Multi-scale mesh saliency with local adaptive patches for viewpoint selection

Anass Nouri, Christophe Charrier, Olivier Lézoray

Multi-scale Mesh Saliency with local adaptive patches for ViewPoint Selection

Anass Nouri, Christophe Charrier, Olivier Lézoray

Normandie Université, UNICAEN, ENSICAEN, GREYC UMR CNRS 6072, Caen, FRANCE

Abstract

Our visual attention is attracted by specific areas into 3D objects (represented by meshes). This visual attention depends on the degree of saliency exposed by these areas. In this paper, we propose a novel multi-scale approach for detecting salient regions. To do so, we define a local surface descriptor based on patches of adaptive size and filled in with a local height field. The single-scale saliency of a vertex is defined as its degree measure in the mesh with edges weights computed from adaptive patch similarities weighted by the local curvature. Finally, the multi-scale saliency is defined as the average of single-scale saliencies weighted by their respective entropies. The contribution of the multi-scale aspect is analyzed and showed through the different results. The strength and the stability of our approach with respect to noise and simplification are also studied. Our approach is compared to the state-of-the-art and presents competitive results.

Keywords: Saliency, 3D Meshes, Graphs, patches.

1. Introduction

In every look thrown at a scene or an object, our attention is attracted by particular regions distinct from the surrounding zones. These striking areas, essentially prominent in the field of 3D objects, are content dependent. However, they are not dependent of the behavior or the experience relative to the human observer [1]. Therefore, saliency computation can allow detecting perceptually important points or regions on the surface of a 3D mesh. The saliency models proposed in the state-of-the-art are inspired from the HVS's (Human Visual System) low-level features. This allows to replace the geo-

metrical attributes used for the computation of saliency by perceptual ones, and as confirmed in [2], these perceptual models reach to model correctly the eye movement of the humain observer.

Many applications in 3D computer vision benefit from visual saliency, we can mention: optimal view point selection [3] where the objective is to generate the most informative viewpoints that capture a maximum of salient regions, and adaptive mesh simplification [4] that aims at maintaining the best perceived quality by performing simplification essentially on regions of low saliency. Likewise, shape matching [5], mesh resizing [6], and face recognition [7] can take advantage from saliency detection.

2. State-of-the-art

Unlike 2D images where saliency was amply dealt with (see [8] and references therein), there is few work on the evaluation of saliency directly on the geometry of the 3D meshes. For example, Lee *et al.* [3] were the first to propose a model based on the computation of curvature [9]. The saliency at a vertex is defined as the absolute difference between the gaussian weighted averages at both fine and coarse scales, respectively defined as σ and 2σ (the bandwidth of the Gaussian filtering scales).

Tal *et al.* [10] detect salient regions of interest of surfaces with the combined use of vertex distinctness and shape extremities. Vertex distinctness is computed from a similarity measure obtained with the diffusion distance between Spin Image 2D histograms [11]. Shape extremities are obtained from extreme geodesic distances on a MDS-transformed mesh.

In [12], Wu *et al.* detect salient regions with a descriptor measuring the local height field into the neighborhood of each vertex; a square map of projection heights [13] is generated to denote its form. Local and global saliencies are computed for each vertex. The final visual saliency of a vertex is computed by combining the global and local saliencies at different scales.

In [14], Zhao *et al.* proposed a saliency detection method based on a sampling for 3D mesh simplification. The approach begins by applying a Gaussian filter to the vertices. Then, parameters representing the mean curvature and the principal curvatures at different scales are computed. Finally, the different maps are filtered by a median filter before being combined to produce the final saliency map.

Acting on the same principle, the same authors Zhao *et al.* proposed in [15] a saliency model based on the index surface diffusion, likewise computed in [14], but with a non-local filter [16]. This model was integrated in applications for mesh registration and mesh simplification.

In [17], Zhao and Liu detect salient regions by diffusing the surface index with a non local filter [16]. This time, the method is based on volumetric 3D patches. The approach begins by filtering the mesh to delete high frequencies and compute similarities between vertices. Afterwards, the mesh is transformed into multi-scale volumetric data. The dissimilarity between two patches located in two sub-voxels allows to generate a dissimilarity map. Finally, the saliency of one patch, which is proportional to its dissimilarity, is defined by the average of the saliency maps across the different scales.

Song *et al.* suggested in [18] to integrate the CRF (Conditional Random Field) framework to detect saliency. The multi-scale representation generated is combined using CRFs in order to label the mesh regions into salient and non-salient areas. Then, the method incorporates the multi-scale information of a mesh into a Conditional Random Field (CRF) framework while introducing a consistency constraint. Finally, to assign a label to each vertex, the CRF is resolved with the Belief Propagation algorithm.

In [19], Zhao *et al.* detect points of interest by estimating the saliency. The Retinex theory [20] is implemented to enhance the local details and estimate the invariant properties of the surface views. After the segmentation of the surface, the saliency estimated is based on the spatial distance between the resulting segments.

Recently, Song et al. suggested in [21] to estimate the saliency in the spectral domain by analyzing the irregularity spectrum of the Log-Laplace operator. First, the spectrum of the Laplacian is calculated, then a logarithmic transformation is applied to the spectrum to amplify variations at low frequencies while removing them from the rest of the spectrum. These deviations represent saliency.

3. Contributions

As described above, most of existing approaches estimating 3D mesh saliency include a simplification step in their process to define a good ratio between Sailency and Noise that is sometimes difficult to precise. This simplification step inevitably removes vertices from the mesh geometry and therefore alters the surface and its initial fluctuations. The result is then

a measure of saliency that does not take into account local variations and exiguous irregularities, yet necessary for the accurate estimation of saliency on surface meshes. Other steps such as smoothing bring back to a very high complexity.

It is commonly accepted that the human visual system is sensitive to large fluctuations surfaces [22]. Thereby, if a mesh vertex stands out strongly from its neighborhood, then this vertex could be considered as salient point. It remains to define the way in which we can locally evaluate the saliency of a vertex within the mesh. For this, local robust descriptors are carried out.

In this paper we propose a new method to define a perceptual multi-scale saliency map for 3D meshes. A synopsis summarizing the proposed approach is shown in Figure 1.

The novelty of the proposed approach relies on several key points. First, patches of adaptive size are used as local vertex descriptors. The dynamic size of the patches offers a better consideration of shape irregularities. Second, the saliency is defined as the vertices's degree measures with edges weights accounting for vertices similarities. The computation of the dissimilarity between patches benefits from a scale parameter. The distance between patches is weighted by the curvature of the target vertex to consider the local curvatures. Third, the weights used to aggregate the saliencies at different scales are based on their respective entropy in order to take into account the disparity of the saliency information in each scale. Other approaches mesure saliency at different scales without considering the entropy criterion [3] [10]. Also note that our approach is independent of any pre-treatment like simplification or smoothing. This autonomy will allow fitting this approach into any mesh processing algorithm. All these factors lead to an accurate estimation of visual saliency. It is also important to note that the proposed approach is an improvement of our recent work [23]. The main differences between the work in the present paper and [23] are:

1. The adaptive patch orientation and its construction are computed according to a spherical neighborhood rather than one-ring neighborhood (improving robustness).
2. The curvature of the mesh surface is taken into account in the computation of visual saliency.
3. The multi-scale aspect is added (improving the measure of visual saliency. See Figures 4 and 5).
4. A comparison of the results of the novel saliency measure with the

pseudo ground truth saliency provided by the database 2007 SHREC Shape-based Retrieval Contest is presented.

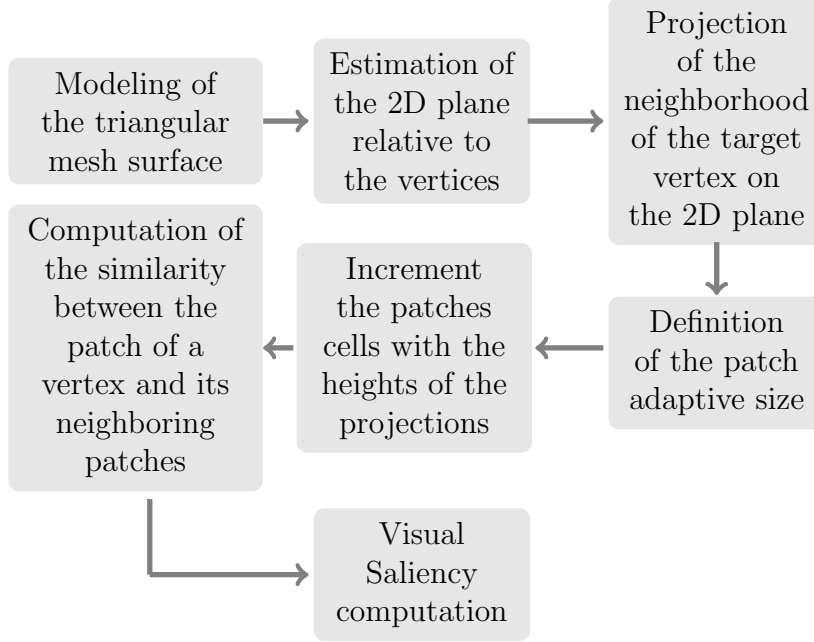5. An application for selecting the most important viewpoints of 3D meshes is provided.

