

# Point Cloud Saliency Detection by Local and Global Feature Fusion

Xiaoying Ding<sup>ID</sup>, Weisi Lin<sup>ID</sup>, *Fellow, IEEE*, Zhenzhong Chen<sup>ID</sup>, *Senior Member, IEEE*,  
and Xinfeng Zhang<sup>ID</sup>, *Member, IEEE*

**Abstract**—Inspired by the characteristics of the human visual system, a novel method is proposed for detecting the visually salient regions on 3D point clouds. First, the local distinctness of each point is evaluated based on the difference with its local surroundings. Then, the point cloud is decomposed into small clusters, and the initial global rarity value of each cluster is calculated; a random walk ranking method is then used to introduce cluster-level global rarity refinement to each point in all the clusters. Finally, an optimization framework is proposed to integrate both the local distinctness and the global rarity values to obtain the final saliency detection result of the point cloud. We compare the proposed method with several relevant algorithms and apply it to some computer graphics applications, such as interest point detection, viewpoint selection, and mesh simplification. The experimental results demonstrate the superior performance of the proposed method.

**Index Terms**—3D point cloud, visual perception, saliency.

## I. INTRODUCTION

**D**UE to the rapid development of technologies, 3D point cloud data is increasingly convenient to obtain, which makes 3D point cloud processing attract more and more attentions [1]. In recent years, some 3D point cloud processing schemes are also proposed for human-centered visual computing applications, for instance, 3D point cloud based object detection [2], recognition [3], localization [4] and registration [5]. However, these methods mainly utilized the signal-based mathematical measures such as the curvatures [6] without considering the human visual system features on 3D models.

Visual saliency is an important feature of the human visual system, it describes the human attention distribution or the eye movements for a given scene [7]. Detecting such visually salient regions is a significant component in computer

vision and computer graphics fields [6]. However, most of the saliency detection works concentrate on 2D images [8] and videos [9]. Only a few works compute saliency for 3D meshes [10] and point clouds. Since 3D models usually consist of large data size and are more likely to be distorted by noise, traditional ways to analyze this kind of data might be computationally expensive, making saliency detection on 3D models more challenging than that on 2D data, and existing saliency detection methods designed on 2D data are also difficult to apply to 3D models straightforwardly. Recently, some algorithms for saliency detection on 3D meshes have been proposed. These methods mainly depend on utilizing topological information such as the spectral properties [11] or the geodesic distances. However, due to the lack of the topological information in the point cloud and its different sampling density, trivially extending these algorithms to unorganized point clouds might be impractical. To address these limitations, we focus on saliency detection for unorganized 3D point clouds.

Since the human visual attention is drawn to regions with variations or changes from the others [1], saliency detection for 3D point clouds can be regarded as finding for perceptually important regions that are unique with respect to their surrounding regions [6]. Finding these perceptually important regions has become a useful tool for many applications such as viewpoint selection [12], point clouds registration [13], simplification [14], segmentation [15] and resizing [16]. For example, when we are doing simplification for the 3D point cloud model, we hope that more details can be preserved in the salient regions, by applying this human perception inspired saliency measure to guide mesh simplification application, more details can be preserved, thus providing better visualization results. This human perception inspired saliency measure can provide more visually appealing results in displaying and processing of 3D point clouds [6] when compared with using purely mathematical measures such as the curvatures [17]. To take the human visual system [18] into consideration, the local distinctness and global rarity cues are considered in the proposed method. The local distinctness accounts for the point's difference from its local surroundings, while the global rarity aims to detect entire unique regions and suppress frequently appeared regions. Since the local distinct region may be of low global rarity if it appears many times around the model, many studies in neuroscience and psychophysics have suggested that both local features and global features should be taken into consideration simultaneously. Both these two characteristics are supported by psychological evidences [19]

Manuscript received May 28, 2018; revised November 21, 2018 and January 25, 2019; accepted May 4, 2019. Date of publication May 30, 2019; date of current version August 22, 2019. This work was supported in part by the Singapore Ministry of Education Tier-2 Fund under Grant MOE2016-T2-2-057(S), in part by the National Natural Science Foundation of China under Grant 61771348, and in part by the China Scholarship Council (CSC). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Wen Gao. (*Corresponding author: Zhenzhong Chen.*)

X. Ding and Z. Chen are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: dingxiaoying@whu.edu.cn; zzchen@ieee.org).

W. Lin is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: wslin@ntu.edu.sg).

X. Zhang is with the School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: zhangxinf07@gmail.com).

Digital Object Identifier 10.1109/TIP.2019.2918735

1057-7149 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.  
See [http://www.ieee.org/publications\\_standards/publications/rights/index.html](http://www.ieee.org/publications_standards/publications/rights/index.html) for more information.

and have been applied to saliency detection in 2D images successfully, such as the saliency detection algorithm presented in [20]. Besides the algorithm presented in [20], many proposed algorithms have also combined these two cues together to obtain satisfying image saliency detection results. For example, Wang *et al.* [21] provided an algorithm utilizing multi-scale local contrast and global spatial distribution to detect saliency. And in [22], Borji and Itti provided a novel saliency detection algorithm which measures patch rarities in different color spaces and combines them into a final saliency detection map. According to these algorithms, the combination of local and global cues can promote the saliency detection accuracy compared with using single cues and is more consistent with the characteristics of the human visual system.

In this paper, a novel point cloud saliency detection algorithm is presented which utilizes both the local distinctness and global rarity cues. To measure the local distinctness, we define it as the distance between a point and its  $k$ -nearest neighboring points in the descriptor space, and utilize the Fast Point Feature Histogram (FPFH) descriptor [23] to extract local geometric features of each point. To measure the global rarity, the point cloud is decomposed into several clusters using supervoxel segmentation method proposed in [24] and the initial global rarity is obtained by calculating the distance between a cluster and the remaining clusters in the descriptor space. Then a random walk ranking method is used to introduce cluster-level global rarity refinement to each point in all clusters by taking the local geometric features of each point into consideration. Since local geometric features represent the relationships between neighboring points, using it to refine initial global rarity will provide more delicate global rarity result. Finally, an effective optimization framework is proposed to integrate both the local distinctness and the global rarity values for each point and the final saliency detection result is obtained. An example of the saliency detection result for the Max Planck model is shown in Fig. 1. According to Fig. 1, our detection result is more concentrated on the facial features such as the eyes when compared with the result reported in [1] and provides better saliency detection result when compared with [25]. Note that the supporting materials of the proposed algorithm is available online.<sup>1</sup>

The main contributions of this paper are two-fold. The first one is to take the local geometric features of each point into consideration when introducing cluster-level global rarity refinement to each point in all clusters. Compared to calculating point-cluster probability matrix in [25] which does not take the influence of local geometric features into consideration, their method will unavoidably ignore some detailed information of the point cloud. To address this limitation, we use local geometric features of each point as a guidance in the random walk process and generate more detailed point-level global rarity result. The second contribution is the proposed optimization framework. Existing saliency detection algorithms usually apply simple summation or linear combination to different saliency cues to obtain the final saliency detection results, which does not take the characteristics of different

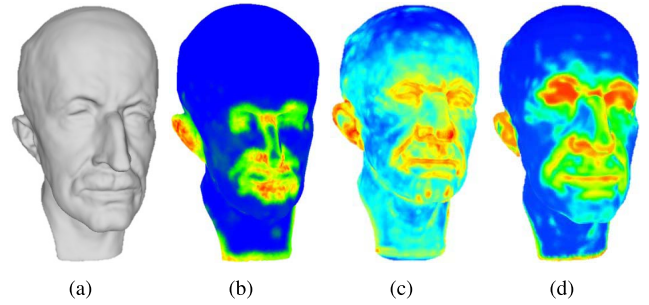


Fig. 1. The saliency detection results for the Max Planck model. (a) The original model. (b) The proposed saliency detection result. (c) and (d) The saliency detection results reported in [1] and [25] respectively. Our saliency detection on this model is more concentrated on facial features and performs better saliency detection result when compared with [1] and [25].

salient features into consideration. The proposed adaptive optimization framework can utilize the inherent relationships between different saliency cues thus achieving more satisfying results.

The remainder of the paper is organized as follows: Section II introduces the related work in saliency detection. In Section III, the proposed saliency detection method is illustrated in detail. Experimental results are described in Section IV with conclusions provided in Section V.

## II. RELATED WORK

In this section, we first summarize the related works about visual saliency detection in images and videos. Then we discuss some works that compute visual saliency for 3D meshes and point clouds.

### A. Visual Saliency on 2D Data

Visual saliency detection is a simulation of the attentional mechanism in the HVS [26]. It describes the human attention distribution [27] or the eye movements for a given scene. Saliency detection can be roughly divided into two categories, the bottom-up saliency detection and the top-down saliency detection. The former one utilizes low-level features [28] such as orientation and texture, e.g., the method proposed in [29]. While the latter one mainly utilizes high-level knowledge [30] such as the depth-from-focus cue used in [31].

Saliency detection problem has been investigated for decades and many saliency detection methods have been proposed [32], especially for saliency detection in 2D images. In [33], a diffusion-based image saliency detection algorithm which first constructs a two-layer sparse graph and then uses the manifold ranking diffusion method is proposed. And in [34] Ishikura *et al.* calculate salient regions by utilizing two global extrema-based features to refine the potential salient regions which are obtained by calculating multi-scale extrema of local perceived color difference.

Besides applying saliency detection to 2D images, it has also been applied to other data such as videos. For instance, in [35], a video saliency detection algorithm is proposed which fuses the motion clues based color saliency to obtain low-rank

<sup>1</sup><http://iip.whu.edu.cn/PCSaliency/>

coherency clues and then uses them to guide the spatial temporal saliency diffusion. In [36], a novel algorithm is proposed by deeply incorporating inter-frame and intra-frame information. Then an energy optimization is applied to improve the saliency detection result.

