

Visual saliency and perceptual quality assessment of 3D meshes

Anass Nouri, Christophe Charrier, Olivier Lézoray

Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, FRANCE

ABSTRACT

This chapter concerns the visual saliency and the perceptual quality assessment of 3D meshes.

Firstly, the chapter proposes a definition of visual saliency and describes the state-of-the-art methods for its detection on 3D mesh surfaces. A focus is made on a recent model of visual saliency detection for 3D colored and non colored meshes whose results are compared with a ground-truth saliency as well as with the literature's methods. Since this model is able to estimate the visual saliency on 3D colored meshes, named colorimetric saliency, a description of the construction of a 3D colored mesh database that was used to assess its relevance is presented. We also describe three applications of the detailed model that respond to the problems of viewpoint selection, adaptive simplification and adaptive smoothing.

Secondly, two perceptual quality assessment metrics for 3D non colored meshes are described, analyzed and compared with the state-of-the-art approaches.

Keywords: Patches, Graphs, Geometric saliency, Colorimetric saliency, A. Nouri, Viewpoint selection, Adaptive compression, Adaptive smoothing, Multi-scale saliency, Full-reference mesh visual quality assessment, No-reference mesh visual quality assessment, visual masking, SVR.

I. VISUAL SALIENCY OF 3D MESHES

1. INTRODUCTION

In every look thrown at a scene or an object, visual attention is fixed on particular regions distinct from their 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 (Itti and Koch, 1998). This faculty of fixation is known as visual attention. The saliency computation would allow detecting these perceptually important regions that guide the visual attention. Visual saliency approaches proposed in the state-of-the-art are inspired from low level features of the Human Visual System (HVS). This permits to replace the geometric attributes used for the saliency computation by perceptual ones, and, as confirmed in Kim et al. (2010), these perceptual models succeed in modeling correctly the eye movements of the human observer.

Many applications in the field of 3D computer vision rely on the detection of visual saliency phenomena such as: 1) Automatic viewpoint selection (Lee et al., 2005) that aims at generating the most informative and attractive viewpoint. 2) Adaptive simplification (Shilane & Funkhouser, 2007) where the goal is to more simplify the non salient areas for the visual quality preservation. Similarly, other applications such as Surface Matching (Gal and Cohen-or, 2006), automatic resizing (Jia, Zhang, and Zhou, 2014), facial recognition (Lee et al., 2004), etc benefit from visual saliency.

In the following, a definition of visual saliency is proposed. Then, a description of the physiological elements that are involved in the sensitivity of the HVS to visual saliency is provided. Afterwards, different approaches proposed in the state-of-the-art that predict visual saliency are analyzed. This chapter focuses also on a recent approach proposed by Nouri et al. (2015a) that is based on local adaptive patches for the detection of visual saliency. A qualitative and quantitative comparison with a pseudo ground truth as well as a comparison with the state-of-the-art methods are achieved to assess its relevance. Finally, three applications guided by the detailed model of saliency are analyzed.

2. VISUAL SALIENCY: PHYSIOLOGICAL EXPLANATION

The colossal amount and the diversity of the visual information conveyed by the world around us is partially processed by the HVS because of its finite capability. For example, the number of photosensitive cells in the retina is limited and non uniformly distributed. Thus, we can only perceive clearly and with a maximal visual acuity a small area of the visual field. This area is placed in the center of the gaze after several ocular movements. This restraint space of the human visual field, processed in detail by our HVS contrary to the rest, is selected using attentional mechanisms. Many complex process are involved in visual attention. This chapter focuses only on the part of visual attention that concerns the saliency of a region. Visual saliency can be described as the perceptual information that permits to some objects or regions in the scene to stand out from their surrounding and thus captures the visual attention of the human observer. Therefore, the degree of saliency of a region depends on the distinction of the target region from its surrounding. In other words, this chapter is interested in attentional mechanisms linked to the visual signal or visual stimuli rather to those associated to human observers looking at the scene or to the tasks entrusted to them before the beginning of the visualization (see Bottom-up and Top-down processes below). Thus, a visual element of the scene is judged salient if it is easily remarkable without any prior provided to the observer or if it stands out first while looking to the scene. This element would tend to capture the attention more importantly than other elements in the scene and therefore further attract the observers' gazes. Visual attention allows us to construct a perception that fits our needs and capacities. The power and the swiftness of this mechanism constitute a major feature for the selection of relevant information in the scene by moving the observer eyes sequentially from a salient region to another salient region while focusing attention.

2.1 ATTENTIONAL PROCESS

There are two attentional processes in human vision influencing the orientation of the visual attention on a particular region of the scene:

1. Bottom-up process: Also called ascending process (from the retina to the brain), are exogenously mechanisms dependent of intrinsic properties of the visual stimuli such as contrast, texture, shape, etc. No willingness from the human observer occurs for placing his eyes on a particular region in the scene. Visual saliency in this context is involuntary and depends only on visual attributes of the region. (we would say that the region stands out from its surrounding or « jumps out » to the observer's eyes. This is called *Pop-out effect*.).
2. Top-down process: Top-down process, also called descending process (from the brain to oculomotor muscles), are endogenous mechanisms dependent on the willingness of the observer and on what he's searching for in the scene. These are influenced by different factors such as the tasks entrusted to the human observer, its experience and the semantic of the stimuli.

Studies of these attentional process have shown that *bottom-up* process are faster and precede the *top-down* process which are more slower to enforce (Wolfe et al. 2000) (Parkhurst et al. 2002) (Tatler et al. 2005).

This chapter is interested on the *Bottom-up* process which is induced by properties of the visual stimuli and that do not depend on the human observer. These will permit to detect perceptually salient regions that are able to guide visual attention of the human observer.

2.2 THEORY OF VISUAL ATTENTION

It's important to present some elements of the visual attention theory in order to understand how are processed some visual attributes and how certain prevail over others. This chapter will present the *Feature Integration Theory (FIT)* proposed by Treisman & Gelade (1980) as well as the Guided Research Model (GS) which is a modification of the *FIT* proposed by Wolfe et al. (1989).

2.1.1 THEORY OF FEATURE INTEGRATION

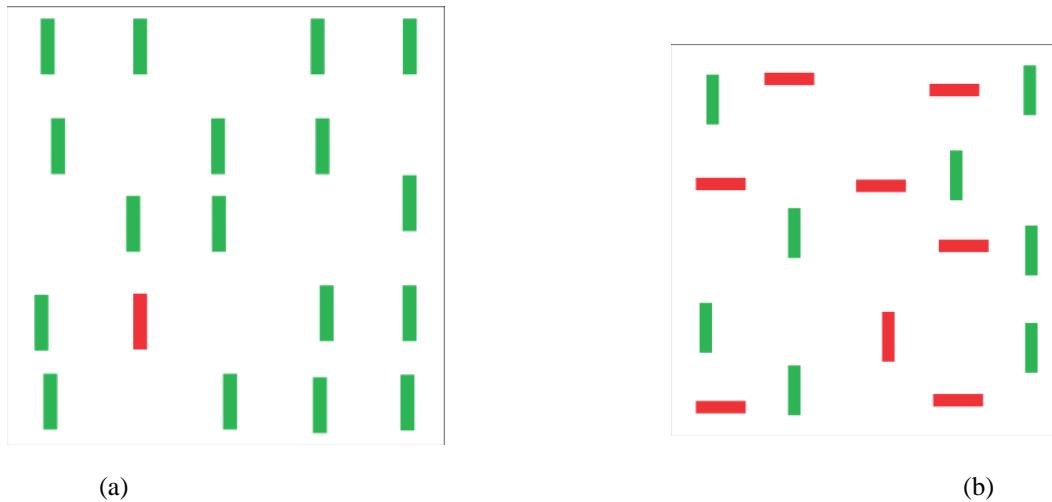


Figure 1. Examples of stimuli used in the research of a target (Wolfe et al. 2010): a) The target has a different color from the distractors. The research in this case is called disjunctive. b) The target shares two visual features with the distractors such as the red colors with the red distractors and the orientation with the green distractors. In this case, the research is called conjunctive.

In order to represent the visual attention process, Treisman analyses the mechanisms involved in the processing of a visual stimuli to verify if its features are processed in parallel or in series. For this, he measures the reaction time to find among the distractors a target that is visually different from all the others and puts forward two suppositions: 1) If the analysis of a visual feature can be done simultaneously in the visual field, then the detection of a target containing this feature will take a constant time regardless the number of distractors. 2) If the research of a target having a particular visual feature is sequential, then, the required time to find this target should increase linearly with the number of the distractors. Consequently, Treisman defines two types of research associated to two experiments that will initiate the *FIT*:

- The disjunctive research: takes place when the target is distinguished from the distractors by a visual feature. Figure 1 presents green distractors oriented vertically and a red target oriented similarly to the distractors. In this case, the reaction time (of the research) is constant regardless of the number of the distractors. The target strongly stands out from the distractors and guides the visual attention. This kind of research occurs in parallel and is qualified as pre-attentive. It can be associated to the Bottom-up mechanisms.
- The conjunctive research: takes place when the target and the distractors share many visual features. Figure 1 (b) presents a target represented by a red bar oriented vertically which has the same orientation of the green distractors and the same red color of the distractors oriented horizontally. In this case, the reaction time increases with the number of the distractors and becomes more important. This kind of research can be considered as sequential and calls in addition a voluntary attention. It can be associated to Top-down mechanisms.

Many experiments of Treisman conclude that the color, the orientation, the size, the shape, the intensity and the motion are part of the elementary attributes for which the research is disjunctive. These are called salient attributes.

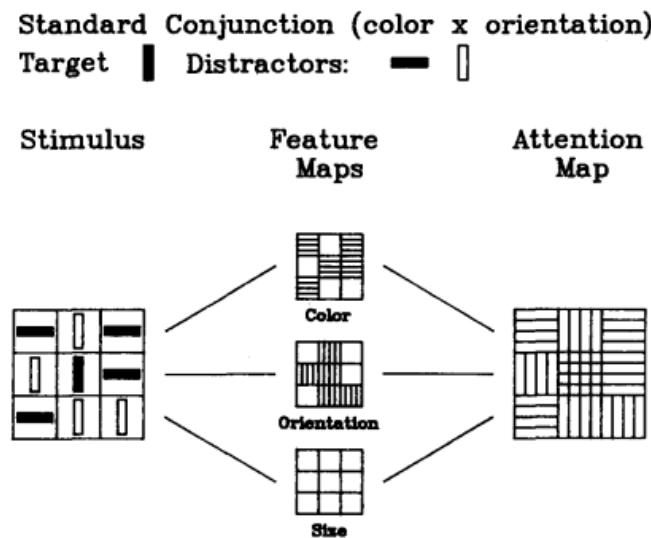


Figure 2. Color features map, orientation features map, size features map and their fusion into an attention map according to the *Guided Research model* in a conjunctive research (Wolf et al. 1989).

The *FIT* theory states that visual information is decomposed into distinct elementary attributes in the visual cortex. These attributes are processed in parallel and are represented in specific maps of attributes (figure 2). The different maps are then merged into a master card called *Master Saliency Map*. Regions that are highlighted on this map refer to salient regions.

2.1.1 THEORY OF FEATURE INTEGRATION

The *Guided Research Model* follows the *FIT*'s principles (elementary attributes extraction into specific attributes maps then grouping into an activation global map (figure 2)) while demonstrating that certain conjunctive researches take place faster than what is proposed in the sequential research of the *FIT* theory (Wolfe et al. 1989). The *Guided Research Model* specifies that during a conjunctive research, the stimuli are classified into distractors and potential targets and are processed in parallel. These targets are then analyzed sequentially which leads to a fast sequential research in comparison to a random sequential research.

In another study (Wolfe and Horowitz 2004), Wolfe and Horowitz inventory perceptual attributes that are able to guide the visual attention. Attributes such as color, motion, orientation have been judged capable of directing the attention.

2.1.2 EYE MOVEMENTS

There are links between the visual attention and ocular movements. This chapter presents the different types of attention as well as the ocular movements which are associated.

2.1.2.1 EYE MOVEMENTS AND VISUAL ATTENTION

There are two types of visual attention:

- **Covert attention:** Studies have established that the attention isn't necessarily linked to eye movements and can move independently (Hederson, 2003) (Liversedge and Findlay 2000). This kind of attention occurs when a distinct region is put forward without moving the eyes to it.

- Overt attention: This kind of attention occurs when the observer moves its eyes toward a region on which the attention is focalized. Deubel et al. (1996), Hoffman and Subramaniam (1995), Kowler et al. (1995) show that the eye movements toward a region are preceded by a movement of the attention in the direction of that region.

Findlay (2004) suggests that the study of attention as Overt or Covert processes in a context that contains complex or ecologic stimuli is wrong. This is in accordance with the study of Rizzolatti et al. (1994) which show that Overt and Covert attentions are controlled by extremely related mechanisms.

This chapter is interested on declared visual attention (Overt attention) which is followed (simultaneously) by ocular movements. With other terms, towards which regions the attention of human observers is moved and then their glance? Visual saliency models answer this question and predict the phenomena of saliency associated to these regions.

2.1.2.2A BETTER ACUITY IN THE CENTER OF THE GLANCE

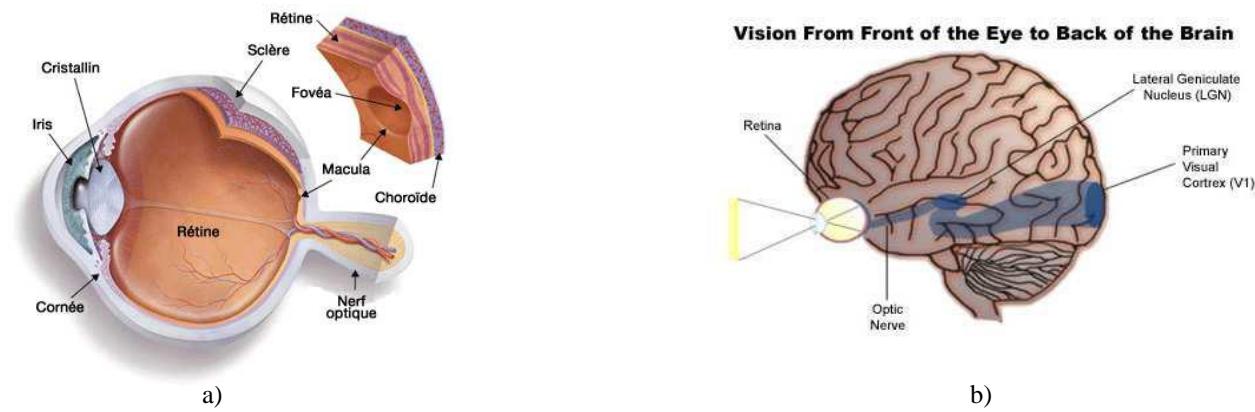


Figure 3. Human visual system: a) Retina (vision du futur). b) Cortex (Human-eye blog).

The HVS can be divided in two principal stages: the stage that relates the eye to the retina and the stage that processes the information coming from the couple eye-retina (figure 3). The first stage captures and pre-treats the stimuli of the surrounding world then sends it to the second stage for decomposition and an analysis.

The Fovea, which constitutes the central area of the retina, contains a large number of the 130 millions photoreceptors distributed non uniformly on the retinal field. This difference in distribution plays a preponderant role in the human visual acuity. Regions that are fixed and placed in the center of the retina (center of the glance) are analyzed more in details and benefit from a better resolution than regions of the scene located in the periphery of the retina. This leads to the need of moving the gaze in order to place the target region we want to analyze minutely in the center of the retina.

The processing of the information by the primary visual cortex (V1) also confirms the necessity of placing the target regions in the center of the gaze. Indeed, the processing of the visual information coming from the Fovea occupies 50% of the area of the primary cortex (V1) while the 50% remaining treat the rest of the visual field.

2.1.2.3 DIFFERENT TYPES OF OCULAR MOVEMENTS

The ocular movements represent the eyes activity and can be measured by oculometry techniques. For this, different approaches exist, however the most used because of its non-invasive character consists in filming the gaze of the observers using a camera placed on the screen. As a consequence, different ocular movements of the eye have been identified (Widdel, 1984):

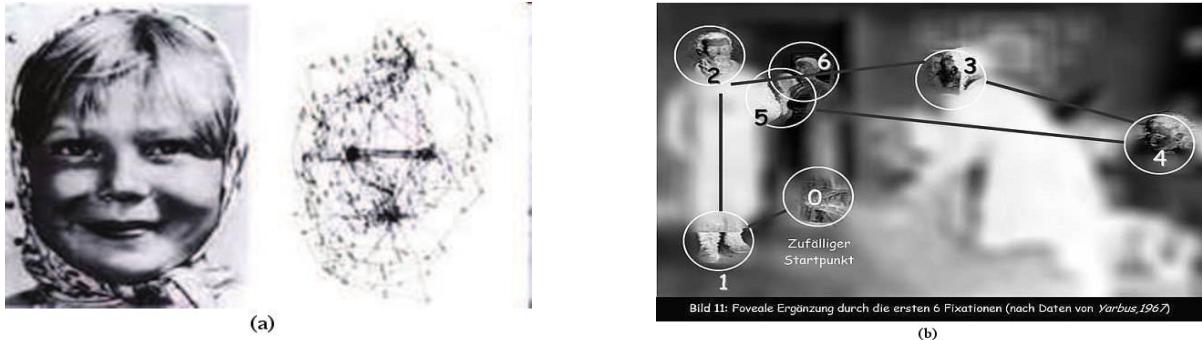


Figure 4. Fixations points and ocular path (Yarbus, 1967): a) Face observation with an eye-tracker and b) illustration of an ocular path.

- Jerks: are very fast movement of the eye (between 30 and 80ms) whose purpose is to place the target region of the scene in the centre of the fovea. Between two jerks, the eye stops for a small period of time between 200 and 250ms on a specific point of the scene. This point is called fixation point. Figure 4 illustrates these fixation points. The net regions refer to the regions that have been fixed by the observers (figure 4 (b)).
- Continuous movements of pursuit: take place when the observer looks to an object in motion. At this moment, the SVH extracts a maximum of informations while placing the object in the center of the fovea. In contrary to jerks, these movements are continuous and are corrected continually.
- Microjerks: are movements carried out without any interruptions. Without these movements, the observed image will not be « refreshed » or « reloaded ». These jerks also avoid the exhaustion of the photoreceptors.

3. VISUAL SALIENCY DETECTION

The saliency prediction models of 2D images outnumber those of 3D meshes (Liu et al., 2014). However, we can notice an evident resemblance between all proposed models. Almost all the approaches are based on the FIT or GS of Wolfe. The reader can refer to two reference models of saliency detection in 2D images: (Itti et al., 1998) and (Achanta et al., 2009).

In the following, a state-of-the-art of saliency models related to 3D non colored meshes is developed.

3.1 VISUAL SALIENCY OF 3D MESHES

Some previous research works on the visual saliency of 3D non colored meshes took back saliency models defined for 2D images and applied them to 2D projections of the target 3D meshes. This approach doesn't take into account sufficiently the relief of the 3D mesh which nevertheless constitutes an important feature in the perception of 3D multimedia contents (Howard, 2002).

This section presents model-based approaches that use the geometry of the 3D mesh for the saliency prediction. More details on the approach that take back the 2D saliency models can be found in (Guy and Mediouni, 1997) and (Yee et al., 2001). The models of (Lee et al., 2005), (Tal et al., 2012), (Song et al., 2014) and (Nouri et al., 2015) will be presented in details as they are used for the comparison with the approach of Nouri et al. (2015a).

3.1.1 DIFFERENT SALIENCY MODELS OF THE STATE-OF-THE-ART

In (Wu et al. 2013), Wu et al. detect salient regions using a descriptor based on the neighborhood's height field of each vertex; a squared map of projection heights is generated to represent its configuration (Maximo et al. 2011). Then, the Zernike moments are extracted from these maps to obtain a representation invariant to rotation. To obtain a multi-scale descriptor, the authors vary the size of the maps. The local saliency is first computed after a segmentation of the surface mesh. The saliency of a vertex is obtained using an interpolation of the saliency of the neighboring patches. Furthermore, the global saliency is computed after a gathering of similar vertices into patches.

The global saliency of each vertex is then computed as the interpolation of the degrees of saliency of the closest patches. The final degree of saliency of a vertex is obtained by combining and normalizing values of global and local saliency.

Zhao et al. (2013) propose a sampling-based saliency detection method for the simplification of 3D non colored meshes. The method begins by applying a Gaussian filter on vertices of the 3D mesh, then, features representing the curvature and its directions are computed on different scales. The obtained maps are filtered with a median filter before being combined to produce the final saliency map.

Acting on same principle, Zhao et al.(2012) propose a saliency model based on the diffusion of the surface index parameter by a non local filter (Buades et al. 2005).This method was used for 3D mesh registration and 3D mesh simplification.

In (Zhao & Liu, 2012), Yitian and Lui detect salient region by transforming the 3D mesh into a 3D volume. The patch based method begins by filtering the 3D mesh in order to suppress high frequencies, then, similarities between vertices are computed. Afterwards the mesh is transformed into 3D volumetric data on different scales. The dissimilarity between two patches localized into two sub-voxels provides the dissimilarity map. Finally, the saliency of a patch proportional to its dissimilarity is defined as the mean of its saliency degree over all considered scales.

With the integration of the CRF (*Conditional Random Field*) in a saliency detection model, Song et al., 2012 propose an approach generating firstly a multi-scale representation of the target 3D mesh by applying a set of Gaussian filters on neighborhoods delimited by a geodesic ray. The difference of Gaussians are computed into each scale and represent the displacement of a vertex after the filtering operation. Subsequently, these are projected on the normal vector of the target vertex to obtain the map of the considered scale. After the computation of the different maps, these are integrated into a CRF with a consistency constraint between the neighboring vertices in order to increase the robustness of the labeling process (labeling of salient regions and non salient regions). Finally, each vertex is labeled in the CRF using the Belief Propagation algorithm.

Zhao et al. (2013) select points of interest using visual saliency. To obtain robust points of interest, noise on the surface mesh is suppressed by applying a bilateral filter on the vertices normals. Then, Retinex theory (Elad, 2005) is performed to strengthen local details and to estimate invariant properties of the the surface's points of views. After the surface segmentation, saliency is estimated according to the spatial distance between the obtained segments.

Tao et al. (2014), compute Zernike coefficients for patches obtained after an over-segmentation of the surface mesh. Then, after measuring the distinctness for each patch, its saliency is estimated based on its relevance to the most unsalient patches via manifold ranking.

3.1.2 REFERENCE SALIENCY MODELS OF THE STATE-OF-THE-ART

Model of Lee et al.

Lee et al. (2005) measure saliency using a center-surround operator on Gaussian curvatures in a *DoG* (*Difference of Gaussian*) scale space. The approach begins by computing a curvature map C and defines a neighborhood $N(v)$ around each vertex:

$$N(v, \sigma) = \{x \mid \|x - v\| < \sigma\}$$

where σ represents the Euclidean distance between the target vertex v and its neighboring vertices x . Then a Gaussian-weighted mean curvature $C(x)$ is computed for each vertex:

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

The mono-scale saliency of a vertex is defined as the absolute differences of the Gaussian-weighted curvature at the fine and coarse scales:

$$S(v) = |G(v, \sigma) - G(v, 2\sigma)|$$

Otherwise, the multi-scale saliency of a vertex v at the scale i is defined as:

$$S_i(v) = |G(v, \sigma_i) - G(v, 2\sigma_i)|$$

where σ_i represents the stand-deviation of the Gaussian filter at the scale i . Five scales are used $\sigma_i \in \{2\delta, 3\delta, 4\delta, 5\delta\}$ where $\delta = 0.3\%$ of the bonding-box's diagonal of the 3D mesh.

Once the saliency maps S_i are computed, a non linear suppression operator is applied in order to only consider the saliency maps with few pics of saliency. Each map is normalized and the maximal saliency value M_i as well as the mean of maxima \bar{m}_i are computed. Finally, S_i is weighted by $(M_i - \bar{m}_i)^2$ and the final saliency map S_{final} is defined as the sum of the saliency maps over all the scales: $S_{final} = \sum_i S_i$.

Figure 5 presents a synopsis of this method.

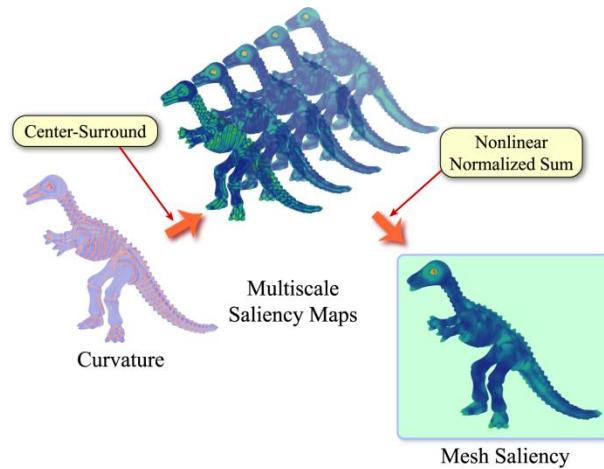


Figure 5. Synopsis (Lee et al., 2005)

Model of Tal et al.

Tal et al. (2012) detect regions of interest of a 3D non colored mesh by the use of the Spin Image descriptor (Johnson and Hebert, 1999) to encode the geometry of the surface mesh for local and global distinctness computation. The extremities of the 3D mesh are also taken into account in the pipeline of this method. The Spin Image descriptor provides a histogram that characterizes the local geometry of each vertex.

To compute the dissimilarity of two histograms $D(h_1, h_2)$, Tal et al., (2012) use the diffusion distance (Ling and Okada, 2006):

$$D(h_1, h_2) = \sum_{l=0}^L k(d_l)$$

where $d_0 = h_1 - h_2$ and $d_l = [d_{l-1} * \phi(\sigma)] \downarrow_2, l = 1, \dots, L$ are the different levels of the Pyramidal Gaussian used for the discretization of the continuous diffusion process, L represents the number of levels of the pyramid, σ is the standard-deviation of the Gaussian filter ϕ , $k(\cdot)$ represents the L_1 norm, and \downarrow_2 refers to a down-sampling by 2.

A vertex v_i is considered as distinct if its similar vertices v_j are close. Otherwise, a vertex is less distinct if similar its vertices are far. Consequently, the dissimilarity measure is inversely proportional to the geodesic distance between these vertices:

$$d(v_i, v_j) = \frac{D(h(v_i), h(v_j))}{1 + c \cdot \text{GeodDist}(v_i, v_j)}$$

where $\text{GeodDist}(v_i, v_j)$ represents the geodesic distance between vertices v_i and v_j and $c = 3$.

The mono-scale distinction of the vertex v_i is then computed as:

$$D(v_i) = 1 - \exp \left\{ -\frac{1}{K} \sum_{k=1}^K d(v_i, v_k) \right\}$$

where K is the number of vertices similar to v_i over the surface mesh.

In order to reduce the distinctiveness of vertices located in a textured region (similar vertices constituting a large region), Tal et al. propose to compute vertices distinctions considering three scales ($F, F/2, F/4$) with F representing the faces number. The multi-scale distinction D is defined as the mean of the distinction values over the three scales.

In regards to the detection of the 3D mesh's extremities, this is done in three steps:

1. Application of the MDS (Multi-Dimensional Scaling) transformation to the 3D mesh in order to equalize the Euclidean distance and the Geodesic distance.

2. The presence of an extremity on the surface mesh is defined by the ratio $\frac{V_{CH}}{V_o}$

where V_{CH} represents the volume of the convex envelope of the transformed 3D mesh and V_o is the volume of the original 3D mesh. If $\frac{V_{CH}}{V_o} > 1.5$, then the 3D mesh contains extremities. Indeed, the

volume of a round object without extremities is equivalent to the volume of its convex envelope. Conversely, if the 3D mesh contains extremities, the volume of the convex envelope of its transformed version is different from its original volume.

3. A vertex v is located at the extremity of a 3D mesh M if $\forall v_n \in N_v$, where N_v represents the neighborhood of v :

$$\sum_{v_j \in M} GeodDist(v, v_j) \geq \sum_{v_j \in M} GeodDist(v_n, v_j)$$

Furthermore, this method considers vertices around which the shape of the 3D mesh is constructed named « centers of attention » (20% of distinct vertices of the mesh are considered). Regions that are close to the centers of attention are considered as regions of interest in comparison to the far ones. A vertex v_i is associated to its center of attention as the following:

$$A(v_i) = D_{foci}(v_i) \exp \left\{ -\frac{GeodFoci^2(v_i)}{2\sigma^2} \right\}$$

where $GeodFoci(v_i)$ represents the geodesic distance between v_i and the closest center of attention, $D_{foci}(v_i)$ is the distinction value of this center of attention and $\sigma = 0.5$.