Figure 1: Synopsis of the mono-scale saliency.

The paper is organized as follows. Section 4 presents our mesh saliency and details: the modeling of the mesh surface, the local adaptive patches construction, the single-scale and multi-scale saliency computation. In section 5, we show the contribution of the single-scale coefficients (distance and curvature) into the saliency map as well as the impact of the key parameters on the rendering of saliency. In the same section, we compare our saliency results with pseudo ground-truth saliencies on several 3D meshes, and make a comparisons with the state-of-the-art. In Section 6, we analyze the robustness and stability of our approach with respect to noise and simplification. In section 7 we present an application for selecting the most informative viewpoints based on our mesh saliency model. Last section concludes and perspectives are discussed.

4. Mesh Saliency

4.1. Modeling the surface

In order to define the local descriptors of a surface mesh, it is necessary to model the surface to be treated. To do so, we represent a mesh M by a non oriented graph $G = (V, E, w)$ where V represents the set of vertices, $E \subset V \times V$ the set of edges, and $w(v_i, v_j)$ the weight of the edge $e(v_i, v_j) \in E$. To each vertex v_i are associated its 3D coordinates $\mathbf{v}_i = (x_i, y_i, z_i)^T$.

For every vertex v_i on the mesh surface, we compute its normal vector $\mathbf{z}(v_i)$ and the directional vectors $\mathbf{x}(v_i)$ and $\mathbf{y}(v_i)$ that correspond to the estimation of the 2D tangent plan at vertex v_i on the mesh. In the first step, we consider a sphere S_ε centered at v_i with radius ε. The vertices located in that sphere will be considered as neighbors to v_i in order to estimate its center of gravity $\hat{\mathbf{v}}_\mathbf{i}$ as:

$$\hat{\mathbf{v}}_i = \frac{1}{|S_\varepsilon(v_i)|} \sum_{j \in S_\varepsilon(v_i)} \mathbf{v}_j \tag{1}$$

together with the associated covariance matrix at v_i defined as:

$$cov(v_i) = \sum_{j \in S_\varepsilon(v_i)} (\mathbf{v}_j - \hat{\mathbf{v}}_\mathbf{i})(\mathbf{v}_j - \hat{\mathbf{v}}_\mathbf{i})^T \in \mathbb{R}^{3 \times 3} \tag{2}$$

where $|S_\varepsilon(v_i)|$ is the cardinality of the neighborhood defined by the vertices belonging to the sphere S_ε. We will use the eigen-values of the covariance matrix to determine both the normal vector $\mathbf{z}(v_i)$ and the 2-directional vectors following the x and y axes. Similar approaches can be found in [24][25].

Thus, the mesh surface will be represented by vertices with their respective normals and tangent planes. However the normal vectors can have different directions (outwards and inwards). To guide these normals outwards, we propagate the orientation of the neighboring normals using the minimum spanning tree of the graph [26].

4.2. Local adaptive patches construction

Once the mesh has been modeled by the procedure described above, we build the local adaptive patches to describe the vertices local surfaces. This extends recents work [13] [25] [24] with an adaptive patch size. First, the nodes contained in a sphere $S_\varepsilon(v_i) = \{v_j \mid \|\mathbf{v}_\mathbf{j} - \mathbf{v}_\mathbf{i}\|_2^2 \leq \varepsilon\}$ centered at v_i within a radius ε are considered. They are projected onto the 2D plan $\mathbf{P}(v_i)$

defined by the associated vectors. From this, 2D projected vectors are obtained $\mathbf{v'_j}$:

$$\mathbf{v'_j} = [(\mathbf{v_j} - \mathbf{v_i}) \cdot \mathbf{x}(v_i), (\mathbf{v_j} - \mathbf{v_i}) \cdot \mathbf{y}(v_i)]^T \tag{3}$$

Then it remains to define the size of the patch. We propose a dynamic configuration based on the distance between the 2D coordinates of the projected vertices from the sphere S_ε. Indeed, it is possible to define the dimensions of the patch according to the abscissa and ordinate axis (respectively denoted $T_x(.)$ and $T_y(.)$) by:

$$\mathrm{T}_d(v_i) = \max_{(\mathbf{v'}_j, \mathbf{v'}_k) \in \mathbf{P}(v_i)} (||\mathbf{v'}_j^d - \mathbf{v'}_k^d||_2^2) \tag{4}$$

where d represents the x or y coordinates, $\mathbf{v'}_j^d$ the d-th coordinate of $\mathbf{v'_j}$, and $||.||_2$ the Euclidean norm.

Thus, the patch at vertex v_i is represented by a rectangle of size $\mathrm{T}_x(v_i) \times \mathrm{T}_y(v_i)$ centered at v_i. It should be noted that usually a patch is represented by a square of fixed size [13] [24] [25] and does not allow to get an adaptative local descriptor depending on the local geometry. This local patch is then divided into $l \times l$ cells in order to specify the indexes (in x and y axis) of the cell in which a vertex $v_j \in S_\epsilon(v_i)$ is projected:

$$\mathrm{index}^d(\mathbf{v'}_j) = \left\lfloor \frac{\mathbf{v'}_j^d}{\mathrm{T}_d(v_i)/l} \right\rfloor \tag{5}$$

where $\lfloor \cdot \rfloor$ denotes the integer rounded.

Finally, each cell P_i^k ($k \in [1, l \times l]$) of the patch P_i (associated to v_i) is filled with the absolute value of the sum of the projections heights:

$$\mathbf{H}(v_i) = \left(\sum_{\mathbf{v'}_j \in P_i^k} ||(\mathbf{v_j} - \mathbf{v'}_j)||_2^2, \forall k \right)^T \tag{6}$$

Figure 2 illustrates the construction of patches.

It is important to note that the adaptive size of patches is not related to the radius ε of the sphere, but instead to the maximum distance between the 2D projections of the vertices along the x axis in one hand, and the y axis in the other hand. This enforces its adaptive aspect.

Since our method is using only coordinates to construct the local adaptive patches, then it could be also used for computing these on 3D unstructured points clouds, as in [24] [25].
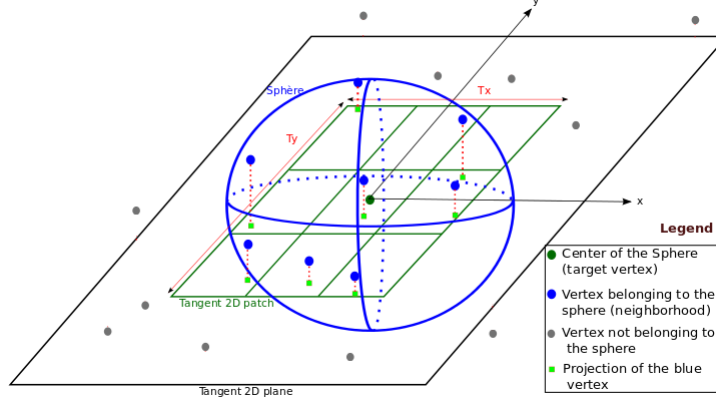
7

Figure 2: Illustration of the adaptive patch construction

4.3. Single-scale saliency computation

To compute the single-scale saliency of a vertex, a measure of similarity between its associated patch and the patches associated to its neighbors is required. We propose to locally compute a scale parameter (standard deviation of Gaussian kernel measuring the dissimilarity). Indeed, using a specific scale parameter for each node allows us to take into account the local distribution around each grid node.

The scale parameter $\sigma(v_i)$ is then determined by:

$$\sigma(v_i) = \max\,_{v_k \sim v_i}(||\mathbf{v_i} - \mathbf{v_k}||_2) \tag{7}$$

Note that we have tested the scale parameter based on the difference between the vectors of accumulated heights (see below) and it has been found that the difference between 3D coordinates vertices leads to better results.

Thus, the similarity is assigned to the weight of the edge (v_i, v_j) is:

$$w(v_i, v_j) = \exp\left[-\frac{\kappa(v_j) * ||\mathbf{H}(v_i) - \mathbf{H}(v_j)||_2^2}{\sigma(v_i) * \sigma(v_j) * ||\mathbf{v}_i - \mathbf{v}_j||_2^2}\right] \text{ with } v_j \sim v_i \tag{8}$$

where $\mathbf{H}(v_i) \in \mathbb{R}^{l \times l}$ is the vector of accumulated heights into the cells of the patch, $\kappa(v_j)$ is the curvature of the vertex v_j, and $||\mathbf{v}_i - \mathbf{v}_j||_2^2$ represents the Euclidean distance between the vertices v_i and v_j. Note that if the Euclidian distance between 2 local adaptive patches $\mathbf{H}(v_i)$ and $\mathbf{H}(v_j)$ is high, then the similarity between the associated vertices v_i and v_j will be 0 (i.e., they are dissimilar).

