

CONDITIONAL RANDOM FIELD-BASED MESH SALIENCY

Ran Song*, Yonghuai Liu, Yitian Zhao

Department of Computer Science,
Aberystwyth University, UK

Ralph R. Martin, Paul L. Rosin

School of Computer Science & Informatics,
Cardiff University, UK

ABSTRACT

We propose a new method for detecting mesh saliency, a reflection of perception-based regional importance for 3D meshes. The basic idea is to incorporate the Conditional Random Field (CRF) framework with a saliency detection process. We first produce a multi-scale representation for a mesh. Then, a CRF is designed to robustly detect salient regions utilising neighbourhood consistency. By inferring the CRF via belief propagation algorithm, we actually make use of the global statistic information in the saliency detection process. Experimental results demonstrate the robustness and the effectiveness of the proposed method.

Index Terms— Saliency, CRF, Mesh Simplification

1. INTRODUCTION

Mesh saliency is a measure that captures the property of a point in a 3D mesh based on human perception rather than local geometry of shape. Saliency can efficiently and effectively reflect perceptual importance of regions for a 3D mesh while curvature is not able to measure. Therefore, mesh saliency has a wide range of applications in the fields of computer vision and graphics, such as mesh simplification, scene rendering, view point selection, point cloud matching, compression and object recognition.

The literatures related to 3D mesh saliency are largely inspired by the correspondent work performed on 2D images [1]. Particularly, the concept of *scale space* has been successfully extended to the 3D domain. We also notice that saliency is a flexible concept that can be defined in line with various tasks. M. Pauly *et al* [2] proposed a multi-scale method to extract line-type features where they introduced a saliency measure of *surface variation* by combining the eigenvalues of the local covariance matrix. In [3], salient geometric features based on curvatures were introduced to improve part-in-whole matching. [4] developed a technique to effectively select 3D shape descriptors of high saliency. [5] proposed a method for detecting and matching salient points from multi-view meshes where the saliency is also estimated by generating multi-scale representation for a mesh in a scale space. The

state-of-the-art method [6] computed mesh saliency using a center-surround operator on Gaussian-weighted curvatures:

$$G(\mathcal{C}(v), \sigma) = \frac{\sum_{x \in N(v, 2\sigma)} \mathcal{C}(x) \exp[-\|x-v\|^2 / (2\sigma^2)]}{\sum_{x \in N(v, 2\sigma)} \exp[-\|x-v\|^2 / (2\sigma^2)]} \quad (1)$$

where $\mathcal{C}(v)$ and $N(v, 2\sigma)$ denote the mean curvature and a neighbourhood region of the vertex v respectively.

Usually, previous methods output a single-saliency map by simply computing the sum or the average of all multi-scale saliency maps to simplify the information and then use thresholding-based methods (often not robust to noise) to determine whether a point is salient or not. These methods are fast but do not make good use of the information embedded in the multi-scale saliency maps. We propose a new method for a more robust saliency detection. Instead of simply combining the multi-scale information, we incorporate such information into a Conditional Random Field (CRF) framework. The detected salient/non-salient locations are the labeling result of the CRF which incorporates information not only associated within a local neighbourhood but also throughout the mesh at all scales. The proposed method involves two stages: multi-scale mesh representation and CRF-based saliency detection.

2. MULTI-SCALE MESH REPRESENTATION

We apply a set of S Gaussian filters on a mesh M to produce a multi-scale representation D_s for M . $G(p, \sigma)$ is a Gaussian kernel with standard deviation σ centred at the vertex $p \in M$. Each Gaussian kernel is applied over a neighbourhood region centred at p with a geodesic radius r . The geodesic distance between two points is computed using the method proposed in [7]. All vertices within this region are viewed as the neighbours of p and involved in the convolution. Compared to the neighbourhood defined in [6] (Eq. 1) using the Euclidean measure, the advantage of using the geodesic measure is the reduction of potential ambiguity. In Fig. 1(a), only surface patch 1 is considered as the neighbourhood of the point P under the geodesic measure, while patch 2 is an ambiguous neighbourhood if the Euclidean measure is used. We set $r = 2.4 * F * \sigma$ in line with the principle of *full width at half maximum* where F is a normalisation parameter related to the scanning resolution R (average interpoint distance) of

*Thanks to the RIVIC project supported by HEFCW/WAG for funding.