The vertices located at extremities of the 3D mesh are considered as centers of attention. Therefore, for each vertex, its geodesic distance to an extremity representing its degree of extremity is defined as:

$$E(v_i) = \exp \left\{ -\frac{GeodExt^2(v_i)}{2\sigma^2} \right\}$$

Finally, the degree of interest of a vertex on the surface mesh is defined as the maximum of its distinction and its degree of extremity:

$$I(v_i) = \max \left(\frac{D(v_i) + A(v_i)}{2}, E(v_i) \right)$$

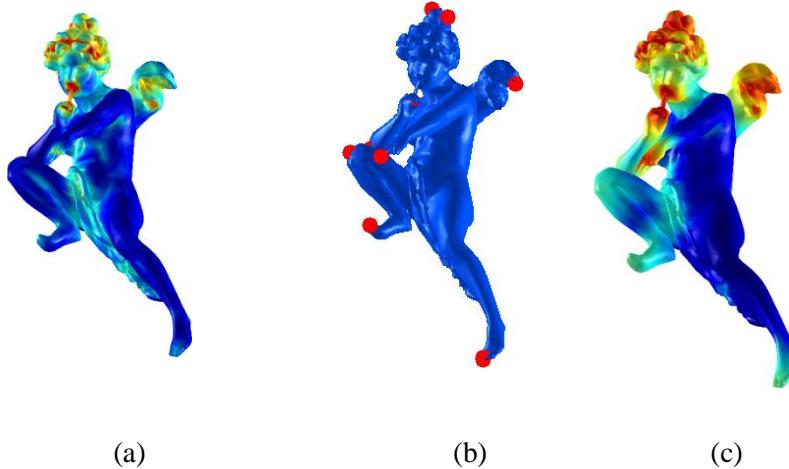


Figure 6. Synopsis (Tal. et al, 2012)

Figure 6 presents a synopsis of the method of (Tal et al. 2012).

Model of Song et al.

Song et al. (2014) predict saliency in the spectral domain using the analysis of the Log-Laplacian spectrum. The underlying hypothesis is that the geometric Log-Laplacian spectrum of a 3D mesh comprises attributes exploitable for saliency estimation.

For a 3D mesh M with m vertices $\{p_1, \dots, p_m\}$, the approach begins by computing a simplified version of the mesh \hat{M} using the QSLIM algorithm (Garland and Heckbert, 1997). Then, a set of smoothed meshes are generated using Gaussian filters of different parameters t_s with $s=1, 2, \dots, 5$:

$$D(p, t_s) = F(p, k t_s) - F(p, t_s)$$

where k is a constant, $p \in M(t_s)$ and F is a linear representation of the scale space defined by the following convolution:

$$F(., t_s) = U(.) \otimes g(., t_s)$$

where U is a signal of dimension d and g is Gaussian kernel with a standard deviation $\sigma = \sqrt{t_s}$ defined as:

$$g(p, t_s) = \frac{1}{(2\pi t_s)^{d/2}} \exp(-p^T p / (2t_s))$$

The neighborhood of the vertex p around which the Gaussian filter is applied contains all vertices located at a distance equal to $2.5\sqrt{t_s}$.

A second set of smoothed meshes $\hat{M}(k(i)t_s)$ is generated using a dynamic scale space:

$$D(p_i, t) = |F(p_i, k(i)t) - F(p_i, t)|$$

where the parameter k depends this time on the vertices density on the surface mesh:

$$k(i) = \frac{cn}{\sum_{j \in N(i)} \|p_i - p_j\|} + 1$$

where n represents the number of adjacent vertices to the vertex i , and c is a normalization constant equal to the mean of the distances between the vertices of the 3D mesh.

Similarly to the approach of (Lee et al., (2005)) for the multi-scale saliency detection, five scales are fixed $t \in \{e^2, 2e^2, 3e^2, 4e^2, 5e^2\}$ where e is equal to 0.2% of the bounding-box diagonal of the mesh.

Once the filter bank of the 3D mesh is constructed, the mono-scale saliency is performed as follows. First, the mesh Laplacian L is computed for a mesh of $\hat{M}(t_s)$:

$$L = A - D$$

where A represents the adjacency matrix between vertices:

$$A(i, j) = \begin{cases} 1 & \text{if } p_i \text{ and } p_j \text{ are adjacent vertices,} \\ 0 & \text{sinon} \end{cases}$$

And D is a diagonal matrix in which D_{ii} represents the degree of the vertex p_i . In order to take into account the local geometric information, the adjacency matrix is weighted by the distance between the adjacent vertices:

$$W(i, j) = \frac{1}{\|p_i - p_j\|^2} A(i, j) \quad W(i, j) = \frac{1}{\|p_i - p_j\|^2} A(i, j)$$

The geometric Laplacian is then defined as:

$$L = W - D$$

and the associated spectrum as:

$$H(f) = \{\lambda_f, 1 \leq f \leq m\}$$

where λ_f represents the eigenvalues (frequencies) of the geometric Laplacian ordered increasingly.

It has been observed that the Laplacian spectrum of similar 3D meshes share strictly similar redundancy (Hou and Zhang, 2007) and that the detection of atypical information on the surface mesh represents saliency. To amplify local variations and global deviations at the level of low frequencies and discard the rest, a logarithmic transformation is applied to the spectrum:

$$\iota(f) = \log(|H(f)|)$$

Once the transformation performed, the goal is to detect less common frequencies in the spectrum. Song et al. (2014) propose to locally average and smooth the spectrum. Then, significant frequencies that are different from the local mean are identified. For this, a local average filter $J_n(f)$ is used:

$$A(f) = J_n(f) * \iota(f)$$

where $J_n(f) = \frac{1}{n}[1 \dots 1]$ is a vector of size $n \times 1$. The spectral deviation is defined by the irregularity R :

$$R(f) = |\iota(f) - A(f)|$$

The irregularity spectrum in the spatial domain is defined as:

$$S = BRB^T W$$

where $R = \text{Diag}\{\exp(R(f)) : 1 \leq f \leq m\}$ is a diagonal matrix where the values represents the exponentials of

Each line of the matrix S is associated to a vertex of surface mesh, consequently, the saliency of a vertex $S(i)$ is defined as the sum of the values of S along each line. This process is used for computing a saliency map $\tilde{S}(i, t)$ at each scale t using the degrees of saliency of scales $k(i)t$ and t :

$$\tilde{S}(i, t) = |S(i, k(i)t) - S(i, t)|$$

The multi-scale saliency map of the simplified 3D mesh is computed by summing all the mono-scale saliency maps. A saliency mapping is performed toward the reference 3D mesh using *K-d-tree*. The final saliency map is finally smoothed to avoid discretization problems related to the simplification step.

3.1.3 DISCUSSION

From the above state-of-the-art description, one can observe that a simplification step is integrated in the pipeline of some saliency models. Despite the swiftness provided by this simplification, the latter removes inevitably vertices of the mesh geometry that could have a high degree of saliency and therefore distorts the mesh surface by deleting initial fluctuations. This results in a saliency measure that does not take into account of all local irregularities and exiguous variations yet necessary for a precise estimation of saliency. Others steps such as smoothing and the segmentation lead to a measure of saliency of high complexity.

We can also highlight from the defined state-of-the-art that only one method (Song et al., 2014) validates qualitatively its saliency detection results with a pseudo ground truth saliency.

3.2 MODEL OF (NOURI ET AL., 2015A): MULTI-SCALE VISUAL SALIENCY OF NON COLORED 3D MESHES WITH LOCAL ADAPTIVE PATCHES

3.2.1 MOTIVATION

Nouri et al. (2015a) propose a new saliency model independent of any pre-treatment or post-treatment such as remeshing, simplification, smoothing or segmentation. This capability provides on the one hand a saliency model measuring precisely the initial degree of saliency and on the other hand a model the complexity of which is not important.

Many studies have shown that the visual saliency is associated to high contrast (Wolf, 1994) and strong discontinuities (Coren et al. 2003). For example, a flat surface would be visually less interesting than a surface with a unique fluctuation. Other studies conclude that the HVS is less sensible and suppress redundant patterns (Koch and Poggio, 1999). Figure 7 represents these two aspects in the context of 3D meshes. Figure 7 (a) and figure 7 (b) respectively present a flat surface and a surface with strong discontinuities. The glance of a human observer will be more attracted by the surface containing a high discontinuity than the flat surface.

Figure 7 (c) presents the paw of a 3D dinosaur mesh. The surface of this paw contains redundant bumps. In this case, the glance of a human observer would be more attracted by regions different from these redundant bumps.

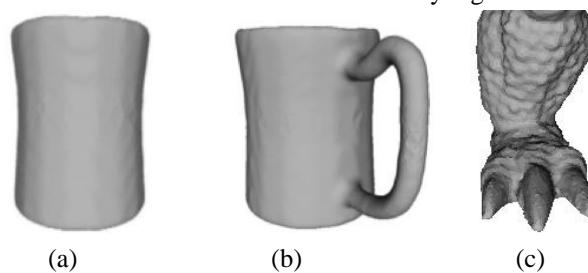


Figure 7. Sensitivity of the HVS to high contrast and redundant patterns: a) Flat surface, b) surface with a strong discontinuity and c) surface with redundant patterns.

Nouri et al. (2015a) propose a saliency model which integrate these SVH features and consider that a vertex on the surface mesh is salient if it stands out from its surrounding and if its local geometric configuration is different from the one of its adjacent vertices. Figure 8. presents a synopsis of this method.

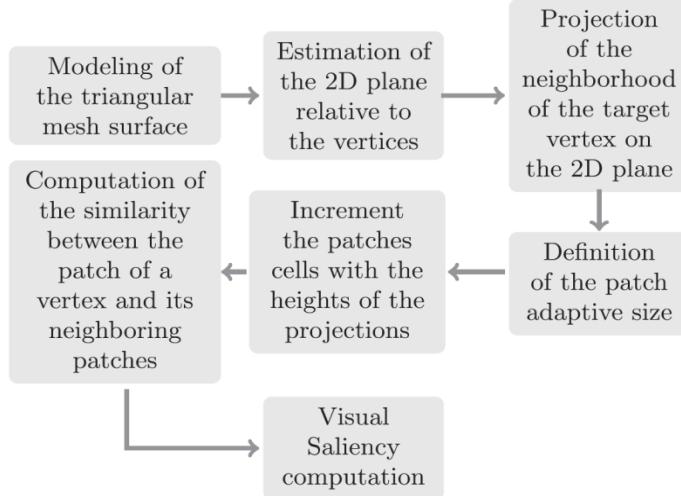


Figure 8. Synopsis of the approach proposed in (Nouri et al., 2015a).

3.2.2 CONCEPTS AND NOTATIONS ON GRAPHS

Nouri et al. (2015a) represent a 3D non colored mesh M by a non oriented graph $G = (V, E, w)$ where $V = \{v_1, \dots, v_N\}$ is a set of N vertices, $E \subset V \times V$ is the set of edges deduces from the triangular faces connecting the vertices and $w: E \rightarrow \mathbb{R}$ represents a symmetric positive weighting function. The concerned graph is non directed (*i.e.*, for each $x, y \in V$, $(x, y) \in E$ if and only if $(y, x) \in E$). For each vertex of mesh surface are associated 3D coordinates $\vec{p}_i = (x_i, y_i, z_i)^T \in \mathbb{R}^3$. The notation $v_i \sim v_j$ refers to two adjacent vertices in G (*i.e.*, $(v_i, v_j) \in E$).

3.2.3 3D SURFACE MODELING

In order to construct local descriptors, Nouri et al., (2005a) model the surface mesh by computing a vector representing the normal $z(v_i)$ on each vertex v_i and its directional vectors $x(v_i)$ and $y(v_i)$ for the tangent plane estimation. The method defines a sphere S_ϵ of ray ϵ centered in v_i . Vertices belonging to this sphere are considered as neighboring vertices of the target vertex v_i . These neighboring vertices are then used for the computation of the center of gravity \hat{v}_i defined as:

$$\hat{v}_i = \frac{1}{|S_\epsilon(v_i)|} \sum_{j \in S_\epsilon(v_i)} \vec{v}_j \quad (1)$$

And used for the computation of the covariance matrix associated to v_i defined as:

$$cov(v_i) = \sum_{j \in S_\epsilon(v_i)} (\vec{v}_j - \hat{v}_i)(\vec{v}_j - \hat{v}_i)^T \in \mathbb{R}^{3 \times 3} \quad (2)$$

where $|S_\varepsilon(v_i)|$ is the cardinality of the spherical neighborhood S_ε . Eigenvectors of the covariance matrix are used to compute the vector representing the normal $z(v_i)$ and the two directional vectors x and y . A similar approach can be found in (Digne et al., 2012).

Hence the surface mesh is represented by vertices with their associated normal vectors and tangent planes. These normal vectors have different directions (outwards and inwards). In order to obtain vectors whose orientations are uniform, the orientation of a vector representing a normal which have been chosen arbitrary is propagated to the neighboring normal vectors. For this, the weight of the edge connecting two vertices v_i and v_j denoted $w(v_i, v_j)$ is defined as:

$$w(v_i, v_j) = 1 - |z_i^T z_j|, v_i \sim v_j \quad (3)$$

Where z_i and z_j represent the normal vectors respectively associated to the vertices v_i and v_j . Afterwards, a Minimum Spanning Tree is generated for the 3D mesh and the normals are re-aligned by going through the MST with a depth first search (Hoppe et al., 1992).

3.2.4 CONSTRUCTION OF THE LOCAL DESCRIPTORS: THE ADAPTIVE LOCAL PATCHES

Once the 3D mesh modeled, the local adaptive patches are constructed. These can be considered as an extension of the works proposed in (Maximo et al, 2011)(Digne et al., 2012) with an adaptive size. For this, the vertices contained in a sphere centered at v_i $S_\varepsilon(v_i) = \{v_j \mid \| \vec{v}_j - \vec{v}_i \|_2^2 \leq \varepsilon\}$ are projected on a 2D plane $\vec{P}(v_i)$ defined by the associated directional vectors. Therefore, 2D vectors are obtained \vec{v}'_j defined as:

$$\vec{v}'_j = [(\vec{v}_j - \vec{v}_i) \cdot \vec{x}(v_i), (\vec{v}_j - \vec{v}_i) \cdot \vec{y}(v_i)]^T \quad (4)$$

To define the patch size, Nouri et al., (2015a) propose a dynamic configuration depending on the distance between the 2D coordinates of the projected vertices \vec{v}'_j . The patch dimensions are defined according to the horizontal and vertical axes (respectively denoted $T_x(\cdot)$ and $T_y(\cdot)$) as:

$$T_d(v_i) = \max_{(\vec{v}'_j, \vec{v}'_k) \in \vec{P}(v_i)} (\| \vec{v}'_j^{d} - \vec{v}'_k^{d} \|_2^2) \quad (5)$$

where d represents the x or y coordinate, \vec{v}'_j^d is the coordinate d of the vector \vec{v}'_j and $\| \cdot \|_2$ is the Euclidean norm. Hence, the patch at v_i is represented by a rectangle of size $T_x(v_i) \times T_y(v_i)$ (note that usually a patch is represented by a square of fixed size (Maximo et al., 2011)(Digne et al., 2012) and does not allow to get an adaptive local descriptor depending on the local geometry). The obtained patch is then divided into $l \times l$ cells in order to precise the index of the cell in which a neighboring vertex is projected:

$$\text{indice}^d = \left\lfloor \frac{\vec{v}'_j^d}{T_d(v_i)/l} \right\rfloor \text{ with } v_j \sim v_i \quad (6)$$

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

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

$$\vec{H}(v_i) = \left(\sum_{\vec{v}'_j \in P_i^k} \|(\vec{v}_j - \vec{v}'_j)\|_2^2, \forall k \right)^T \quad (7)$$

where $\vec{H}(v_i)$ represents the vector of accumulated heights in the patch cells.

Figure 9 illustrates the construction process of the local adaptive patches and figure 10 presents three patches related to 3 vertices belonging to different regions: 1) Salient (finger of the 3D mesh) 2) Moderately salient (located at the neck or the shoulder of the 3D mesh) 3) Non salient (the flat region situated on the tibia of the 3D mesh). The red pixels in the constructed patch refer to important projections heights of the neighboring vertices associated to the target vertex. One can remark that the adaptive size of patches depends on the local configuration of the target vertex.

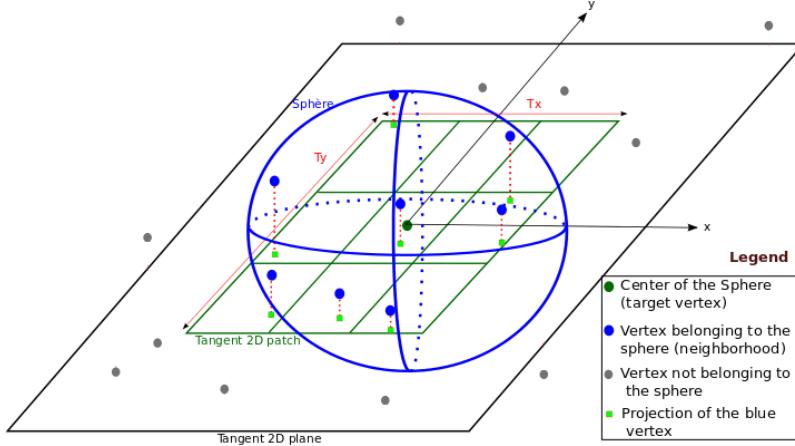


Figure 9. Local adaptive patch: blue points represent vertices located in the sphere S_ϵ where the center is represented in green. Vertices in gray don't belong to the sphere S_ϵ . Hence, only blue vertices are projected on the 2D plan. Projections in the form of a green square represent 2D projected vectors.

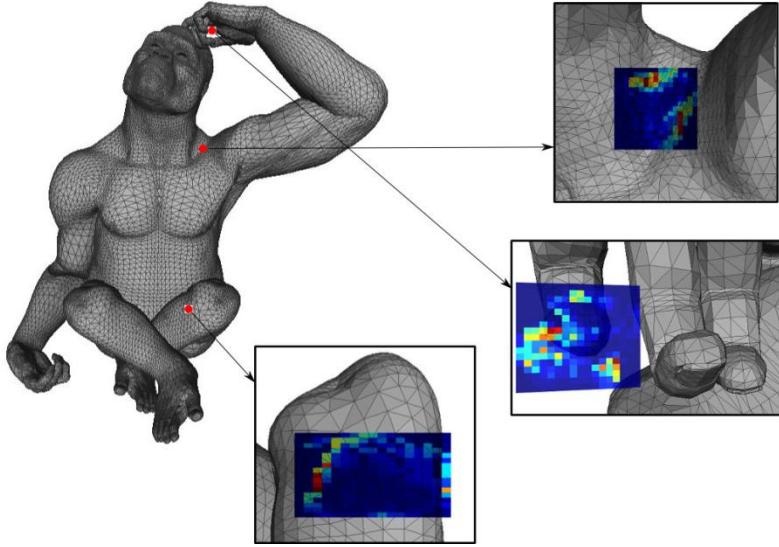


Figure 10. Constructed local patches associated to three vertices (red points) belonging to different regions on the surface of the 3D mesh: 1) salient region (finger of the 3D mesh) 2) moderately salient (located at the neck or the shoulder of the 3D mesh) 3) non salient region (the flat region situated on the tibia of the 3D mesh).

It's important to note that the adaptivity of the patches size isn't linked to the ray ε of the sphere, but rather to the maximal distance between 2D projections of the neighboring vertices along the x and y axes. This strengthens their adaptivity. This method also works for 3D point clouds instead of 3D meshes since the mesh topology isn't taken into account.

3.2.5 MONO-SCALE SALIENCY

To compute the mono-scale saliency of a target vertex, Nouri et al. (2015a) define a similarity measure between the patch of a vertex and the patches associated to its neighboring vertices. A scale parameter (standard deviation of a Gaussian kernel measuring the dissimilarity) is computed locally. Indeed, the use of a specific scale parameter for each vertex permits to take into account the local distribution. The scale parameter is defined as:

$$\sigma(v_i) = \max_{v_k \sim v_i} (\| \vec{v}_i - \vec{v}_k \|_2) \quad (8)$$

Note that a scale parameter computed using the Euclidean norm between patches instead of coordinate vertices was tested. However, this has led to less interesting results. Hence, the similarity affected to the weight of the edge $e(v_i, v_j)$ is defined as:

$$w(v_i, v_j) = \exp \left[-\frac{\kappa(v_j)^* \|\vec{H}(v_i) - \vec{H}(v_j)\|_2^2}{\sigma(v_i)^* \sigma(v_j)^* \|\vec{v}_i - \vec{v}_j\|_2^2} \right] \text{ with } v_j \sim v_i \quad (9)$$

where $\kappa(v_j)$ represents the curvature of the vertex v_j computed with the method of (Rusinkiewicz, 2004) and $\|\vec{v}_i - \vec{v}_j\|_2^2$ is the Euclidean distance between the vertices v_i and v_j . When the Euclidean distance between the

patches $\vec{H}(v_i)$ and $\vec{H}(v_j)$ is important, the similarity between the associated vertices v_i and v_j tends towards 0 (*i.e.*, they are dissimilar).

Finally the mono-scale saliency of a vertex v_i is defined by its mean degree:

$$\text{Mono-scale-saliency}(v_i) = \left(\frac{1}{|v_j \sim v_i|} \right) \sum_{v_i \sim v_j} w(v_i, v_j) \quad (10)$$

where $|v_j \sim v_i|$ represents the cardinality of the neighborhood containing adjacent vertices and $w(v_i, v_j)$ is the weight of the edge $e(v_i, v_j)$. This mono-scale saliency is defined in $[0 \ 1]$ where 0 refers to a high saliency degree (very dissimilar from its neighborhood) and 1 refers to a weak degree of saliency (very similar to its neighborhood).

3.2.6 THE CONTRIBUTION OF THE CURVATURE AND THE DISTANCE-COEFFICIENT WEIGHTS

Figure 11 presents the mono-scale saliency of the 3D mesh Gorilla with the method of Nouri et al.(2015a). One can remark that this method accurately highlights regions associated to the paws and toes and judge them as salient regions. Exiguous details such as the eyes, the nose, the mouth and the ears are well detected but with a certain imperfection at the level of the eyes. The saliency of the eyes is saturated. This will be corrected by the use of the multi-scale aspect.

It's interesting to analyze here the contribution of the curvature and the distance coefficient weights on the single-scale saliency rendering. Figure 12 presents a comparison between the detected saliency using the coefficient weights and the saliency without these. One can easily notice that the saliency estimated using the coefficient weights is widely finer than that one detected without. The surplus of saliency at the level of the ribs, the chest, the knees, the forearms and the paws has been corrected. Moreover, salient regions such as the nose, the eyes and the mouth have been preserved. Indeed, using vertex curvature (equation 9) permits to estimate the discontinuity between the target vertex and its adjacent vertices. A strong discontinuity will more contribute to the saliency detection than a weak one. Therefore, this parameter helped to eliminate the surplus of saliency.

The distance coefficient in the denominator of the equation 9 reduces the contribution of the remote vertices in the computation of patch similarities which also explain the deletion of excellent saliency.

3.2.7 THE INFLUENCE OF THE NUMBER OF CELLS AND THE NEIGHBORHOOD IN THE SALIENCY DETECTION

Two parameters affect the detection of saliency: the number of patch cells and the ray of the sphere . Table 1 presents the amplitude of the detected saliency according to the and parameters. One can remark that saliency detection depends strongly on both the neighborhood delimited by the sphere and the number of patch cells . In this table, a radius equal to 1 provides the detection of very small salient details. Remark that in this case, increasing the number of cells doesn't affect the saliency estimation. In contrast, a sphere with a radius equal to three greatly expands the neighborhood, which lead to an over description of the surface mesh and thus detecting large salient regions. Note that the number of salient vertices increases proportionally to the number of cells. Hence, the neighborhood have to be specified according to the application using saliency. Finally, these three radii of the sphere will permit to design a multi-scale saliency detection model based on the local adaptive patches.

3.2.8 MULTI-SCALE SALIENCY COMPUTATION

In order to enhance the quality of their saliency measure, Nouri et al. (2005a) propose to compute saliency at different scales. Saliency at primary scales will detect finest and cramped details, while higher scales will highlight large regions (see table 1). Another interest of the multi-scale aspect consists in the robustness of the approach to the noise since the latter is perceptible only on some scales. For this, three scales will highlight large regions (see Table 1).

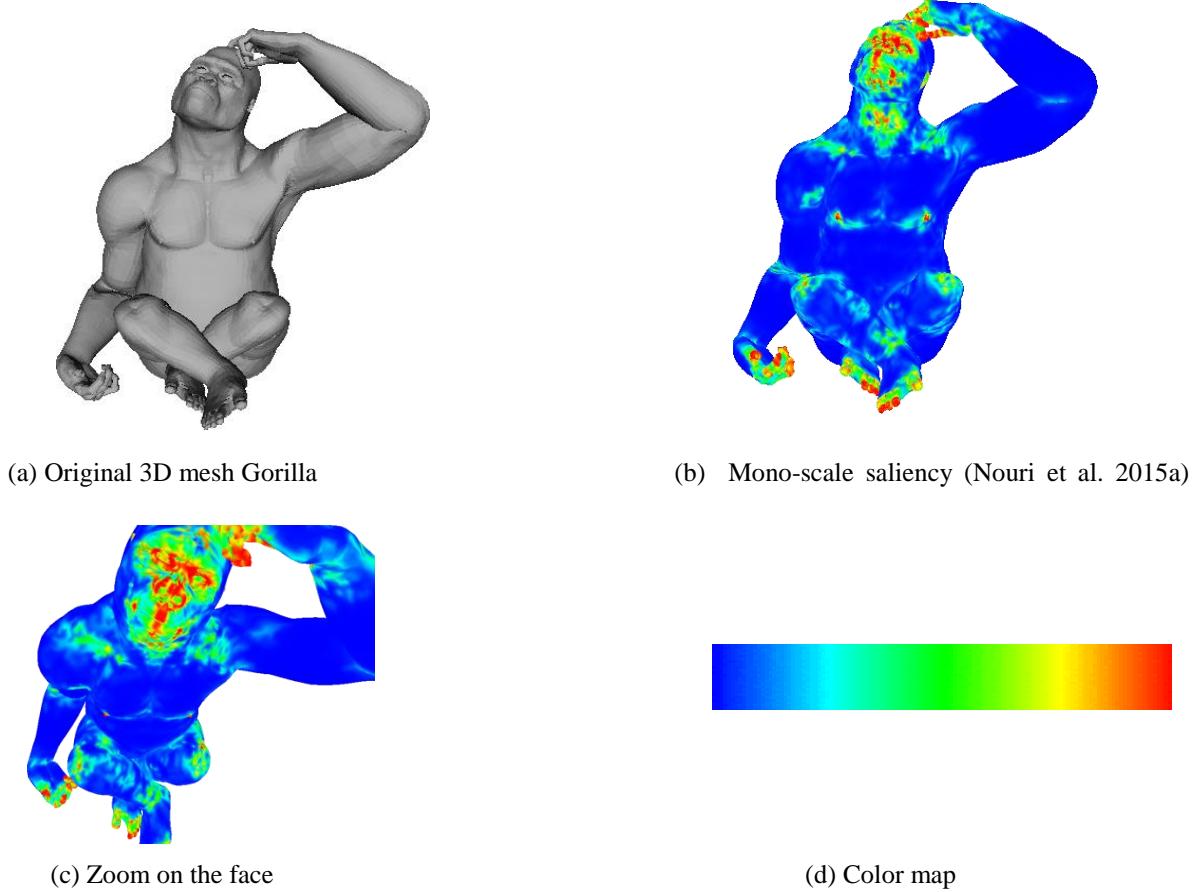


Figure 11. Mono-scale saliency (Nouri et al., 2015a). Red areas refer to very salient regions. Those in blue are not salient.