8

Finally, the visual saliency of the vertex v_i on the 3D mesh is defined by its degree as:

$$\text{single-scale-saliency}(v_i) = \left(\frac{1}{|v_j \sim v_i|} \right) \sum_{v_i \sim v_j} w(v_i, v_j) \qquad (9)$$

where $|v_j \sim v_i|$ is the cardinality of the neighborhood representing the adjacent vertices, and $w(v_i, v_j)$ is the weight of the edge between vertices v_i and v_j. In fact, since the weight of the edge $e(v_i, v_j)$ represents the similarity between 2 adjacent vertices v_i and v_j (Eq.(8)), the computation of the degree (Eq.(9)) of a vertex will define its single-scale saliency compared to its neighborhood. This single-scale saliency is defined in [0,1] where 0 refers to high saliency (very dissimilar from its neighborhood) and 1 refers to very low saliency (very similar from its neighborhood).

4.4. Multi-scale saliency

To enhance the quality of the measured saliency, we compute it at different scales similarly to [3]. Saliency at primary scales will detect finest and cramped details, while higher scales will highlight large regions (see Table 1). The aim is to detect saliency at different scales to cope with noise, since it is only noticeable at some scales. To do so, we vary the radius ε of the sphere S_ε to define the local patch descriptor, and we consider three different radius ε, 2ε, 3ε with ε to be fixed in accordance to the mesh. Then we calculate the single-scale map saliency for each considered neighborhood. Before combining the obtained saliency maps, we calculate the entropy which measures the disorder and the disparity of the saliency information on each map. To do so, at a given scale k, we compute a histogram h_k of the saliency values of the vertices to obtain the probability to have the saliency value i:

$$Pr_{i,k} = h_k^i / |V| \qquad (10)$$

where h_k^i gives the number of vertices of saliency i at the scale k. Then, the entropy at scale k is defined by:

$$\text{entropy}_k = -\sum Pr_{i,k} * log_2 Pr_{i,k} \qquad (11)$$

By weighting the saliency of each node by the scale-entropy, in the combination formula of the different scales, we can obtain a robust multi-scale

9

saliency map that considers the disparity of the saliency at each scale. It is computed as follows:

$$\text{Multi-scale-saliency}(v_i) = \frac{\sum_{k=1}^{3} \text{Single-scale-saliency}_k(v_i) * \text{entropy}_k}{\sum_{k=1}^{3} \text{entropy}_k} \quad (12)$$

where k is the scale index and $\text{Single-scale-saliency}_k(v_i)$ corresponds to single scale saliency (9) with the corresponding radius to the k-th scale.

5. Results and analysis

5.1. The contribution of the curvature and the distance-coefficient weights

Figure 3 presents the single-scale saliency computed on the 3D mesh Gorilla by our approach. We can notice that the regions associated to the paws and toes are accurately highlighted and are judged as salient. Also, one can easily see that tight details such as eyes, mouth, nose, and ears are well detected but with a certain imperfection at the level of the eyes. These are saturated. This behavior will be corrected by the multi-scale saliency (see Figure 5).
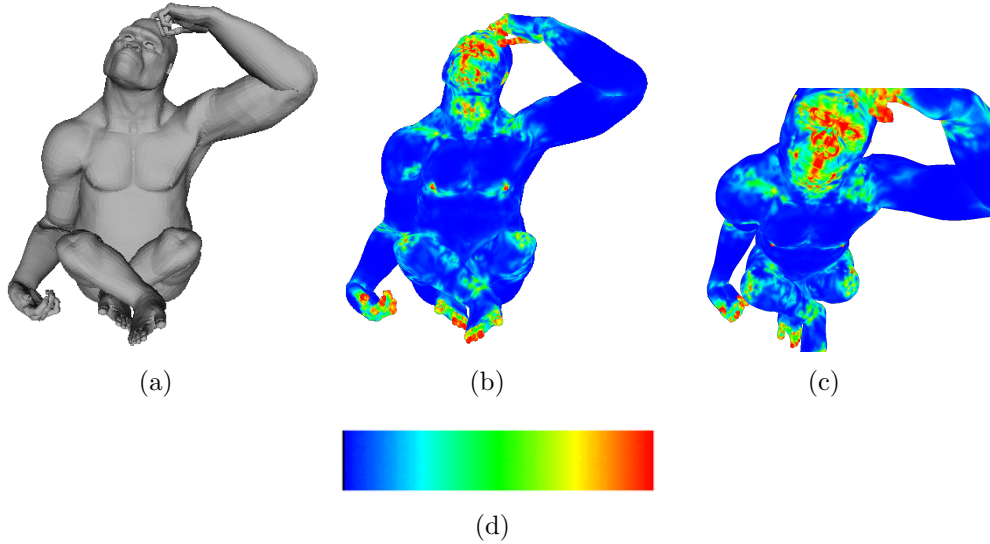

(a) (b) (c)

(d)

Figure 3: Single-scale saliency with our approach: image (a) shows the original 3D mesh Gorilla, image (b) shows its mesh saliency with our method($\varepsilon = 2$ and l=17), image (c) shows a zoom on the face of the gorilla mesh and image (d) shows the colormap. The red areas on the 3D mesh represent the most salient regions. Those in blue are not salient.

What is interesting here is the contribution of the curvature and the distance coefficient weights on the single-scale saliency rendering. Figure 4 shows a comparison of the saliency detected using these weights (in equation 8) and the saliency without these [23]:
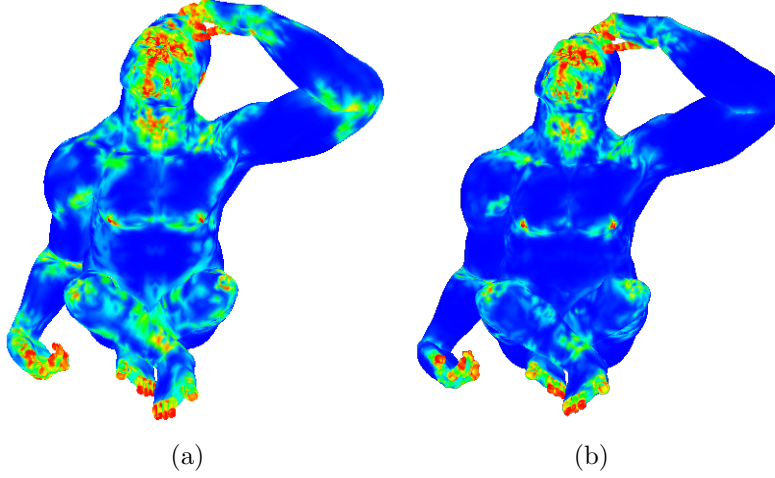

(a)   (b)

Figure 4: The contribution of the weighting parameters: image (a) shows the mesh saliency without the curvature and the distance coefficient weights [23] (25 438 vertices)($\varepsilon = 2$ and l=17) and image (b) shows the mesh saliency using the curvature and the distance weight ($\varepsilon = 2$ and l=17).

It can be seen that the saliency estimated with these curvature and distance weights is widely finer than that one detected without. The surplus of saliency at the level of the ribs, the chest, the shoulders, the forearms, and the paws have been corrected. Moreover, the salient regions like the nose, the eyes, the mouth and, the toes have been preserved.

Indeed, using vertex curvature in equation (8) permits to estimate the discontinuity between the target vertex and its neighbors. A strong discontinuity will more contribute to the saliency than a weak one. This parameter therefore helped to eliminate the surplus of saliency.

Also, the distance-coefficient in the denominator of equation (8) reduces the influence of the remote vertices in the computation of patch similarities which also explains the deletion of excedent saliency.

5.2. The influential parameters

Two parameters affect the rendering of the saliency: the number of patch cells l and the radius of the sphere S_ε containing the vertices to be projected. Table 1 presents the amplitude of the detected saliency according to the l and ε parameters.

We can notice on Table 1 that saliency depends strongly on both the neighborhood delimited by the sphere S_ε and the number of cells of the adaptative patch. In Table 1, a radius equal to 1 provides the detection of very small salient details. Also note that in this case, increasing the number of cells does not affect the estimate of saliency. In contrast, a sphere of radius equal to 3 greatly expands the neighborhood, which brings to an extreme description of the mesh surface, and thus taking into account the extreme fluctuations on the surface which brings to detect the large salient regions. We can also notice that the number of salient points increases proportionally to the number of cells. These 3 radii of the sphere S_ε will permit to design a multi-scale saliency based on the local adaptive patches.