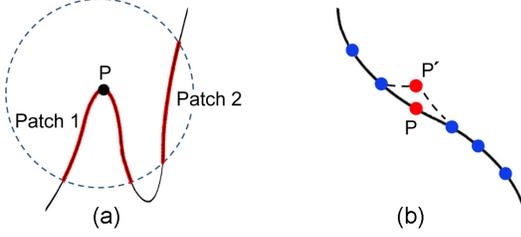


Fig. 1. (a) Geodesic vs Euclidean. Patch 1 is the neighbourhood of P defined by geodesic measure; Patch 2 is an ambiguous neighbourhood produced by Euclidean measure. Ambiguous neighbourhood usually leads to ambiguous saliency. (b) A perturbation can only change P 's position while the curvatures of P and its neighbours are all changed.

the mesh M and we choose $F = 2R$. We propose a new algorithm to do a rank-based Gaussian filtering:

- (1) For a point p , find all of its k_p neighbours (including itself) within a distance equal to r from all of the points in M .
- (2) Sort the k_p neighbours in the descending order of the distance to p . So the first point in the neighbourhood sequence v_p is p itself and the last one is the point furthest from p .
- (3) Construct a discrete Gaussian kernel with standard deviation σ sampled as a k_p -dimensional vector.
- (4) Sort the elements in this Gaussian kernel in descending order, yielding an a Gaussian kernel G_p . Thus in the following convolution, nearer neighbours have larger weights.
- (5) Do convolution using v_p and G_p .
- (6) Repeat the steps listed above for all points in M .

The Difference-of-Gaussians (DoG) scale space is constructed by computing the difference of two Gaussians at scale s :

$$D_s(p) = G_p(p, \sigma_s) - G_p(p, \eta\sigma_s), \quad s = 1, 2, \dots, S \quad (2)$$

where G_p denotes the Gaussian applied to the point p . η is set as 1.6, which makes the DoG a good approximation of the Laplacian of Gaussian (LoG). By 'approximation', we mean $DoG(x)/LoG(x) \approx \text{constant}$ or the DoG is approximately equal to the scale-normalised LoG which can achieve true scale invariance. We use four scales ($S = 4$) of filtering with $\sigma_s \in \{0.6, 1.2, 1.8, 2.4\}$ for a reasonable balance between reliability of saliency detection and computational cost.

$D_s(p)$ is a 3D vector representing the displacement of the point p from its original position in M after the filtering. We define the multi-scale representation of a point on its Gaussian-weighted position rather than mean curvature as in [6] for higher robustness against scanning noise. Scanning noise can be viewed as a perturbation or a displacement added to a point to move it away from its correct position. As illustrated in Fig. 1(b), once the point P suffers from a perturbation, only its position has been changed while all its neighbours' 3D positions remain unchanged. In contrast, the curvature-based computation in Eq. (1) is less robust because the curvatures of P 's neighbouring points have

also been changed. Furthermore, instead of directly using the interpoint distances to calculate the weights in the Gaussian convolution in Eq. (1), we determine the weight of a neighbouring point by its distance-based rank among all neighbouring points. A small perturbation at one point caused by scanning noise is highly likely to change its distance to the centred point but less likely to change its distance-based rank among all neighbouring points. Therefore, in most cases, salient locations can still be correctly detected in the presence of scanning noise by using our method.

To reduce $D_s(p)$ in a scalar quantity, we project it onto the normal $n(p)$ at the point p to obtain the scale maps M_s . Then, we employ the method proposed in [1] to normalise the scale maps. The normalisation is designed to globally promote maps in which a small number of strong peaks are present, while globally suppressing maps which contain numerous comparable peaks. To further enhance the difference between potential salient and non-salient locations, we apply an arc-tangent operation to each scale map to produce the final multi-scale representation \hat{M}_s (see Fig. 2).

3. CRF-BASED SALIENCY DETECTION

Then we incorporate the multi-scale information of a mesh into a Conditional Random Field (CRF) framework. The reason that we apply a CRF rather than a simple summation over the multi-scale saliency maps employed by previous methods [6, 5] is to increase the robustness of our saliency detection method by introducing a consistency constraint.