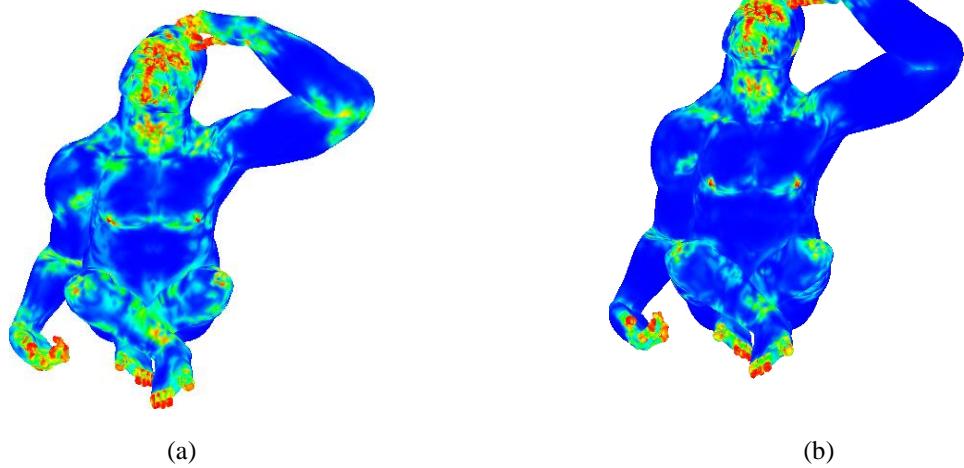
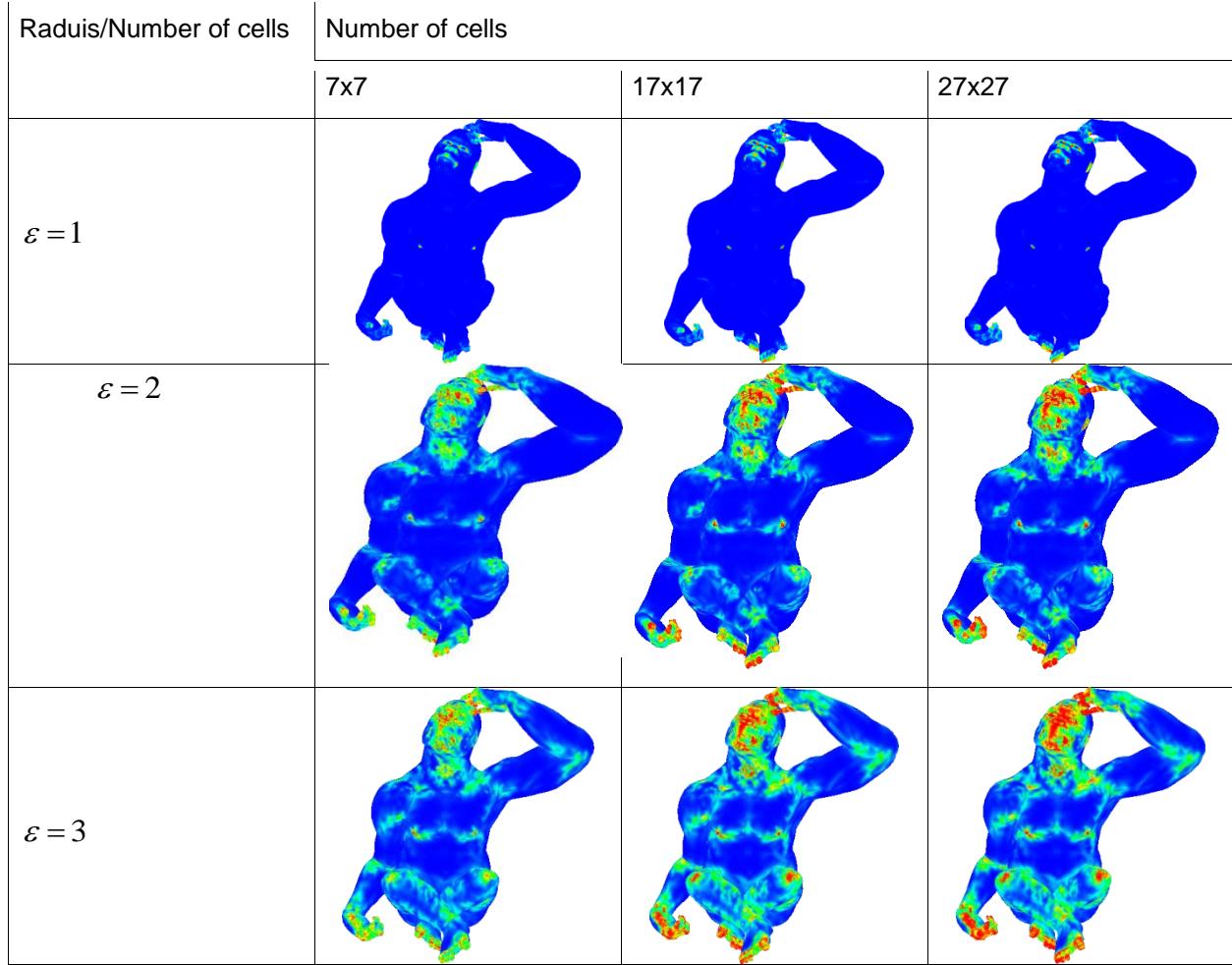


Figure 12. The contribution of the weighting parameters: (a) Saliency without the curvature and the distance coefficient weights ($\varepsilon = 2$ and $l = 17$) b) Saliency using the curvature and the distance weights ($\varepsilon = 2$ and $l = 17$).

Table 1. Influence of parameters ε and N in the saliency detection



Another interest of the multi-scale aspect consists in the robustness of the approach to the noise since the latter is perceptible only on some scales. For this, three rays are considered in the construction of the local adaptive patches. Three saliency maps are hence generated and merged taking into account their respective entropy. Indeed, the entropy weight permits to evaluate the disorder and the disparity of each generated map. For a defined scale k , histogram of saliency h_k is computed in order to obtain the probability of having a saliency value i :

$$Pr_{i,k} = h_k^i / |V| \quad (11)$$

where h_i^k represents the number of vertices of degree of saliency equal to i at a scale k . Hence, the entropy at a scale k is defined as:

$$\text{Entropy}_k = - \sum Pr_{i,k} * \log_2 Pr_{i,k} \quad (12)$$

By weighting the saliency degree of each vertex while merging the different scales with their respective entropy, a robust multi-scale saliency map is obtained which considers the saliency disparity on the different scales. The multi-scale saliency is defined as:

$$\text{Multi-scale-saliency}(v_i) = \frac{\sum_{k=1}^3 \text{Mono-scale-saliency}_k(v_i) * \text{Entropy}_k}{\sum_{k=1}^3 \text{Entropy}_k} \quad (13)$$

where k is the scale index.

Figure 13(b) presents the multi-scale saliency of the 3D Gorilla mesh. One can remark that the proposed multi-scale saliency model brings out finely the salient regions on the 3D mesh surface. Figure 13(c) shows the saliency detected at the level of the face. In comparison with the saliency result presented in figure 11, one can notice that the over estimation of saliency at the level of the eyes and nose has been corrected. Only the eyes are now judged as salient and not their peripheral zones (figure 13 (d)). Note that a precise detection of the eyes in scenes containing faces is mandatory since the eyes are a highly regarded parts (Henderson et al., 2005).

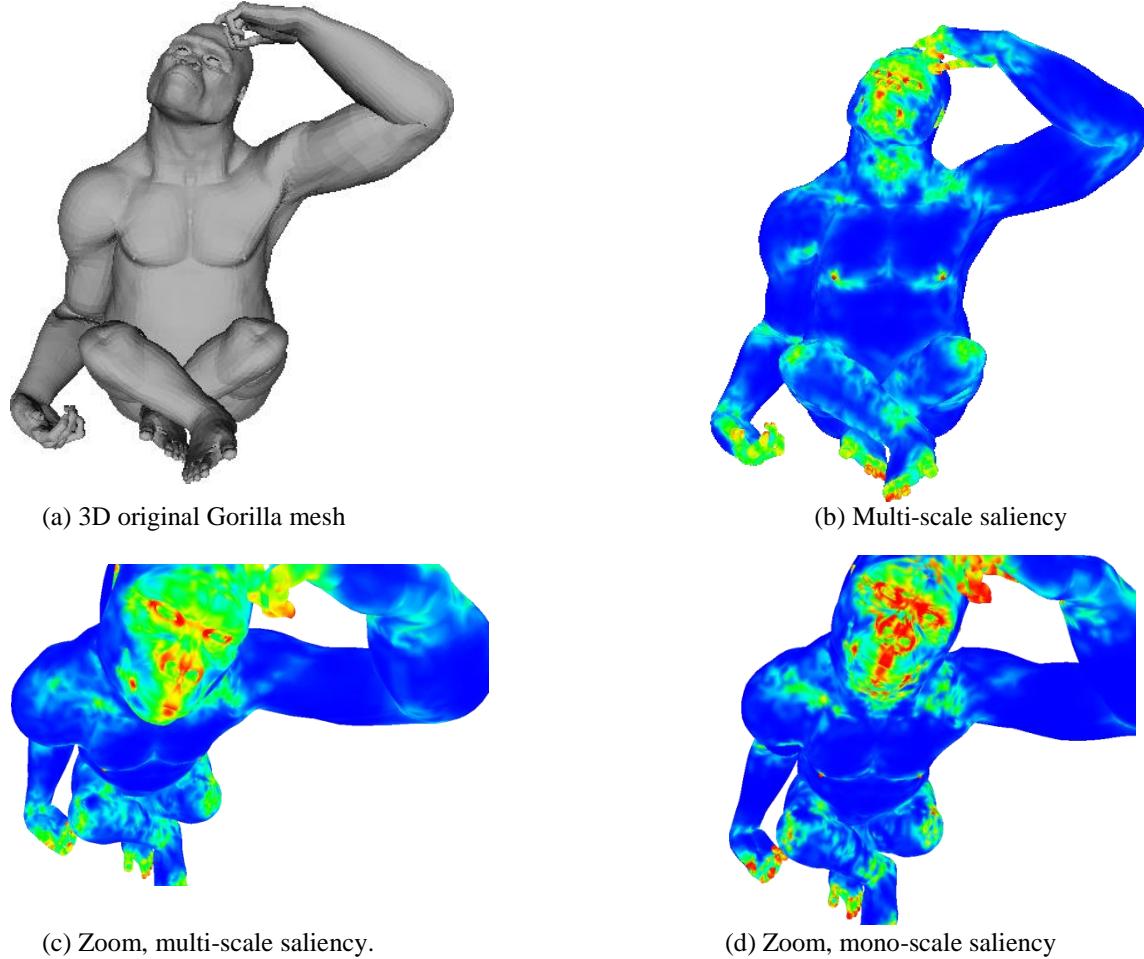


Figure 13. Multi-scale saliency detected with approach of (Nouri et al., 2015a).

3.3 MODEL OF (NOURI ET AL., 2015B): EXTENSION TO VISUAL SALIENCY OF 3D COLORED MESHES

Recent studies on the contribution of color in the orientation of attention have shown that colorimetric information modifies the eye movements while visualizing a multimedia content (Shahrbabaki, 2015). Nouri et al. (2015b) have proposed to extend their multi-scale saliency model to 3D colored meshes whose saliency detection has never been studied before. Indeed, 3D colored meshes are great of interest for many applications such as video games, civil engineering, 3D impression, etc. With the recent development of 3D scanners, it is possible now to acquire simultaneously the geometry and colors of an object or a scene. A color \vec{c}_i is provided for each vertex in the form of an RGB vector for the colorimetric description.

In their previous saliency model, Nouri et al. 2015a have only considered 3D coordinates \vec{p}_i of vertices. In order to extend their multi-scale saliency model to 3D colored meshes, they proceed as follows: a local adaptive patch is constructed similarly to the one constructed for 3D non colored meshes, however, its cells P_i are filled with the mean *RGB* colors of the projected vertices $\frac{1}{|\vec{c}'_j \in P_i|} \sum_{\vec{c}'_j \in P_i} \vec{c}'_j$ thus defining a color vector $\vec{C}(v_i)$ representing the local patch on each vertex. Edges are then weighted by the following similarity:

$$w_C(v_i, v_j) = \exp \left[-\frac{\|\vec{C}(v_i) - \vec{C}(v_j)\|_2^2}{\sigma_C(v_i) * \sigma_C(v_j) * l^2} \right] \quad (14)$$

Where $\sigma_C(v_i) = \max_{v_k \sim v_i} (\|\vec{c}_i - \vec{c}_k\|_2)$. The mono-scale colorimetric saliency at a scale k is defined similarly to the mono-scale geometric saliency (mean degree of the target vertex):

$$\text{Mono-scale-colorimetric-saliency}_k(v_i) = \frac{1}{|\{v_j \sim v_i\}|} \sum_{v_i \sim v_j} w_C(v_i, v_j) \quad (15)$$

The multi-scale colorimetric saliency is defined as:

$$\text{Multi-scale-colorimetric-saliency}(v_i) = \frac{\sum_{k=1}^3 \text{Mono-scale-colorimetric-saliency}_k(v_i) * \text{Entropy}_k}{\sum_{k=1}^3 \text{Entropy}_k} \quad (16)$$

The results of the colorimetric saliency of 3D colored meshes are presented and analyzed in section 3.5.2.

3.4 RESULTS AND VALIDATION OF THE DETECTED VISUAL SALIENCY OF 3D NON COLORED MESHES

3.4.1 COMPARISON WITH A PSEUDO GROUND TRUTH

Figure 14 and figure 15 present the saliency results of the approach of (Nouri et al., 2015a) as well as a subjective saliency (detected by human observers) of 3D non colored meshes belonging to the 2007 SHREC Shape-based Retrieval Contest database (Chen et al., 2012) which constitutes a pseudo ground truth relative to visual

saliency of non colored meshes. These subjective results were acquired in an online experimentation where the observers were asked to select 3D vertices which are likely to be selected by other observers. From the collected informations, the authors perform a regression analysis to produce an analytical model that localizes salient vertices on the surface mesh. Figure 14 and figure 15 show that the saliency detected by the approach of (Nouri et al., 2015a) corresponds well to the regions that have guided the visual attention of the human observes. The low values of the mean square errors (NMSE - Normalized Mean Square Error) confirm the accuracy of the proposed saliency model.

Despite this precision in the estimation of saliency, the approach of (Nouri et al., 2015a) doesn't perform well for some non « optimized » 3D meshes (figure 16). The surface of a non optimized 3D mesh is constituted from regular triangles (triangle with a constant area) and edges of the same length (figure 17). Indeed, one can remark in figure 16 (a) and figure 16 (c) that all regions of the 3D mesh (regions with few details and regions containing lot of details) have similar number of triangles (for example the palm of the hand and the areas between the fingers). However, for an « optimized » 3D mesh (figure 16 (b) and figure 16 (d)), when some regions contain lot of details (for example the hollow at the top of the bird's back), the latter requires a high number of triangles. For regions with less details (the bottom of the bird's back), these need less number of triangles. Consequently, triangles on the surface of a 3D optimized mesh will have different areas and de facto edges with different lengths.

While computing similarities (equation 9), a weight representing the Euclidean distance between the target vertex and its neighbor is used. This coefficient allows reducing the influence of far vertices in the computation of the saliency degree of a target vertex and vice versa. For non optimized meshes, all neighboring vertices contribute with the same weight to the saliency degree of the target vertex since the distance weight is constant. This explains the surplus of red vertices on the saliency map (figure 16).

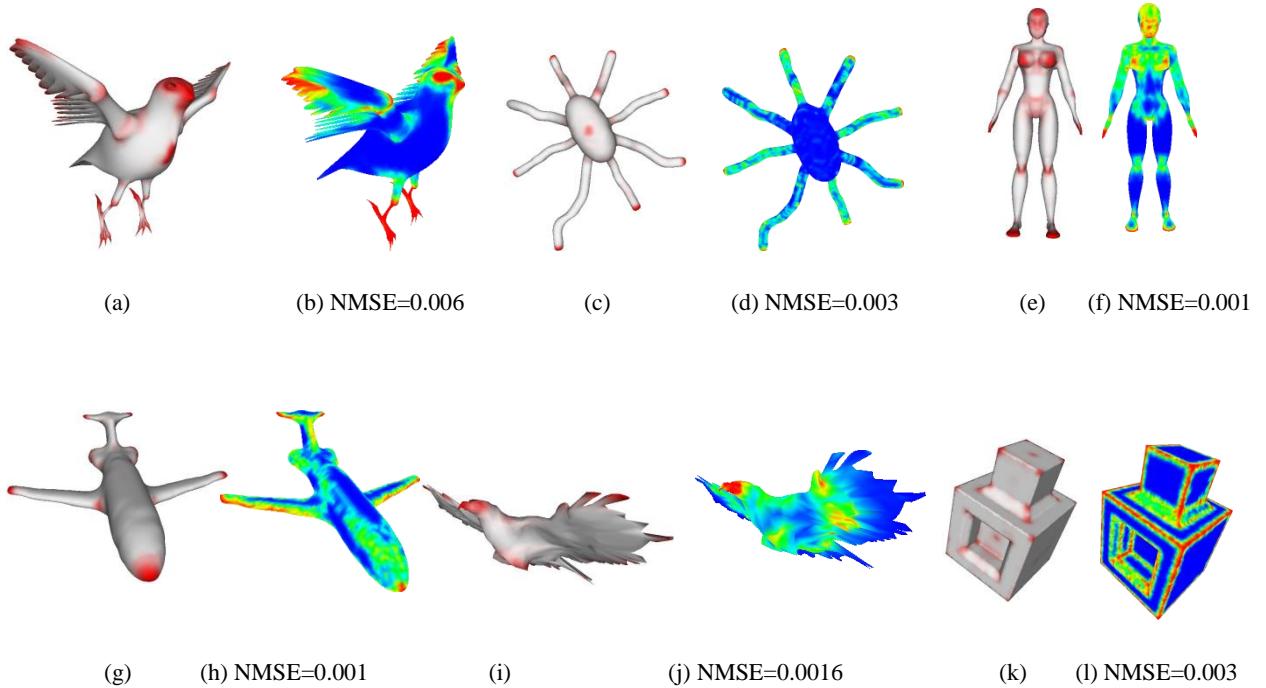


Figure 14. Comparison of the predicted multi-scale saliency (Nouri et al., 2015) with a pseudo ground truth saliency (Chen et al., 2012). Images (a), (c), (e), (g), (i) and (k) represent the pseudo ground truth saliency for various non colored 3D meshes. Images (b), (d), (f), (h), (j), (l) and (n) represent the predicted multi-scale saliency.

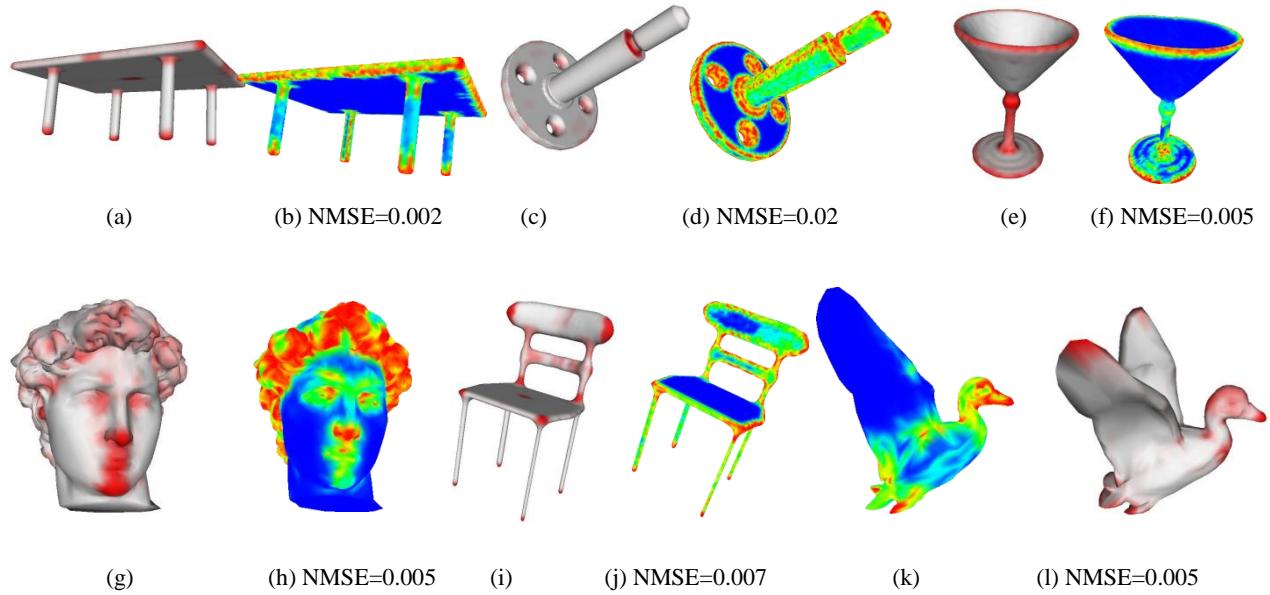


Figure 15. Comparison of the predicted multi-scale saliency (Nouri et al., 2015) with a pseudo ground truth saliency (Chen et al., 2012). Images (a), (c), (e), (g), (i) and (k) represent the pseudo ground truth saliency for various non colored 3D meshes. Images (b), (d), (f), (h), (j) and (l) represent the predicted multi-scale saliency.

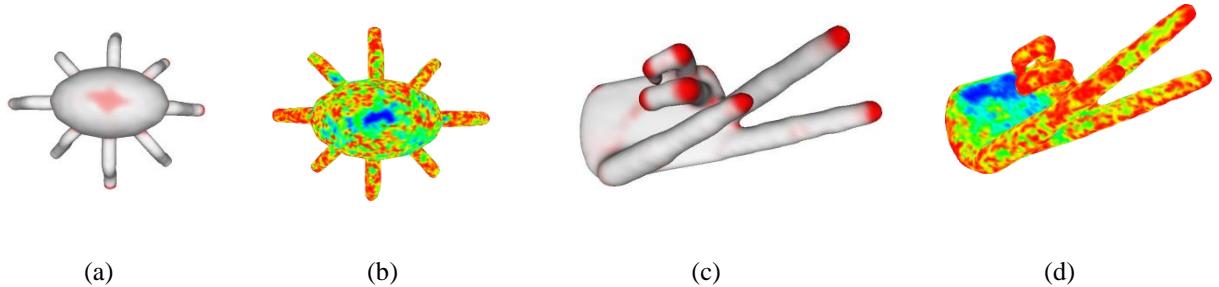


Figure 16. Inaccurate detection of saliency: a) - c) pseudo ground saliency and b) d) predicted multi-scale saliency

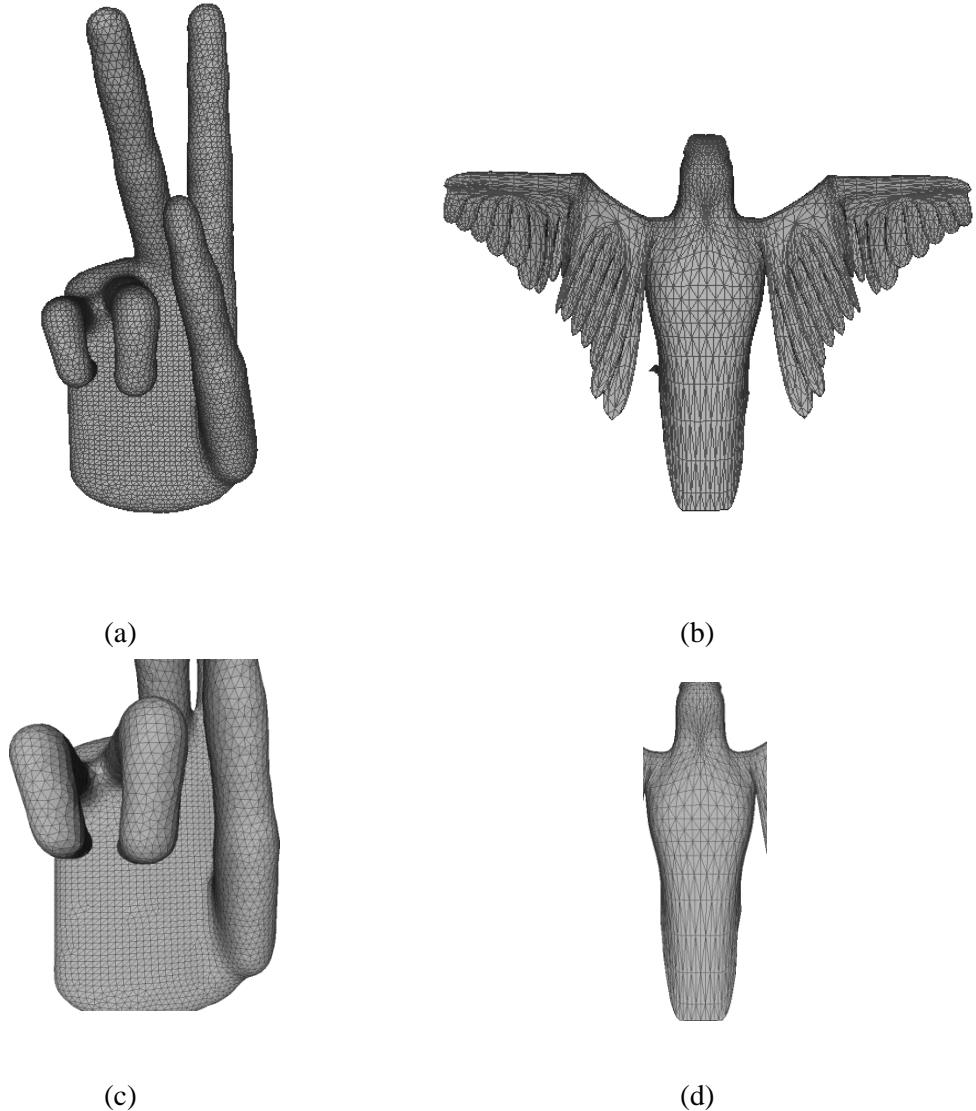


Figure 17. Non « optimized » and « optimized » 3D meshes: a) 3D non « optimized » mesh », b) 3D « optimized » mesh, c) zoom on (a). and d) zoom on (b). In optimized meshes, the number of triangles varies according to the amount of the details.

3.4.2 ANALYSIS AND COMPARISON OF THE STATE-OF-THE-ART

As the source codes associated to the different approaches of the state-of-the-art are not available, the 3D meshes used for the following analysis are similar to those of the state-of-the-art. In figure 18, Nouri et al., (2015) considers a 3D non colored mesh representing a Dinosaur on which they compute the multi-scale saliency for the comparison with the methods proposed in (Tal et al., 2012), (Song et al., 2014) and (Lee et al., 2005). Tal et al., (2012) judge the ribs of the 3D Dinosaur model located on the back and the stomach as non-salient regions (Figure18(c)). Yet, these areas fluctuate enormously and contain high discontinuities in the surface. In figure 8(b), the method of (Nouri et al., 2015) assesses ribs of the mesh Dinosaur as salient regions given their high discontinuities, and contrary to the approach of (Tal et al., 2012) (see figure18(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 18 (d), one can notice that the method of (Lee et al., 2005) weakly detect the ribs of the mesh Dinosaur (colored in light green). Locally finest details are not taken into account, contrary to the approach of (Nouri et al., 2015). However, in figure 18(d), the approach of Song et al., 2014 provide a saliency map similar to one of (Nouri et al., 2015). It can differentiate between areas with high discontinuities and

flat

ones.

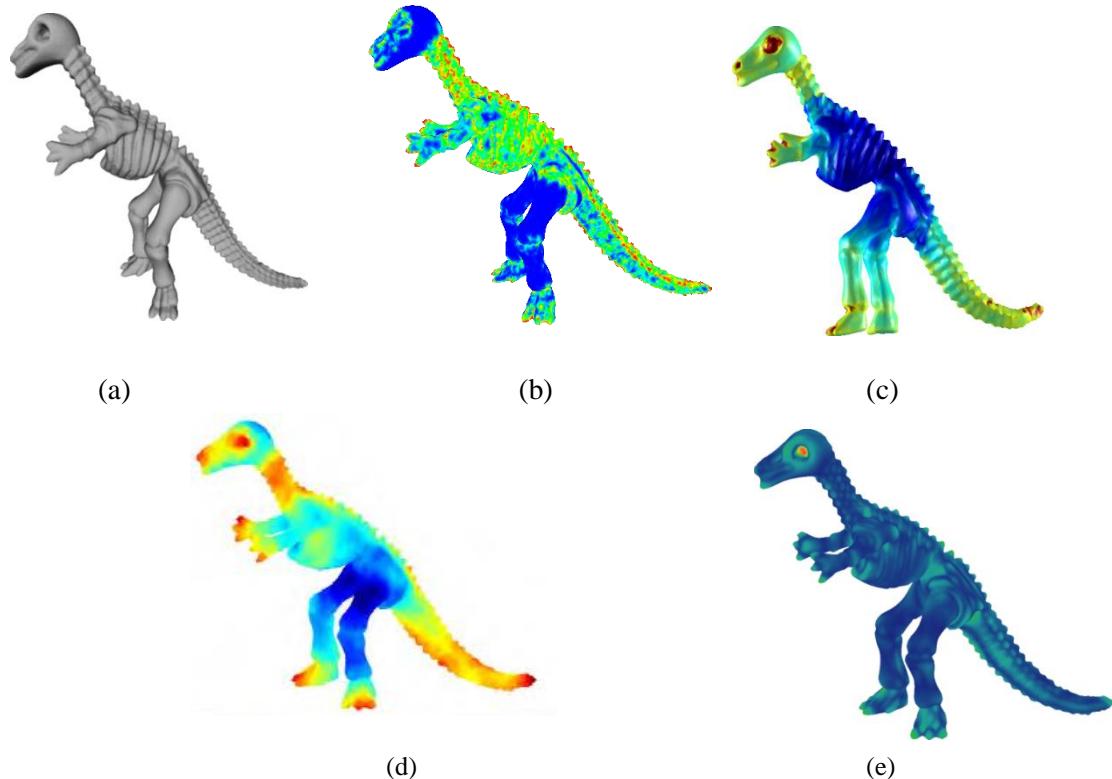


Figure 18. comparison with the state-of-the-art: a) Original 3D mesh Dinosaur, b) Multi-scale saliency detected by (Nouri et al., 2015), c) Saliency detected by (Tal et al., 2002), d) Saliency detected by the (Song et al., 2014) and e) Saliency detected by (Lee et al., 2005).