5.3. Experimental results

Figure 5 presents the multi-scale saliency estimated by our approach on the 3D mesh Gorilla (Figure 5(a)). Figure 5(b) shows that our multi-scale method brings out finely the salient regions on the 3D mesh surface. Figure 5(c) shows a zoom on the face of the 3D mesh Gorilla. We can see that compared to the single-scale saliency in Figure 3(c), the saturation of the saliency of the eyes and nose has been corrected and now we see that the eyes are gently highlighted. This was expected, since as previously mentioned, our method of combining the different scales includes an entropy weighting parameter.
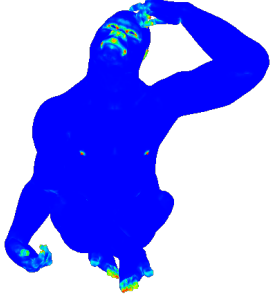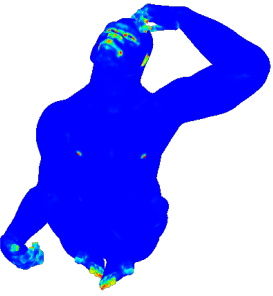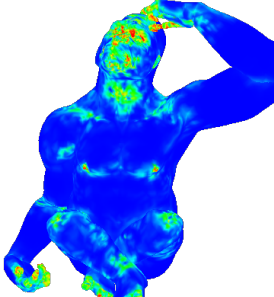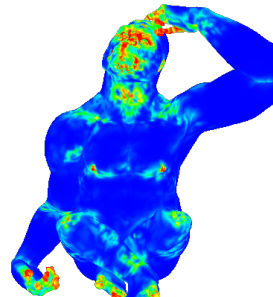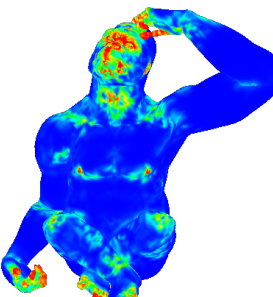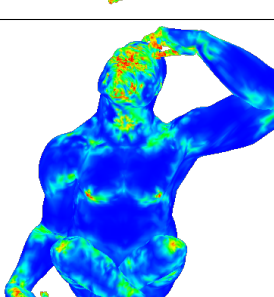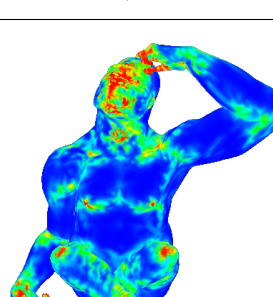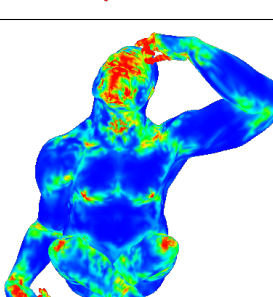
	Number of cells l		
	7x7	17x17	27x27
$\varepsilon=1$			
$\varepsilon=2$			
$\varepsilon=3$			

Table 1: Influence of parameters ε and l on the detection of saliency.
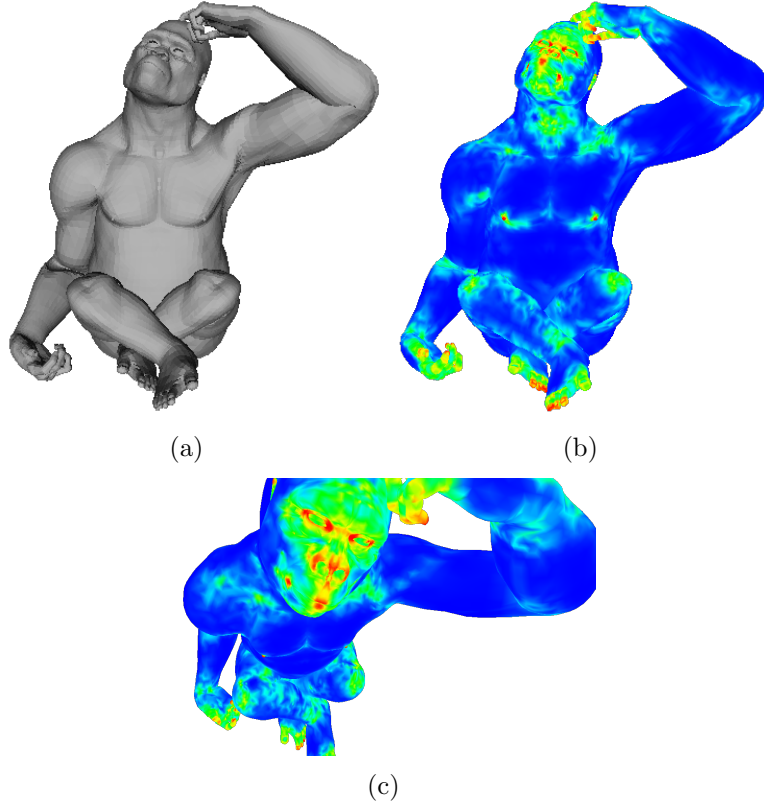
(a)

(b)

(c)

Figure 5: Multi-scale saliency with our approach: image (a) shows the original 3D mesh Gorilla, image (b) shows its multi-scale mesh saliency with our method and image (c) shows a zoom on the multi-scale saliency detected on the face of the Gorilla mesh. This has to be compared with Figure 3.

5.4. Comparisons with pseudo-ground truth saliency

Figure 6 presents a comparison between the saliency detected on different 3D meshes from the 2007 SHREC Shape-based Retrieval Contest by our approach and their pseudo-truth saliency obtained from [27]. These latter were elaborated by designing an online experiment that asked people to select 3D points that they expect to be selected by other people. Based on these observations, they used a regression analysis to produce an analytical model that predicts where salient points are likely to be. Figure 6 shows that our detected saliency in the various 3D meshes corresponds well to the pseudo ground-truth saliency. For some unoptimized meshes, our approach for measuring saliency might not perform well (Figure 7). An unoptimized

mesh contains triangles of constant surfaces and edges of constant lengths. Indeed, when some regions of the mesh exhibit a lot of details, this ones require more vertices and therefore edges of constant lengths and triangles of constant surfaces (i.e., distances between vertices are approximatively equal). On the contrary, regions with less details, like flat surfaces, require less vertices and therefore edges of different lengths (i.e., distances between vertices are different). In the computation of the similarity in Equation (8), we use a distance-weight coefficient $||\mathbf{v}_j - \mathbf{v}_i||_2^2$. This coefficient permits to reduce the influence of far direct neighbors and vice-versa. For unoptimized meshes, all direct neighbors of a vertex contribute to the saliency since the distance-coefficient is constant. This explains the exceed of red dots on the obtained saliency in Figure 7.
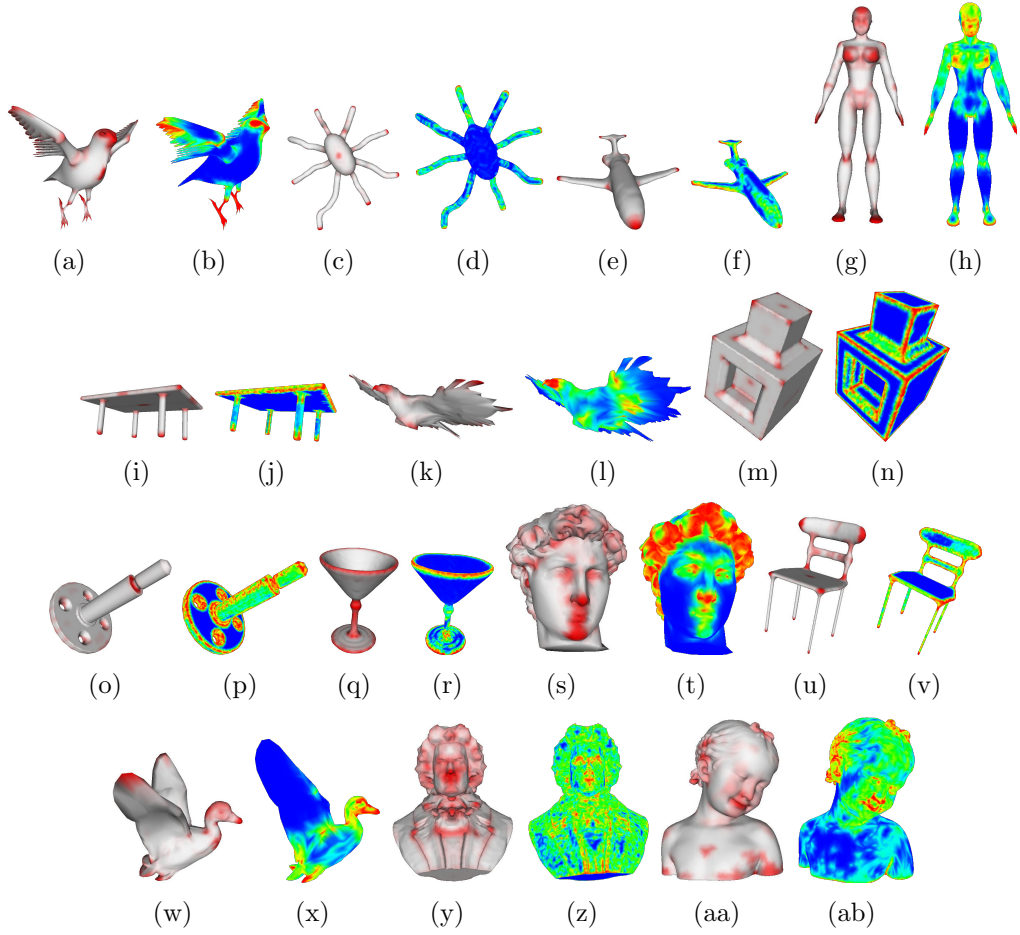
Figure 6: Comparison with the pseudo ground-truth saliency: images (a),(c),(e),(g),(i),(k),(m),(o),(q),(s),(u),(w),(y), and (aa) show the pseudo ground-truth saliency on divers 3D meshes. Images (b),(d),(e),(h),(j),(l),(n),(p),(r),(t),(v),(x),(z), and (ab) show the multi-scale saliency detected with our approach.
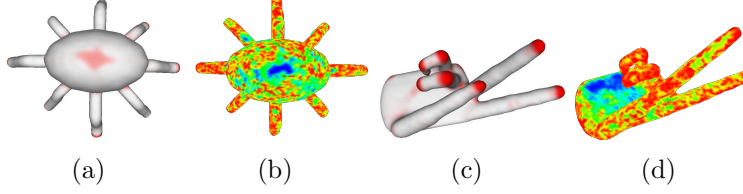
Figure 7: Failure cases: images (a) and (c) show the pseudo ground-truth saliency on divers 3D meshes. Images (b) and (d) show the multi-scale saliency detected with our approach.

