ground-truth. In addition, we tested the repeatability of interest points of our and three other state-of-the-art detectors. The results show that the proposed method yielded the best scores in response to viewpoints changing and noise corruption. Future research will include applying the detected saliency for segmentation and automatically fine-tuning parameters in the proposed method.

References

1. Dutagaci, H., Godil, A.: Evaluation of 3D interest point detection techniques via human-generated ground truth. Vis. Comput. 28(9), 901–917 (2012)
2. Gelfand, N., Ikemoto, L.: Geometrically Stable Sampling for the ICP Algorithm. In: Proc. of 3D Digital Imaging and Modeling (2003)
3. Lee, C., Varshney, A., Jacobs, D.: Mesh saliency. ACM Transactions on Graphics 24(3), 659–666 (2005)
4. Castellani, U., Cristani, M., Fantoni, A., Murino, V.: Sparse points matching by combining 3D mesh saliency with statistical descriptors. Comput. Graph. Forum 27(2), 643–652 (2008)
5. Sun, J., Ovsjanikov, M., Guibas, L.: A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion. Comput. Graph. Forum 28(5), 1383–1392 (2009)
6. Mian, A., Bennamoun, M., Owens, R.: On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes. Int. J. Comput. Vis. 89(2-3), 348–361 (2010)
7. Sipiran, I., Bustos, B.: Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes. Vis. Comput. 27(11), 963–976 (2010)
8. Godil, A., Wagan, A.I.: Salient local 3D features for 3D shape retrieval. In: Proc. of 3DIP (2011)
9. Taubin, G.: A signal processing approach to fair surface design. In: 22nd Annual Conference on Computer Graphics and Interactive Techniques, pp. 351–358. ACM (1995)
10. Land, E., McCann, J.: Lightness and Retinex Theory. JOSA 61(1), 1–11 (1971)
11. Zhong, G., Hou, X., Cheng, L.: Relative Distance-based Laplacian Eigenmaps. In: CJKPR, Nanjing, China (2009)
12. Zhao, Y., Liu, Y., Song, R., Zhang, M.: Extended non-local means filter for surface saliency detection. In: Proc. of ICIP, pp. 633–636 (2012)
13. Zhao, Y., Liu, Y.: Patch based saliency detection method for 3D surface simplification. In: Proc. of ICPR, pp. 845–848 (2012)
14. Achanta, R., Estrada, F., Wils, P., Strunk, S.: Frequency tuned Salient Region Detection. In: Proc. of CVPR (2009)