Figure 19 shows a comparison of the estimated saliency on the 3D mesh Angel with the approaches of (Nouri et al., 2015) , (Tal et al., 2012) and (Song et al., 2014). The 3D mesh 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 (Song et al., 2014) and the approach of (Nouri et al., 2015) while the approach of (Tal et al., 2012) considers it as non salient. The discontinuities on the eyes, the arms, the hip and the stomach are represented as salient regions by the approaches of (Song et al., 2014) and (Nouri et al., 2015) contrary to the method of (Tal et al., 2012).

Figure 20 shows a comparison between the saliency maps of the 3D mesh Horse. One can notice that the methods of (Song et al., 2014) and (Nouri et al., 2015) are able to detect the eyes, highly regarded parts in scenes or meshes containing faces, and judge them as very salient regions while the method of (Tal et al., 2012) fails in this regard. Also, the horse's back has muscled parts. These areas are also assessed as salient regions with the methods of (Song et al., 2014) and (Nouri et al., 2015) contrary to the method of (Tal et al., 2012).

Note that the saliency map provided by the approach of (Nouri et al., 2015) isn't smoothed in contrary to other approaches. This aspect allows the detection of exiguous salient details which are of great interest for applications such as adaptive compression and smoothing.

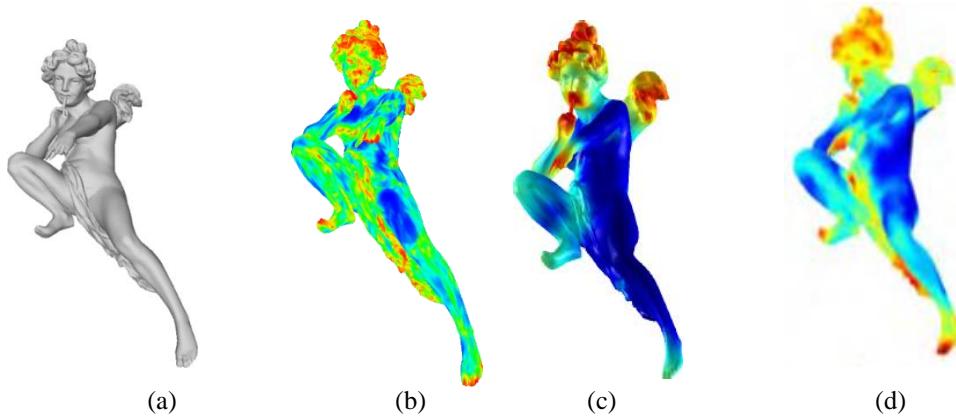


Figure 19. Comparison with the state-of-the-art: a) Original 3D mesh Angel, b) Multi-scale saliency detected by (Nouri et al., 2015), c) Saliency detected by (Tal et al., 2002) and d) Saliency detected by the (Song et al., 2014).

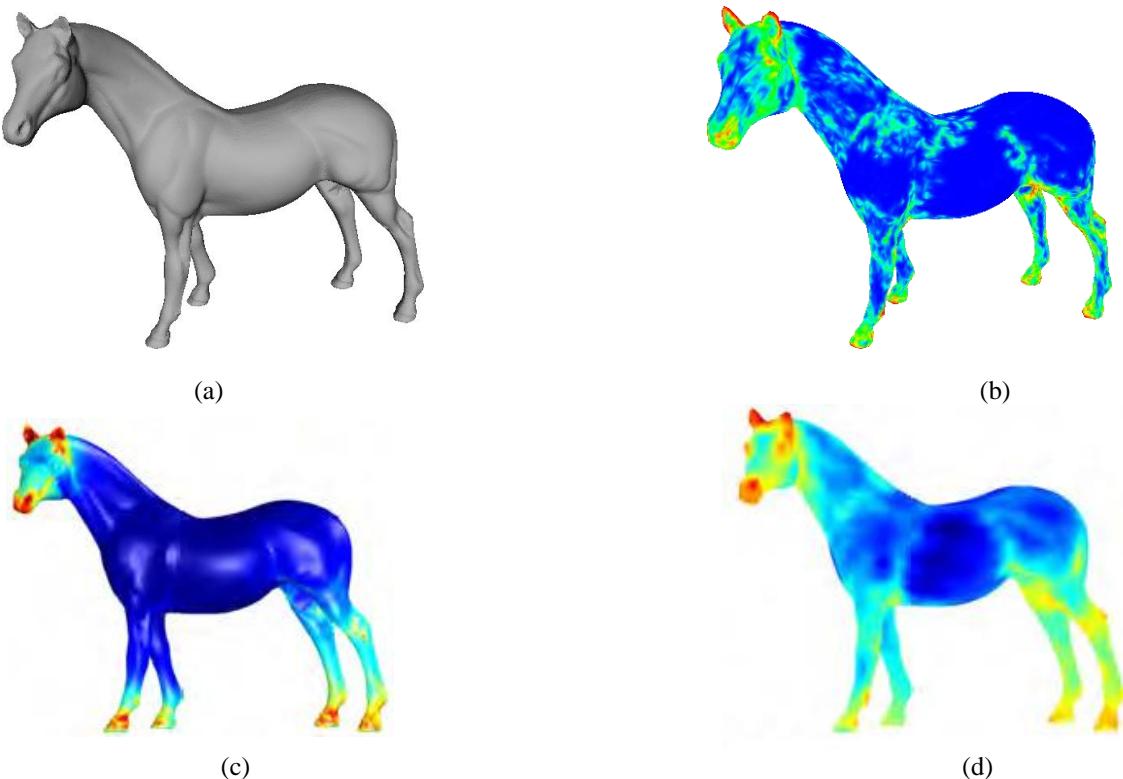


Figure 20. Comparison with the state-of-the-art: a) Original 3D mesh Horse, b) Multi-scale saliency detected by (Nouri et al., 2015), c) Saliency detected by (Tal et al., 2002) and d) Saliency detected by the (Song et al., 2014).

Figure 21 shows the multi-scale saliency maps of various 3D non-colored meshes provided by the approach of (Nouri et al., 2015).

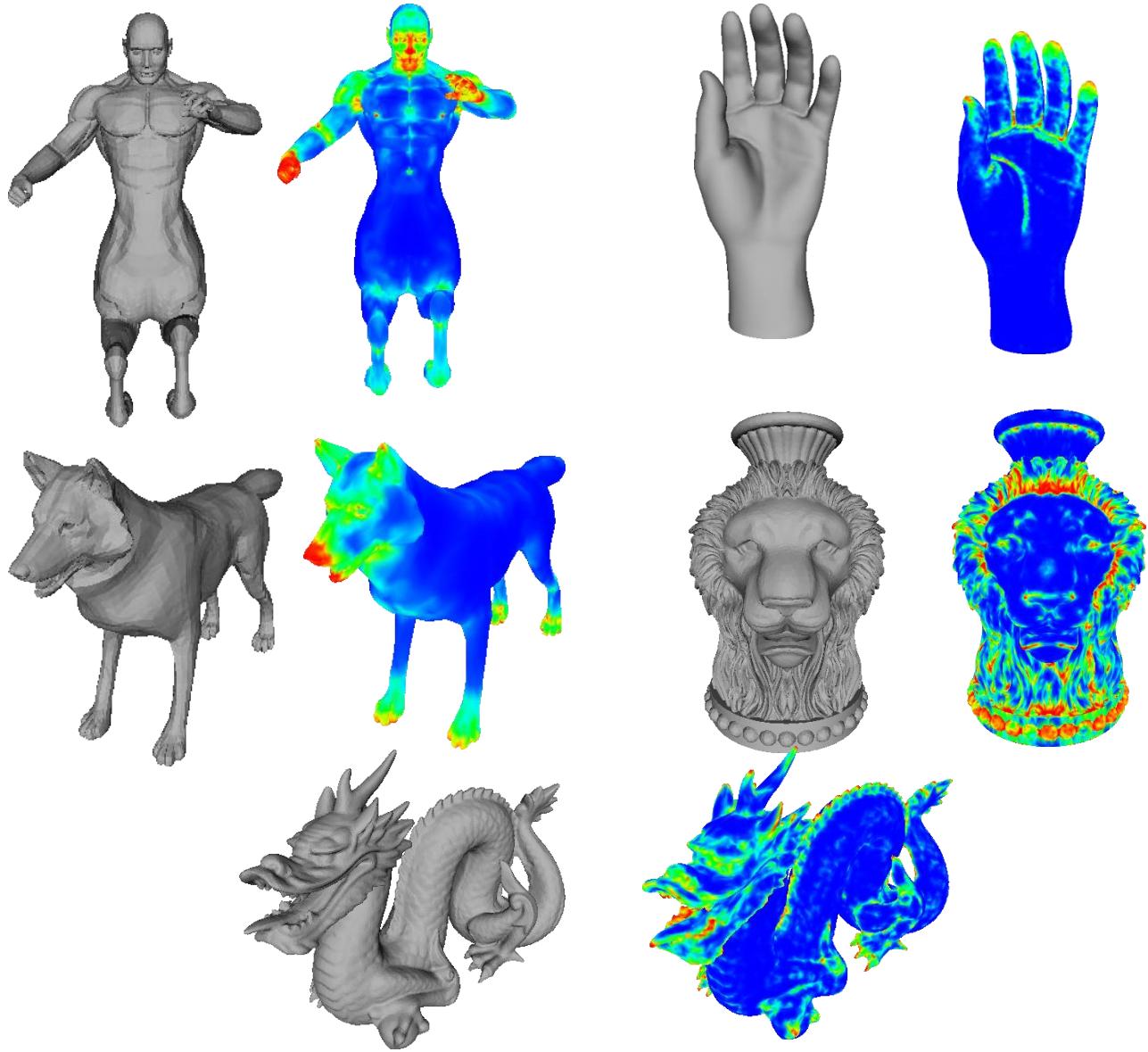


Figure 21. Multi-scale saliency detection on various 3D meshes with the approach of (Nouri et al., 2015).

3.4.3 ROBUSTNESS AND STABILITY

To attest the robustness of their approach, Nouri et al. (2015) distort a target 3D mesh (a 3D mesh representing a Centaur is used in this case) with noise by randomly displacing the positions of its vertices according to two levels of noise. Then, they apply their multi-scale saliency model for saliency estimation. Figure 22 shows this experimentation. One can see that the method of (Nouri et al., 2015) always succeeds to detect the same salient regions despite the noise affected to the mesh surface. 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 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 their approach is robust to noise.

Another experiment that demonstrate the stability consists in measuring saliency on simplified meshes. Simplification was operated using the algorithm proposed in (Garland et al., 1997). Figure 23 shows the behavior of

the method of (Nouri et al., 2015) on simplified meshes. Firstly, the mesh was simplified to 25% (in figure 23(a), 25% of vertices were deleted). One can notice that the initially salient regions are always detected (see figure 23(c)). In figure 23(b), the mesh is simplified to 50%. The same regions remain salient (figure 23 (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.

3.5 RESULTS OF THE DETECTED VISUAL SALIENCY OF 3D COLORED MESHES

3.5.1 CONSTRUCTION OF A 3D COLORED MESH DATABASE

Currently, there's no 3D colored mesh database in the state-of-the-art which allow the evaluation of saliency models or quality assessment algorithms. Consequently, a 3D colored mesh database have been designed by (Nouri et al., 2017a) to evaluated their colorimetric saliency model (Nouri et al., 2015b). 15 real 3D objects have been scanned using a 3D laser scanner *NextEngine* provided with a rotating plate. This scanner is able to acquire both geometric and colorimetric properties of a target 3D object (figure 24). Figure 25 presents some acquired 3D meshes.

In a context of visual quality assessment of 3D meshes, algorithms, named metrics, are designed for scoring the visual quality of a distorted 3D mesh in a similar way to human beings. The distorted versions of the acquired meshes have to reflect as fully as possible the degradations that may occur during common treatments. For this, (Nouri et al., 2017a) consider several distortions (table 2) which are applied to the original corpus according to three intensities: weak, medium and high. In order to take into account the visual masking effect that may occur on some textured or rough regions, the selected distortions are applied according to four situations: 1) Uniformly on the surface mesh .2) On rough areas/ 3) On smooth areas. 4) On the colors of the 3D mesh.

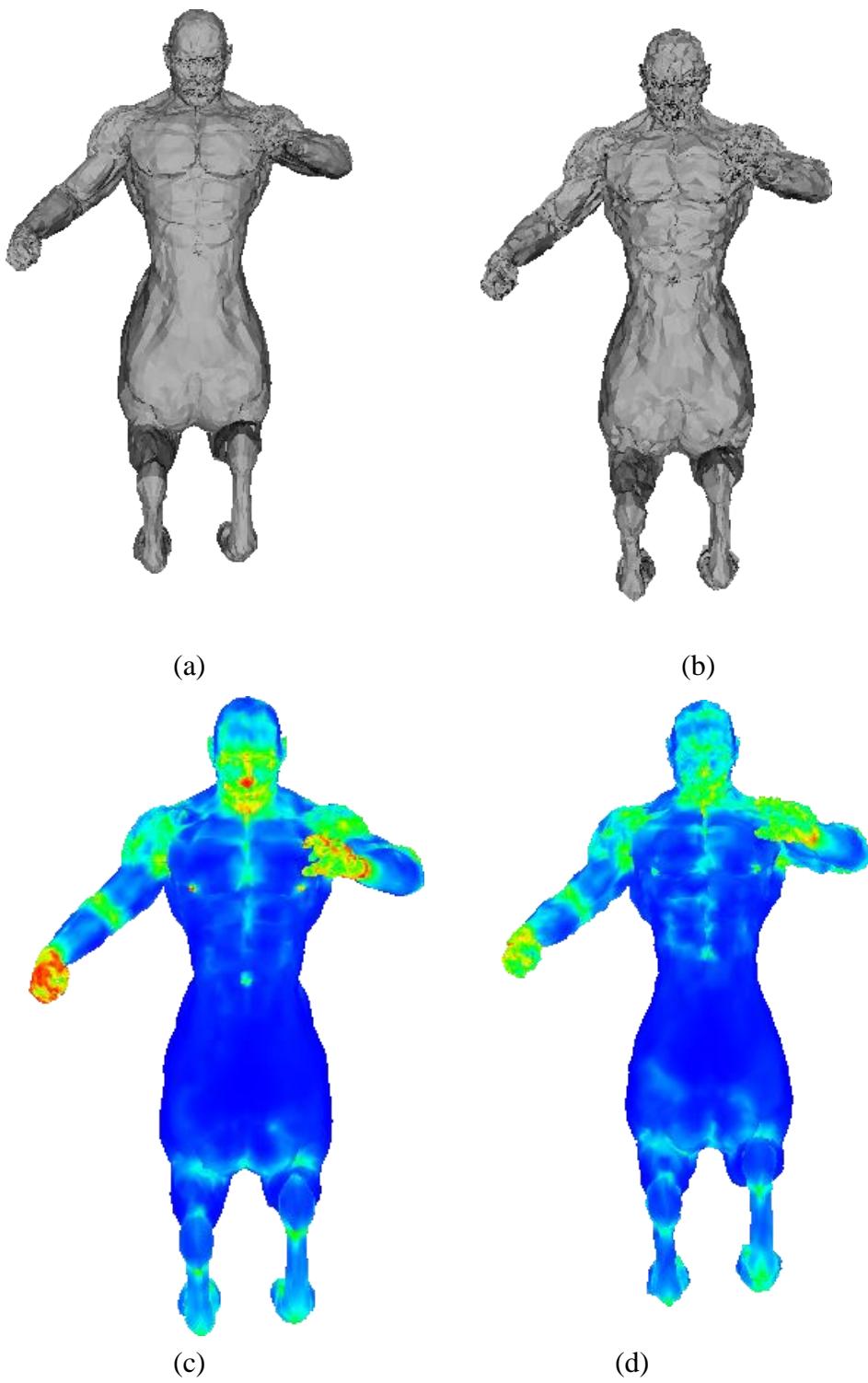
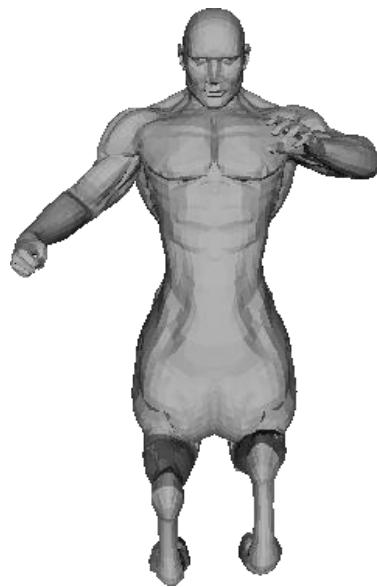


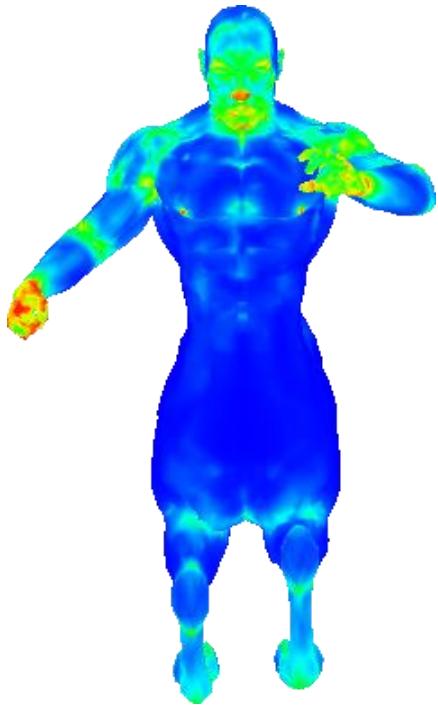
Figure 22. Robustness to noise: a) Original 3D mesh Centaur noised (displacement= -0.1%), b) Original 3D mesh Centaur noised (displacement= -0.2%), c) Multi-scale saliency of (a) (Nouri et al., 2015) and d) Multi-scale saliency of (b) (Nouri et al., 2015). Note how the original salient regions still remain salient despite the sever noise.



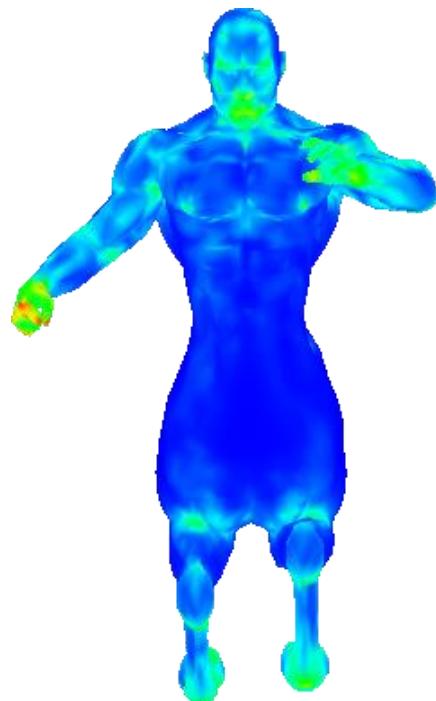
(a)



(b)



(c)



(d)

Figure 23. Robustness to simplification: a) 3D mesh Centaur simplified to 25%, b) 3D mesh Centaur simplified to 50%, c) Multi-scale saliency of (a) (Nouri et al., 2015) and d) Multi-scale saliency of (b) (Nouri et al., 2015). Note how the original salient regions still remain salient despite the sever noise.



Figure 24: A 3D object being acquired.

Type of degradation
Gaussian noise on 3D coordinates
Simplification
Gaussian noise on RGB colors
Smoothing on 3D coordinates
Smoothing on RGB colors

Table 2. Distortions used in the proposed 3D colored mesh database

In order to distinguish between rough and smooth regions, Nouri et al., (2017a) have implemented the method of (Wang et al., 2012) to generate a roughness map which associates a roughness value to each vertex in [0 1]. To distort a rough or smooth region of the surface mesh, it's necessary to modify the vertex coordinates belonging to this region:

$$\left\{ \begin{array}{ll} \vec{p}_i = \vec{p}_i + noise \times \text{roughness}(v_i) & \text{distortion of rough regions} \\ \vec{p}_i = \vec{p}_i + noise \times (1 - \text{roughness}(v_i)) & \text{distortion of smooth regions} \end{array} \right\}$$

The same process is used to distort the colors of the 3D mesh. Therefore the constructed colored mesh database contains 425 colored 3D meshes (17 reference meshes + 17x3x8 distorted meshes). Figure 26 presents examples of distorted 3D meshes from the proposed database.



Figure 25: Some of the constructed 3D colored mesh database.



(a) Original 3D colored mesh



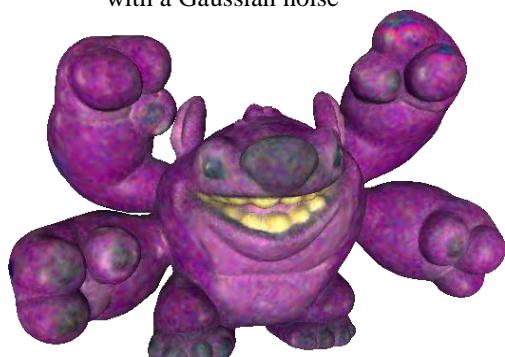
(b) Coordinates affected with a Gaussian noise



(c) Coordinates of rough regions affected with a Gaussian noise



(d) Coordinates of smooth regions affected with a Gaussian noise



(e) RGB colors affected with a uniform Gaussian noise



(f) Coordinates affected uniformly with an isotropic smoothing



(g) Coordinates of rough regions affected with an isotropic smoothing



(h) RGB colors affected with an isotropic smoothing

Figure 26: Some distorted 3D colored meshes of the constructed 3D mesh database.

3.5.2 Results and analysis

In order to illustrate the difference between the colorimetric saliency and the geometric saliency of 3D meshes, figure 27(a) presents a 3D scan of a human being. Figure 27 (b) shows the geometric saliency which takes into account only coordinated of vertices during the construction of the local adaptive patches. One can remark that the flat regions are judged as non salient while fluctuating regions not redundant are considered as salient. Furthermore, figure 27 (c) presents the colorimetric saliency based only on the RGB colors of vertices. This result is very different from the geometric saliency map. Indeed, regions with important color variations such as the eyes and eyebrows are now considered as salient. White areas on the shirt that were considered non salient on the geometric saliency map are put forward on the colorimetric saliency map. Likewise, the varying colors of the skin are well detected in the colorimetric saliency map while they were not in the geometric saliency map.

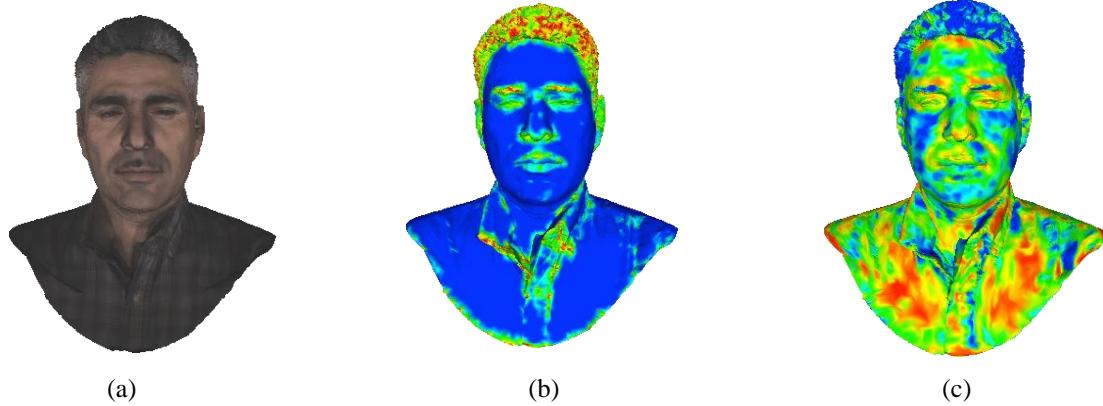


Figure 27. Multi-scale saliency of a 3D colored mesh representing head of a human being: a) 3D colored mesh, b) Multi-scale geometric saliency (Nouri et al., 2015) and c) Multi-scale colorimetric saliency (Nouri et al, 2015b).

The same contribution of colors in the detection of saliency can be observed in figure 28. The white collar of the duck is very contrasted due to the colors delimiting it (red and white colors). The latter is judged very salient in the colorimetric saliency map whereas it was considered to be moderately salient in the geometric saliency map.

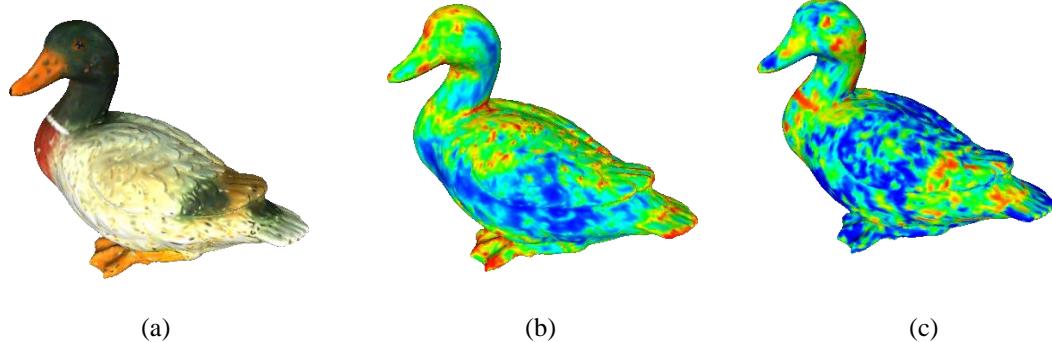


Figure 28. Multi-scale colorimetric saliency of a scanned 3D colored mesh representing a stuffed duck: a) 3D colored mesh, b) Multi-scale geometric saliency (Nouri et al., 2015) and c) Multi-scale colorimetric saliency (Nouri et al, 2015b).

Figure 29 presents the colorimetric saliency of various 3D meshes belonging to the constructed 3D colored mesh database. One can remark that the colorimetric saliency proposed in (Nouri et al., 2015) predicts saliency precisely. Note that all the colorimetric saliency results were computed in the RGB color-space. Figure 30 presents the colorimetric saliency of the 3D meshes representing a duck and a human being head in different color-spaces.

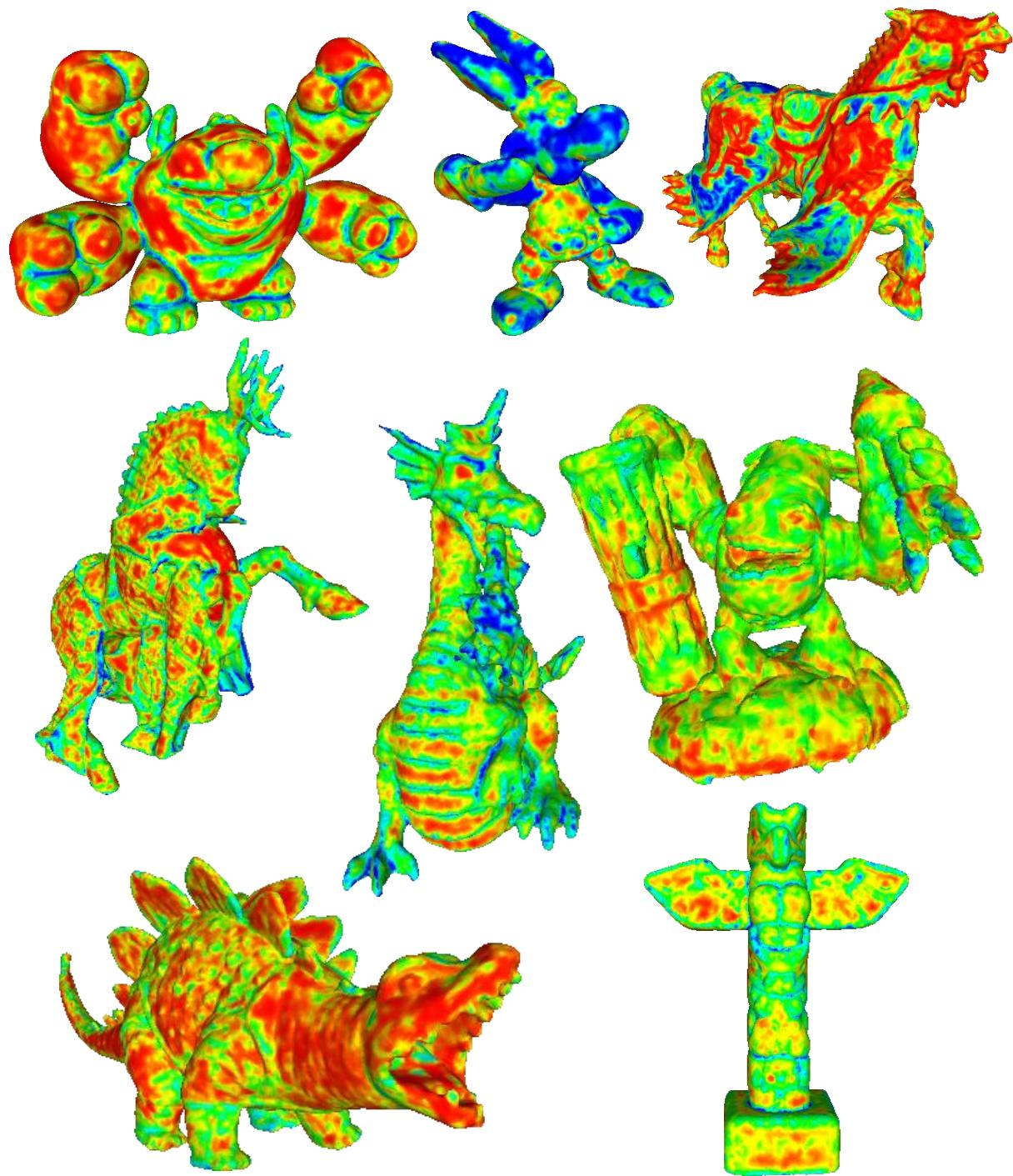


Figure 29. Multi-scale colorimetric saliency of the 3D colored mesh presented in figure 25.