5.5. Comparisons with the state-of-the-art

Since the sources/codes associated to the different approaches of the state-of-the-art are not available, we evaluate saliency on the same 3D meshes to compare the results. Figure 8 shows a comparison of the detected saliency on the 3D mesh Dinosaur with the state-of-the-art. As outlined above, our approach considers out-standing vertices (in relief) in a flat surface as salient points (discontinuities on surfaces naturally attract the attention of the human observer). Tal *et al.*'s approach [10] judges the ribs of the 3D Dinosaur model located on the back and the stomach as non-salient regions (Figure 8(c)). Yet, these areas fluctuate enormously and contain high discontinuities in the surface. In Figure 8(b), our method assesses ribs of the mesh Dinosaur as salient regions given their high discontinuities, and contrary to the approach of [10] Figure 8(c), the relative area of skull (except the eye and some curvatures) is not considered as a completely salient region. This also means that at the first glance in the direction of the mesh Dinosaur, visual attention will be placed firstly on a part of the fluctuating ribs or the neck, rather than on the surface of its skull. In Figure 8(d) [3], we can notice again that the saliency on the ribs is weakly detected (colored in light green). Locally finest details are not taken into account, contrary to our approach. However, in Figure 8, Song *et al.*'s approach [21] has a similar saliency as ours. It can differentiate between areas with high discontinuities and flat ones.

Figure 9 shows a comparison of the estimated saliency on the 3D mesh Angel with the approaches [10] and [21]. 3D object Angel's surface is complex as it contains many extremities. It has also both rough and smooth surfaces. One can see that the extremity of the scarf presents lots of fluctuations. This one is considered as salient by the approach of [21] and ours while the approach of [10] considers it as non salient. The discontinuities on the eyes,
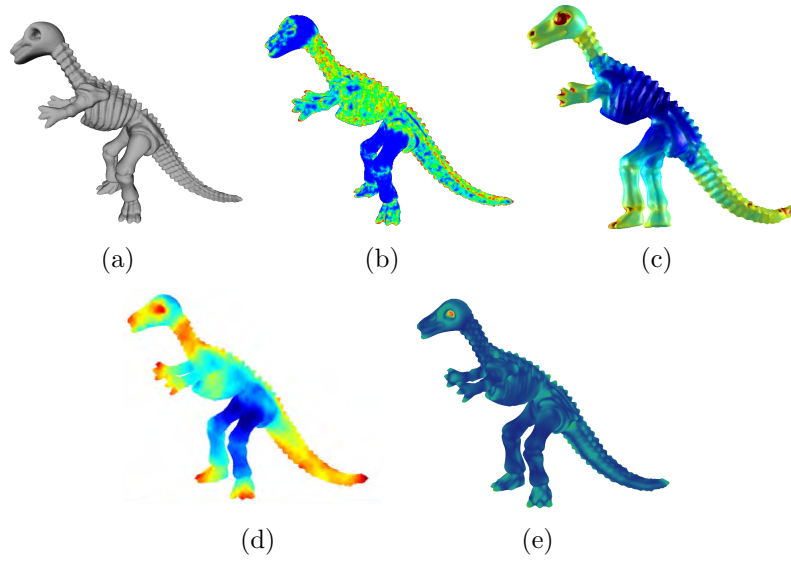
17

Figure 8: Comparison with the state-of-the-art: image (a) shows original 3D mesh dinosaur (21777 vertices), image (b) shows its multi-scale mesh saliency with our method (l=17), image (c) shows the saliency detected with the approach of [10], image (d) shows the saliency detected with the approach of [21] and image (e) shows the saliency detected with the approach of [3]

the arms, the hip and the stomach are represented as salient by the approach of [21] and ours, contrary to the method of [10].

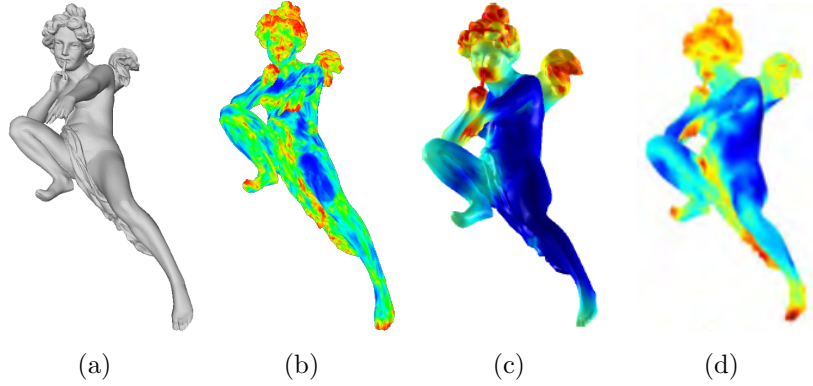<div style="text-align:center">(a)   (b)   (c)   (d)</div>

Figure 9: Multi-scale saliency on the 3D mesh Angel: image (a) shows the original 3D mesh Angel (14227 vertices), image (b) shows its multi-scale mesh saliency with our method (l=14227), image (c) shows the saliency detected with the approach of [10] and image (d) shows the saliency detected with the approach of [21].

Figure 10 compares the salient regions estimated on the 3D mesh horse with our approach and those estimated by approaches of [10] and [21]. We can notice that our method and the method of [21] are able to detect the eyes, highly regarded parts in scenes or meshes containing faces [28], and judge them as very salient regions while the approach of [10] fails in this regard. The horse's back has muscled parts like the Centaur mesh (Figure 11). These areas are also assessed as salient regions with our approach and the approach of [21] contrary to the approach of [10].

From the comparisons, we can notice that the compared methods generate saliency with a more smooth quality than our approach. This is due to the fact that we aim at detecting a fine estimation of saliency that is important for many applications like mesh compression and the viewpoint selection.

Figure 11 presents the saliency detected on other 3D meshes. We can notice the same behavior of our method, which allows a precise estimate of the saliency.
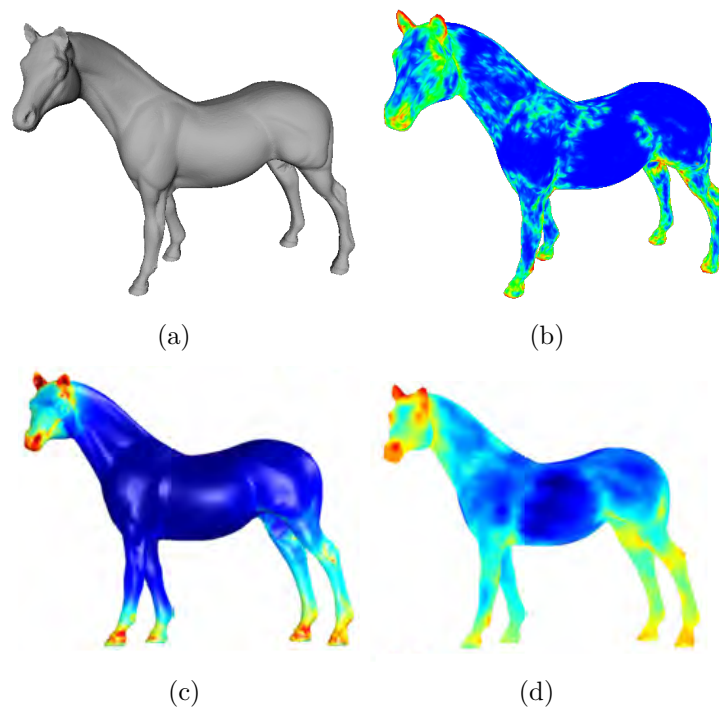
(a)   (b)