We define a label assignment $\mathbf{s} = \{s_p, \forall p \in M\}$ and the label set comprises the scale indices $\{s\} = \{1, 2, 3, 4\}$. For a point p , each label corresponds to a scale saliency $\hat{M}_s(p)$ for the scale s . In line with the standard CRF formulation where the points/sites are written as subscripts and the labels assigned to the points are the real variables that we try to figure out, in the rest of this paper, we rewrite $\{\hat{M}_s(p)\}$ as $\{\hat{M}_p(s)\}$. If we use $\mathbf{s}' = \{s'_p, \forall p \in M\}$ to denote a known observation field, a CRF can be defined as:

$$Pr(\mathbf{s}|\mathbf{s}') = \frac{1}{Z(\mathbf{s}')} \exp \left\{ - \sum_{c \in C} (\lambda_c \cdot \psi_c(s_c | s'_c)) \right\} \quad (3)$$

where λ_c is a weighting parameter and $Z(\mathbf{s}')$ is a normalising constant. The factors ψ_c are potential functions of the random variables s_c within a clique $c \in C$. The Gibbs energy of a pairwise CRF is expressed as

$$\begin{aligned} E(\mathbf{s}|\mathbf{s}') &= -\ln Pr(\mathbf{s}|\mathbf{s}') - \ln Z(\mathbf{s}') = \sum_{c \in C} \lambda_c \cdot \psi_c(s_c | s'_c) \\ &= \sum_{p \in M} \psi_p(s_p | s'_p) + \lambda \sum_p \sum_{q \in \mathcal{N}_p(s)} \psi_{pq}(s_p, s_q | s'_p, s'_q) \end{aligned} \quad (4)$$

where $\mathcal{N}_p(s)$ denotes the neighbourhood of p at the scale s as used in Gaussian filtering but without p itself. To simplify notations, in the rest of this paper, we use shorthand $\psi_p(s_p)$ for $\psi_p(s_p | s'_p)$ and $\psi_{pq}(s_p, s_q)$ for $\psi_{pq}(s_p, s_q | s'_p, s'_q)$.

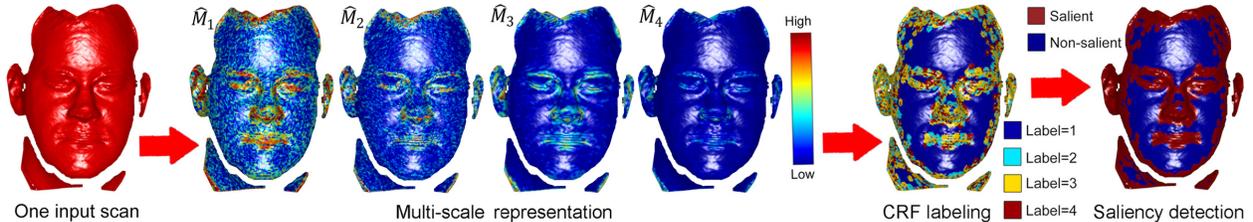


Fig. 2. The pipeline of our saliency detection method with the result of each stage.

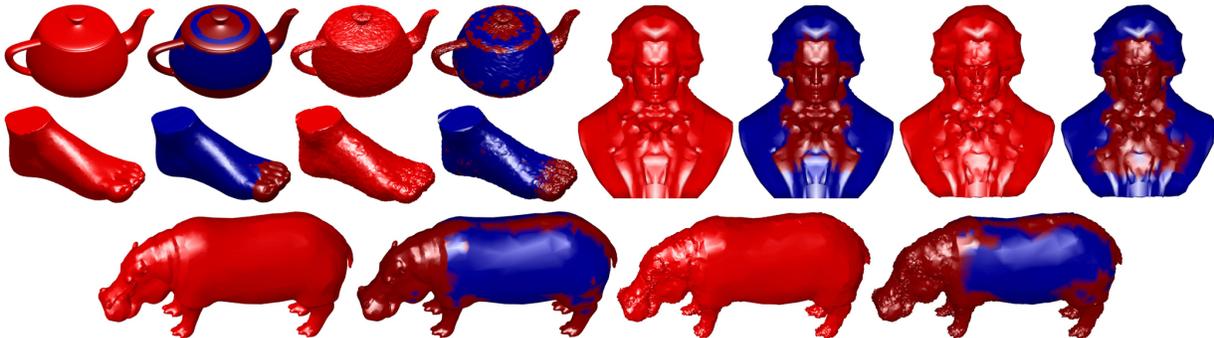


Fig. 3. Saliency detection results of complete 3D surface models with and without noise. Each quaternary subfigure consists of the original model, the saliency detection for the original model, the noisy model and its saliency detection result.