In addition to video saliency detection, recent works also focus on saliency detection for images that contain 3D information, for example, RGB-D image and panoramic image. The former one represents an RGB image with its corresponding depth image while the latter one records the whole scene in the 3D world. For RGB-D image saliency detection, Cong *et al.* [37] presented a method which first measures the confidence of the depth information and then computes the compactness saliency by integrating color and depth information. Later different saliency cues are integrated to acquire the final saliency detection results. In [38], Niu *et al.* proposed an algorithm utilizing global disparity contrast and domain knowledge to calculate RGB-D saliency and in [39], a learning-based algorithm is proposed for RGB-D saliency detection. In their paper, the monocular features are first used to capture the inner characteristics of individual saliency and then they utilize the compensatory features to maintain the consistency between stereo matching regions. For saliency detection in panoramic image, Sitzmann *et al.* [40] captured and analyzed gaze and head orientation data to help get better understanding of how people explore the virtual environment. In addition, they also explored several applications related to saliency prediction in virtual reality, and in [41], Zhu *et al.* proposed an algorithm to help predicting the head movement of the viewers while they are viewing the panoramic images. In [42], Hu *et al.* proposed an online agent which can navigate the 360 degree videos by selecting the viewing angle focusing on the most salient object.

### B. Visual Saliency on 3D Data

Although saliency detection has been applied to various kinds of 2D data, few approaches have been proposed to process 3D data. Point clouds are the raw output data of many sensors, such as the Lidar or the Microsoft Kinect sensors and 3D meshes can be reconstructed by evaluating a known mathematical description or by interpolating a given set of unorganized 3D point clouds. For example, in [43], Hoppe *et al.* presented a novel algorithm to reconstruct the surface from unorganized point clouds without assuming any structure on the data points and Mitra and Nguyen [44] presented a method which estimates the surface normal at all sample points using local least square fitting method. Directly detecting salient regions on 3D point clouds rather than 3D meshes will greatly improve the computational efficiency since it avoids the time-consuming mesh generation using the point clouds. Moreover, 3D meshes usually have bigger data size due to the storage of the surface information, detecting salient regions on point clouds will greatly reduce the storage capacity and eliminate the errors introduced by doing surface reconstruction.

Mesh saliency detection has been investigated for a long period and many algorithms have been proposed. Early mesh

saliency detection algorithms usually evaluate the saliency of a model by applying saliency detection in its 2D projection. For instance, Yee *et al.* [45] applied saliency detection to a reference image which projects the 3D dynamic scene and used the saliency map to reduce computational load for calculating global illumination solution. Frintrop *et al.* [46] provided a method which can accelerate 3D object detection and classification by feeding the rendering images of the 3D scene into attention system to detect the regions of potential interest. And Howlett *et al.* [47] captured the human gaze data of a 2D image which presents a 3D object, and used the human gaze data to help determine the salient features of the 3D object.

Recently, many mesh saliency detection algorithms have been proposed which are inspired by works of saliency detection in 2D images [48]. For example, Lee *et al.* [6] calculated mesh saliency in a scale-dependent manner which uses the center-surround mechanism, since it has the intuitive appeal of identifying regions that are different from their surroundings. This center-surround operation is first presented by Itti *et al.* [48] which is an useful technique for image saliency detection. Besides, Castellani *et al.* [49] presented a novel mesh saliency detection method using the joint multi-scale mechanism which is also proposed by Itti *et al.* [48]. Wu *et al.* [50] provided a mesh saliency detection algorithm based on global rarity, which has also been used as a saliency detection principle by Cheng *et al.* [20]. Leifman *et al.* [12] calculated the mesh saliency both locally and globally, and took the surrounding context into consideration. This is motivated by the context-aware algorithm proposed in [51].

Despite the various algorithms proposed for saliency detection in 3D meshes, few saliency models can handle 3D point clouds. Reference [1] proposed a robust algorithm which can help saliency detection in point clouds. Their method takes both the low- and high-level distinctness into consideration and has been evaluated by various applications. And Tasse *et al.* [25] provided a cluster-based algorithm for detecting 3D point clouds saliency which utilizes cluster uniqueness and spatial distribution to formulate the final saliency detection result.

## III. THE PROPOSED METHOD

In this paper, we focus on saliency detection for 3D point clouds. To address this problem, we take the local distinctness and global rarity into account, which are two basic characteristics of the human visual system supported by psychological evidences [19]. Since these two characteristics have achieved satisfying results in 2D saliency detection, we utilize them to improve saliency detection for point clouds. Fig. 2 presents the proposed framework. The local distinctness of each point is firstly evaluated, capturing local salient features such as the contour of ears and nose. Then the global rarity finds larger salient regions such as the whole ears. Finally, an optimization framework is proposed to integrate both the local distinctness and global rarity values to obtain the final saliency detection result for the point cloud.



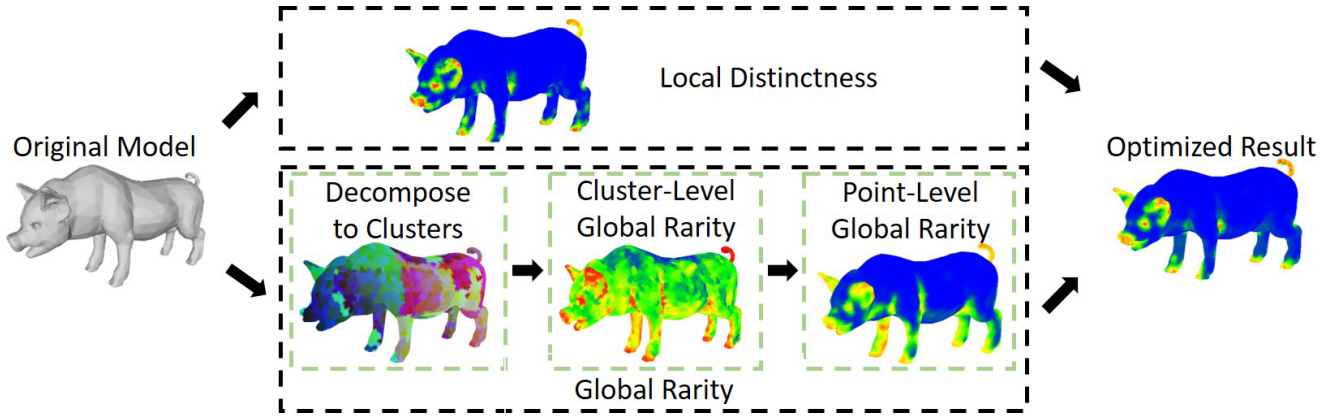


Fig. 2. Algorithm framework. First the local distinctness of each point is evaluated. Then the point cloud is decomposed into several clusters and the global rarity value of each point is obtained. Finally, an optimization framework is proposed to integrate both the local distinctness and global rarity values to obtain the final saliency value of each point in the point cloud.

#### A. Local Distinctness

To calculate the local distinctness of each point, a descriptor is needed to characterize the local geometric features. The descriptor should be robust to noise and efficient to compute, as well as having good expressive power of the local geometric features [1]. We have experimented with several descriptors such as Zernike coefficients [52], SHOT [53] descriptor, and found that FPFH [23] descriptor is more efficient in harmony with our local geometric features of the point cloud.

To compute the FPFH descriptor for a query point  $p_i \in P$  with  $P$  representing the point cloud, its  $k$ -neighboring points in the sphere with radius  $r$  are selected and a Darboux  $uvn$  frame ( $u = n_i, v = (p_j - p_i) \times u, w = u \times v$ ) is defined below as in [23].

$$\begin{aligned} \alpha &= v \cdot n_j \\ \phi &= (u \cdot (p_j - p_i)) / \|p_j - p_i\| \\ \theta &= \arctan(w \cdot n_j, u \cdot n_j) \end{aligned} \quad (1)$$

where  $p_j$  is a point belonging to the  $k$ -neighborhood of point  $p_i$ .

Using the quantized angles described above, the relationships between a query point and its neighbors can be calculated, which is called the Simplified Point Feature Histogram (SPFH) [23]. Later for each query point, the SPFH values are used to obtain the final FPFH descriptor for that point, which is specified as:

$$FPFH(p_i) = SPFH(p_i) + \frac{1}{k} \sum_{j=1}^k \frac{SPFH(p_j)}{\|p_i - p_j\|} \quad (2)$$

where  $k$  is the number of the nearest neighboring points of point  $p_i$  and  $\|\cdot\|$  represents the  $L2$  distance between the point  $p_i$  and a  $k$ -nearest neighboring point  $p_j$  in a given metric space as described in [23].

Due to the influence of noise and different sampling density in 3D point clouds, the Euclidean distance is not suitable for measuring the dissimilarity between two points. Following the method in [1], the Chi-Square distance is used to measure the dissimilarity between points. Given two points  $p_i$  and  $p_j$  in

$P$ , the Chi-square distance  $\chi^2(p_i, p_j)$  is calculated as:

$$\chi^2(p_i, p_j) = \sum_{n=1}^N \frac{(FPFH_n(p_i) - FPFH_n(p_j))^2}{FPFH_n(p_i) + FPFH_n(p_j)} \quad (3)$$

where  $N$  represents the number of bins in the FPFH descriptor.

Since a point  $p_i$  is distinct if it differs from its surrounding points, the local distinctness can be defined as the weighted sum of dissimilarities with its surrounding points. The formula is defined as:

$$D(p_i) = 1 - \exp\left(-\frac{1}{R} \sum_{j=1}^R \frac{\chi^2(p_i, p_j)}{1 + \|p_i - p_j\|}\right) \quad (4)$$

where  $D(p_i)$  represents the local distinctness value of point  $p_i$ ,  $R$  is the number of the surrounding points taken into consideration for point  $p_i$ . Note that  $R$  is set according to the sampling density of the point cloud.

To prevent labeling too many distinct points, a simple but effective piecewise suppression function is designed in Eq.(5) to reduce the number of distinct points.

$$F(p_i) = \begin{cases} D(p_i) & D(p_i) \geq h \\ D(p_i) * D(p_i) & \text{otherwise} \end{cases} \quad (5)$$

where  $F(p_i)$  represents the amplified local distinctness value for point  $p_i$  and  $h$  denotes the threshold controlling the range of the distinct points to be suppressed. In our experiment, we used an adaptive threshold  $h$  which equals to the top 20% local distinctness values in the point cloud. We use this threshold since it achieves better local distinctness calculation results when compared with other thresholds and many other algorithms in image saliency detection have also used this threshold, such as the algorithms proposed by Borji [54] and Kanan *et al.* [30].

#### B. Global Rarity

To compute the global rarity for the point cloud, we first decompose the point cloud into small clusters to reduce the number of regions that needs to be taken into consideration. The voxel cloud connectivity segmentation algorithm proposed

in [24] is used to decompose the point cloud. Papon's method uses the voxel relationships to produce oversegmentations which are fully consistent with the spatial geometry of the scene in three dimensions, so it obtains the minimal loss of information and improves the calculation efficiency.