(c)   (d)

Figure 10: Saliency detected on the 3D horse : image (a) shows the original 3D mesh horse (20871 vertices), images (b) shows its multi-scale mesh saliency with our method (l=27), image (c) shows the saliency detected with the approach of [10], image (d) shows the saliency detected with the approach of [21].
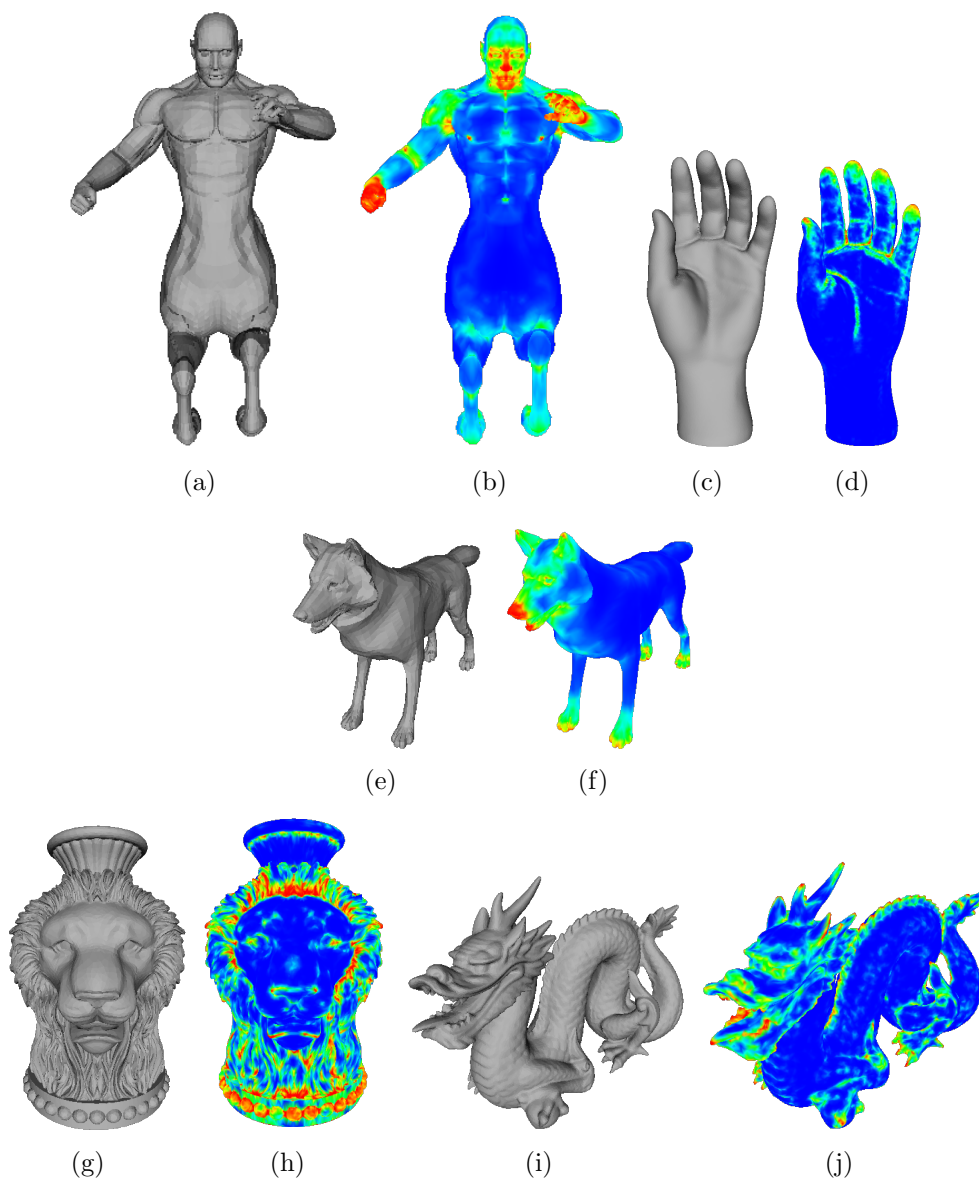
(a)　　　　(b)　　　　(c)　　　　(d)

(e)　　　　(f)

(g)　　　　(h)　　　　(i)　　　　(j)

Figure 11: The multi-scale saliency detected with our approach on other 3D meshes

6. Robustness and Stability

To attest the robustness of our approach, we put noise on the 3D mesh Centaur by randomly displacing the positions of its vertices according to 2 levels of noise. Then, we applied our multi-scale saliency measure.
Figure 12 shows that visually, our method always succeeds to detect the same salient regions despite the fact that the mesh surface was severely noised. Areas like the eyes, the mouth, the nose, the paws still are considered salient and can be differentiated from the other flat regions. Also, the NMSE (Normalized Mean Squared Error) was computed between the saliency associated to the original mesh and the saliency on its noisy version. The low value of the NMSE confirms that our approach is robust to noise.

(a)   (b)

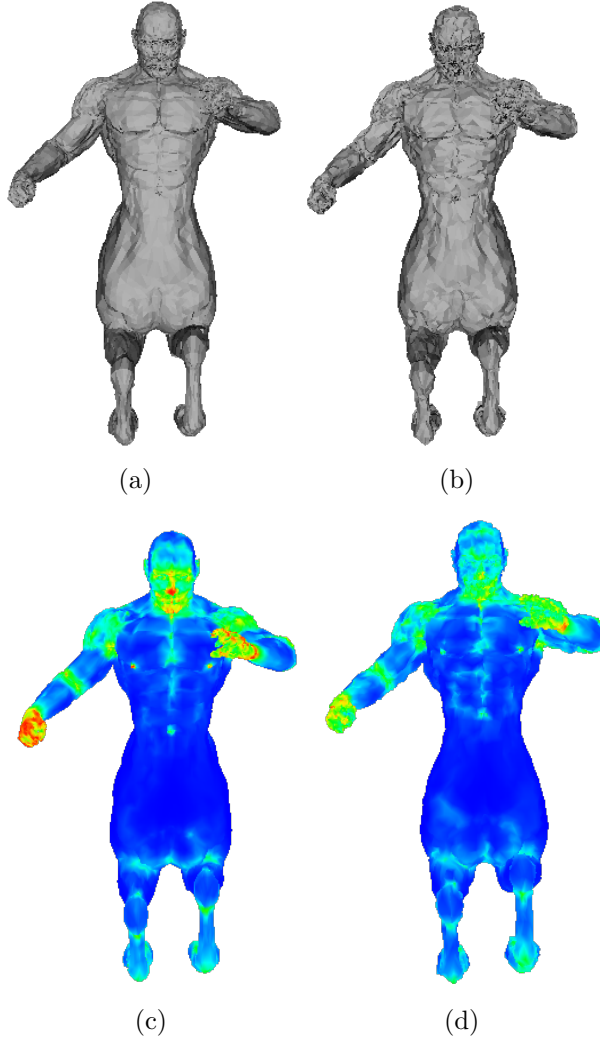(c)   (d)

Figure 12: Robustness to noise: image (a) shows the original 3D mesh Centaur noised
(displacement= -0.1%), image (b) shows the original 3D mesh Centaur noised (displace-
ment= -0.2%), image (c) shows the multi-scale saliency of the Centaur noised with -0.1%
NMSE=3.27098e-06 and image (d) shows the multi-scale saliency of the Centaur noised
with -0.2% NMSE=1.1485e-05. Note how the original salient regions still remain salient
despite the sever noise.