In Eq. (4), $\psi_p(s_p)$ is the data term associated to the state that we observe or most likely observe at point p , defined as:

$$\psi_p(s_p) = \left| \hat{M}_p(s_p) - \max_s \hat{M}_p(s) \right|, \quad s = 1, 2, 3, 4. \quad (5)$$

The compatibility term $\psi_{pq}(s_p, s_q)$ captures the consistency between two neighbouring points. It can be regularised by the general and scene-specific knowledge. For instance, the smoothness prior, essentially an intensity consistency applied to a neighbourhood, is widely used in 2D applications such as image segmentation, restoration and depth estimation. Generally, consistency constraints are based on the nature that neighbouring points are more likely to have the same properties (e.g., whether they are salient). Therefore, such constraints usually increase the robustness of labeling. Scanning noise, occlusions, outliers and unreliable triangulation can lead to unreliable saliency detection, turning a point in non-saliency region into a salient point. But the consistency constraint is highly likely to correct that by investigating its non-saliency neighbours. We use a consistency constraint which encourages two neighbouring points to take the same scale:

$$\psi_{pq}(s_p, s_q) = \begin{cases} 0, & \text{if } s_p = s_q, \\ \left| \hat{M}_p(s_p) - \hat{M}_q(s_q) \right|, & \text{otherwise.} \end{cases} \quad (6)$$

We solve this CRF in accordance with the *maximum a posteriori probability* criterion via belief propagation (BP). It requires the minimisation of the energy function in Eq. (4). The solution assigns a label to each point on the mesh M . Typically, most points are assigned the same principal scale as shown in Fig. 2, and these points comprise the non-saliency regions. All other points comprise the salient regions.

4. EXPERIMENTS

We tested our method using both range scans and complete 3D surface models. Fig. 2 takes a range scan of a human face as an example to show the pipeline of our CRF-based saliency detection method. It is consistent with human perception that the eyebrows, the eyes, the nose, the mouth, the ears and the contour of the face are detected as salient regions.

To demonstrate the robustness of our method, we added some Gaussian noise to the original data sources and then implemented our method using these noisy data. The results are shown in Fig. 3. It can be seen that our method can still distinguish salient and non-saliency regions in the presence of considerable amount of noise. For the *teapot*, the detected salient regions such as the lid, the nozzle and the handle remain intact. For the *Beethoven*, the facial region, as a whole, are detected as salient region despite the presence of noise. For the *foot*, the toes are robustly recognised as salient regions. For the *hippo*, the head, the limbs, the spine and the tail of the hippo are detected as salient regions even if some local details are significantly (e.g. the ears) corrupted by noise.

We demonstrate the effectiveness of our method through an application: mesh simplification. We have modified the QSlim method [9] by guiding the order of simplification contractions using a weight map derived from the saliency map. [6] reported that using the simplification weights based on a nonlinear amplification of the saliency gives good results where they simply amplified the saliency values greater than a threshold. In our method, the CRF has robustly partitioned a mesh into salient and non-saliency regions. So we just amplified the saliency values of the points in salient regions.

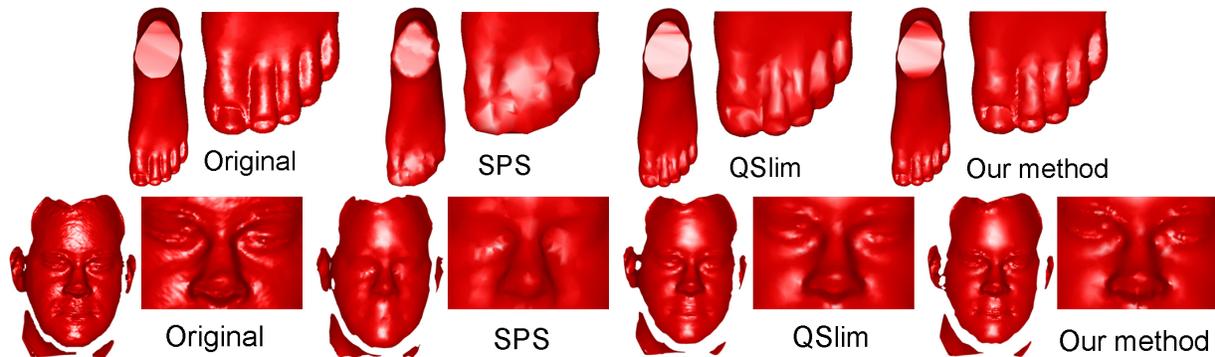


Fig. 4. Simplification by using SPS [8], QSlim [9] and our method with simplification rate 80% on *foot* and *face*.