To characterize the geometric features of each cluster, the mean FPFH descriptor  $FPFH(c_i)$  is defined for cluster  $c_i$ . Note that  $FPFH(c_i)$  is the mean value of the FPFH descriptors for all the points in cluster  $c_i$ . Since the global rarity is regarded as the inverse of a region's probability of happening over the entire point cloud [22], the initial global rarity for cluster  $c_i$  is calculated below by taking different weight of each cluster into consideration:

$$G(c_i) = 1 - \exp\left(-\frac{1}{N} \sum_{j=1}^N \frac{\chi^2(c_i, c_j)}{1 + \|c_i - c_j\|}\right) \quad (6)$$

where  $G(c_i)$  is the initial global rarity value for cluster  $c_i$ ,  $N$  is the number of clusters in the point cloud, and  $\|c_i - c_j\|$  represents the  $L_2$  distance between the centers of cluster  $c_i$  and  $c_j$ . Note that the Chi-square distance  $\chi^2(c_i, c_j)$  is calculated using Eqn. (3).

Using the formula in Eq. (6), the initial global rarity value for each cluster is obtained. Then we need to introduce cluster-level global rarity to each point in all clusters. Tasse *et al.* [25] defined a point-cluster probability matrix to represent the probability of a point  $p_i$  belongs to a cluster  $c_j$ . Although the spatial distribution and the difference in the descriptor space of each cluster are taken into consideration to build the probability matrix, their method does not consider influence of the local geometric features of each point which has shown impressive results on applications such as object categorization [55]. Calculating global rarity without using local geometric features of each point will unavoidably ignore some detailed information from the original model and decrease the accuracy of global rarity calculation.

Random walk is an algorithm that calculates the probability of a random walker starting at each unseeded point will first reach each of the seeded points, and assigns predefined values to the unseeded points with the greatest probability. It has been widely used in image processing field such as image segmentation and image saliency detection. Motivated by [56] which uses random walk ranking method to introduce the prior saliency estimation to each pixel in the image, we apply the random walk ranking algorithm to the proposed method to help introduce the cluster-level global rarity to point-level global rarity for each point in all clusters by taking the local geometric features of each point into consideration. Note that this introduction process is quite different from the binarization approach used in some image processing algorithms such as [57], [58] which works like an optimization framework to help eliminating non-salient regions while highlighting the salient regions. Instead, this process helps to assign global rarity values to each unseeded point by calculating its probability of first reaching each of the seeded point. Fig. 3 shows an example of using random walk ranking algorithm to introduce cluster-level global rarity refinement to each point in all clusters. Model (a) represents the initial cluster-level global

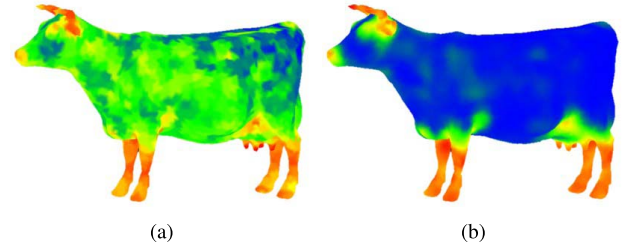


Fig. 3. Comparison between the initial (a) cluster-level global rarity result and (b) the point-level global rarity refinement which is obtained by using the random walk ranking method.

rarity result calculated using formula in Eq. (6) and model (b) shows the point-level global rarity refinement obtained using the random walk ranking method. Compared to model (a), model (b) achieves more visual appealing result and the details of the model such as the mouth and horn of the cow are better preserved while the inaccurate detection results on the body of the cow are eliminated efficiently.

Using the initial global rarity values calculated for each cluster in Eq. (6), we can mark some clusters in the point cloud as the salient clusters or non-salient clusters according to their cluster-level global rarity values. Note that since a cluster contains many points, the point nearest to the center of the salient or non-salient cluster will be labeled as the salient or non-salient seed point for later calculation in the proposed algorithm. For example, if a cluster's global rarity value is higher than threshold  $th_1$  or lower than threshold  $th_2$ , the point nearest to the center of this cluster will be labeled as the salient seed point or non-salient seed point while the other points belonging to this cluster will be labeled as the unseeded points. For clusters whose global rarity value is neither higher than threshold  $th_1$  nor lower than threshold  $th_2$ , all the points in the cluster will be categorized into unseeded points. The threshold of  $th_1$  and  $th_2$  used in the proposed algorithm are described as:

$$\begin{aligned} th_1 &= \text{mean}(G) + \frac{1}{2}\Delta \\ th_2 &= \text{mean}(G) \end{aligned} \quad (7)$$

where  $\Delta = \max(G) - \text{mean}(G)$ .

When the seed points are chosen, we give a random walker starting at each of the unseeded point and calculate the probability of this random walker first reaches each of the seed point. For example, if the number of the seed points is set to  $K$ , a  $K$ -tuple vector will be built for each unseeded point to record the probability of the random walker starting at this location first reaches each of the  $K$  seed points. Then for each unseeded point, we select the most probable destination according to its  $K$ -tuple vector to obtain the point-level global rarity refinement. The details of applying random walk ranking algorithm to help introduce cluster-level global rarity refinement to each point in all clusters are described thereafter.

After seed points are chosen according to the cluster-level global rarity results, a graph  $G = (V, E)$  is subsequently built with nodes  $V$  and edges  $E$ , where  $V$  corresponds to all the points in the point cloud and  $E$  represents the connections between two points. Different from 2D images which can

build the connection between any two pixels according to their features and distance [56], in 3D models, the connection between any two points is much more complicated due to the difference of the data structure [59]. For example, given a 3D point cloud  $P$ , there always exists a point  $p_i$  and a point  $p_j$  in  $P$  which cannot be seen from the same viewpoint, this indicates that the global rarity value of point  $p_i$  is less likely to have influence on point  $p_j$ , so we need to bias the random walker to avoid moving along the connection between point  $p_i$  and point  $p_j$ . Following the method proposed in [60], we set the weight zero for this connection to inhibit the random walker moving along the edge. Considering the complexity of judging whether the connection between two different points should be set to zero, we simplify the process by just calculating the connections between a point  $p_i$  and its  $k$ -nearest neighboring points and assign weight zero for the connections between the point  $p_i$  and the other points in the point cloud  $P$ . By choosing a smaller  $k$ , we can assume that the point  $p_i$  will have strong influence to its  $k$ -nearest neighboring points.

The connection between the point  $p_i$  and a  $k$ -nearest neighboring point  $p_j$  is quantified by a weight matrix calculated using the local geometric features of each point as in Eq. (8).

$$w_{ij} = \exp\left(-\frac{\|FPFH(p_i) - FPFH(p_j)\|^2}{\sigma_1^2}\right) \quad (8)$$

where  $w_{ij}$  represents the connection between the point  $p_i$  and  $p_j$ .  $\sigma_1$  is a controlling constant which is set to 3.7 in our experiment. Note that the local geometric features of each point represented by the FPFH descriptor is used for calculating the weight matrix in the proposed method, which is different from using the CIELab color descriptor in 2D images [56]. After the weight matrix is built, the Laplacian matrix  $L$  can be built as described in [56] in further.

Let  $P^s$  denotes the probability vector for points belonging to class  $s$ , and it consists of two subset  $P_u^s$  and  $P_{seeds}^s$ .  $P_u^s$  represents the probability vector for the unseeded points while the  $P_{seeds}^s$  shows the seed points in the point cloud with a fixed value 0 or 1. The optimized  $P^s$  can be achieved by minimizing the Dirichlet integral [56]. Following the method proposed in [61], a fitting constraint as in Eq.(9) is added to help restrict the Dirichlet integral to be close to the initial cluster-level global rarity values,

$$Dir[p^s] = \frac{1}{2}(p^s)^T L(p^s) + \frac{\eta}{2}(p^s - Y)^T (p^s - Y) \quad (9)$$

where  $\eta$  is a controlling parameter and  $Y$  represents the point-wise indication vector inheriting the initial cluster-level global rarity values.

By setting the differentiation of  $Dir[p^s]$  with respect to  $P_u^s$  as zero, the optimized solution of  $P_u^s$  for the points in all clusters excluding the seed points can be calculated as in [61]:

$$p_u^s = (L_u + \eta I)^{-1}(-B^T p_{seeds}^s + \eta Y_u^s) \quad (10)$$

Using the random walk ranking method mentioned above, we introduce the initial cluster-level global rarity refinement to each point in all clusters by taking the local geometric features of each point into consideration.

### C. Optimization Framework

After the local distinctness value and the global rarity value for each point are obtained, we need to integrate these two cues to obtain the final saliency detection result. Existing saliency detection algorithms usually apply simple summation or linear combination of different saliency cues to obtain final saliency detection results [65]. However, linear combination cannot provide a proper saliency detection results for all the data since the weight coefficients are usually fixed. A combination of weight coefficients might work well for data  $A$  but fail to provide a good result for data  $B$ , thus a more adaptive integration mechanism is required. To address this problem, a novel optimization framework is proposed by utilizing the inherent relationships between different saliency cues.

Motivated by the method proposed in [66], we model the problem of integrating different saliency cues as an optimization of the saliency values of all points in the point cloud. The loss function in Eq.(11) assigns the salient point value 1 and non-salient point value 0, with a smooth term added to the loss function since the human visual system tends to group items together and obtain a uniform saliency detection results [67], rather than an isolated salient point,

$$L = \sum_{k=1}^2 \sum_{i=1}^N \left( \frac{s_i^2}{Z_k(p_i)} + Z_k(p_i)(1 - s_i)^2 \right) + \sum_{j \in R} W_{ij}(s_i - s_j)^2 \quad (11)$$

where  $Z \in \{F, G\}$  with  $Z_1(p_i)$  represents the amplified local distinctness value  $F(p_i)$  and  $Z_2(p_i)$  denotes the point-level global rarity refinement  $G(p_i)$ .  $N$  equals to the number of the points in the point cloud while  $R$  represents the neighboring points of point  $p_i$ . Note that the denominator may be equal to zero in the actual operation, so a small constant 0.1 is added to  $F(p_i)$  and  $G(p_i)$  if its value is close to zero.