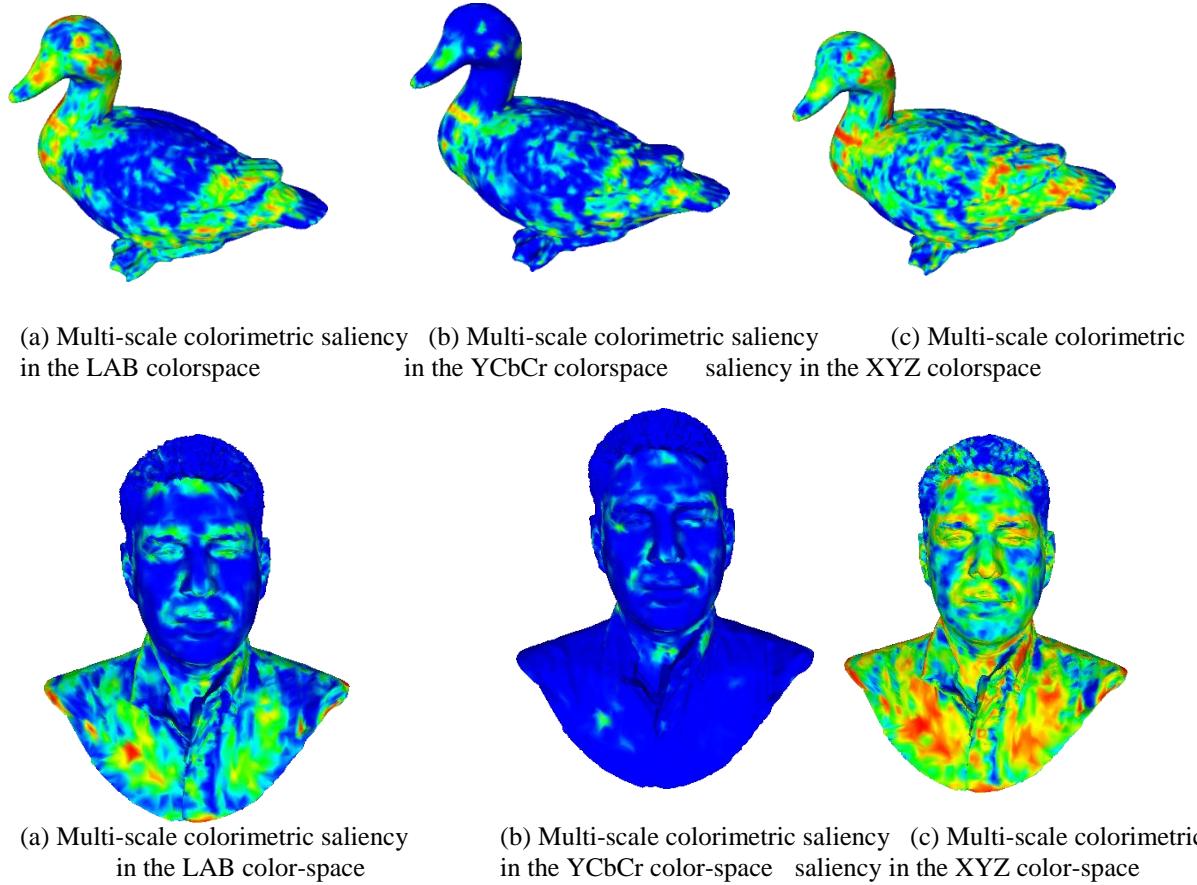


Figure 30. Multi-scale colorimetric saliency computed in LAB, YCbCr and XYZ colorspaces

3.5.3 ROBUSTNESS TO NOISE ON GEOMETRIC AND COLORS

Nouri et al., (2015b) assesses the robustness of they colorimetric saliency models on distorted colored meshes from the constructed colored mesh database. Two distortions have been selected for this test: 1) Uniform Gaussian on the coordinates and 2) Gaussian noise on RGB colors. Figure 31 presents the results of the detected saliency on the distorted colored meshes. One can remark that despite the distortion affected to the coordinates or colors, the initial detected salient regions (figure 31 (b)) are still considered as salient (figure 31 (d) and (f)). This attests that the extended colorimetric saliency model (Nouri et al., 2015b) is as stable as the geometric saliency model (Nouri et al., 2015a).

4. FIELDS OF APPLICATIONS OF VISUAL SALIENCY

In this chapter, three saliency-guided applications are presented. The developed applications concern the automatic viewpoint selection of 3D non colored meshes, the adaptive simplification of 3D non colored meshes and the adaptive smoothing of 3D colored meshes. The saliency models of (Nouri et al., 2015a) and (Nouri et al., 2015b) are used to guide these applications.

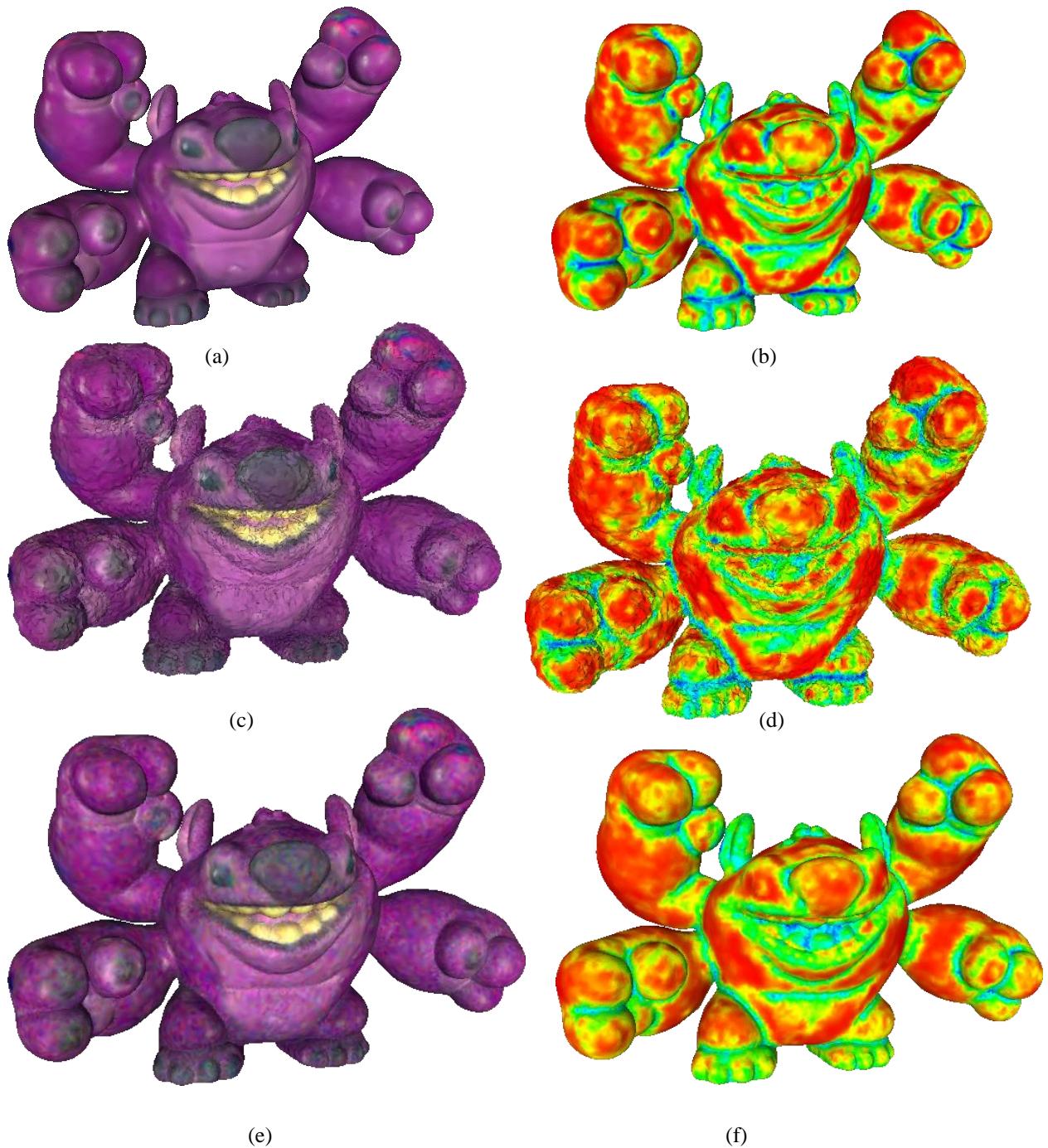


Figure 31. Robustness of the multi-scale colorimetric model to Gaussian noise on 3D coordinates and RGB colors:
 a) Original 3D mesh, b) multi-scale colorimetric saliency of (a), c) distorted colored mesh with Gaussian noise on 3D coordinates, d) multi-scale colorimetric saliency of (c), e) distorted colored mesh with Gaussian noise on RGB colors and f) multi-scale colorimetric saliency of (e).

4.1 INTRODUCTION

For optimization and execution time purposes, the technologic advances associated to applications interacting with human beings and particularly with their vision are increasingly taken into account the capacities and limitation of the HVS.

Selective visual attention represents a major mechanism in the human perception. The latter significantly saves time when viewing the surrounding environment by selecting the visual information potentially interesting while ignoring the rest. Sections below explain how a saliency map allows to distinguish regions that can orient the visual attention of a human observer. We can denote applications such as the optimal viewpoint selection (Tal et al., 2012) (Nouri et al., 2015a) where the goal is to select the most informative and attractive viewpoint and the adaptive simplification (Shilane and Funkhouser, 2007) which aims to simplify the 3D mesh while preserving the most important feature for quality purposes. Similarly, others application take advantage from the intake of visual saliency such as the Surface Matching (Gal and Cohen-or, 2006), the automatic resizing (Jia et al., 2014), the facial recognition (Jinho et al., 2004), the icon generation (Shilane and Funkhouser, 2007), mesh demonizing (Mao et al., 2006), etc.

In order to evaluate the intake of their saliency model, Nouri et al., 2015a have proposed to select the optimal viewpoint selection of a 3D non colored mesh based on its visual saliency. Another application of this model is described in this section and is related to the adaptive simplification of 3D non colored meshes (Nouri et al., 2016a).

To validate the colorimetric saliency extension proposed in (Nouri et al., 2015b), Nouri et al., (2016a) propose to smooth and denoise 3D colored meshes adaptively.

4.2 APPLICATION TO THE AUTOMATIC OPTIMAL VIEWPOINT SELECTION OF 3D NON COLORED MESHES

Nouri et al., (2015a) propose to select automatically the optimal viewpoint with perceptually important regions to the human observer. The principal criterion of the proposed approach is to distinguish regions with high visual saliency degree. For this, a saliency map is computed, then, a viewpoint maximizing saliency is selected the abscissa axis. This is done by informally sampling a sphere that bounds 3D mesh. Let vp the viewpoint selection along the abscissa axis and $surface(vp)$ the visible vertices from this point of view. The total saliency of this surface is defined as:

$$\text{Total-saliency}_{axe_x}(vp) = \sum_{v \in surface(vp)} \text{Multi-scale-saliency}(v)$$

Hence the optimal viewpoint along the abscissa axis is defined as:

$$vp_x = \max(\text{Total-saliency}_{axe_x}(vp_i))$$

where vp_i represents the different viewpoint along the x-axis. From this viewpoint, the same processing is performed to select the optimal viewpoint maximizing saliency along the ordinate axis vp_y (see figure 32). Once this viewpoint selected, a gradient descent is executed to search for the optimal viewpoint along the three axis at the same time. Figure 33 presents the most informative viewpoints of some 3D non colored meshes. One can remark that the generated viewpoints correspond well to the criterion of the proposed approach selecting the optimal viewpoints, and also, these viewpoints correspond to the most likely views of the associated 3D objects.

It is interesting to note that the proposed approach of viewpoint selection can be used in a heritage valorization context. Indeed, to more attract the glance of a future visitor, it would be relevant to present the most informative and attractive viewpoint. Figure 34 shows an example of selected optimal viewpoints in a patrimonial context.

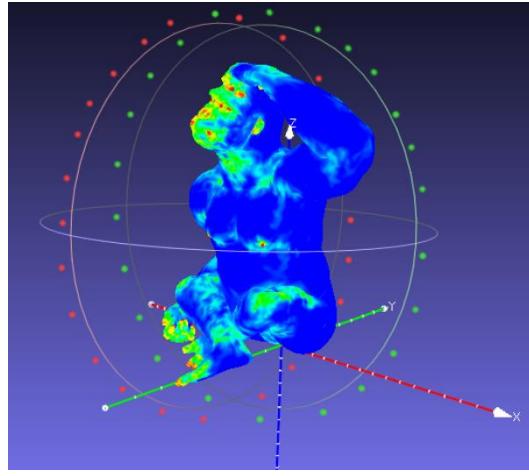


Figure 32. Illustration of the sampling of a sphere bounding a 3D non colored mesh along the abscissa and ordinate axis: red and green dots around the 3D mesh refer to positions from which the viewpoints are generated. A selection of the viewpoint maximizing saliency is then performed. Note that 3D mesh is a non colored mesh. Colors in the figure refer to saliency degrees.

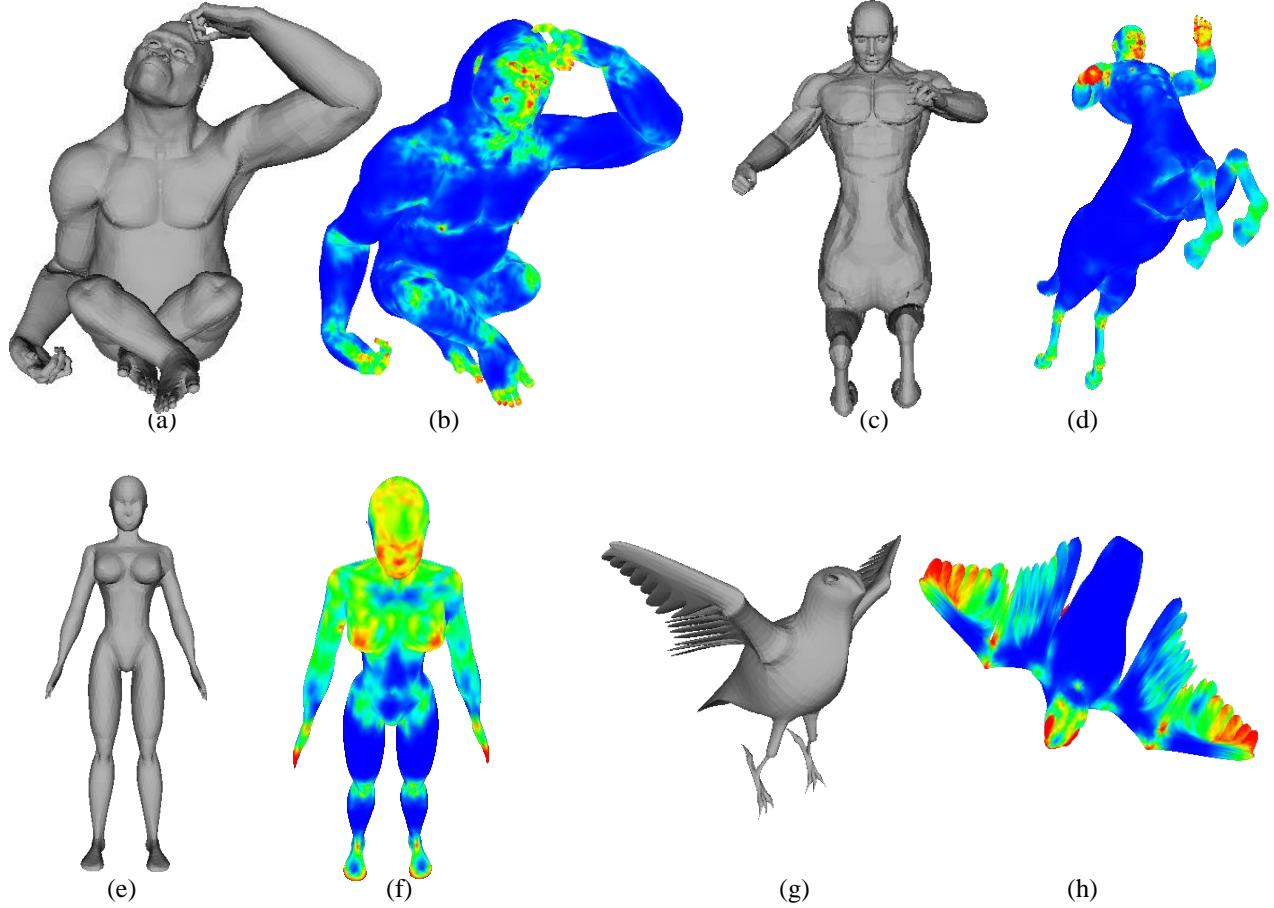


Figure 33. Optimal viewpoints generated by the approach of (Nouri et al., 2015). Images (a), (c), (e) and (g) present the initial viewpoints. Images (b), (d), (f) and (h) present the most informative viewpoints selected automatically.

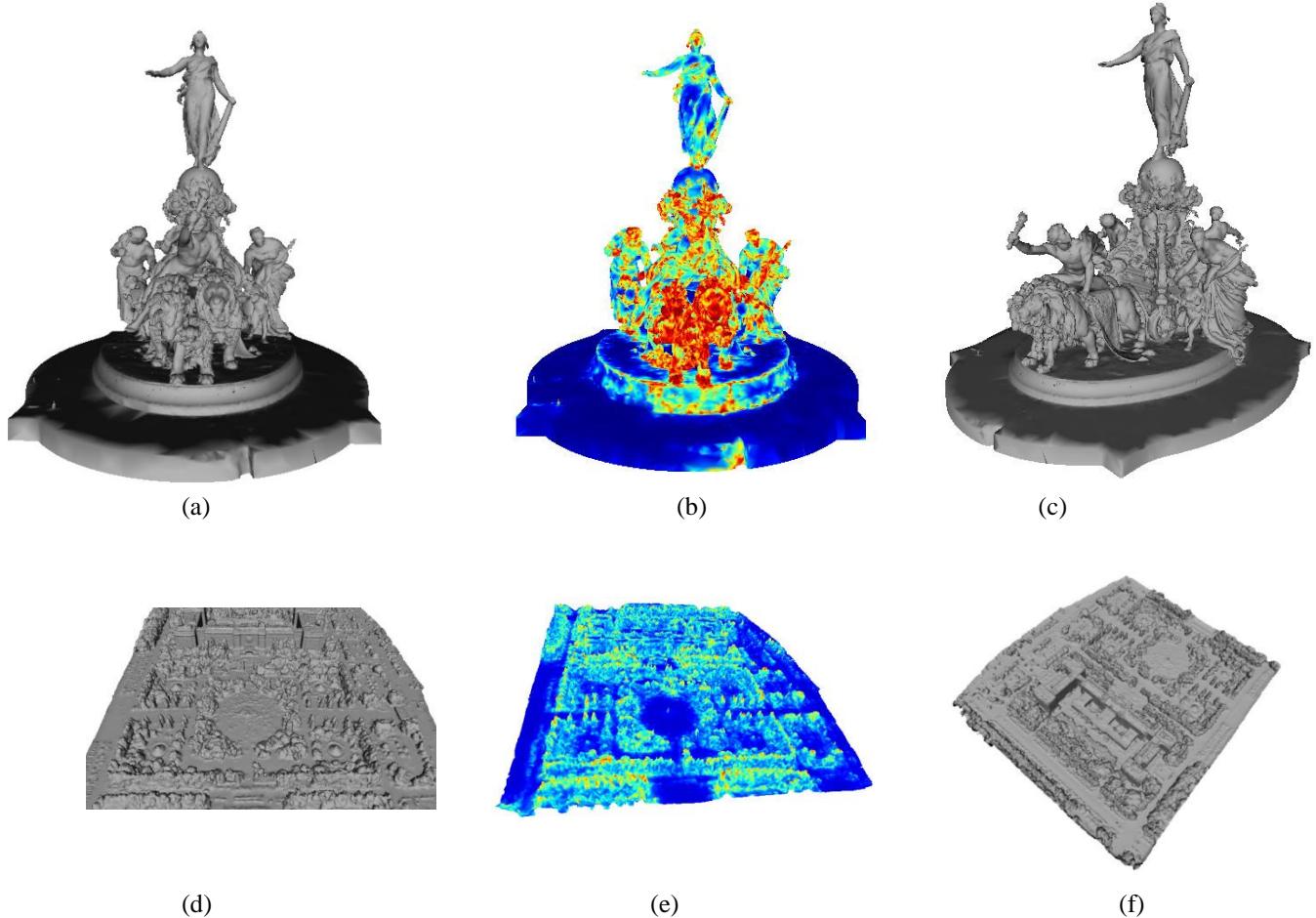


Figure 34. Results of the optimal viewpoints selection of two architectural heritage: (a) Initial viewpoint of a 3D mesh representing the statue « Le triomphe de la république » located at Place De La Nation in Paris-France. The scan was obtained with the Photogrammetry process from 500 high definition pictures. b) Multi-scale saliency of (a). c) Selected Optimal viewpoint. d) Initial viewpoint of a 3D mesh representing an Russian city. e) Multi-scale saliency of (d). f) Selected optimal viewpoint.

Figure 35 presents a comparison between the selected optimal viewpoints by the approach of (Nouri et al., 2015a) and the ones obtained by the approach of (Tal et al., 2012). One can remark that the method of (Nouri et al., 2015a) is very competitive with the method of (Tal et al., 2012). However, some limitations exist and are presented on figure 35 (g) and figure 35 (i).

To select the most informative viewpoints, the method of (Tal et al., 2012) begins by generating the candidates viewpoints by uniformly sampling a sphere that bounds the 3Dmesh .From the viewpoint maximizing saliency, a gradient-descent optimization is applied to define the most informative viewpoint. On the contrary, to make the processing faster, the method of (Nouri et al., 2015a) generates first the candidates viewpoints by sampling the sphere bounding the3Dmesh along the x-axis and the y-axis, which leads to a fewer viewpoints candidates. Second, similarly to the method of (Tal et al., 2012), 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, the proposed 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 adaptive patches while computing the saliency. Besides, the viewpoints generated for the Car and the Piano seem to be logic insofar that the

method of (Nouri et al. 2015a) 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).

4.3 APPLICATION TO ADAPTIVE SIMPLIFICATION OF 3D NON COLORED MESHES

In order to more evaluate the effectiveness of the saliency model, Nouri et al., (2016a) modify the quadric-based simplification approach of (Garland & Heckbert, 1997) by integrating weights obtained from the multi-scale saliency map. The goal is to preserve (less simplify) salient regions of a non colored 3D mesh while the simplification process.

Garland & Heckbert, (1997) simplify a 3D mesh by contracting vertices ranked in an ascending order according to their quadric error. Let p a plane belonging to the set of planes P associated to the triangles incident to the vertex v . The latter is defined by the equations $ax + by + cz + d = 0$, $a^2 + b^2 + c^2 = 1$ and is represented by the vector $(abcd)^T$. The quadric of the plan p is defined as $Q_p = pp^T$. Garland & Heckbert, (1997) define the error of a vertex v according to the plane p as $v^T Q_p v$. The quadric Q of the vertex v is computed by the sum of all the quadrics related to the neighboring planes: $Q = \sum_{p \in P} Q_p$. Once the quadrics have been computed for all vertices, the method computes the position of the vertex \bar{v} related to the optimal contraction of each pair (v_i, v_j) in order to minimize the quadric error $\bar{v}^T (Q_i + Q_j) \bar{v}$ where Q_i and Q_j represent respectively the quadrics of v_i and v_j . Afterwards, the method contracts iteratively all pairs of vertices having a minimal contraction cost $\bar{v}^T (Q_i + Q_j) \bar{v}$. After a contraction of a pair of vertices (v_i, v_j) , the quadric of the resulting vertex \bar{v} is defined as $Q_i + Q_j$.

Nouri et al., (2016a) modify the simplification process of Garland et Heckbert, (1997) by integrating weights from the multi-scale saliency map MS . Each quadric associated Q each vertex v is multiplied by the weight $MS(v)$. In the same way, after a contraction of a pair of vertices (v_i, v_j) , the weight of the quadric associated to the resulting vertex \bar{v} is obtained by summing the degrees of saliency of the contracted vertices:

$$MS(v_i) + MS(v_j)$$

Figure 35 presents the contribution of the saliency weighting in the simplification results of a 3D mesh representing head of a human being. One can remark that the proposed adaptive simplification approach based on visual saliency preserves salient regions such as the eyes, the fore head ,the eyebrows and the mouth in the case of a strong simplification (figure 35 (d)). Indeed, further triangles are maintained at the level of these regions in comparison with the method of (Garland and Heckbert, 1997). Figure 35 (e) and Figure 35 (f) respectively show these results in a wireframe representation.

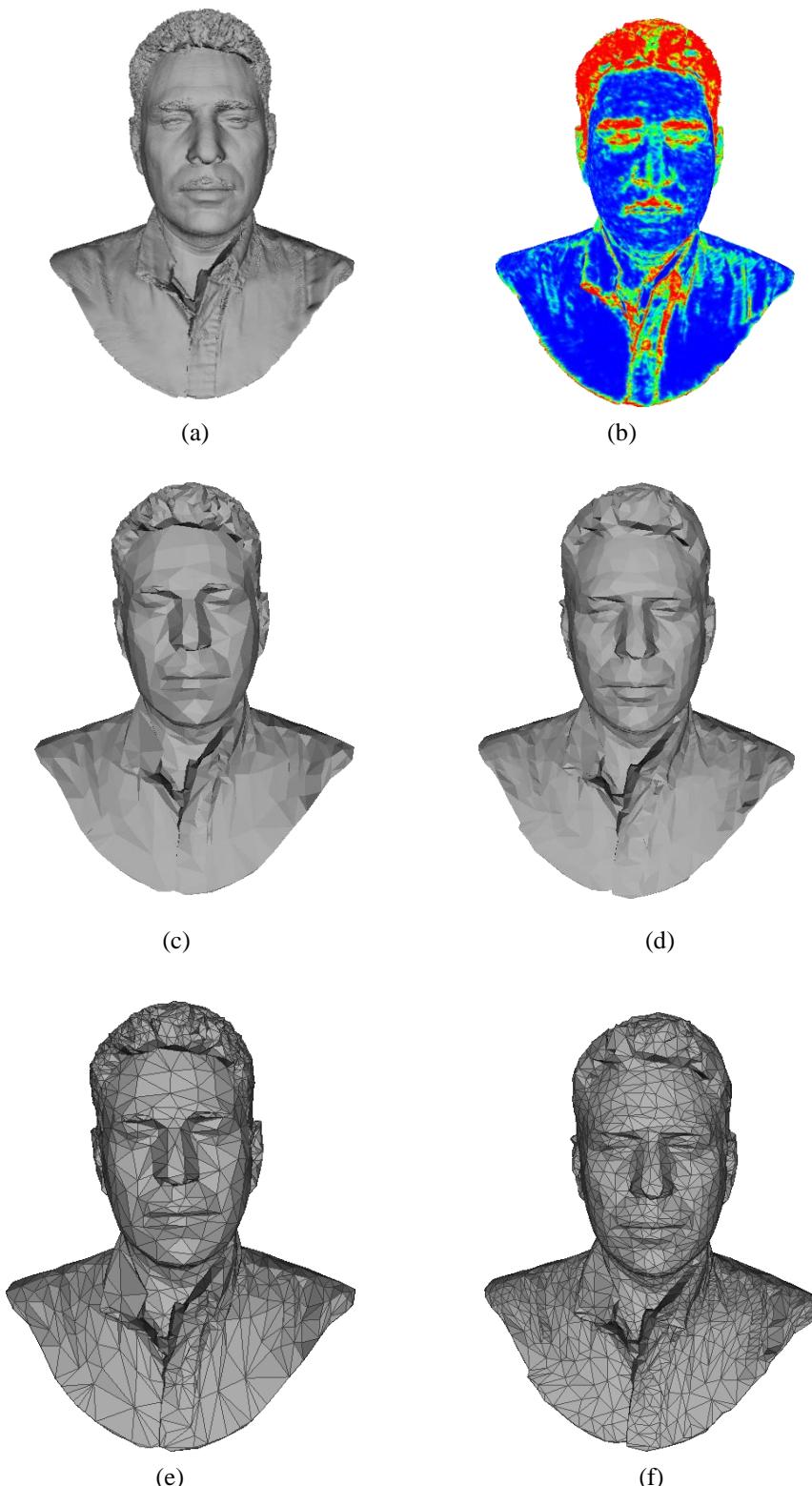


Figure 35. Simplification results of a 3D mesh representing head of a human being with and without the use of a saliency map: a) Original 3D mesh (276402 faces), b) geometric saliency map of (a), c) 3D mesh simplified with the method of (Garland & Heckbert, 1997) (5000 faces), d) 3D mesh simplified with the proposed saliency-based method of (Nouri et al., 2016a) (5000 faces), e) wireframe representation of (c), and f) wireframe representation of (d).