Model	Simplification rate	RMSE		MESH	
		Lee	Our method	Lee	Our method
Beethoven	50%	0.0096	0.0081	0.0068	0.0056
	80%	0.0325	0.0311	0.0253	0.0242
	95%	0.1125	0.0964	0.0854	0.0748
Foot	50%	2.18×10^{-4}	1.81×10^{-4}	1.76×10^{-4}	1.42×10^{-4}
	80%	6.33×10^{-4}	5.92×10^{-4}	5.26×10^{-4}	4.92×10^{-4}
	95%	0.0022	0.0020	0.0018	0.0016
Hippo	50%	0.0033	0.0022	0.0017	0.0012
	80%	0.0115	0.0103	0.0087	0.0077
	95%	0.0589	0.0540	0.0616	0.0429
Pat-face	50%	0.0282	0.0278	0.0233	0.0214
	80%	0.0771	0.0766	0.0632	0.0630
	95%	0.2440	0.2327	0.1848	0.1789
Teapot	50%	3.62×10^{-5}	3.13×10^{-5}	3.10×10^{-5}	2.63×10^{-5}
	80%	1.12×10^{-4}	1.07×10^{-4}	9.55×10^{-5}	9.11×10^{-5}
	95%	3.92×10^{-4}	3.50×10^{-4}	3.16×10^{-4}	2.87×10^{-4}

Table 1. RMSE and MESH errors measured with different simplification rates using our method and Lee’s method [6].

We compared our method with the stratified point sampling (SPS) [8] and the QSlim. Fig. 4 presents the simplification results where the simplification rate is 80% (80% points are removed). It can be observed that local details are better preserved with our saliency-based simplification. In particular, it is consistent with our saliency detection results that the toes of the *foot* and the eyebrows, the eyes, the nose and the mouth of the *face* are significantly better preserved.

We also compare Lee’s method [6] with our method by computing the root mean square error (RMSE) and the MESH error [10] between the original mesh and the simplified mesh. As shown in Table. 1, if the simplification rate increases, the RMSE and MESH errors become larger for both methods as expected. Even though, the errors of our method are lower than those of Lee’s method, which means that the simplified meshes produced by our method are the better approximations of the original meshes than those produced by Lee’s method.

5. CONCLUSIONS

In this paper, we present a novel method for robust mesh saliency detection. The multi-scale representation of a mesh is derived via a rank-based Gaussian filtering algorithm where we use geodesic measure to define neighbourhood. The fol-

lowing saliency detection is based on a CRF labeling enforcing neighbourhood consistency. It is thus more robust than the methods based on saliency thresholding. As demonstrated by the experimental results, continuous surface regions are detected as salient regions, which largely complies with human perception on regional importance for 3D meshes.

6. REFERENCES

- [1] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *PAMI*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [2] M. Pauly, R. Keiser, and M. Gross, “Multi-scale feature extraction on point-sampled surfaces,” *Computer Graphics Forum*, vol. 22, no. 3, pp. 281–289, 2003.
- [3] R. Gal and D. Cohen-Or, “Salient geometric features for partial shape matching and similarity,” *ToG*, vol. 25, no. 1, pp. 130–150, 2006.
- [4] P. Shilane et al, “Selecting distinctive 3d shape descriptors for similarity retrieval,” in *Proc. SMI*, 2006.
- [5] U. Castellani et al, “Sparse points matching by combining 3d mesh saliency with statistical descriptors,” in *Proc. Eurographics*, 2008.
- [6] C.H. Lee, A. Varshney, and D.W. Jacobs, “Mesh saliency,” in *Proc. SIGGRAPH*, 2005.
- [7] Y. Sun and M.A. Abidi, “Surface matching by 3d point’s fingerprint,” in *Proc. ICCV*, 2001.
- [8] D. Nehab and P. Shilane, “Stratified point sampling of 3d models,” in *Eurographics symposium on point-based graphics*, 2004.
- [9] M. Garland et al, “Surface simplification using quadric error metrics,” in *Proc. SIGGRAPH*, 1997.
- [10] N. Aspert, D. Santa-Cruz, and T. Ebrahimi, “Mesh: Measuring errors between surfaces using the hausdorff distance,” in *Proc. ICME 2002*, 2002.