Since both the local distinctness and the global rarity are important for the human visual system, the loss function takes the symmetry of these two characteristics into consideration. It can be proved that the loss function is a convex function with  $s_i$ , so the loss function can be minimized by taking the derivative of the loss function with respect to  $s_i$  and letting it equal to zero. By minimizing the loss function, the first term in the loss function encourages the point with higher local distinctness value and higher global rarity value to obtain higher saliency value and suppress the point with lower local distinctness value and lower global rarity value. The second term encourages the smoothness of the final saliency detection results. For the point  $p_i$  with higher local distinctness value and lower global rarity value or the point  $p_j$  with lower local distinctness value and higher global rarity value, their saliency values will be calculated by considering the impact of the other points.

The weight coefficient  $W_{ij}$  in the loss function represents the weighted influence of the  $k$ -nearest neighboring point  $p_j$  towards the point  $p_i$ , which is defined as

$$W_{ij} = \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_2^2}\right) \quad (12)$$



where  $\sigma_2$  is a controlling constant which is set to 0.02 and  $\|\cdot\|$  is the  $L_2$  distance between the point  $p_i$  and  $p_j$ .

#### IV. EXPERIMENTAL RESULTS

In this section, we provide comparison of our method with several state-of-the-art mesh-based and point cloud-based saliency detection algorithms to demonstrate our performance. Besides, an ablation study is also provided to show the effectiveness of the local distinctness calculation, global rarity calculation and the proposed optimization framework. Finally, we discuss how the proposed algorithm can be applied to graphics applications such as interest point detection, view-point selection and mesh simplification.

##### A. Evaluation of Saliency Detection Results

Due to the fast development of technology, many saliency detection algorithms have been presented. For example, Yun and Sim [68] presented a novel algorithm for colored 3D point cloud data and Nouri *et al.* [69] proposed a method for 3D colored mesh data. Besides, Jeong and Sim [10] come up with an algorithm which deals with saliency detection for semi-regular meshes while Lau *et al.* [70] designed an algorithm for predicting the tactile saliency for meshes. Although these algorithms all achieved quite satisfying saliency detection results, the input data types are quite different from the proposed algorithm, so we do not compare the proposed algorithm with these methods. In this section, the proposed algorithm is compared with several state-of-the-art mesh-based or point cloud-based saliency detection algorithms to demonstrate its performance.

The comparison of the proposed method against the point cloud-based algorithms proposed in [25], [62], mesh-based algorithms proposed in [11], [63] and the ground-truth data collected in [64] is shown in Fig. 4. According to the figure, the method proposed in [25] is more useful in detecting salient regions such as the ends of octopus, but it fails at the body of octopus and the hand palm. This might be caused by the substantially changed cluster descriptor from the flat surface due to the use of Euclidean space for clustering. The method proposed in [62] produced quite correct saliency results, but some non-salient regions such as the contour of the teddy model and the hand palm are also highlighted. The method proposed in [63] captures good saliency detection result on the vase model but fails to detect salient regions on the human model. This might be caused by the selection of the background queries, since patches on the chest are quite similar to each other. If a patch belonging to these regions is selected as the background, the other patches with higher relevance to the selected patch are more likely to be regarded as the background. For the method proposed in [11], it detects the salient regions quite well but is more likely to provide similar regions with different saliency values, which might be explained by the lack of spatial symmetry of the model. Compared with the methods mentioned above, the proposed method performs a more uniform saliency detection results and in most cases detects better salient regions of the model, such as the face of the Igea model as well as the limbs and chest

TABLE I  
QUANTITATIVE EVALUATION OF THE ABLATION STUDY

	Column (a)		Column (b)		Column (c)		Column (d)	
	SE	CC	SE	CC	SE	CC	SE	CC
Dolphin	17.1	0.44	31.4	0.40	23.9	0.43	<b>11.0</b>	<b>0.45</b>
Man	36.6	0.15	73.0	<b>0.22</b>	54.3	<b>0.22</b>	<b>32.0</b>	0.19
Teapot	31.0	0.34	28.1	0.35	29.4	0.35	<b>28.0</b>	<b>0.36</b>
Igea	23.7	0.25	27.3	<b>0.30</b>	25.2	<b>0.30</b>	<b>16.9</b>	0.29

Column (a) to (d) represent the column (a) to (d) in Fig. 5.

for human model. However, the proposed algorithm also needs further improvement since it often fails to detect the correct salient regions in cylindrical shapes, for instance, the legs of the chair model and the feet of the octopus model.

##### B. Ablation Study

To provide a better illustration about the effectiveness of both the local distinctness and global rarity features, the ablation study is conducted where we report the results of only using local distinctness feature or global rarity feature to help point cloud saliency detection. Besides, we also provided the comparison between using the linear combination and the proposed optimization framework to show the better integration results using the optimization framework. Fig. 5 shows the saliency detection results of only using local distinctness feature and global rarity feature, as well as the integration results using the linear combination and the proposed optimization framework.

According to Fig. 5, the saliency detection results using local distinctness feature highlight smaller regions when compared with global rarity feature. For example, in the fourth row of the figure, the local distinctness feature mainly highlights the mouth, eyes and the contour of the nose region and fails to highlight the whole nose, while the global rarity feature highlights the whole facial region. But in the second row, the global rarity feature also highlights the facial region including the cheek, which is not salient in the ground truth data. This demonstrate that using local distinctness feature or global rarity feature alone can hardly obtain satisfying saliency detection results and these two features are quite complementary to each other. According to the experimental results, the local distinctness feature fails to detect the head of the dolphin while the global rarity feature detects it correctly. If the linear combination is used to integrate the local distinctness and global rarity features, it will combine two features without bias. For instance, in the first and third rows of the third column, it can be observed that linear combination cannot eliminate the incorrect salient regions detected by the local distinctness feature, such as the head of the dolphin and the spot in the body of the teapot. It just adds the saliency values of the corresponding regions directly so these incorrect salient regions will remain in the integration results and influence the visual performance. When we compare the integration results obtained by the optimization framework with linear combination, for example, the first and third rows of the fourth column, it can be observed that the optimization framework preserves the correct salient regions such as the head of the



Fig. 4. Comparison of the different saliency detection results. From top to bottom: saliency detection results calculated using the proposed method, point cloud-based saliency detection results reported in [25] and [62], mesh-based saliency detection results reported in [11], [63] and the groundtruth of the models reported in [64]. Note that blue color represents low saliency value and red color shows high saliency value.

dolphin and eliminates the incorrect salient regions, providing better visual performance. The experimental results above demonstrate the superiority of combining local distinctness and global rarity features and proved the effectiveness of utilizing the proposed optimization framework.

Besides qualitative analysis, we also provide quantitative evaluation to make clearer illustration. We first use the Saliency Error (SE) proposed in [71] to help the evaluation. This metric is quite intuitive and lower SE represents better saliency detection results. The metric is defined as:

$$SE = \sum_{i=1}^N |S(i) - S_{gt}(i)|$$

where  $N$  represents the total number of points in the model while  $S(i)$  denotes the normalized saliency value of point  $i$  calculated using the algorithm, and  $S_{gt}(i)$  is the normalized Schelling distribution value reported in [64] which is used as the groundtruth.

We also employed the evaluation metric CC score which is widely used in 2D image saliency evaluation. Higher CC score represents better detection accuracy of the algorithm. The quantitative evaluation results are presented in Table. I.

According to Table. I, the proposed optimization framework achieved the lowest SE score among all the four models. In the Dolphin model and the Teapot Model, the optimization framework also obtains the highest CC score. For the Man model and the Igea model, the CC scores are ranked the third



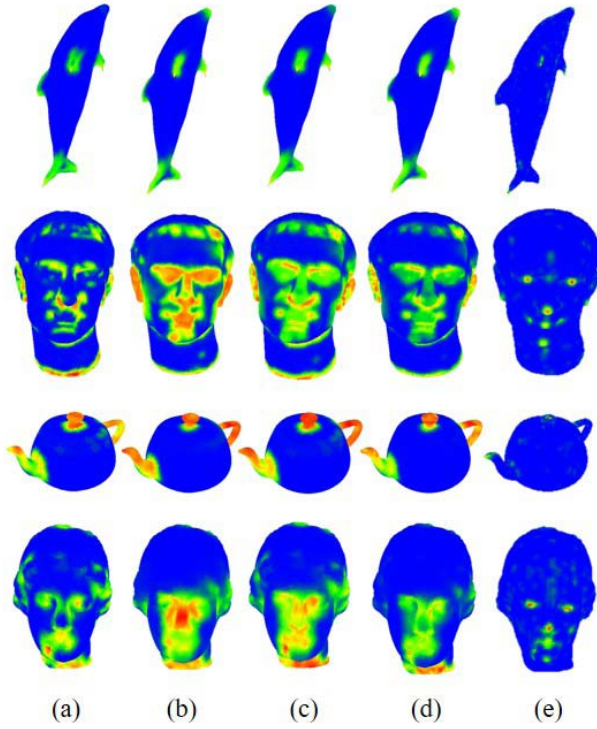


Fig. 5. Saliency detection results using different features and different integration algorithms. Column (a) shows the saliency detection results using local distinctness feature while column (b) represents the saliency detection results utilizing global rarity feature. Column (c) presents the saliency integration results using linear combination and column (d) shows the saliency integration results using the proposed optimization framework. Column (e) presents the groundtruth of the models reported in [64].

place but are quite close to 0.22 and 0.30 respectively. This might be explained as the local distinctness results of these two models are quite unsatisfactory, making it hard for the optimization framework to obtain better CC score.

### C. Interest Point Detection

3D interest point refers to the points that are distinctive in their locality [76]. It can provide local features that are semantically significant and is essential for many graphics applications, such as mesh segmentation [77] and registration [78]. To provide better illustration about the performance of the proposed algorithm, we apply it to help 3D interest point detection and evaluate the interest point detection results against the 3D interest point detection benchmark proposed in [76]. The benchmark consists of two datasets. Dataset A includes 24 triangular meshes which uses the interest point data labeled by 23 participants while dataset B contains 43 triangular meshes (including all the models in dataset A) and uses the interest point data labeled by 16 participants, and Dutagaci *et al.* developed an application that asks human participants to label the interest points on the triangular meshes. Using the human-labeled points, Dutagaci *et al.* constructed the ground-truth data and compared it with six different interest point detection techniques based on the False negative error (FNE), False positive error (FPE) and Weighted miss error (WME). The six interest point detection techniques include: Salient points [49], Mesh

saliency [6], SD-corners [75], 3D-Harris [72], 3D-SIFT [74] and HKS-based interest point detection method [73]. The computation of the evaluation matrices are briefly summarized below:

1) *False Negative Error*: Let  $G$  denotes the ground-truth points for a model  $M$ . For an interest point  $g \in G$ , the geodesic neighborhood of point  $g$  is calculated as,

$$C_r(g) = \{p \in M | d(g, p) \leq r\}, \quad (13)$$

where  $d(g, p)$  denotes the geodesic distance between the point  $g$  and the point  $p$  while  $r$  represents the radius of the geodesic neighborhood.