4.4 APPLICATION TO ADAPTIVE SMOOTHING AND DENOISING OF 3D COLORED MESHES

In 3D mesh processing, smoothing and denoising are of major importance for noise suppression which may affect geometric or colorimetric properties of a 3D mesh. However, the smoothing process induces undoubtedly a loss of details on the surface mesh which may distort its visual rendering. Nouri et al., (2016a) propose to modify the diffusion process defined in (Elmoataz et al., 2008) which performs an isotropic smoothing:

$$\begin{cases} f^{(0)} = f \\ f^{(t+1)}(u) = \frac{\sum_{v \sim u} w(u, v) f^{(t)}(v)}{\sum_{v \sim u} w(u, v)} \quad \forall u \in V \end{cases}$$

where f is a function associating a set of RGB color vectors \vec{c}_i to vertices of the colored mesh M

$f : G \rightarrow c \subset \mathbb{C}^3$ and $w(u, v)$ is the edge weight between the two vertices v_i and v_j .

In order to preserve details and salient features of the surface mesh while the smoothing process, Nouri et al., (2016a) propose to modify the diffusion process of (Elmoataz et al., 2008) with a multi-scale saliency weight:

$$\begin{cases} f^{(0)} = f^0 \\ f^{(t+1)}(u) = \frac{\sum_{v \sim u} w(u, v) f^{(t)}(v) MS(u) MS(v)}{\sum_{v \sim u} w(u, v)} \quad \forall u \in V \end{cases}$$

where $MS(u)$, $MS(v)$ represent the multi-scale colorimetric saliency of vertices u and v , and $w(u, v)$ is the colorimetric similarity defined as:

$$w(u, v) = \exp\left(\frac{\|\vec{c}(v) - \vec{c}(u)\|_2^2}{\sigma_{\vec{c}}(u) \sigma_{\vec{c}}(v)}\right)$$

Where $\sigma_{\vec{c}}(v) = \max_{v \sim u} (\|\vec{c}_u - \vec{c}_v\|_2)$.

Figure 36 presents the denosing results of a 3D colored mesh whose RGB colors have been affected with a Gaussian noise (figure 36 (b)). It's easy to remark that saliency weights integrated in the diffusion process allows the suppression of noise while preserving the contrast at the level of eyes, the mouth and the stomach of the 3D colored mesh (figure 36 (d)) contrary to the result provided by the method of (Elmoataz et al., 2008) (figure 36 (c)).



Figure 36. Denoising of 3D colored meshes: a) Original colored 3D mesh, b) 3D noised colored mesh (Gaussian noise on RGB colors), c) 3D denoised colored mesh with method of (Elmoataz et al., 2008) and d) 3D denoised colored mesh with the method of (Nouri et al., 2016a).

5. CONCLUSION AND PERSPECTIVES

In this chapter, a complete and detailed state-of-the-art related to the visual saliency of 3D meshes has been presented. Limits of the state-of-the-art approaches have been identified and analyzed. Afterwards, a recent saliency model for non colored meshes as well as a pioneering approach for 3D colored meshes have been presented and analyzed. These recent approaches associated to the works of (Nouri et al., 2015a, Nouri et al., 2015b) provide respectively a geometric and colorimetric saliency map which point out regions with high potential of saliency. Local descriptors, in the form of patches of adaptive size filled with a local height field depending on the heights of the spherical neighborhood's projection associated to a target vertex or on its mean RGB values, are constructed in order to characterize the surface of a 3D mesh. The analysis of the model of (Nouri et al., 2015a) have shown that a spherical neighborhood with an important ray allows a more accounted detection of saliency (detection of large salient regions on the surface mesh), while a restraint ray leads to detect only finest salient areas related to small details. The generated saliency maps presented in the above sections depend of a ray specified empirically for each 3D mesh. For applications requiring a saliency map corresponding to a local analysis of the surface mesh, these should choose a reduced spherical neighborhood (small ε). Otherwise, applications requiring a saliency map associated to a less local analysis of the surface mesh will consider a larger ε . However, for applications whose need is a multi-scale saliency map computed in real time, an automatic ε has to be computed automatically. The analysis of this model has shown that the ε ray is inversely proportional to the number of vertices. Therefore, a first

solution was to consider a coefficient of proportionality depending of the number of vertices of the 3D mesh. The associated results have not been very satisfying. Another solution would consist of taking into account the areas of the mesh face in order to precise ε automatically.

In the above sections, results of the detected geometric and colorimetric saliency have allowed to distinguish the difference between them. These multi-scale saliency maps (geometric and colorimetric) can be merged in order to produce a final saliency map which take into account both the geometric and colorimetric saliency (HVS is sensitive to both the shape and colors). Figure 32 presents the fusion result of the two saliency maps obtained from a simple multiplication. The final saliency map (figure 32 (e) and figure 32 (j)) appear more informative than the previous saliency maps (geometric and colorimetric). Obviously, the two saliency maps could be merged in a more effective way. The aim of the proposed fusion is to show the intake of the colorimetric saliency to the geometric saliency map.

Currently, there's no ground truth related to the colorimetric saliency of 3D meshes. Future works will aim to lead various experimentations with an eye tracker in order to produce a ground truth associated to the constructed colored mesh database. This ground truth will permit to validate the colorimetric saliency model proposed in (Nouri et al., 2015b).

Also, a novel ground truth associated to the saliency of non colored meshes will be of great interest. Indeed, the pseudo ground truth saliency (Chen et al., 2012) used in the work of (Nouri et al., 2015a) and another one proposed in (Dutgaci et al., 2011) are not fully related to the *Bottom-up* attentional process. In online experimentations associated to (chen et al., 2012), observers were asked to choose 3D points that might be selected by others observers. In the second ground truth (Dutgaci et al., 2011), observers were manipulating the 3D meshes while selecting 3D points which seemed interesting to them. This takes the form of a task provided to an observer which influence the orientation of his visual attention and therefore the detection of saliency (Yarbus, 1967). In addition, no eye-tracker was used to track the ocular movements that are sensitive to the intrinsic features of the visual stimuli. Consequently, a new ground truth related to the visual saliency of 3D non colored meshes associated to the bottom process is mandatory. This constitutes one goal of our future works.

The first publicly available 3D colored mesh database in the state-of-the-art have been described in details in this chapter. Examples of acquired meshes, as well as the characteristics of this database such as the size, the distortions selected and the process of acquisition have been analyzed.

This chapter have also presented three applications of the saliency models proposed in (Nouri et al., 2015a) and (Nouri et al., 2015b) in the processing of 3D non colored and colored meshes. The contribution of the saliency information is notable in all presented applications. However, some improvements can be developed:

Application to the optimal viewpoint selection: a first improvement would be to perform a local search around the generated optimal viewpoint. This will permit to refine the viewpoint maximizing saliency.

Application to the adaptive simplification: in the process of the adaptive simplification, salient vertices are preserved according to their degree of saliency. One improvement would be to more preserve these salient vertices by amplifying their saliency degree if the latter is superior to a defined threshold.

II. PERCEPTUAL QUALITY ASSESSMENT OF NON COLORED 3D MESHES

1. INTRODUCTION

We live in a digital world where 3D data are ever more present. With the development of 3D acquisition techniques large quantities of 3D objects are represented mostly in the form of triangular 3D meshes and are used in several human centered applications like compression, watermarking ,medical imaging, content enhancement and so forth. This progress, coupled with the fact that the decisions of human beings are strongly based on their vision, require that the 3D meshes representing the targets to analyze are of high quality.

A 3D mesh may be subject to various processing before being presented to a human observer. This can induce distortions that may affect its visual rendering. The first distortion that may affect a 3D mesh while its acquisition is the sensor noise. Then, for an optimal transmission, the 3D mesh can be compressed for the purpose of reducing size and lightening bandwidth. Also, for copyrights and intellectual property protection, a watermarking process may be performed to the 3D mesh. This process can also distort the surface mesh if it's not well performed. Thenceforth, assuming that one or more distortions previously listed are applied, a perceptual quality assessment becomes necessary in order to quantify the visual impact of these distortions on the geometry of the 3D mesh presented to the final consumer and which is generally the human observer.

In the following, we begin by briefly describing a feature of the HVS influencing the human perception and which is often used in the objective quality assessment of 2D images and 3D meshes. Afterwards, we present the two approaches for the quality assessment: Subjective and Objective quality assessment. While the first calls on human observers to evaluate and provide a quality score to a 3D mesh, the second aims at designing automatic quality assessment algorithms of the perceived quality called also *metrics*. These must provide a quality score that is correlated to the quality score provided by human observers.

2. THE HUMAN VISUAL SYSTEM AND THE PERCEPTION OF QUALITY

2.1 VISUAL MASKING EFFECT

Visual masking is one of the principal features of the HVS which have been strongly used in the quality assessment of 2D images and 3D meshes. The latter can be defined as the visibility reduction of a stimuli (masked signal) due to the simultaneous presence of another signal (masking signal). The degree of the visual masking effect is measured by the variation of the visibility of the masked signal while the presence or the absence of the masking signal. Generally, the visual masking effect is important when the spatial localization, the frequency content and the orientation of the masking and masked signals are similar.

In the context of 3D meshes, the visual masking effect can occur when a rough region is able to hide a distortion. Figure 37 illustrates this case. One can remark that the added noise is visible when it is applied on a smooth surface in contrary to a rough surface.

2.2 THE PERCEPTION OF 3D MESHES QUALITY

Whereas objective quality assessment of 2D images has experienced notable progress (Wang & Bovic, 2006) (Wang and Bovik, 2011), that of 3D meshes still is in its infancy (Corsini et al., 2013) (Guo et al., 2016). Several authors have chosen to use 2D images metrics with a good comprehension of psychovisual and physiological factors in the quality assessment of 3D meshes (Qu & Meyer, 2008) (Bolin & Meyer, 1998) (Zhu et al., 2010). Generally, 2D views of the target 3D mesh are taken, then, a metric specific for 2D images is applied in order to assess the quality of these views. However, this type of metrics does not appear to be adequate to assess the quality of 3D meshes as confirmed in (Rogowitz et al., 2001). In this study, authors show that the depth, the animation and the orientation of illumination of 3D meshes affect the perception of quality. For example, the simplification distortion is less visible when a 3D mesh is animated and illuminated from face. The study concludes that model-based metrics operating on the geometry of a 3D mesh predict better the quality of 3D meshes.

Recently, the study of (Lavoué et al., 2016) compares performances of six based-image metrics with four model-based metrics. The results of this study show that 2D image metrics assess quality of 3D meshes with precision when only one distortion is considered in the corpus. In other complex cases where different types of distortions are present, the model-based metrics outperform 2D image metrics in term of correlation with human score quality. Therefore, all the metrics considered in the following are model-based metrics operating directly on mesh geometry.

2.3 SUBJECTIVE QUALITY ASSESSMENT

Despite the proliferation of 2D image quality assessment algorithms and the advances currently being made in quality assessment approaches of 3D meshes, human beings remain the only reliable judges of any multimedia content. As a consequence, objective quality score which are provided by quality assessment metrics must correlate with human scores of quality obtained in the psychovisual experiments. However, this method of evaluation of quality remains inappropriate due to the huge amount of data obtained while psychovisual experiments and the time consuming associated (ITU-T Recommendation P.910, 2008). Nevertheless, this method allows to compute mean score of quality denoted *MOS* (*Mean Opinion Score*) for each considered 3D mesh. Therefore, the *MOS* values represent the ground truth that any measure must reproduce as rigorously as possible.

In contrary of 2D images (ITU-R BT.500-11, 2002) (ITU-T Recommendation P.910, 2008), there is no standard for the subjective quality assessment of 3D meshes. Notwithstanding such deficiency, researchers have been inspired from 2D images standards in order to design psychovisual campaigns for the quality assessment of 3D meshes and hence generate ground truths available for the scientific community (Lavoué et al., 2006) (Lavoué, 2009) (Silva et al., 2009).

2.4 OBJECTIVE QUALITY ASSESSMENT

In addition to the disadvantage of the subjective quality assessment approach described above, is added the impossibility of integrating this latter into 3D mesh processing algorithms for purposes of optimization and visual rendering. These algorithms need a high execution speed as well as an optimal complexity, conditions that are not covered by the subjective quality assessment approach. The alternative approach to these problems consists in assessing the quality objectively using algorithms (or metrics) that predict the perceived quality. The objectives to be achieved by while designing a quality assessment metric must can be summarized as:

- 1) Reliability of predicted scores: predicted *MOS* values of quality scores provided by the metric must correlate with the quality scores provided by human observers.
- 2) Reproducibility: the source code of a proposed metric must be stable and able to provide identical results on the same data regardless of the used CPU.

In the state-of-the-art, quality assessment metrics are classified into three categories:

1) Full Reference metrics (*FR-Full Reference*): this type of metrics necessitates that the reference version (renowned perfect, without distortions) of the 3D mesh is present in its entirety. This reference version will be used to quantify the degree of distortion associated to the degraded version. Note that in this context, it is more convenient to use the term « fidelity measure or similarity measure » instead of quality measure or quality metric since the goal is to measure the degree of conformity of the distorted mesh to the reference mesh. This category of metrics is used in the comparison of compression or enhancement algorithms of 3D meshes. Another goal of these metrics correlated to human perception is to replace geometric metrics such as the Haussdorf distance or the mean squared error (MSE) which don't take into account of any feature of the SVH nor a visual correlated attribute. These geometric metrics fails to assess the quality similarly to the human visual perception as showed in (Lavoué & Corsini, 2010) (Bulbul et al., 2011) (see figure 37).

2) Reduce Reference metrics (*RR-Reduced Reference*): Reduced reference quality assessment algorithms aims at assessing the visual quality of a 3D distorted mesh by having only partial information about the reference 3D mesh. The visual quality score is predicted using features extracted from both the reference and the distorted meshes.

2) No Reference metrics (*NR-No Reference*): No reference quality assessment metrics are considered as the most interesting insofar as the availability of the reference 3D mesh isn't necessary to assess the quality of a 3D distorted mesh. Hence, this type of metrics can be integrated in all processing algorithms of 3D meshes. The

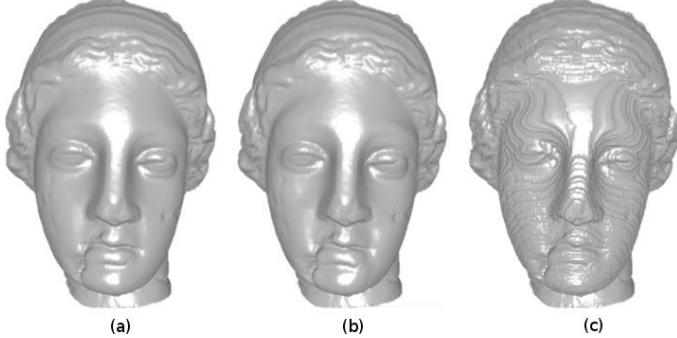


Figure 37. Geometric metric and correlation with human perception: a) Venus reference 3D mesh, b) Venus 3D mesh watermarked with the method of (Wang et al., 2011) and c) Venus 3D mesh watermarked with the method of (Cho et al., 2007). The two 3D meshes obtain a similar quality score computed with the MRMSE (Maximum Root Mean Squared Error) distance despite their different visual rendering (Wang et al., 2012).

application fields of no reference metrics is very wide; they can be used in systems of acquisition, compression and communication, visualization, rendering, watermarking, etc.

3. RELATED WORKS

Research area of perceptual quality assessment metrics of 3D meshes has raised the interest of many researchers especially for its considerable use in various applications, but also because of its lack of maturity. Primary quality assessment metrics were purely geometric due to their simplicity and fast execution time. We can cite the Hausdorff distance (Aspert et al., 2002), the Root Mean Squared error (RMS) (Cignoni et al., 1998) and the Maximal Root Mean Squared error (MRMS). These metrics provide quality score not correlated to the human visual perception. To fix this problem, researchers have integrated some HVS features. In the following a brief state-of-the-art of model-based quality assessment metrics (also called viewpoint independent metrics) is described.

Karni and Gotsman, (2000) proposed a metric that combines the RMS geometric distance between corresponding vertices with the Laplacian coordinate error in order to evaluate their compression method. In the same context, Sorkine et al., (2000) improve this metric by increasing the weights associated to the Laplacian coordinate error. Gelasca et al., (2005) proposed a metric based on the variation of the global roughness measure. The roughness is computed as the difference between the 3D mesh and its smoothed version. Acting on the same principle, Corsini et al., (2007) proposed a metric based on the variation of the global roughness measure. The roughness is computed through the variance of the dihedral angles. Lavoué et al., (2011) proposed an extension of the SSIM index developed for 2D images to the quality assessment of 3D meshes (called Mesh Structural Distortion Measure: MSDM). Differences of statistics are computed on the curvature maps of the two meshes being compared. Brian et al., (2009) measure the quality of a 3D mesh using the concept of strain energy. Strain energy refers to the energy that causes the deformation of the mesh geometry. Lavoué (2011) proposed an improvement of MSDM called MSDM2. This metric takes into account the multi-scale aspect and can perform on 3D meshes with different connectivities. Vásá & Rus (2012) suggest to evaluate the quality of a 3D mesh by measuring the local changes of oriented dihedral angles. Wang et al., (2012) proposed a metric based on the variation of local roughness that is derived from the Laplacian of the discrete Gaussian curvature. Torkhani et al., (2014) proposed a metric based on the comparison of the curvature tensors and the roughness character of the geometry. One can remark that the presented metrics model principally one feature of the human visual system: visual masking. Another major aspect of human perception is related to visual saliency. This feature is necessary to human perception insofar that the human glance is attracted by distinctive regions having high saliency potential belonging to the surrounding environment. Nouri et

al., (2016b) propose the first full-reference metric using both visual saliency and visual masking properties for the objective quality assessment of 3D meshes.

Another remark can be made from the state-of-the-art description. No approach was proposed to assess the quality of a 3D distorted mesh without the availability of its reference version. Indeed, assessing the quality of a 3D distorted mesh is a new challenging problem since no information of the original 3D version is available. To address this problem, Nouri et al., (2017) propose a no-reference algorithm which permits to assess the perceptual quality of a 3D mesh without any reference to its reference version neither to the type of distortion.

4. FULL-REFERENCE SALIENCY-BASED 3D MESH QUALITY ASSESSMENT INDEX (NOURI ET AL., 2016b)

4.1 Synopsis

Nouri et al., (2016b) propose a novel metric named SMQI (Saliency Mesh Quality Index) for the full-reference quality assessment of 3D non colored meshes. The proposed metric is inspired from the approach of (Wang et al., 2004) named SSIM related to the quality assessment of 2D images and the approach of (Lavoué et al., 2006) named MSDM related to the quality assessment of 3D meshes. However, instead of computing local statistics reflecting structural informations using a curvature map as MSDM, Nouri et al., 2016(b) propose to generate a saliency map that is used as basis for computing compute local statistics over corresponding neighborhoods on the two compared 3D meshes. Indeed, the authors make the assumption that the perceived quality of a mesh is strongly related to the modification of local and global saliency of the mesh surface. That is to say, if a specified region on the reference mesh is attractive in a certain degree and if the latter on the distorted mesh is no longer attractive (or is more attractive), hence, this region will have been distorted. Additionally, for the two compared 3D meshes, we use a roughness map on which we compute the differences of mean local roughness of each node. This allows us to capture the visual masking effect that may occur while a rough region is able to hide a geometric distortion. This roughness information is combined with the statistics computed from the saliency map in order to provide a quality score of the target 3D mesh. Figure 38 presents a synopsis of the approach of (Nouri et al., 2016b).

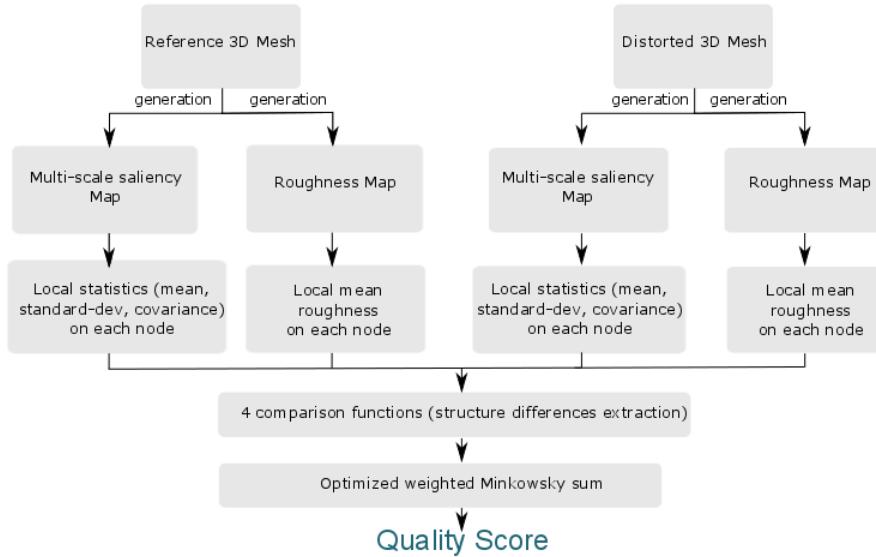


Figure 38. Block diagram illustrating the pipeline of the approach of (Nouri et al., 2016b).

4.2 Motivation: Visual quality assessment and visual saliency

Selective visual attention is a crucial characteristic in the HVS. We are constantly swallowed into a huge amount of information which cannot be neither processed in sum nor in details. Therefore, the natural solution is the real time selection of fractions of the available informations for detailed processing while the rest is discarded. Given a 3D object, human visual attention is attracted by particular regions on the surface object that are distinct from their 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 (Itti & Koch, 1998). Therefore, a saliency map can point out where the human visual attention is maximal or minimal. As said above, a distortion is much more perceived when it is located in a perceptually salient region (high scalar value on the saliency map) where the human visual attention is maximal and vice-versa. This has been proved by a series of subjective experiments where the outcome confirmed that a perceived degradation is the highest when a distortion affects a salient region in the content (Boulos et al., 2009) (Engelke et al., 2010). The same result can be seen in figure 39. Figure 39 (a) and figure 39 (b) respectively represent the reference 3D mesh and its multi-scale saliency map. Figure 39 (c) and figure 39 (d) represent respectively the noisy versions in more salient and less salient areas. We can easily notice that a distortion is much more annoying when it is located in salient areas (figure 39 (c)) than in other areas (figure 39 (d)). Also, the global perceived quality of the distorted 3D mesh in salient regions (figure 39 (c)) is more affected than the one distorted in less salient regions (figure 39 (d)). The reference perceptual metrics TPDM (Fakhri et al., 2014), MSDM2 (Lavoué et al., 2011) FMPD (wang et al., 2012) of the state-of-the-art are limited by this case. They provide a higher objective score quality for the 3D mesh distorted in less salient areas (figure 39 (d)) $TPDM=0.14$ $FMPD=0.54$ $MSDM2=0.41$ and a lower one for the 3D mesh distorted in salient areas (figure 39 (c)) $TPDM=0.008$ $FMPD=0.15$ $MSDM2=0.36$. Note that for those metrics, a higher objective score denotes a poor quality and vice versa. This shows the most outstanding metrics of the state-of-the-art fails in assessing the perceptual quality similarly to the human perception when distortions are located on visual salient regions.

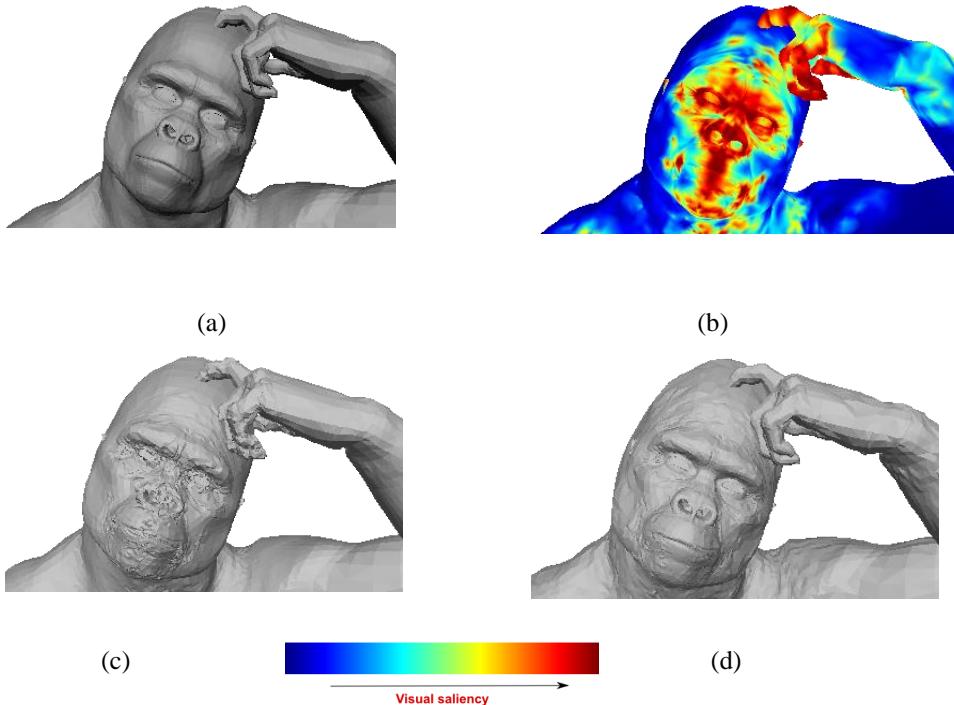


Figure 39. Comparison of 3D meshes with different perceptual qualities. (a) Original Gorilla 3D mesh. (b) Saliency map of (a) with the method of (Nouri et al., 2015a). (c) Gorilla 3D mesh noised in more visual attention areas. (d) Gorilla 3D mesh noised in less visual attention areas. (e) color-map.

4.3 The visual saliency map for structural informations extraction

To compute the local statistics (mean, standard-deviation and covariance) which reflect the structural information of a 3D mesh, we use the multi-scale saliency map as a basis. For a local neighborhood $N(v_i)$ representing the adjacent vertices of v_i on the mesh surface, we define the local mean saliency and the standard deviation respectively denoted $\mu_{N(v_i)}$ and $\sigma_{N(v_i)}$ as:

$$\mu_{N(v_i)} = \frac{1}{|N(v_i)|} \sum_{v_j \in N(v_i)} MS(v_j) \quad \sigma_{N(v_i)} = \sqrt{\frac{1}{|N(v_i)|} \sum_{v_j \in N(v_i)} (MS(v_j) - \mu_{N(v_i)})^2}$$

Where $MS(v_j)$ represents the multi-scale saliency of vertex v_j and $|N(v_i)|$ is the cardinality of the neighborhood of v_i .