Another experiment that demonstrate the stability of our method is mea-
suring saliency on simplified meshes. Simplification was operated using Gar-
land *et al*'s algorithm [29]. Figure 13 shows the behavior of our method on
simplified meshes. Firstly, the mesh was simplified to 25% (In Figure 13(a),

25% of vertices were deleted). One can notice that the initially salient regions are always detected (see Figure 13(c)). In Figure 13(b), the mesh is simplified to 50%. It's clear that the same regions remain salient (see Figure 13(d)) but with less intensity. This is due to the strong simplification that tends to delete discontinuities and therefore flattens the vertices. Flat surfaces present naturally a low visual saliency.
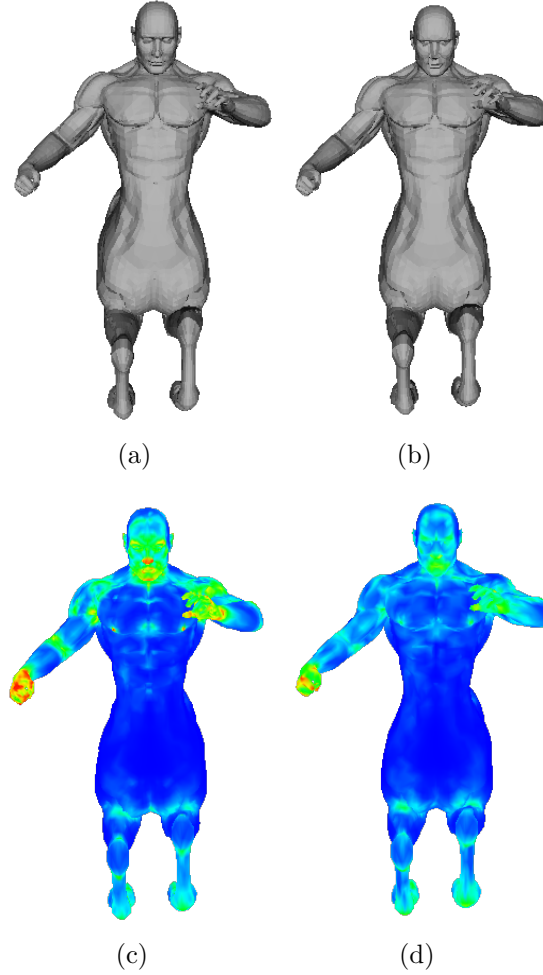

(a) (b)

(c) (d)

Figure 13: Robustness to simplification: image (a) shows the original 3D mesh Centaur simplified to 25%, image (b) shows the 3D mesh Centaur simplified to 50%, image (c) shows the multi-scale saliency of the Centaur simplified with 25% and image (d) shows the multi-scale saliency of the Centaur simplified to 50%.

7. Viewpoint selection

In order to applicate and evaluate the benefit of our mesh saliency model, we have elaborated a method selecting areas that are the most perceptually important for the observer. The main criterion of our viewpoint-selection approach is to distinguish regions that maximize saliency. To do this, we begin by selecting the best viewpoint maximizing saliency along the abscissa axis. This is done by uniformly sampling a sphere that bounds the object. Let vp be a given viewpoint along the abscissa axis and $surface(vp)$ the vertices visible from vp. The global saliency associated to these vertices is defined as:

$$\text{Global-saliency}_{axis_x}(vp) = \sum_{v \in surface(vp)} \text{Multi-scale-saliency}(v) \qquad (13)$$

Thus the best viewpoint along the abscissa axis is defined by $max(GlobalSaliency_{axis_x}(vp_i))$ where vp_i are the different viewpoints along the x-axis. From this last viewpoint, we do the same processing to select a new viewpoint maximizing saliency along the ordinate axis. From the latter, we use a gradient-descent optimisation along the three axes $(\mathbf{x}, \mathbf{y}, \mathbf{y} - 45°)$ at the same time to perform the same processing for selecting the viewpoint that maximize saliency. Figure 14 shows the obtained results.

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

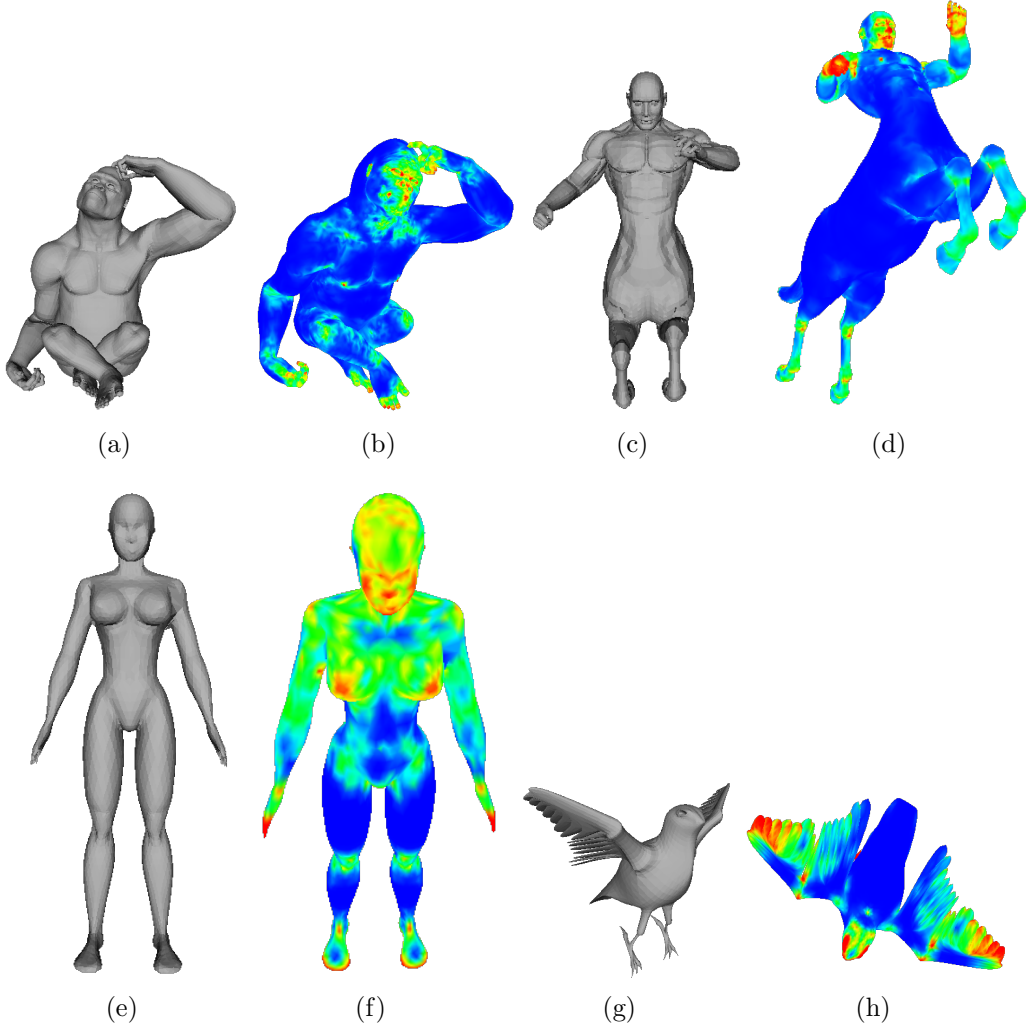(e)   (f)   (g)   (h)

Figure 14: Most informative viewpoints selected by our approach: images (a), (c), (e) and (g) show the originals 3D objects. Images (b), (d), (f), and (h) show their respectives perceptually important viewpoints.

One can see that the generated viewpoints correspond well to the criterion of our approach selecting the optimal viewpoints, and also, these viewpoints correspond to the most likely views of the 3D objects.

Figure 15 shows a comparison between the optimal selected viewpoints with our approach and the optimal ones generated with the approach of Tal *et*

al.[10]. We can observe that our method for selecting the salient viewpoints is competitive with the approach of [10]. However, some limitations are shown on images (g) and (i).

For selecting the most informative viewpoints, the method of Tal *et al.*[10] begins by generating the candidates viewpoints by uniformly sampling a sphere that bounds the 3D mesh. From the viewpoint maximizing saliency, a gradient-descent optimization is applied to define the most informative viewpoints. On the contrary, to make the processing faster, our method generates first the candidates viewpoints by sampling the sphere bounding the 3D mesh only along the x-axis and the y-axis, which leads to a fewer viewpoints candidates. Second, similarly to the method of Tal *et al.*[10], a gradient-descent optimization is applied.