We denote  $D$  as the interest points set detected by the evaluated method, the point  $g$  is detected correctly if there exists a point  $a \in D \cap C_r(g)$  which is not closer to any other points in  $G$  [76]. Let  $N_c$  be the correctly detected points' number in  $G$  and  $N_g$  be the total number of the points in  $G$ . The *FNE* is computed as:  $FNE = 1 - N_c/N_g$ .

2) *False Positive Error*: For each correctly detected point  $g \in G$ , there is a corresponding point  $a \in D$ . So all the points in  $D$  without a corresponding point in  $G$  are seen as false positive. Let  $N_d$  denotes the amount of the detected interest points by the evaluated method, the *FPE* is defined as:  $FPE = 1 - N_c/N_d$ .

3) *Weighted Miss Error*: For a geodesic neighborhood with radius  $r$ , a ground truth point might be marked by  $n_i$  human participants, so the *WME* is defined as:

$$WME = 1 - \frac{1}{\sum_{i=1}^{N_g} n_i} \sum_{i=1}^{N_g} n_i \delta_i \quad (14)$$

with

$$\delta_i = \begin{cases} 1 & \text{if } g_i \text{ is detected} \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

where  $g_i$  belongs to the ground-truth point.

Both the *FNE*, *FPE* and *WME* are measures between 0 and 1. The lower *FNE*, *FPE* and *WME* an algorithm performs, the better performance in interest point detection the algorithm achieves.

Following [25] and [1], our interest points are defined as the local maxima over a threshold which is computed as the average value over all local maxima values. Fig. 6 presents the evaluation results of our method against the interest point detection benchmark using the *FNE*, *FPE* and *WME* as the radius  $r$  increases, and compared with the six other interest point detection algorithms mentioned in [76]. According to Fig. 6, The HKS [73] based method shows significantly lower *FPE* when compared with other methods since the HKS [73] based interest point detection method finds very few interest points but has higher accuracy. The results in Fig. 6 also show that our method achieves lower *FPE* than the other methods except for HKS [73], while the *FNE* is slightly higher than some methods but still remains comparable to the 3D-SIFT [74] algorithm in both of the two datasets. Since a method can hardly achieve lower *FPE* and lower *FNE* simultaneously, the balance between false positive and false negative values

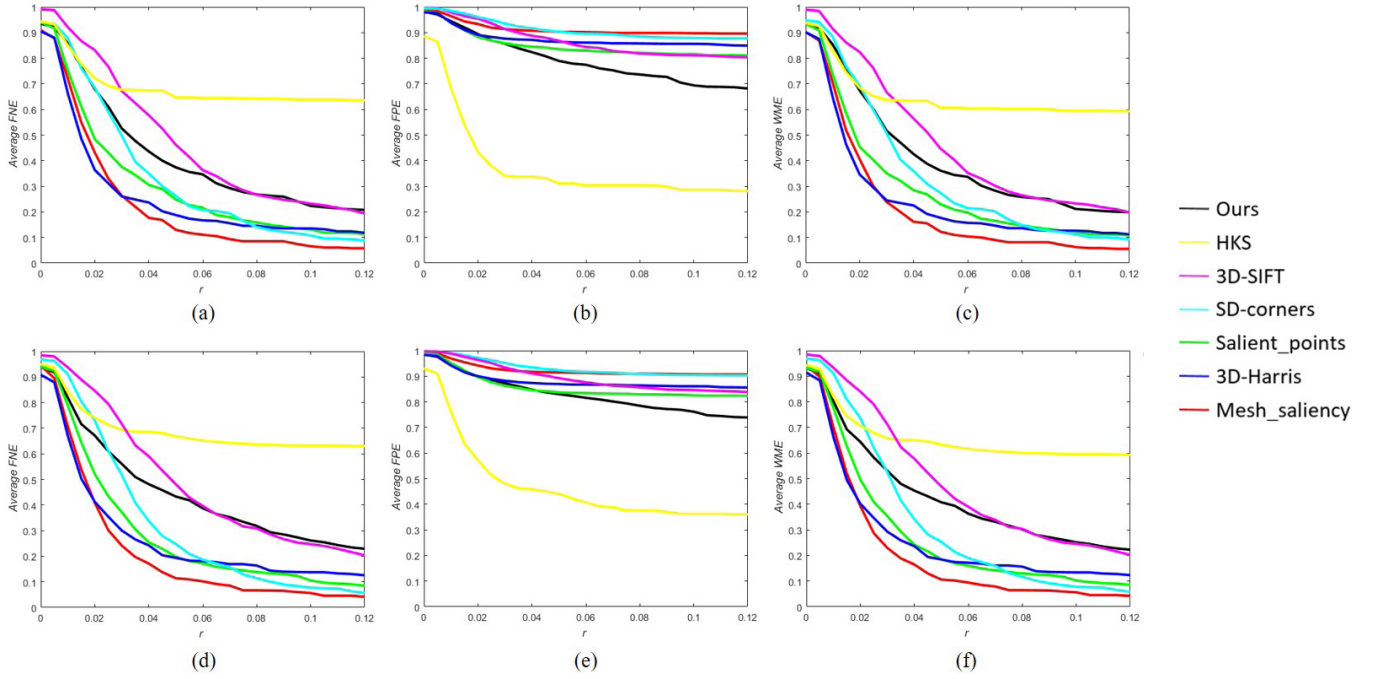


Fig. 6. Comparison results using the FNE, FPE and WME to evaluate the performance of our algorithm against the other interest point detection methods on the interest point detection benchmark. The top three figures provide the evaluation results for dataset A while the bottom three figures present the evaluation results for dataset B. The lines in different colors represent the evaluation score of each interest point detection algorithm with respect to the ground truth interest points with  $\sigma = 0.03$  and  $n = 8$ .

TABLE II  
 $F_\beta$  VALUE FOR INTEREST POINT DETECTION

	Dataset A						Dataset B					
	$r=0.03$		$r=0.06$		$r=0.09$		$r=0.03$		$r=0.06$		$r=0.09$	
	$F_{1.0}$	$F_{2.0}$	$F_{1.0}$	$F_{2.0}$	$F_{1.0}$	$F_{2.0}$	$F_{1.0}$	$F_{2.0}$	$F_{1.0}$	$F_{2.0}$	$F_{1.0}$	$F_{2.0}$
Ours	0.25	0.36	0.33	<b>0.47</b>	0.40	<b>0.54</b>	0.20	0.30	0.28	0.42	0.35	<b>0.50</b>
Salient Points [49]	0.23	<b>0.37</b>	0.28	0.46	0.30	0.49	0.22	<b>0.36</b>	0.27	<b>0.46</b>	0.28	0.48
3D-Harris [72]	0.21	<b>0.37</b>	0.24	0.42	0.25	0.43	0.20	0.35	0.22	0.40	0.23	0.42
HKS [73]	<b>0.43</b>	0.35	<b>0.47</b>	0.39	<b>0.48</b>	0.40	<b>0.38</b>	0.33	<b>0.44</b>	0.38	<b>0.46</b>	0.40
Mesh Saliency [6]	0.15	0.29	0.17	0.34	0.18	0.35	0.13	0.27	0.15	0.31	0.16	0.32
3D-SIFT [74]	0.13	0.21	0.25	0.39	0.30	0.47	0.10	0.17	0.20	0.34	0.25	0.41
SD-corners [75]	0.11	0.21	0.18	0.34	0.21	0.39	0.08	0.17	0.15	0.29	0.17	0.33

$\sigma = 0.03$  is used to compute the  $F_\beta$  value.

needs to be quantified. Following [25], we quantify the balance using the formula below:

$$F_\beta = (1 + \beta^2) \frac{(1 - FPE)(1 - FNE)}{\beta^2(1 - FPE) + (1 - FNE)} \quad (16)$$

where  $(1 - FPE)$  represents the precision and  $(1 - FNE)$  is the recall.  $F_\beta$  represents the quantified balance between the  $FPE$  and  $FNE$ .  $\beta$  is used as the weight coefficient between the precision and the recall.

Using the formula in Eq.(16), we compute the quantified balance values for all the interest point detection methods mentioned previously. The quantified balance values are shown in Table. II. According to Table. II, it can be also observed that the HKS algorithm achieves the best performance when the recall is as important as the precision, and the proposed algorithm performs the best experimental value when recall is two times as important as precision and the tolerance radiuses

are set to  $r = 0.06$  and  $r = 0.09$  for dataset A and  $r = 0.09$  for dataset B. Although the HKS algorithm achieves quite satisfying results, it is at the expense of the higher false negative error and weighted miss error, which might not be suitable for some applications which requires higher recall rate. Besides performing interest point detection evaluation, an example of the interest point detection results obtained by the proposed method and the other algorithms is provided in Fig. 7. When compared with the cluster saliency based interest point detection result reported in [25], the proposed method detects better interest points at the knee and the tale of Armadillo model

#### D. Viewpoint Selection

3D models are more convenient to obtain with the development of the remote sensing technologies. The increasing number of 3D models attaches more importance to automatic