For two corresponding local neighborhoods of the two compared meshes (the reference mesh M_1 and the distorted mesh M_2) $N_1(v_i) = N_{M_1}(v_i)$ and $N_2(v_i) = N_{M_2}(v_i)$, Nouri et al., (2016b) define the covariance $\sigma_{N_1(v_i)N_2(v_i)}$ as:

$$\sigma_{N_1(v_i)N_2(v_i)} = \frac{1}{|N_1(v_i)|} \sum_{v_j \in N_1(v_i), N_2(v_i)} (MS_{M_1}(v_j) - \mu_{N_1(v_i)})(MS_{M_2}(v_j) - \mu_{N_2(v_i)})$$

Where MS_{M_1} and MS_{M_2} represent respectively the multi-scale saliency maps of the two compared meshes M_1 and M_2 . Afterwards, similarly to (Lavoué et al., 2006), Nouri et al., (2016b) define three comparison function between two corresponding neighborhoods $N_1(v_i)$ and $N_2(v_i)$ in order to quantify the deformation that have affected the structural informations of the distorted 3D mesh:

$$L(N_1(v_i), N_2(v_i)) = \frac{\|\mu_{N_1(v_i)} - \mu_{N_2(v_i)}\|_2}{\max(\mu_{N_1(v_i)}, \mu_{N_2(v_i)})} \quad (17)$$

$$C(N_1(v_i), N_2(v_i)) = \frac{\|\sigma_{N_1(v_i)} - \sigma_{N_2(v_i)}\|_2}{\max(\sigma_{N_1(v_i)}, \sigma_{N_2(v_i)})} \quad (18)$$

$$S(N_1(v_i), N_2(v_i)) = \frac{\|\sigma_{N_1(v_i)}\sigma_{N_2(v_i)} - \sigma_{N_1(v_i)N_2(v_i)}\|_2}{\sigma_{N_1(v_i)}\sigma_{N_2(v_i)}} \quad (19)$$

where L, C and S refer respectively to the saliency comparison, the contrast comparison and the structure comparison.

4.4 The roughness map for taking into account the visual masking effect

Once the saliency comparison functions defined, Nouri et al., (2016b) noticed that the visual masking effect on the 3D meshes isn't captured well by these functions when a rough region is present on the reference surface mesh. Indeed, given a rough and a smooth region, a distortion will be much more visible on the smooth region than on the rough one. To deal with this problem, we implemented the work of (Wang et al., 2012) that provides a roughness map of a 3D mesh based on the Laplacian of the Gaussian curvature. Consequently, we introduced a fourth function based on the comparison of the mean local roughness. The aim of this function is to induce a large difference when a smooth region becomes a rough region and is defined as follows:

$$R(N_1(v_i), N_2(v_i)) = \frac{\|\delta_{N_1(v_i)} - \delta_{N_2(v_i)}\|_2}{\max(\delta_{N_1(v_i)}, \delta_{N_2(v_i)})} \quad (20)$$

where $\delta_{N_1(v_i)} = \frac{1}{|N_1(v_i)|} \sum_{v_j \in N_1(v_i)} LRF(v_j)$ and $LRF(v_j)$ represents the roughness value of the vertex v_j . It's

important to note that a roughness map is different from a saliency map since only novel non-redundant information are putted forward in a saliency map in contrary to a roughness map. (See figure 40).

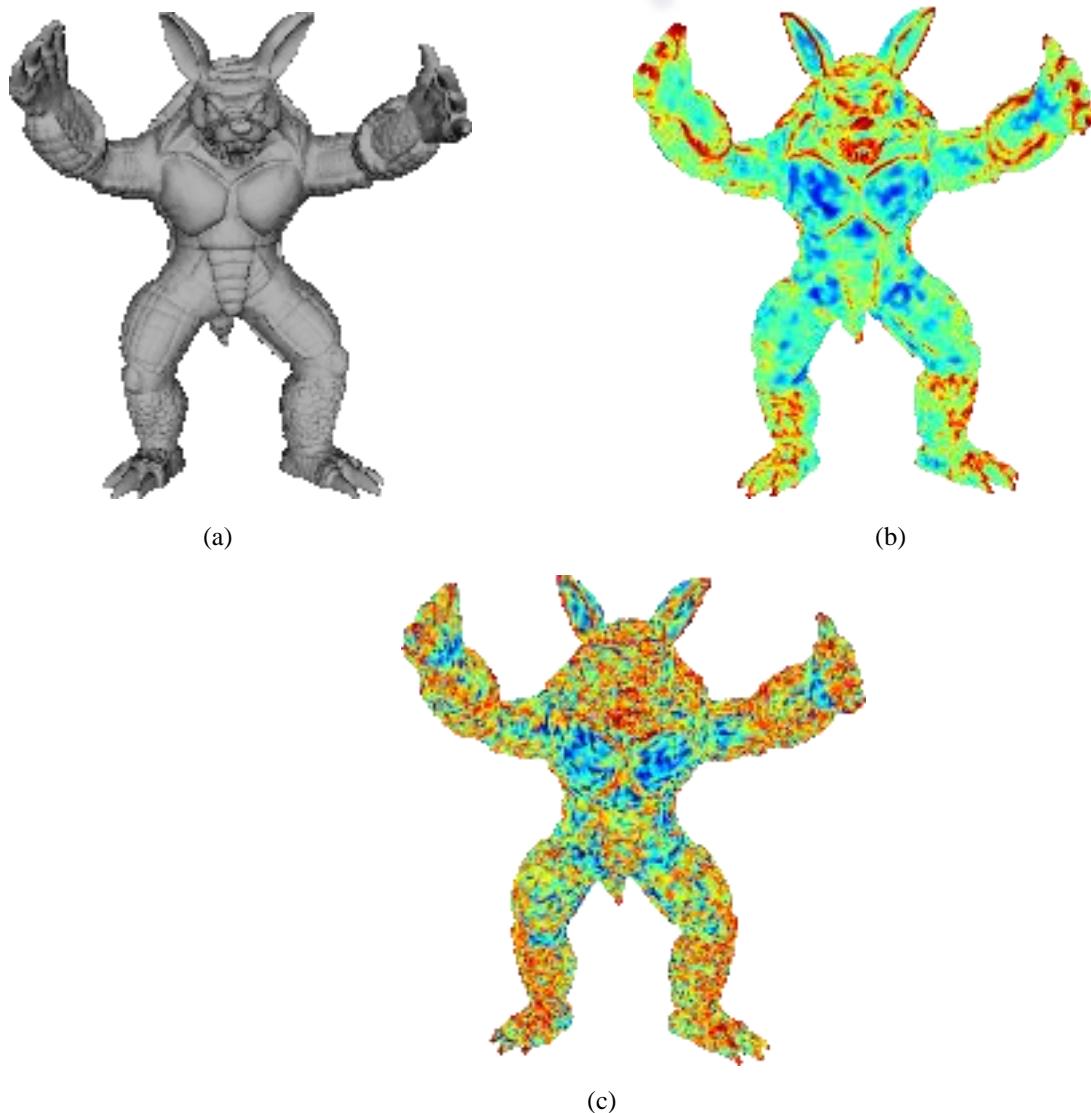


Figure 40. Difference between a multi-scale visual saliency map and a roughness map: (a) The original 3D mesh Armadillo, (b) its multi-scale saliency map with the method of (Nouri et al., 2015a) and (c) its roughness map

which the method of (Wang et al., 20120). Note that only novel and non-redundant informations are pointed on the saliency map

4.4 Learning and quality score prediction

Once the four functions are defined, extracted features from saliency and roughness maps are combined. Nouri et al., (2016b) have chosen one of the most used combinations in the field of quality assessment: the Minkowski sum. Hence, the quality measure SMQI between two 3D meshes M_1 and M_2 is defined as the weighted Minkowski sum of their local distances:

$$SMQI(M_1, M_2) = \left(\frac{1}{|V|} \sum_i^{|V|} L(N_1(v_i), N_2(v_i)) \right)^\alpha + \left(\frac{1}{|V|} \sum_i^{|V|} C(N_1(v_i), N_2(v_i)) \right)^\beta + \left(\frac{1}{|V|} \sum_i^{|V|} S(N_1(v_i), N_2(v_i)) \right)^\gamma + \left(\frac{1}{|V|} \sum_i^{|V|} R(N_1(v_i), N_2(v_i)) \right)^\delta \quad (21)$$

where coefficients α, β, γ and δ are weights obtained from an optimization based on genetic algorithms. Indeed, the proposed perceptual distance depends of four independent parameters of which the manual tuning will be difficult and probably ineffective. It's important to note that the number of 3D meshes contained in the two datasets described in next sections is small in comparison with available datasets of 2D images. Consequently, to deal with this weakness, we opted for a Leave-One-Out training on a corpus constructed from the two considered databases. The goal of this approach is to perform the learning of the model on $k - 1$ observations and to validate it on the k^{th} one. This process is repeated $k \times 999$ times. In this case, an observation refers to the MOS values of a reference 3D mesh and its distorted versions. The fitness function used to perform the genetic optimization is defined as:

$$f(\alpha, \beta, \gamma, \delta) = \sqrt{\sum_{i=0}^{k-1} (MOS_i - SMQI_i(M_1, M_2))^2}$$

Where MOS_i represents the vector of MOS values of the observation i and $SMQI(M_1, M_2)$ is the perceptual distance computed with equation (21). After genetic optimization, we obtain: $\alpha = 23.63$, $\beta = 3.26$, $\gamma = 5.04$ and $\delta = 0.77$. Note that in figure 39, the Gorilla 3D mesh was assessed with these parameters. The values of these parameters have been fixed for all quality assessment experimentations in this chapter.

4.5 Results and analysis

4.5.1 3D mesh databases and subjective scores of quality

To compare the proposed full reference metric with state-of-the-art methods, two publicly available subject-rated databases are used: 1) The Liris/Epf1 General-Purpose database (Lavoué et al., 2011) and 2) the Liris-Masking database (Lavoué, 2011). The first database contains 4 reference 3D meshes. They are affected by two types of distortions: Noise addition and Smoothing. These distortions are applied in 3 different strengths either uniformly over the 3D mesh surface, specifically to rough or smooth regions (for simulating the masking effect) and to transitional areas between rough and smooth regions. In total, 22 distorted 3D meshes of each reference mesh are generated and evaluated by 12 human observers. Figure 41 shows some 3D meshes from the Liris/Epf1 General-Purpose database with their associated normalized MOS values.

The Liris Masking Database consists of 4 reference 3D meshes which are distorted by adding noise of three different strengths in either rough and smooth areas to generate 6 degraded versions of each 3D mesh. 12 human observers have evaluated the database. Figure 42 presents some 3D meshes of this database. The performance of the proposed method is measured by the Spearman Rank Ordered Correlation Coefficient (SROOC) between the subjective values and the objective scores obtained from the proposed metric.

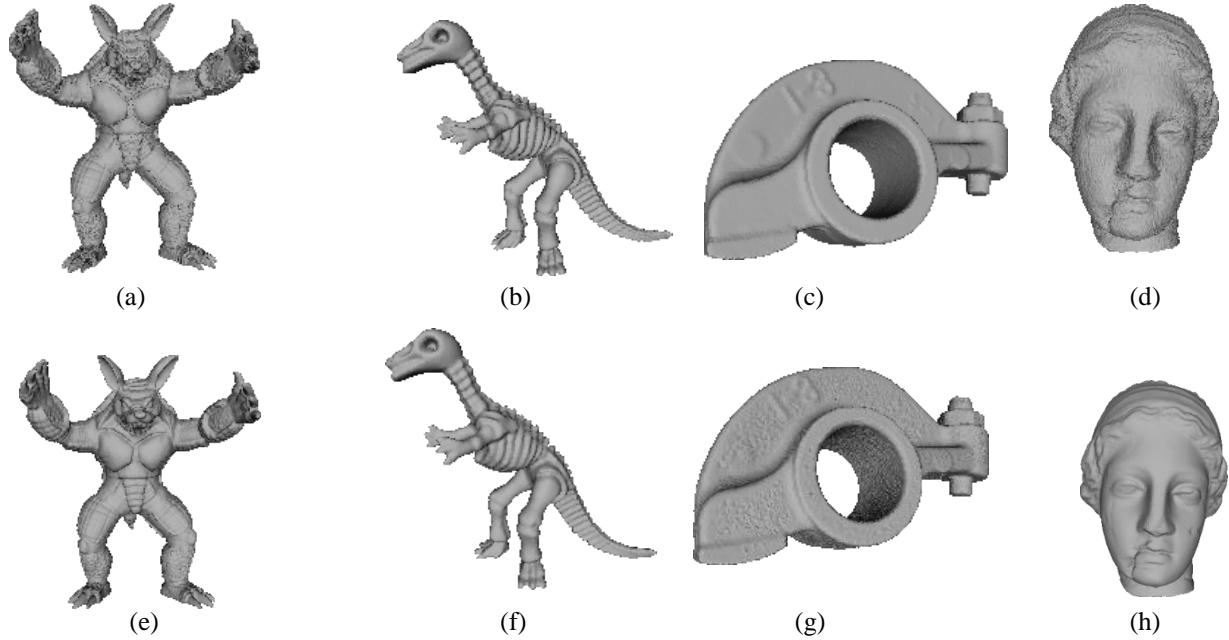


Figure 41. Example of 3D meshes from the Liris/Epfl General-Purpose database. In the top row are the four reference 3D meshes. In the bottom row are four examples of deformed 3D meshes, from left to right are respectively: Armadillo with noise on rough regions (MOS=0.84), Dinosaur after a uniform smoothing (MOS=0.43), RockerArm with noise on smooth regions (MOS=0.75) and Venus with uniform noise (MOS=1).

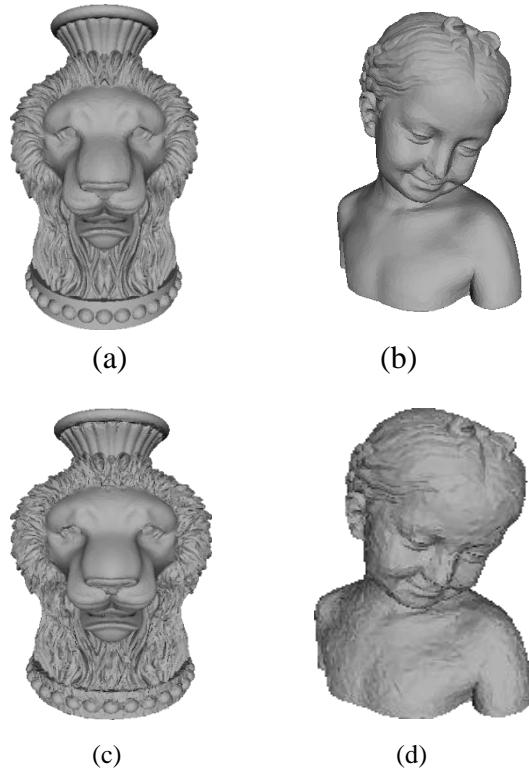


Figure 42. Example of 3D meshes from the Liris Masking database. In the top row are two reference 3D meshes. In the bottom

row are two examples of their distorted versions from left to right are respectively: Lion 3D mesh with noise on rough regions (MOS=0.20), Bimba 3D mesh with noise in smooth regions (MOS=1.0).

4.5.2 Results and comparison with the state-of-the-art

To evaluate the performance of their proposed full-reference metric, Nouri et al., (2016b) use the Spearman correlation coefficient (SROOC: Spearman Rank Ordered cOrrelation Coefficient) between the subjective scores of quality provided by the human observers and the objective scores of quality provided by the proposed metric. Commonly used, this term of correlation is used to define a link between two variables. In the field of statistics, the term « correlation » allows to quantify the link between two quantitative variables. This junction could be either symmetric (possibility to switch the quantitative variables x and y) or asymmetric (one of the two variables depends of the other). Hence, the two variable x and y cannot be switched). This measure of correlation is characterized by the fact that the only the ranks of the observations values are used instead of their real values. The Spearman correlation is defined between 1 and -1:

1: strong positive correlation.

0: no correlation.

-1: strong negative correlation.

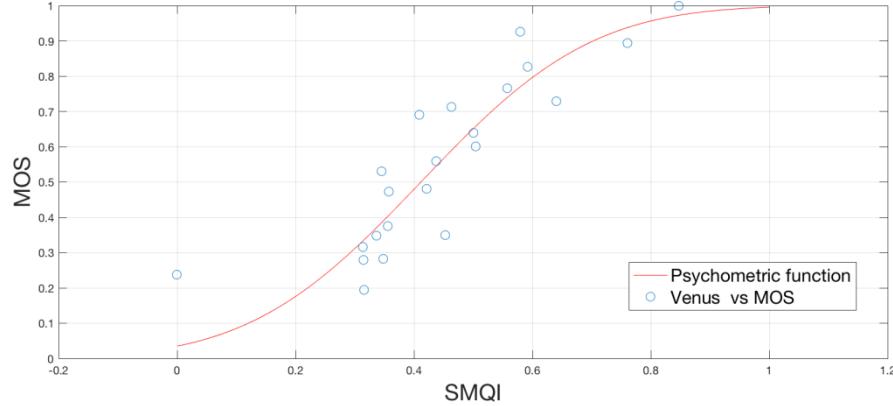
Table 2 presents the performance of the proposed metric and the state-of-the-art metrics in term of the Spearman correlation with the subjective scores provided by the Liris/Epf1 General-Purpose. One can notice that SMQI provides important correlation values for all the 3D meshes and particularly for the Venus 3D mesh where the SROOC values are the highest among the reminder values. The result of a psychometric fitting using a cumulative Gaussian psychometric function between the objective scores and the subjective ones is presented in Figure 43 (top) and confirms this result for the Venus 3D mesh. This fitting was performed by choosing the cumulative Gaussian psychometric function (Engeldrum, 2000):

$$g(m, n, R) = \frac{1}{\sqrt{2\pi}} \int_{m+nR}^{\infty} e^{-t^2} dt$$

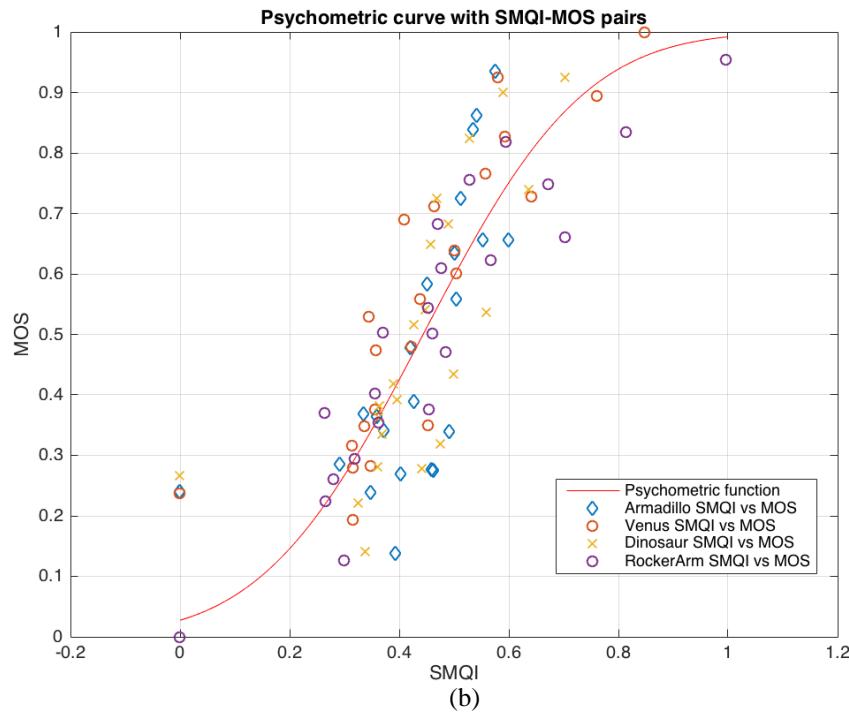
Where m and n are estimated with a non-linear least squares fitting based on the Levenberg-Marquardt algorithm and R is the objective distance. Moreover, it appears that the proposed metric SMQI is the second best metric after TPDM over the entire database (the SROOC values are 89.6% for TPDM, 84.6% for SMQI and 80.4% for FMPD). It's important to note that the SROOC value over the entire database isn't an average of the SROOC values associated to the sub-databases but it's a correlation between all the human quality scores referring to each 3D mesh and the objective provided by the proposed metric. This high correlation with the subjective rates in the LIRIS/EPFL General-Purpose database demonstrates the capacity of the generalization over all the 3D meshes and can be confirmed by the curve fitting presented in figure 43 (bottom) where the SMQI-MOS points are very close to the psychometric curve.

SMQI was also tested and compared with state-of-the-art metrics on the Liris-Masking database. Table 3 provides the Spearman correlation values of the different metrics on this database. From these results, three main observations can be made. The first one is that SMQI is very competitive with TPDM and MSDM2 and succeed in capturing the masking effect. The second observation deals with the correlation value associated to the Lion3D mesh that is slightly lower in comparison with the values of TPDM and MSDM2. This can be explained by the fact that the multi-scale saliency map of the distorted 3Dmesh on which are computed the statistics doesn't reflect well the distorted salient areas. Indeed, the reference 3D mesh and the distorted 3D use the same ray (defined empirically) for the spherical neighborhood considered while the construction of the local patches in the saliency model (Nouri et al., 2015a). The authors think that a ray specified properly to each 3D mesh (the reference and the distorted 3D meshes) will lead to a better estimation of the multi-scale saliency map. The third observation is related to the

correlation values over all 3D meshes of the database Liris-Masking database. Correlation rates over all this database are not presented since as confirmed in (Lavoué & Corsini, 2010),



(a)



(b)

Figure 43. Psychometric curve fitting between SMQI and MOS values: (top) the fitting performed using SMQI and MOS values from the Venus 3Dmesh; (bottom) the psychometric curve plotted with the SMQI-MOS pairs from the whole LIRIS/EPFL General-Purpose database.

the used protocol while the subjective evaluations of quality have specified a different referential for each 3D, and consequently, values of correlation over the whole corpus of 3D meshes are not significative.

Liris/Epf General- Purpose	HD	RMS	3DWPM1	3DPWPM 2	MSDM2	FMPD	TPDM	SMQI
Armadillo	69.5	62.7	65.8	74.1	81.6	75.4	84.9	77.5
Venus	1.6	90.1	71.6	34.8	89.3	87.5	90.6	91.6
Dinosaur	30.9	0.3	62.7	52.4	85.9	89.6	92.2	84.8
RockerAr m	18.1	7.3	87.5	37.8	89.6	88.8	92.2	91.8
Entire database	13.8	26.8	69.3	49.0	80.4	81.9	89.6	84.6

Table 2. SROOC values (%) of different viewpoint-independent metrics on the LIRIS/EPFL General Purpose database

Table 3. SROOC values (%) of different viewpoint-independent metrics on the LIRIS-Masking database

Liris- Masking	HD	RMS	3DWPM1	3DPWPM 2	MSDM2	FMPD	TPDM	SMQI
Armadillo	48.6	65.7	58.0	48.6	88.6	88.6	88.6	88.6
Lion-vase	71.4	71.4	20.0	34.3	94.3	94.3	82.9	83.00
Bimba	25.7	71.4	20.0	37.1	100.0	100.0	100.0	100.0
Dinosaur	48.6	71.4	66.7	71.4	100.0	94.3	100.0	100.0

From the above results and comparisons, it appears that the SMQI metric is strongly correlated to the human perception due to the integration of visual saliency. Additionally, SMQI is the second best metric in the Liris/Epf General-Purpose Database and is very competitive on the Liris-Masking database. Two interesting questions need to be raised after we presented the limitation of the three state-of-the-art best metrics so far and the success of the proposed metric SMQI while assessing the quality of a 3D mesh distorted in more perceptually important areas: 1) Does the performance of state-of-the-art perceptual metrics in term of correlation with human perception will still be the same when they perform on a 3Dmesh corpus distorted in more perceptual important regions than less important ones ? 2) Does the design of subjectrated databases must account for new distortions targeting further visual salient regions ? These important questions will be investigated in future works.

5.3D BLIND MESH QUALITY ASSESSMENT INDEX (MODEL OF NOURI ET AL., 2017B)

5.1 Introduction

In most of applications that manipulate 3D meshes, the reference version of 3D mesh isn't available which makes the quality assessment of the target 3D mesh more difficult. This capacity of assessing objects without their reference version is an easy task for humans but this is far from being the case for machines and algorithms. Many quality assessment metrics for 3D meshes were proposed in the state-of-the-art, however they are still limited due to their dependence to the reference version of the 3D mesh. To resolve this problem, Nouri et al. (2017b) investigate the use of saliency and visual masking properties for the blind quality assessment of 3D non colored meshes and propose a novel no-reference metric named BSMQI (Blind Saliency-based Mesh Quality assessment Index). Given a distorted non colored 3D mesh, the proposed metric is able to assess its perceived quality without referring to its reference version similarly to the human perception. No assumption on the degradation to evaluate is required for this metric, which makes it powerful and usable in any context requiring quality assessment of 3D meshes.

5.2 Proposed approach

5.2.1 Visual saliency, roughness and quality assessment

The principal challenge met while designing this quality assessment metric was to select visual features that have the capability to quantify the structural deformation that the 3D mesh undergo and that are correlated with human perception. To do this, a multi-scale visual saliency and a roughness map are used. Visual saliency is an important characteristic for human visual attention. Its use in computer graphics applications like mesh quality assessment (Nouri et al., 2016b), optimal view point selection (Nouri et al., 2015a) (Tal et al., 2012) and simplification (Lee et al., 2005) has proven beyond any doubt its correlation with human visual perception. We suppose as in (Nouri et al., 2016b) that visual quality of a 3D mesh is more affected when salient regions are affected rather than less or not salient regions. Likewise, variations of 3D mesh roughness appear to be correlated with human perception (Wang et al., 2012). Indeed, a roughness map points regions that expose a strong visual masking effect. Regions with high roughness magnitude expose an important degree of visual masking effect since distortions are less visible on these ones. Nouri et al., (2017b) show that local variations of saliency and roughness combined together succeed in assessing the visual quality of a distorted 3D mesh without the need of its referenceversion.

5.2.2 Method

Given a 3D distorted mesh, the method of Nouri et al., (2017b) begins by computing a multi-scale saliency map MS and a roughness map R respectively with the methods of (Nouri et al., 2015a) and (Wang et al., 2012). Then, the method of (Simari et al., 2014) is adapted to segment the 3D mesh into a number of Superfacets N_{SF} . In this context, these Superfacets will play the role of local patches since the HVS locally processes information. Once the segmentation is performed, values of saliency $MS(v_i)$ and roughness $R(v_i)$ associated to each vertex v_i of a Superfacet SF_j are assigned to it. Afterwards, a feature vector of four attributes for each Superfacet SF_j is constructed:

$$\phi_j = \left[\mu_{SF_j}, \sigma_{SF_j}, \delta_{SF_j}, \gamma_{SF_j} \right] \text{ with } j \in [1, N_{SF}]$$

where μ_{SF_j} and σ_{SF_j} represent respectively the local mean saliency and local standard deviation saliency of the Superfacet SF_j and are defined as:

$$\mu_{SF_j} = \frac{1}{|SF_j|} \sum_{v_i \in SF_j} MS(v_i)$$

$$\sigma_{SF_j} = \sqrt{\frac{1}{|SF_j|} \sum_{v_i \in SF_j} (MS(v_i) - \mu_{SF_j})^2}$$

where $|SF_j|$ represents the cardinality (*i.e.*, the number of vertices) of the Superfacet SF_j .

δ_{SF_j} and γ_{SF_j} denote respectively the local mean roughness and the local standard deviation roughness and are defined as:

$$\delta_{SF_j} = \frac{1}{|SF_j|} \sum_{v_i \in SF_j} LRF(v_i)$$

$$\gamma_{SF_j} = \sqrt{\frac{1}{|SF_j|} \sum_{v_i \in SF_j} (LRF(v_i) - \delta_{SF_j})^2}$$

Finally a learning step using the constructed feature vector is performed. This is done using the Support Vector Regression (SVR) (Vapnik, 1995) which is also used for scoring the visual quality of the 3D mesh. Figure 44 presents the block-diagram of the proposed approach.

5.2.3 Segmentation, learning and regression

5.2.3.1 Segmentation into Superfacets

One of the novelties of the proposed approach falls within the use of the Superfacets - the result of an over-segmentation of the mesh surface into regions whose borders fit well the semantic entities of the mesh - into the pipeline of a mesh quality assessment metric. To segment the mesh, Nouri et al., (2017b) modified the approach of (Simari et al., 2014) which, for a 3D mesh and a number of desired Superfacets, execute the following steps based on the *farthest point* principle:

Initialization: The method begins by associating the center of the first Superfacet to the triangle of which the centroid is the nearest to the global centroid of the mesh. Then, each center of a new Superfacet is affected to the triangle with the euclidian distance to the latest considered triangle is the highest.

Update of the centers: Once the triangles have been affected to different Superfacets, it's necessary to compute the new center of each Superfacet. For this, the method computes the mean area of all triangles belonging to a Superfacet and associates the new center to the triangle of which the area is the nearest to the computed mean area. If the new center is different from the prior one, the algorithm stops. Otherwise , the classification step is computed.

Classification: For each triangle, the method computes, using the Dijkstra Algorithm, the shortest paths between the centers of the defined superfacets and the triangles of the mesh. When a

triangle is considered while computing the shortest path from a Superfacet center and if the current computed distance is less than the prior stored one (obtained from the initialization step or from an expansion that started from a different center) then both distance and label associated to the considered triangle are updated (the Superfacet that contains the triangle is fixed).

Geodesic weight: Given two adjacent faces f_i and f_j sharing an edge $e(i, j)$ with a median point $m_{i,j}$ and two respective centroids c_i and c_j , the geodesic weight is defined as $geo(f_i, f_j) = \|c_i - m_{i,j}\| + \|m_{i,j} - c_j\|$. This latter is affected to the weight $w(f_i, f_j)$ of the edge $e(i, j)$ as follows:

$$w(f_i, f_j) = \frac{geo(f_i, f_j)}{d}$$

where d is the length of the diagonal of the bounding box including the 3D mesh.