The observed limitations on the 3D meshes representing the Piano and the Car are due to the nature of the used 3D meshes that contain a small number of vertices (we can see the large faces on the surface of these). Indeed, our approach is based on the construction of adaptive patches that are filled with a local height-field of the mesh vertices. Thereby, if the number of vertices is low, this will lead to using mostly empty adaptative patches while computing the saliency. Besides, the viewpoints generated for the Car and the Piano seem to be logic insofar that our method selects the regions that are highly salient according to the used saliency model (i.e., the underside of the Piano contains much more discontinuities than its topside).
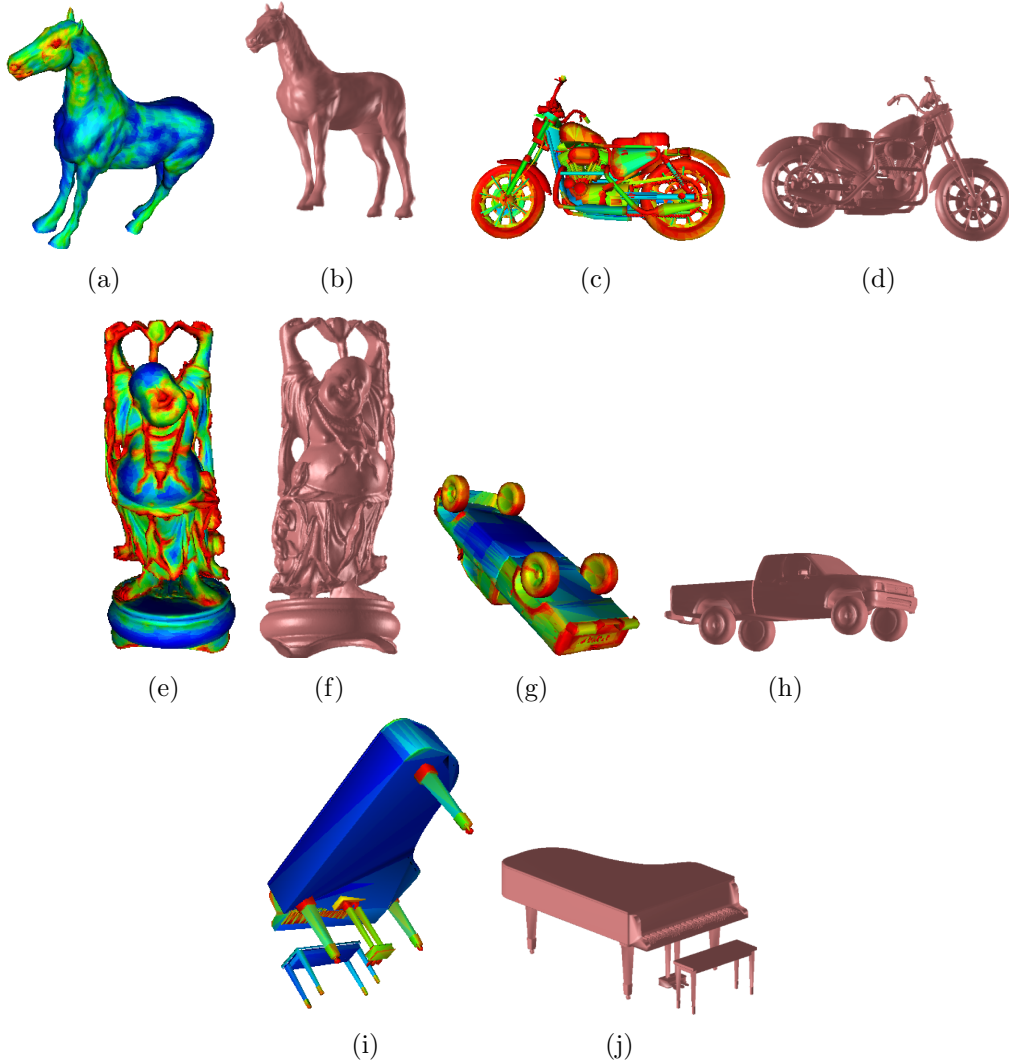
Figure 15: Most informative viewpoints selected by our approach (top) and by the approach of [10]: images (a), (c), (e), (g) and (i) show the most salient viewpoints. Images (b), (d), (f), (h) and (j) show the most salient viewpoints selected by the approach of [10].

8. Conclusion

In this paper we have presented an approach simulating the visual attention in order to detect salient regions on 3D meshes. Our method is based on the projection of the prominent vertices (that stand out from a local neighborhood on the surface) on local adaptive patches. This vertices exhibit

28

discontinuities that attract visual attention.

In order to detect these salient regions, local adaptive patches accumulating the heights projections of the neighborhood are computed at each vertex. This permits to characterize the mesh surface and are used as local descriptors. Subsequently, a similarity measure between patches leads to quantify the perceptual saliency of a vertex relative to another, and then to obtain the saliency at each vertex of the mesh. We have also presented an application of our 3D saliency model that selects the optimal viewpoints exposing the most salient regions.

Future works will aim at improve the saliency map by integrating new perceptual attributes in measuring the degree of vertices that express saliency, and analyze their respective contributions. To ease the comparison of future 3D saliency works with our proposed approach, the saliencies we obtained in Figures 5-11 can be downloaded at https://nouri.users.greyc.fr/3DSaliency.html

9. Acknowledgments

[1] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.

[2] Y. Kim, A. Varshney, D. W. Jacobs, F. Guimbretière, Mesh saliency and human eye fixations, ACM Trans. Appl. Percept. 7 (2) (2010) 12:1–12:13.

[3] C. Lee, A. Varshney, D. W. Jacobs, Mesh saliency, ACM Trans. Graph. 24 (3) (2005) 659–666.

[4] P. Shilane, T. Funkhouser, Distinctive regions of 3D surfaces, ACM Trans. Graph. 26 (2).

[5] R. Gal, D. Cohen-Or, Salient geometric features for partial shape matching and similarity, ACM Trans. Graph. 25 (1).

[6] S. Jia, C. Zhang, X. Li, Y. Zhou, Mesh resizing based on hierarchical saliency detection, Graph. Models 76 (5) (2014) 355–362.

[7] L. Jinho, M. Baback, P. Hanspeter, M. R. Machiraju, Finding optimal views for 3D face shape modeling, in: Proc. International Conference on Automatic Face and Gesture Recognition, 2004, pp. 31–36.

[8] Z. Liu, W. Zou, O. Le Meur, A novel saliency detection framework, IEEE Transactions on Image Processing 23 (5) (2014) 1937 – 1952.

[9] G. Taubin, Estimating the tensor of curvature of a surface from a polyhedral approximation, in: Proceedings of the Fifth International Conference on Computer Vision, IEEE Computer Society, 1995, pp. 902–.

[10] A. Tal, E. Shtrom, G. Leifman, Surface regions of interest for viewpoint selection, IEEE Conference on Computer Vision and Pattern Recognition (2012) 414–421.

[11] A. E. Johnson, M. Hebert, Using spin images for efficient object recognition in cluttered 3d scenes, IEEE Trans. Pattern Anal. Mach. Intell. 21 (5) (1999) 433–449.

[12] J. Wu, X. Shen, W. Zhu, L. Liu, Mesh saliency with global rarity, Graph. Models 75 (5) (2013) 255–264.

[13] A. Maximo, R. Patro, A. Varshney, R. Farias, A robust and rotationally invariant local surface descriptor with applications to non-local mesh processing, Graph. Models 73 (5).

[14] Y. Zhao, Y. Liu, R. Song, M. Zhang, A saliency detection based method for 3d surface simplification, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, Kyoto, Japan, March 25-30, 2012, 2012, pp. 889–892.

[15] Y. Zhao, Y. Liu, R. Song, M. Zhang, Extended non-local means filter for surface saliency detection, in: 19th IEEE International Conference on Image Processing, 2012, pp. 633–636.

[16] A. Buades, B. Coll, J.-M. Morel, A non-local algorithm for image denoising, in: Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition Volume 02, IEEE Computer Society, 2005, pp. 60–65.

[17] Y. Zhao, Y. Liu, Patch based saliency detection method for 3d surface simplification, in: Proceedings of the 21st International Conference on Pattern Recognition, 2012, pp. 845–848.

[18] R. Song, Y. Liu, Y. Zhao, R. R. Martin, P. L. Rosin, Conditional random field-based mesh saliency, in: 19th IEEE International Conference on Image Processing, 2012, pp. 637–640.

[19] Y. Zhao, Y. Liu, Z. Zeng, Using region-based saliency for 3d interest points detection., in: CAIP (2), 2013, pp. 108–116.

[20] M. Elad, Retinex by two bilateral filters, in: Proceedings of the 5th International Conference on Scale Space and PDE Methods in Computer Vision, 2005, pp. 217–229.

[21] R. Song, Y. Liu, R. R. Martin, P. L. Rosin, Mesh saliency via spectral processing, ACM Trans. Graph. 33 (1) (2014) 6.

[22] S. Coren, L. M. Ward, J. T. Enns, Sensation and perception, Wiley, 2003.

[23] A. Nouri, C. Charrier, O. Lézoray, Mesh saliency with adaptive local patches, in: Proceeding of SPIE Electronic Imaging, Three-Dimensional

Image Processing Measurement (3DIPM) and Applications, Vol. to appear, 2015.

[24] F. Lozes, A. Elmoataz, O. Lézoray, Nonlocal processing of 3d colored point clouds, in: 21st International Conference on Pattern Recognition, 2012, pp. 1968–1971.

[25] D. Julie, C. Raphaelle, V. Sebastien, Self-similarity for accurate compression of point sampled surfaces, Eurographics 33 (2).

[26] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, W. Stuetzle, Surface reconstruction from unorganized points, SIGGRAPH Comput. Graph. 26 (2) (1992) 71–78.

[27] X. Chen, A. Saparov, B. Pang, T. Funkhouser, Schelling points on 3d surface meshes, ACM Trans. Graph. 31 (4) (2012) 29:1–29:12.

[28] J. Henderson, C. Williams, R. Falk, Eye movements are functional during face learning, Memory and cognition 33 (1) (2005) 98–106.

[29] M. Garland, P. S. Heckbert, Surface simplification using quadric error metrics, in: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, 1997, pp. 209–216.