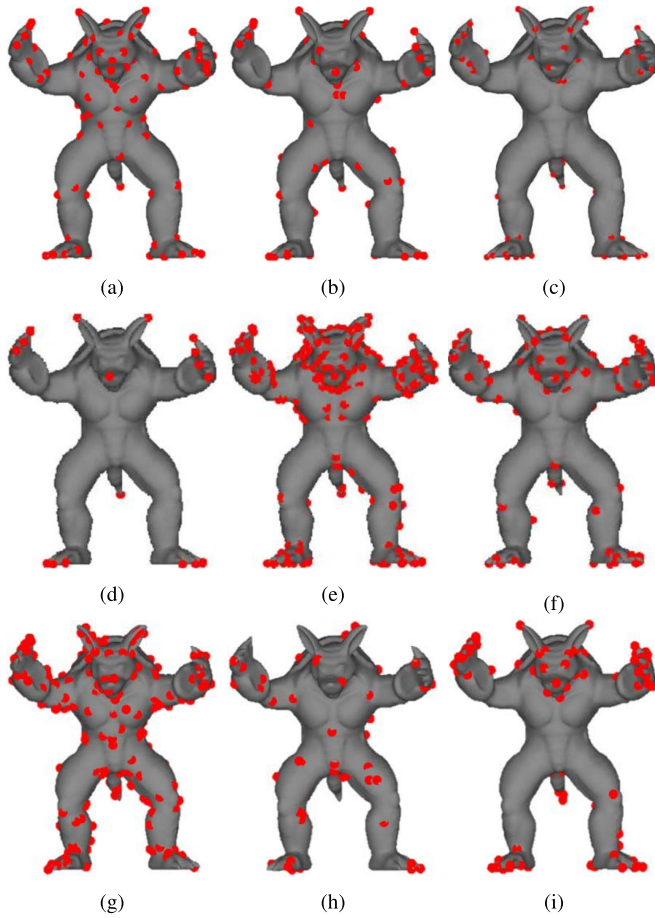


Fig. 7. Examples of the interest point detection result calculated by the proposed method and the other algorithms. Note that (a) is presented based on  $\sigma = 0.03$  and  $n = 2$ . (a) Ground truth. (b) Our results. (c) Cluster saliency [25]. (d) HKS [73]. (e) Mesh saliency [6]. (f) Salient points [49]. (g) Scale dependent [75]. (h) 3D-SIFT [74]. (i) 3D Harris [72].

selection of best viewpoint which can maximally elucidate the most important features of the model. Recently, many papers have addressed this problem but most of these algorithms cannot handle the point clouds. So we apply the proposed method to help automatic viewpoint selection of the point cloud.

For a given point cloud, viewpoint selection is to automatically determine the most informative viewpoint that can best describe the point cloud [12]. So the aim of viewpoint selection is to find a viewpoint that can maximize the summed saliency values for visible points on the point cloud. We will describe the whole process of viewpoint selection thereafter.

Initially, candidate viewpoints are generated by uniformly sampling a sphere that bounds the point cloud model [79]. Following [12], the radius is set to be twice the length of the tight bounding sphere. The number of the candidate viewpoints for a point cloud model is about 2600 viewpoints. For a point cloud model  $P$ ,  $S$  represents the saliency detection result for each point in the model  $P$ . Given a viewpoint  $v$ , let  $F(v)$  represents the point sets that are visible from viewpoint  $v$ . Following [6], we define the summed saliency value from viewpoint  $v$  as:

$$U(v) = \sum_{p_i \in F(v)} S(p_i) \quad (17)$$

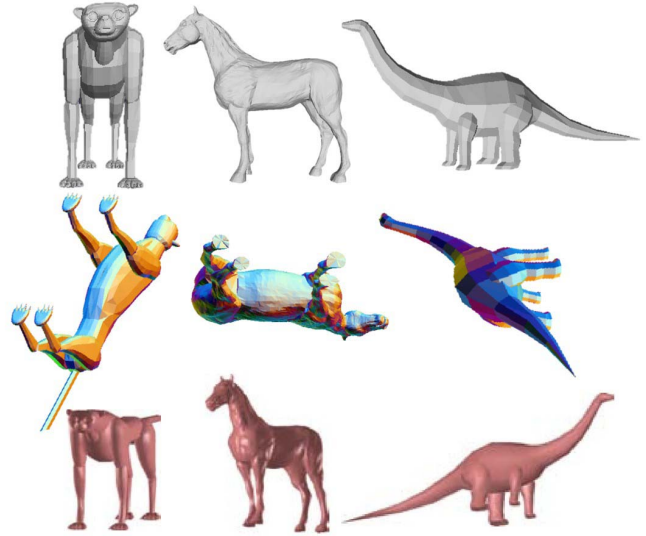


Fig. 8. Comparison of the viewpoint selection results. From top to bottom, the results from the proposed method, the results reported in [80] and the results reported in [12].

where  $U(v)$  represents the summed saliency value for viewpoint  $v$  and  $p_i$  denotes a point visible from viewpoint  $v$ .

For each candidate viewpoint, the  $U(v)$  is calculated. Then we find the best viewpoint  $v_p = \arg \max_v U(v)$ . Note that we do not perform gradient-descent method to find local maxima as described in [6] because the number of the candidate viewpoints is quite sufficient. Fig. 8 presents the comparison between the viewpoint selection results based on the proposed method, the semantic-driven viewpoint selection results reported in [80] and the viewpoint selection results based on surface region of interest reported in [12].

According to Fig. 8, the proposed method detects the face of the dog as the most salient region on the model, so the selected viewpoint mainly focuses on the facial region. Though our selected viewpoint fails to show the tale of the model when compared with the results reported in [80] and [12], it is very likely to be adopted by the human visual system since human is likely to be attracted by the face of the animals. As for the horse model, the viewpoints selected by the proposed algorithm and [12] are more helpful for recognizing the model since both of these two algorithms provide a better description of the model's outline. For the dinosaur model, the viewpoints selected by the three algorithms are quite comparable.

### E. Mesh Simplification

High detailed models present more realistic-looking scenes and are commonly appeared in games and videos. However, they usually require bigger storage space and increase the complexity of the computation [81]. Since simplifying high detailed 3D models is of vital importance to computer graphics field, we apply the proposed method to help mesh simplification.

Saliency detection result identifies the visually salient regions on the 3D model, so it can guide the mesh simplification process to help preserve the important details.



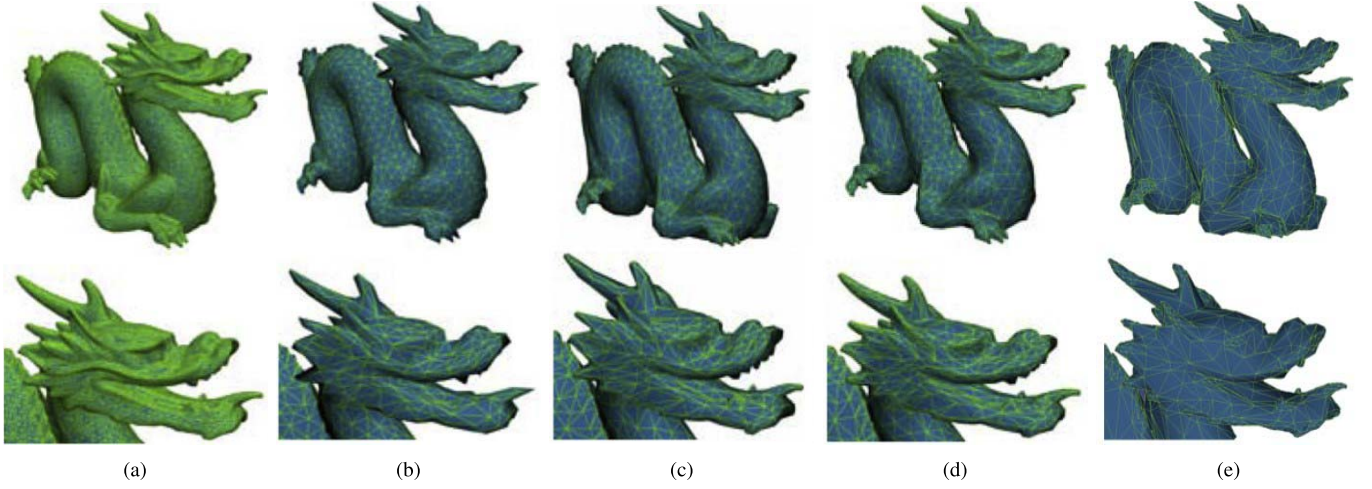


Fig. 9. Comparison of the simplification results. The left four columns are reported in [50]. From left to right: the dragon model (30k vertices), the simplified result generated using the QEM [82] algorithm (2500 vertices), the simplified results guided by [6] (2500 vertices) and [50] (2500 vertices), the simplified result using the proposed method (2500 vertices). (a) Original model. (b) Qslim [82]. (c) Mesh saliency [6]. (d) Global rarity [50]. (e) Our result.

Many mesh saliency detection algorithms have been applied to help mesh simplification. For example, Limper *et al.* [83] presented a mesh saliency detection algorithm using the local curvature entropy and applied their algorithm to help mesh simplification. In [11], a novel algorithm is proposed which uses spectral attributes of the mesh data and then uses the detected salient regions to guide the simplification process. Zhao *et al.* [14] also proposed a simplification algorithm based on saliency detection which assigns more points to the salient regions. Although the papers mentioned above all applied mesh saliency estimation to help mesh simplification and achieved quite good results, their codes are seldom available and the models they used are quite different from each other, making it hard to make comparison of the mesh simplification results between our algorithm and their methods. To evaluate the effectiveness of our method, we modified the quadric-based simplification method (QEM) proposed by Garland and Heckbert [82] by weighting the quadrics with the detected saliency values. The QEM method simplifies a mesh by repeatedly contracting vertex pairs with the least quadric error [6]. The whole process of mesh simplification is described below.

Given a vertex  $v = [v_x, v_y, v_z, 1]^T$  on the mesh model  $M$ , let  $F$  denotes the set of triangle planes incident at the vertex  $v$  and  $S$  be the saliency detection results for mesh model  $M$ . For a plane  $f \in F$ , which is defined by the equation  $ax + by + cz + d = 0$  with  $a^2 + b^2 + c^2 = 1$ , the plane  $f$  can be described as  $f = [a, b, c, d]^T$ . Following the method in [6], the saliency result  $S$  of the mesh model  $M$  is used to guide the simplification contraction process. The error of the vertex  $v$  is defined as:

$$\Delta(v) = \Delta([v_x, v_y, v_z, 1]^T) = W(v) \sum_{f \in F} (f^T v)^2, \quad (18)$$

where  $W(v)$  is the amplified saliency value  $S(v)$  at the vertex  $v$  with the amplification operator specified as:

$$W(v) = \begin{cases} \lambda S(v) & \text{if } S(v) \geq \alpha \\ S(v) & \text{otherwise} \end{cases} \quad (19)$$

where  $\lambda$  is the amplifying parameter and  $\alpha$  is the threshold controlling the range of saliency points to be amplified. Note that in this paper, the  $\lambda$  is set to 100 and  $\alpha$  is set to the top 30% saliency values as in paper [6].

For each vertex in the model, we compute its error using the formula in Eq.(18). Then we select point pairs at initialization stage for contraction if the two points  $(v_1, v_2)$  is an edge. During the contraction of each pair, the cost of contraction needs to be considered. For a given contraction  $(v_1, v_2) \rightarrow \bar{v}$ , we define the contraction error as the summation of the errors at vertex  $v_1$  and vertex  $v_2$ . To perform this contraction, a position for  $\bar{v}$  which minimizes the contraction error needs to be found. By solving  $\partial \Delta / \partial x = \partial \Delta / \partial y = \partial \Delta / \partial z = 0$ , the optimal position for  $\bar{v}$  can be obtained.

After the optimal contraction target  $\bar{v}$  for each pair is calculated as well as the cost of the contraction, all the pairs are sorted in ascending order using the cost of the contraction values and the pair  $(v_1, v_2)$  with the least cost is contracted. Then the cost of all pairs involving vertex  $v_1$  is updated and the whole contraction process is repeated until the number of the remaining vertices satisfies the need. Fig. 9 presents the comparison of the simplification results using different methods. From the first row, it can be seen that the simplified results guided by saliency (the third, fourth and fifth column) preserve better details on the face and the feet of the dragon when compared with the QEM [82] method. More specifically, our method preserves better details on the face and the feet with the least number of vertices distributed on the body of the dragon. According to the second row, the proposed method preserves more vertices on the eyes, mouth and horn of the dragon when compared with mesh saliency algorithm proposed in [6] and achieves quite comparable result with [50] without using the topological structure of the model.

## V. CONCLUSION

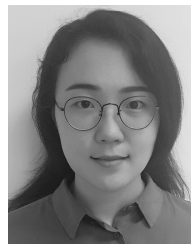
In this paper, a novel method has been proposed for detecting point clouds saliency by jointly utilizing the local

distinctness and global rarity cues which are supported by psychological evidences. Comparing with other point cloud saliency detection algorithms, the main contribution is that we take the local geometric features of each point into consideration when using random walk ranking method to introduce cluster-level global rarity refinement to each point in all clusters. Due to the noise in the point cloud, supervoxel segmentation might be inaccurate. The random walk ranking method is independent of supervoxel segmentation so it can help to reduce the impact of imprecise segmentation towards global rarity calculation. Since the local geometric features represent the relationships between neighboring points, it can be used as a guidance in the random walk process to generate more delicate global rarity result. Moreover, we have also proposed an adaptive optimization framework which can effectively combine different saliency cues by taking the inherent relationships of different saliency cues into consideration and achieve state-of-the-art saliency detection results. To demonstrate the utility, we have applied it to three computer graphics applications including: interest point detection, viewpoint selection and mesh simplification. The experiments demonstrated our performance.

## REFERENCES

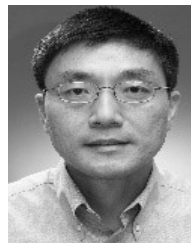
- [1] E. Shtrom, G. Leifman, and A. Tal, "Saliency detection in large point sets," in *Proc. ICCV*, 2013, pp. 3591–3598.
- [2] G. Pang and U. Neumann, "3D point cloud object detection with multi-view convolutional neural network," in *Proc. ICPR*, 2016, pp. 585–590.
- [3] U. Shafiq, M. Taj, and M. Ali, "More for less: Insights into convolutional nets for 3D point cloud recognition," in *Proc. ICIP*, 2017, pp. 1607–1611.
- [4] W. Cheng, W. Lin, X. Zhang, M. Goesele, and M.-T. Sun, "A data-driven point cloud simplification framework for city-scale image-based localization," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 262–275, Jan. 2017.
- [5] J. Ma, J. Zhao, and A. L. Yuille, "Non-rigid point set registration by preserving global and local structures," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 53–64, Jan. 2016.
- [6] C. H. Lee, A. Varshney, and D. W. Jacobs, "Mesh saliency," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 659–666, Jul. 2005.
- [7] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," in *Matters of Intelligence. Synthese Library* (Studies in Epistemology, Logic, Methodology, and Philosophy of Science), vol. 188, L. M. Vaina, Ed. Dordrecht, The Netherlands: Springer, 1987.
- [8] F. Huang, J. Qi, H. Lu, L. Zhang, and X. Ruan, "Salient object detection via multiple instance learning," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1911–1922, Apr. 2017.
- [9] W. Wang, J. Shen, and L. Shao, "Video salient object detection via fully convolutional networks," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 38–49, Jan. 2018.
- [10] S. Jeong and J.-Y. Sim, "Saliency detection for 3D surface geometry using semi-regular meshes," *IEEE Trans. Multimedia*, vol. 19, no. 12, pp. 2692–2705, Dec. 2017.
- [11] R. Song, Y. Liu, R. Martin, and P. Rosin, "Mesh saliency via spectral processing," *ACM Trans. Graph.*, vol. 33, no. 1, p. 6, 2014.
- [12] G. Leifman, E. Shtrom, and A. Tal, "Surface regions of interest for viewpoint selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 12, pp. 2544–2556, Dec. 2016.
- [13] G. Elbaz, T. Avraham, and A. Fischer, "3D point cloud registration for localization using a deep neural network auto-encoder," in *Proc. CVPR*, 2017, pp. 4631–4640.
- [14] Y. Zhao, Y. Liu, R. Song, and M. Zhang, "A saliency detection based method for 3D surface simplification," in *Proc. ICASSP*, 2012, pp. 889–892.
- [15] X. Jiao, T. Wu, and X. Qin, "Mesh segmentation by combining mesh saliency with spectral clustering," *J. Comput. Appl. Math.*, vol. 329, pp. 134–146, Feb. 2018.
- [16] S. Jia, C. Zhang, X. Li, and Y. Zhou, "Mesh resizing based on hierarchical saliency detection," *Graph. Models*, vol. 76, no. 5, pp. 355–362, 2014.
- [17] R. Gal and D. Cohen-Or, "Salient geometric features for partial shape matching and similarity," *ACM Trans. Graph.*, vol. 25, no. 1, pp. 130–150, Jan. 2006.
- [18] K. Grill-Spector and R. Malach, "The human visual cortex," *Annu. Rev. Neurosci.*, vol. 27, pp. 649–677, Jul. 2004.
- [19] J. M. Wolfe, "Guided search 2.0 a revised model of visual search," *Psychonomic Bull. Rev.*, vol. 1, no. 2, pp. 202–238, 1994.
- [20] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015.
- [21] P. Wang, Z. Zhou, W. Liu, and H. Qiao, "Salient region detection based on local and global saliency," in *Proc. ICRA*, 2014, pp. 1546–1551.
- [22] A. Borji and L. Itti, "Exploiting local and global patch rarities for saliency detection," in *Proc. CVPR*, 2012, pp. 478–485.
- [23] R. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proc. ICRA*, 2009, pp. 3212–3217.
- [24] J. Papon, A. Abramov, M. Schoeler, and F. Worgatter, "Voxel cloud connectivity segmentation—Supervoxels for point clouds," in *Proc. CVPR*, 2013, pp. 2027–2034.
- [25] F. P. Tasse, J. Kosinka, and N. Dodgson, "Cluster-based point set saliency," in *Proc. ICCV*, 2015, pp. 163–171.
- [26] S. Frintrop, E. Rome, and H. I. Christensen, "Computational visual attention systems and their cognitive foundations: A survey," *ACM Trans. Appl. Perception*, vol. 7, no. 1, p. 6, 2010.
- [27] B. Ni *et al.*, "Touch saliency: Characteristics and prediction," *IEEE Trans. Multimedia*, vol. 16, no. 6, pp. 1779–1791, Oct. 2014.
- [28] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior," in *Proc. BMVC*, 2011, pp. 1–12.
- [29] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. NIPS*, 2007, pp. 1–8.
- [30] C. Kanan, M. H. Tong, L. Zhang, and G. W. Cottrell, "SUN: Top-down saliency using natural statistics," *Vis. Cognit.*, vol. 17, nos. 6–7, pp. 979–1003, 2009.
- [31] H. Tian, Y. Fang, Y. Zhao, W. Lin, R. Ni, and Z. Zhu, "Salient region detection by fusing bottom-up and top-down features extracted from a single image," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4389–4398, Oct. 2014.
- [32] R. Cong, J. Lei, H. Fu, M. Cheng, W. Lin, and Q. Huang, "Review of visual saliency detection with comprehensive information," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.
- [33] L. Zhou, Z. Yang, Z. Zhou, and D. Hu, "Salient region detection using diffusion process on a two-layer sparse graph," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5882–5894, Dec. 2017.
- [34] K. Ishikura, N. Kurita, D. M. Chandler, and G. Ohashi, "Saliency detection based on multiscale extrema of local perceptual color differences," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 703–717, Feb. 2018.
- [35] C. Chen, S. Li, Y. Wang, H. Qin, and A. Hao, "Video saliency detection via spatial-temporal fusion and low-rank coherency diffusion," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3156–3170, Jul. 2017.
- [36] W. Wang, J. Shen, and L. Shao, "Consistent video saliency using local gradient flow optimization and global refinement," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4185–4196, Nov. 2015.
- [37] R. Cong, J. Lei, C. Zhang, Q. Huang, X. Cao, and C. Hou, "Saliency detection for stereoscopic images based on depth confidence analysis and multiple cues fusion," *IEEE Signal Process. Lett.*, vol. 23, no. 6, pp. 819–823, Apr. 2016.
- [38] Y. Niu, Y. Geng, X. Li, and F. Liu, "Leveraging stereopsis for saliency analysis," in *Proc. CVPR*, 2012, pp. 454–461.
- [39] F. Guo, J. Shen, and X. Li, "Learning to detect stereo saliency," in *Proc. ICME*, 2014, pp. 1–6.
- [40] V. Sitzmann *et al.*, "Saliency in VR: How do people explore virtual environments?" *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 4, pp. 1633–1642, Apr. 2018.
- [41] Y. Zhu, G. Zhai, and X. Min, "The prediction of head and eye movement for 360 degree images," *Signal Process., Image Commun.*, vol. 69, pp. 15–25, Nov. 2018.
- [42] H. Hu, Y.-C. Lin, M. Liu, H. Cheng, Y. Chang, and M. Sun, "Deep 360 pilot: Learning a deep agent for piloting through 360° sports video," in *Proc. CVPR*, 2017, pp. 1396–1405.
- [43] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle, "Surface reconstruction from unorganized points," *ACM SIGGRAPH Comput. Graph.*, vol. 26, no. 2, pp. 71–78, 1992.

- [44] N. Mitra and A. Nguyen, "Estimating surface normals in noisy point cloud data," in *Proc. SoCG*, 2003, pp. 322–328.
- [45] H. Yee, S. Pattanaik, and D. P. Greenberg, "Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments," *ACM Trans. Graph.*, vol. 20, no. 1, pp. 39–65, Jan. 2001.
- [46] S. Frintrop, A. Nüchter, and H. Surmann, "Visual attention for object recognition in spatial 3D data," in *Proc. WAPCV*, 2004, pp. 168–182.
- [47] S. Howlett, J. Hamill, and C. O'Sullivan, "An experimental approach to predicting saliency for simplified polygonal models," in *Proc. APGV*, 2004, pp. 57–64.
- [48] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
- [49] U. Castellani, M. Cristani, S. Fantoni, and V. Murino, "Sparse points matching by combining 3D mesh saliency with statistical descriptors," *Comput. Graph. Forum*, vol. 27, no. 2, pp. 643–652, Apr. 2008.
- [50] J. Wu, X. Shen, W. Zhu, and L. Liu, "Mesh saliency with global rarity," *Graph. Models*, vol. 75, no. 5, pp. 255–264, 2013.
- [51] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, Oct. 2012.
- [52] A. Maximo, R. Patro, A. Varshney, and R. Farias, "A robust and rotationally invariant local surface descriptor with applications to non-local mesh processing," *Graph. Models*, vol. 73, no. 5, pp. 231–242, 2011.
- [53] F. Tombari, S. Salti, and L. Di Stefano, "Unique signatures of histograms for local surface description," in *Proc. ECCV*, 2010, pp. 356–369.
- [54] A. Borji, "Boosting bottom-up and top-down visual features for saliency estimation," in *Proc. CVPR*, 2012, pp. 438–445.
- [55] M. Stark and B. Schiele, "How good are local features for classes of geometric objects," in *Proc. ICCV*, 2007, pp. 1–8.
- [56] Y. Yuan, C. Li, J. Kim, W. Cai, and D. D. Feng, "Reversion correction and regularized random walk ranking for saliency detection," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1311–1322, Mar. 2018.
- [57] J. Krause, H. Jin, J. Yang, and L. Fei-Fei, "Fine-grained recognition without part annotations," in *Proc. CVPR*, 2015, pp. 5546–5555.
- [58] Y. Peng, X. He, and J. Zhao, "Object-part attention model for fine-grained image classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1487–1500, Mar. 2018.
- [59] Y.-K. Lai, S.-M. Hu, R. R. Martin, and P. L. Rosin, "Fast mesh segmentation using random walks," in *Proc. SPM*, 2008, pp. 183–191.
- [60] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.
- [61] C. Li, Y. Yuan, W. Cai, Y. Xia, and D. D. Feng, "Robust saliency detection via regularized random walks ranking," in *Proc. CVPR*, 2015, pp. 2710–2717.
- [62] Y. Guo, F. Wang, and J. Xin, "Point-wise saliency detection on 3D point clouds via covariance descriptors," *Vis. Comput.*, vol. 34, no. 10, pp. 1325–1338, 2010.
- [63] P. Tao, J. Cao, S. Li, X. Liu, and L. Liu, "Mesh saliency via ranking unsalient patches in a descriptor space," *Comput. Graph.*, vol. 46, pp. 264–274, Feb. 2015.
- [64] X. Chen, A. Saparov, B. Pang, and T. Funkhouser, "Schelling points on 3D surface meshes," *ACM Trans. Graph.*, vol. 31, no. 4, p. 29, 2012.
- [65] B. Han, L. Tcheang, V. Walsh, and X. Gao, "A novel feature combination methods for saliency-based visual attention," in *Proc. ICNC*, 2009, pp. 18–22.
- [66] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. CVPR*, 2014, pp. 2814–2821.
- [67] Y. Yeshurun, R. Kimchi, G. Sha'shoua, and T. Carmel, "Perceptual objects capture attention," *Vis. Res.*, vol. 49, no. 10, pp. 1329–1335, 2009.
- [68] J. Yun and J.-Y. Sim, "Supervoxel-based saliency detection for large-scale colored 3D point clouds," in *Proc. ICIP*, 2016, pp. 4062–4066.
- [69] A. Nouri, C. Charrier, and O. Lezoray, "Global visual saliency: Geometric and colorimetric saliency fusion and its applications for 3D colored meshes," in *Proc. IPTA*, 2017, pp. 1–6.
- [70] M. Lau, K. Dev, W. Shi, J. Dorsey, and H. Rushmeier, "Tactile mesh saliency," *ACM Trans. Graph.*, vol. 35, no. 4, p. 52, 2016.
- [71] R. Song, Y. Liu, R. R. Martin, and K. R. Echavarria, "Local-to-global mesh saliency," *Vis. Comput.*, vol. 34, no. 3, pp. 323–336, 2018.
- [72] I. Sipiran and B. Bustos, "Harris 3D: A robust extension of the harris operator for interest point detection on 3D meshes," *Vis. Comput.*, vol. 27, no. 11, pp. 963–976, 2011.
- [73] J. Sun, M. Ovsjanikov, and L. Guibas, "A concise and provably informative multi-scale signature based on heat diffusion," *Comput. Graph. Forum*, vol. 28, no. 5, pp. 1383–1392, 2010.
- [74] A. Godil and A. I. Wagan, "Salient local 3D features for 3D shape retrieval," *Proc. SPIE*, vol. 7864, Jan. 2011, Art. no. 78640S.
- [75] J. Novatnack and K. Nishino, "Scale-dependent 3D geometric features," in *Proc. ICCV*, 2007, pp. 1–8.
- [76] H. Dutagaci, C. P. Cheung, and A. Godil, "Evaluation of 3D interest point detection techniques via human-generated ground truth," *Vis. Comput.*, vol. 28, no. 9, pp. 901–917, 2012.
- [77] X. Chen, A. Golovinskiy, and T. Funkhouser, "A benchmark for 3D mesh segmentation," *ACM Trans. Graph.*, vol. 28, no. 3, p. 73, 2009.
- [78] F. Bogo, J. Romero, M. Loper, and M. J. Black, "FAUST: Dataset and evaluation for 3D mesh registration," in *Proc. CVPR*, 2014, pp. 3794–3801.
- [79] J. C. Mitchell, "Sampling rotation groups by successive orthogonal images," *J. Sci. Comput.*, vol. 30, no. 1, pp. 525–547, 2008.
- [80] H. Laga, "Semantics-driven approach for automatic selection of best views of 3D shapes," in *Proc. 3DOR Workshop*, 2010, pp. 15–22.
- [81] M. Franc and V. Skala, "Fast algorithm for triangular mesh simplification based on vertex decimation," in *Proc. ICCS*, 2002, pp. 42–51.
- [82] M. Garland and P. S. Heckbert, "Surface simplification using quadric error metrics," in *Proc. SIGGRAPH*, 1997, pp. 1–8.
- [83] M. Limper, A. Kuijper, and D. Fellner, "Mesh saliency analysis via local curvature entropy," in *Proc. EUROGRAPHICS*, 2016, pp. 1–4.



**Xiaoying Ding** received the B.E. degree in remote sensing science and technology from Wuhan University, China, in 2016, where she is currently pursuing the Ph.D. degree.

From 2017 to 2018, she was a Visiting Student with the School of Computer Science and Engineering, Nanyang Technological University. Her research interests include computational vision, image processing, and 3D point cloud saliency detection. She received the National Scholarship.



**Weisi Lin** (M'92–SM'98–F'16) received the Ph.D. degree from King's College London, U.K. He is currently a Professor with the School of Computer Science and Engineering, Nanyang Technological University. His research interests include image processing, perceptual signal modeling, video compression, and multimedia communication, in which he has published over 200 journal papers, over 230 conference papers, filed seven patents, and authored two books. He is a fellow of the IET and an Honorary Fellow of the Singapore Institute of Engineering Technologists. He was the Technical Program Chair of the IEEE ICME 2013, PCM 2012, QoMEX 2014, and the IEEE VCIP 2017. He has been an Invited/Panelist/Keynote/Tutorial Speaker at over 20 international conferences and was a Distinguished Lecturer of the IEEE Circuits and Systems Society from 2016 to 2017 and the Asia-Pacific Signal and Information Processing Association (APSIPA) from 2012 to 2013. He has been an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, and the IEEE SIGNAL PROCESSING LETTERS.





**Zhenzhong Chen** (S'02–M'07–SM'15) received the B.Eng. degree in electrical engineering from the Huazhong University of Science and Technology and the Ph.D. degree in electrical engineering from The Chinese University of Hong Kong. He is currently a Professor with Wuhan University (WHU). His current research interests include image and video processing, computer vision, multimedia data mining, human–computer interaction, photogrammetry, and remote sensing. He has been selected for the Thousand Talents Plan for Young Professionals and

was a recipient of the CUHK Young Scholars Dissertation Award, the CUHK Faculty of Engineering Outstanding Ph.D. Thesis Award, and a Microsoft Fellowship. He has been a VQEG Board Member, an Immersive Media Working Group Co-Chair, a Selection Committee Member of ITU Young Innovators Challenges, an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (TCSVT), the *Journal of the Association for Information Science and Technology* (JASIST), and the *Journal of Visual Communication and Image Representation*, and an Editor of the *IEEE IoT Newsletter*.



**Xinfeng Zhang** (M'16) received the B.S. degree in computer science from the Hebei University of Technology, Tianjin, China, in 2007, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2014. From 2014 to 2017, he was a Research Fellow with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore. From 2017 to 2018, he was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, USA. From 2018 to 2019, he was a Research Fellow with the Department of Computer Science, City University of Hong Kong. He is currently an Assistant Professor with the School of Computer Science and Technology, University of Chinese Academy of Sciences. He has submitted over 20 technical proposals to ISO/MPEG, ITU-T, and AVS standards and authored more than 100 refereed journal/conference papers. His research interests include video compression, image/video quality assessment, and image/video analysis. He received the Best Paper Award of the IEEE Multimedia 2018, the Best Paper Award at the 2017 Pacific-Rim Conference on Multimedia (PCM), and is the coauthor of a paper that received the Best Student Paper Award at the IEEE International Conference on Image Processing 2018.