In all the following experimentations, the 3D meshes have been segmented to 450 Superfacets.

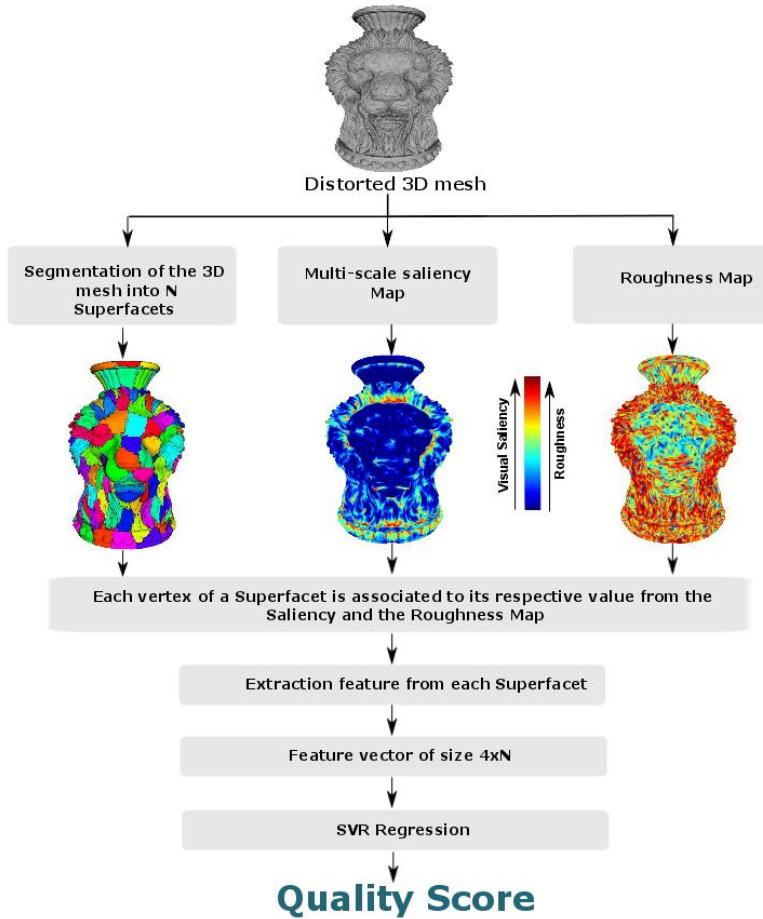


Figure 44. Block diagram of BMQI (Nouri et al., 2017b)

5.2.3.2 Learning and regression

Even if it's unusual that a human observer associates a quality score in the form of a scalar to a 3D mesh but rather proceeds to a classification of the quality according to the perceived sensation(for example: « good » or « bad » quality), nevertheless, the applications context of the quality metrics forces to provide a single scalar reflecting the perceived quality. For this, the extension of Support Vector Machines (SVM) to regression, named SVR, is used.

The aim is to estimate a function f presenting at most a maximal deviation ϵ reflecting the dependence between a features vector x_i and an affiliation class y_i . Thus, for a feature vector x_i of a distorted 3D mesh M_i with a subjective quality score y_i , the regression function of an observation x to classify is defined as :

$$f_{SVR} = \sum_{x_i \in V_s} \alpha_i y_i K(x_i, x) + b$$

where V_s are the support vectors, (x_i, y_i) is the learning set, α is the Lagrange coefficient obtained from a minimization process and $K(x_i, x)$ represents the RBF (Radial Basis Function) kernel defined as:

$$K(x_i, x_j) = \exp(\gamma \|x_i - x_j\|^2)$$

Indeed, the RBF function is often used as a kernel function due to its resemblance to a similarity measure between 2 examples to classify. Also, the motivations related to the use of the SVR areas follows:

1. The regression solution includes a small number of examples x_i (rapidity and efficiency).
- 2 .Results of the regression depend of the used kernel. The test of different kernels is beneficial in so far as the correlation rate between the objective scores and subjective ones may depends on the chosen kernel.

In order to compare the proposed no-reference metric BMQI with the state-of-the-art approaches, Nouri et al., (2017b) use the same publicly available subject-rated mesh databases as for their proposed full reference metric SMQI (Nouri et al., 2016b): 1) Liris/EpfIGeneral-Purpose database (Lavoué et al., 2006) and 2) Liris-Masking database(Lavoué et al., 2009). A learning step is carried onto the Liris-Masking database in order to determine the RBF kernel's parameters (γ and C which represents the error penalty coefficient) with a 4 parts cross-validation. Each of these parts represents the distorted versions of one of the four reference meshes associated to their subjective quality score. The SVR regression was performed using the LIBSVM (Chang & Lin, 2001) library. The selected parameters of the RBF kernel for the Liris-Masking database are: $\gamma = 0.002$ and $C = 32$. For the Liris-EpfI General Purpose database, the selected parameters are: $\gamma = 0.005$ and $C = 2$.

5.3 Results and analysis

5.3.1 Performance and comparison with the state-of-the-art

Table 4 presents the SROOC correlation values of the proposed no-reference metric and the correlation values of 7 full-reference metrics from the state-of-the-art associated to the Liris-Maskingdatabase. We can notice that the metric BMQI produces important correlation values for all the 3D meshes without the need of their reference version on the contrary of the full-reference metrics.These results confirm that the metric of (Nouri et al., 2017) succeeds very well in taking into account the visual masking effect.

Table 5 shows the correlation values of BMQI associated to the Liris/EpfI General Purpose database. One can notice that the performances of BMQI on this database are not as good as those on Liris-Masking database.

Indeed, distortions on the Liris-Epfl General Purpose database (noise addition and smoothing) are applied on four distinct regions of the surface mesh (uniform regions, rough regions, smooth regions, and transitional regions). This aims at reflecting the distortions associated to common mesh processing methods like simplification, compression and watermarking (Lavoué et al., 2006) which makes the quality assessment more difficult. From the results presented in table 5, it seems that BMQI assesses the visual quality in a multi-distortion context with less precision than in a mono-distortion context even if the correlation values of the three groups of meshes (Dinosaur, Venus and RockerArm) are important. This is mainly due to the correlation value of the sub-corpus Armadillo which is lower in comparison to the other correlation values. This can be explained by the generated multi-scale saliency map that may not reflect well the distorted salient regions. Thus, the objective quality scores aren't consistent with the quality scores of human observers. The number of Superfacets and their size are two parameters that could influence the performance of the proposed metric. A precise definition of theses parameters may improve the results. Finally, when the whole corpus of the Liris-Epfl General Purpose database is considered, the metric BMQI provides a correlation value relatively low in comparison to the full-reference metrics of the state-of-the-art. This is related on one hand to the low correlation value of the Armadillo sub-corpus and on the other hand to the number of 3D meshes considered in the learning step that is very small. Indeed, a corpus consisting in 88 meshes with their associated MOS wouldn't allow the design of an effective quality metric in a multi-distortion context.

In the light of these results, and given the capacity of the proposed approach to assess the perceived quality of a distorted mesh without the need of its reference version, BMQI seems nevertheless competitive with the full-reference methods. For example, BMQI obtains better correlation rates associated to the Liris-Masking database in comparison to the full-reference metric SMQI proposed in (Nouri et al., 2016b). This could be explained by the contribution of the segmentation of the mesh into Superfacets.

Table 4. SROOC values (%) of different viewpoint-independent metrics on the LIRIS-Masking database.

Mesh database	Full-reference								No-reference <u>BMQI</u>
	HD	RMS	3DWPM 1	3DPWP M2	MSDM2	FMPD	TPDM	SMQI	
Liris-Masking	48.6	65.7	58.0	48.6	88.6	88.6	88.6	88.6	94.3
Lion-vase	71.4	71.4	20.0	34.3	94.3	94.3	82.9	83.00	94.3
Bimba	25.7	71.4	20.0	37.1	100.0	100.0	100.0	100.0	100.0
Dinosaur	48.6	71.4	66.7	71.4	100.0	94.3	100.0	100.0	83.0

5.3.2 Performance on indépendant 3D meshes

Nouri et al., (2017b) have also tested the no-reference metric BMQI for the quality assessment of 3D meshes not belonging to any database. This permits to analyze the behavior of this metric when assessing the visual quality of any 3D mesh. Figure 44 presents two reference 3D meshes with their distorted versions. The distortions considered are: additive noise and simplification. It's important to note that the simplification distortion wasn't considered in the learning process for selecting the parameters of the RBF kernel since both subject rated databases doesn't include this type of distortion. In these experimentations, the selected parameters associated to the Liris/Epfl General Purpose are used. From the top row of figure 45, one can notice that BMQI provides coherent scores of quality in accordance with human perception. The reference mesh (figure 4(a)) obtains a perceived quality score equal to 6.13. Its noised version obtains a quality score equal to 6.25 and its simplified version (more visually distorted) obtains a quality score equal to 6.74 (note that a low score signifies a good quality score and vice versa). The same remarks could be made to the second row of figure 45.

Table 5. SROOC values (%) of different viewpoint-independent metrics on the LIRIS/EPFL General Purpose database.

Mesh database	Full-reference								No-reference <u>BMQI</u>
	HD	RMS	3DWPM 1	3DPWP M2	MSDM2	FMPD	TPDM	SMQI	
Liris/Epf G. Purpose	HD	RMS	3DWPM 1	3DPWP M2	MSDM2	FMPD	TPDM	SMQI	<u>BMQI</u>
Armadillo	69.5	62.7	65.8	74.1	81.6	75.4	84.9	77.5	20.1
Venus	1.6	90.1	71.6	34.8	89.3	87.5	90.6	91.6	88.9
Dinosaur	30.9	0.3	62.7	52.4	85.9	89.6	92.2	84.4	83.5
RockerArm	18.1	7.3	87.5	37.8	89.6	88.8	92.2	91.8	92.7
Entire database	13.8	26.8	69.3	49.0	80.4	81.9	89.6	84.6	78.1

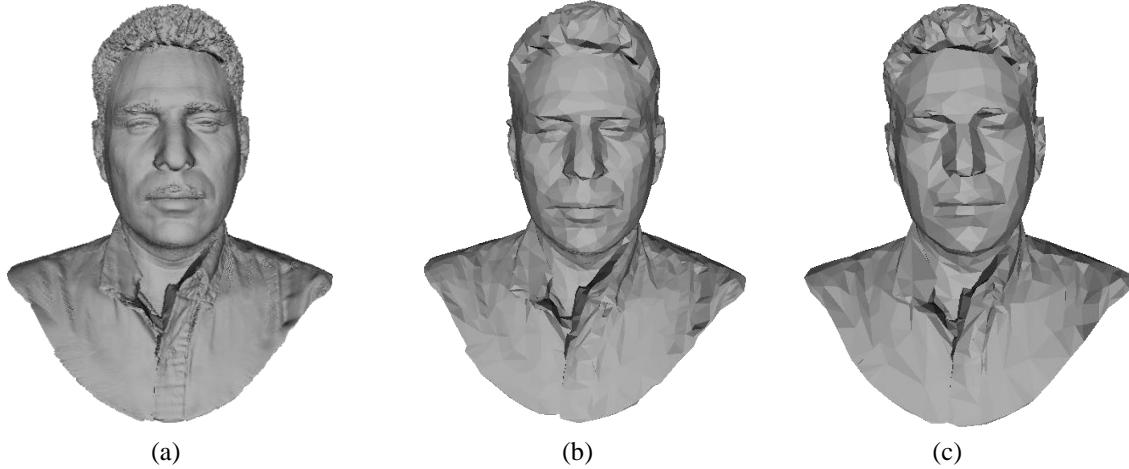


Figure 45. Results of BMQI on independent 3D meshes not belonging to any database used in the learning or the regression step: a) Reference 3D mesh (276402 faces), b) 3D mesh simplified adaptively with saliency information (Nouri et al., 2016a) (5000 faces) and c) 3D mesh simplified with the method of (Garland & Heckbert, 1997) (5000faces).

6. CONCLUSION AND PERSPECTIVES

In this second section (II. Perceptual quality assessment of non colored 3D meshes), a full-reference perceptual quality assessment metric SMQI proposed in (Nouri et al., 2016b) was described. The latter compare

structural informations of a reference 3D mesh with the ones of its distorted version. For this, a multi-scale saliency map is used from which local statistics reflecting the structures of the 3D mesh are computed. In order to take into account the visual masking effect, a roughness map is computed to measure means of roughness. Four comparisons functions are combined using a weighted Minkowski sum in order to provide an objective quality score which quantify the visual similarity between the reference 3D mesh and its distorted version. Experimental results as well as well a comparison with the state-of-the-art approaches show the strong correlation of the full-reference metric SMQI with scores of quality provided by human observers. This has attested its effectiveness. However, as explained in the results analysis, an automatic ray - which serve to consider a spherical neighborhood while the construction of local adaptive patches for the saliency map generation - proper to the geometry of the target meshes can lead to better correlation results.

SMQI don't allow assessing the fidelity of a distorted 3D mesh having a different connectivity from its reference version (the case of a simplified mesh). An improvement would be to add a matching step from vertices of the distorted mesh towards vertices of the reference 3D mesh.

Sensitivity to the modifications of mesh frequencies can be taken into account since distortions are much more visible on high frequencies than lower frequencies. A first initiative proposed by (Corsini et al., 2013) suggests to obtain frequencies of the 3D mesh from the eigenvectors associated to its Laplacian matrix and to link these eigenvectors to perceptible frequencies of the HVS. Therefore, the modification of 3D mesh components will be linked to the sensitivity to the frequencies of HVS.

The size of the considered subject rated databases play an important role in the performance of quality assessment metrics having a learning step in their pipeline. Existing 3D mesh databases have a small size. This would not allow designing a perceptual metric which is able to assess any time of distortion (Lavoué et al., 2006) (Lavoué, 2009) (Silva et al, 2009).

Also, another kind of perceptual quality assessment metrics was studied in this section. This one concerns the blind quality assessment metrics. Nouri et al., (2017b) have proposed to combine visual saliency and visual masking effect to design a no-reference quality assessment metric. This metric, BMQI, uses simple features (mean and standard deviation) computed on a multi-scale saliency map and a roughness map in order to assess the quality of a distorted 3D mesh without the need of its reference version. Correlation rates with subjective scores of quality show that BMQI is very competitive with full-reference metrics of the state-of-the-art. Two improvements would be to ameliorate the multi-scale saliency map as well as the regression used process. Indeed, a learning performed on a corpus of important size would lead to a better prediction of quality in a multi-distortion context. Another amelioration would be to considerate the multi-scale aspect from the size of the Superfacets.

REFERENCES

- Achanta, R., Hemami, S. S., Estrada, F. J. & Süstrunk, S. (2009). Frequency-tuned salient region detection.. CVPR (p./pp. 1597-1604), : IEEE Computer Society. ISBN: 978-1-4244-3992-8
- Aspert, N., Cruz, D. S. & Ebrahimi, T. (2002). MESH: measuring errors between surfaces using the Hausdorff distance.. ICME (1) (p./pp. 705-708), : IEEE Computer Society. ISBN: 0-7803-7304-9
- Bian, Z. & Tong, R. (2011). Feature-preserving mesh denoising based on vertices classification.. Computer Aided Geometric Design, 28, 50-64.
- Boulos, F. & Parrein, B., Le Callet P. & Hands D. (2009). Perceptual effects of packet loss on H.264/AVC encoded videos. In VPQM workshop.

Buades, A., Coll, B. & Morel, J.-M. (2005). A Non-Local Algorithm for Image Denoising. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2 - Volume 02 (p./pp. 60–65).

Engelke, U., Pepion, R., Le Callet, P., Zepernick, H. (2010). Linking distortion perception and visual saliency in h.264/AVC coded video containing packet loss. Proc. SPIE.

Bulbul, A., Çapin, T. K., Lavoué, G. & Preda, M. (2011). Assessing Visual Quality of 3-D Polygonal Models.. IEEE Signal Process. Mag., 28, 80-90.

Cignoni, P., Rocchini, C. & Scopigno, R. (1998). Metro: Measuring Error on Simplified Surfaces.. Comput. Graph. Forum, 17, 167-174.

Coren, S., Ward, L. M., & Enns, J. T. (2003). *Sensation and Perception*. Wiley & Son.

Corsini, M., Larabi, C., Lavoué, G., Petrík, O., Vása, L. & Wang, K. (2012). Perceptual Metrics for Static and Dynamic Triangle Meshes.. In M.-P. Cani & F. Ganovelli (eds.), Eurographics (STARs) (p./pp. 135-157), : Eurographics Association.

Corsini, M., Larabi, C., Lavoué, G., Petrík, O., Vása, L. & Wang, K. (2012). Perceptual Metrics for Static and Dynamic Triangle Meshes.. In M.-P. Cani & F. Ganovelli (eds.), Eurographics (STARs) (p./pp. 135-157)

Cho, J.-W., Prost, R. & Jung, H.-Y. (2007). An Oblivious Watermarking for 3-D Polygonal Meshes Using Distribution of Vertex Norms.. IEEE Trans. Signal Processing, 55, 142-155.

Bolin, M. R. & Meyer, G. W. (1998). A Perceptually Based Adaptive Sampling Algorithm.. In S. Cunningham, W. Bransford & M. F. Cohen (eds.), SIGGRAPH (p./pp. 299-309), : ACM. ISBN: 0-89791-999-8.

Corsini, M., Gelasca, E. D., Ebrahimi, T. & Barni, M. (2007). Watermarked 3-D Mesh Quality Assessment.. IEEE Trans. Multimedia, 9, 247-256.

Deubel, H., Schneider, W., X.& Bridgeman, B. (1996). Postsaccadic target blanking prevents saccadic suppression of image displacement. Vision Research, 36, 985 - 996.

Digne, J., Chaine, R. & Valette, S. (2014). Self-similarity for accurate compression of point sampled surfaces.. Comput. Graph. Forum, 33, 155-164.

Dutagaci, H., Cheung, C. P. & Godil, A. (2011). Evaluation of 3D Interest Point Detection Techniques.. In H. Laga, T. Schreck, A. Ferreira, A. Godil, I. Pratikakis & R. C. Veltkamp (eds.), 3DOR (p./pp. 57-64), : Eurographics Association. ISBN: 978-3-905674-31-6

Findlay, J. M. (2004). Eye scanning in visual search. In J. Henderson & F. Ferreira (Eds.), The interface of language, vision and action : eye movements and the visual world,(pp.135–159). NY: Psychology press.

Gal, R. & Cohen-or, D. (2006). Salient geometric features for partial shape matching andsimilarity. ACM Transactions on Graphics, 25, 130–150.

Garland, M. & Heckbert, P. S. (1997). Surface simplification using quadric error metrics.. In G. S. Owen, T. Whitted & B. Mones-Hattal (eds.), SIGGRAPH (p./pp. 209-216), : ACM. ISBN: 0-89791-896-7.

Gelasca, E. D., Ebrahimi, T., Corsini, M. & Barni, M. (2005). Objective evaluation of the perceptual quality of 3D watermarking.. IEEEInternational Conference On Image Processing (p./pp. 241-244).

Guo, J., Vidal, V., Cheng, I., Basu, A., Baskurt, A. & Lavoué, G. (2017). Subjective and Objective Visual Quality Assessment of Textured 3D Meshes.. TAP, 14, 11:1-11:20.

Hou, X. & Zhang, L. (2007). Saliency Detection: A Spectral Residual Approach.. CVPR, : IEEE Computer Society. ISBN: 1-4244-1179-3

Guy, G. & Medioni, G. G. (1997). Inference of Surfaces, 3D Curves, and Junctions From Sparse, Noisy, 3D Data.. IEEE Trans. Pattern Anal. Mach. Intell., 19, 1265-1277.

Henderson, J.M. (2003). Human gaze control during real-world scene perception. Trends in cognitive sciences, 7 11, 498-504.

Henderson, J., Williams, C. & Falk R. (2005). Eye movements are functional during face learning. Mem. Cogn., 133, 98-106.

Hoffman, J.E.& Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. Perception and Psychophysics, 787-795.

Howard, J. (2002). *Seeing in depth*. Toronto: university of Toronto press.

Human-eye blog. Eye colour. Retrieved from <http://the--human-eye.blogspot.com/>.

Itti, L., & Koch, E.N. (1998). A model of saliency-based visual attention for rapid science analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20 (11), 1254 -1259.

ITU-RBT.500-11 (2002). *Méthodologie d'évaluation subjective de la qualité des images de télévision. Technical Report. International Telecommunication Union*

ITU-T Recommendation P.910(2008). *Subjective video quality assessment methods for multimedia applications. International Telecommunication Union*

Jia, S., Zhang, C., Li, X. & Zhou, Y. (2014). Mesh resizing based on hierarchical saliency detection. Graphical Models, 76, 355-362.

Johnson, A. E. & Hebert, M. (1999). Using Spin Images for Efficient Object Recognition in Cluttered 3DScenes. IEEE Transactions on Pattern Analysis and Machine Intelligence(PAMI) (p./pp. 433--449) .

Karni, Z. & Gotsman, C. (2000). Spectral Compression of Mesh Geometry. EuroCG (p./pp. 27-30),

Koch, C.& Poggio, T.(1999). Predicting the visual world: silence is golden. Nature Neuroscience, 2, 9-10.

Kowler, E., Anderson, E., Dosher, B. & Blaser, E. (1995). The role of visual attention in the programming of saccades. Vision research, 35, 1897-1916.

Kim, Y., Varshney, A., Jacobs, D. W. & Guimbretière, F. (2010). Mesh saliency and human eye fixations. Transactions on Applied Perception, 7 (2), 1-13.

Lavoué, G.& Corsini, M. (2010).A Comparison of Perceptually-Based Metrics for Objective Evaluation of Geometry Processing. IEEE Transactions on Multimedia,12, 636-649.

Lavoué, G., Drelie Gelasca, E., Dupont, F., Baskurt, A. & Ebrahimi, T. (2006).Perceptually driven 3D distance metrics with application to watermarking. Proceedings of SPIE, Applications of Digital Image Processing, 6312, 63120L-63120L12.

Lavoué, G. (2009).A local roughness measure for 3D meshes and its application to visual masking. ACM Transactions on Applied Perception, 21:1-21:23.

Lavoué, G. (2011). A Multiscale Metric for 3D Mesh Visual Quality Assessment.. Comput. Graph. Forum, 30, 1427-1437.

Lee, J., Moghaddam, B., Pfister, H. & Machiraju, R. (2004). Finding Optimal Views for 3D Face Shape Modeling. FGR (p./pp. 31-36), : IEEE Computer Society.

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

Liu, Z, Zou, Wenbin & Le Meur, Olivier (2014). A Novel Saliency Detection Framework. IEEE Transactions on Image Processing, 23, 1937 - 1952.

Ling, H. & Okada, K. (2006). Diffusion Distance for Histogram Comparison. IEEE CVPR (p./pp. 246--253), . ISBN: 0-7695-2597-0

Liversedge, S. P.& Findlay, J. M. (2000). Saccadic eyemovements and cognition. Trends in Cognitive Sciences, 4, 6–14.

Maximo, A., Patro, R., Varshney, A. & Farias, R. C. (2011). A robust and rotationally invariant local surface descriptor with applications to non-local mesh processing.. Graphical Models, 73, 231-242.

Nouri, A., Charrier, C. & Lézoray, O. (2015). Multi-scale mesh saliency with local adaptive patches for viewpoint selection. *Signal Processing: Image Communication*, 38, 151-166.

Nouri, A., Charrier, C. & Lézoray, O. (2015b). Multi-scale saliency of 3D colored meshes, International Conference on Image Processing (IEEE), 2820-2824.

Nouri, A., Charrier, C. & Lézoray, O. (2016a). Cartes de saillance et évaluation de la qualité des maillages 3D. (Doctoral dissertation). Retrieved from HAL ARCHIVES OUVERTES. (Accession No. tel-01418334).

Nouri, A., Charrier, C. & Lezoray, O. (2016b). Full-reference saliency-based 3D mesh quality assessment index. ICIP (p./pp. 1007-1011).

Nouri, A., Charrier, C. & Lezoray, O. (2017a). Greyc 3D colored Mesh Database. Technical report. <https://nouri.users.greyc.fr/ColoredMeshDatabase.html>

Nouri, A., Charrier, C. & Lezoray, O. (2017b). 3D Blind mesh quality assessment index. Proc. of IS&T Electronic Imaging. Three-Dimensional Image Processing, Measurement (3DIPM), and Applications(p./pp. 9-26).

Parkhurst, D., Law, K. & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Reresearch*, 42, 107-123.

Rogowitz, B. E. & Rushmeier, H. E. (2001). Are image quality metrics adequate to evaluate the quality of geometric objects?. In B. E. Rogowitz & T. N. Pappas (eds.), *Human Vision and Electronic Imaging* (p./pp. 340-348), : PROC. SPIE. ISBN: 978-0-8194-3977-2

Rizzolatti, G., L. Riggio& B. M. Shelia (1994). Space and selectiveattention. In C. Ulmità & M. Moscovitch (Eds.), *Attention and performance*,(pp.231–265). Cambridge:Mit Press.

Rusinkiewicz, S. (2004). Estimating Curvatures and Their Derivatives on Triangle Meshes.. 3DPVT (p./pp. 486-493), : IEEE Computer Society. ISBN: 0-7695-2223-8.

Song, R., Liu, Y., Martin, R. R. & Rosin, P. L. (2014). Mesh saliency via spectral processing.. ACM Trans. Graph., 33, 6:1-6:17.

Sorkine, O., Cohen-Or, D. & Toledo, S. (2003). High-Pass Quantization for Mesh Encoding.. In L. Kobbelt, P. Schröder & H. Hoppe (eds.), *Symposium on Geometry Processing* (p./pp. 42-51), : Eurographics Association.

Shahrbabaki, S., T. (2015).Contribution de la couleur dans l'attention visuelle et un modèle de saillance visuelle. (Doctoral dissertation). Retrieved from HAL ARCHIVES OUVERTES. (Accession No. tel-01241487).

Shilane, P. & Funkhouser, T. (e 07). Distinctive regions of 3D surfaces. ACM Trans. Graph., 26, 7. doi: <http://doi.acm.org/10.1145/1243980.1243981>

Silva, S., Santos, B. S., Ferreira, C.&Madeira, J. (2009). A perceptual data repository for polygonal meshes. Second International Conference in Visualisation, 207-212.

Simari, P. D., Picciau, G. & Floriani, L. D. (2014). Fast and Scalable Mesh Superfacets.. Comput. Graph. Forum, 33, 181-190.

Song, R., Liu, Y., Martin, R. R. & Rosin, P. L. (2014). Mesh saliency via spectral processing.. ACM Trans. Graph., 33, 6:1-6:17.

Song, R., Liu, Y., Zhao, Y., Martin, R. R. & Rosin, P. L. (2012). Conditional random field-based mesh saliency.. ICIP (p./pp. 637-640), : IEEE. ISBN: 978-1-4673-2534-9

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

Tao, P., Cao, J., Li, S., Liu, X. & Liu, L. (2015). Mesh saliency via ranking unsalient patches in a descriptor space.. Computers & Graphics, 46, 264-274.

Tatler, B.W, Baddeley, R.J&Glichrist, I.D. (2005). Visual correlates of fixation selection: effects of scale and time. Vision reresearch, 45, 643-659.

Fakhri, T., Kai, W. & Jean-Marc, C. (2014).A curvature-tensor-basedperceptual quality metric for 3d triangular meshes. Machine Graphics and Vision, pp,1-25.

Treisman, A. M. & Gelade, G. (1980). A Feature-Integration Theory of Attention.. Cognitive Psychology, 12, 97-136.

Vásá, L. & Rus, J. (2012). Dihedral Angle Mesh Error: a fast perception correlated distortion measure for fixed connectivity triangle meshes.. Comput. Graph. Forum, 31, 1715-1724

Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. Springer.

Vision du futur. La rétine. Retrieved from <http://lavisiondufutur.e-monsite.com/pages/l-oeil/la-retine.html>

Wang, K., Lavoué, G., Denis, F. & Baskurt, A. (2011). Robust and blind mesh watermarking based on volume moments.. Computers & Graphics, 35, 1-19.

Wang, K., Torkhani, F. & Montanvert, A. (2012). A fast roughness-based approach to the assessment of 3D mesh visual quality.. Computers & Graphics, 36, 808-818.

Wang, Z. & Bovik, A. C. (2011). Reduced- and No-Reference Image Quality Assessment.. IEEE Signal Process. Mag., 28, 29-40.

Wang, Z., Bovik, A. C. (2006). Modern Image Quality Assessment. Morgan & Claypool Publishers.

Widdel, H. (1984). Operational problems in analyzing eye movements. Theoretical and applied aspects of eye movement research, 22, 21 -29.

Wolfe, J. M., Alvarez, G. A. & Horowitz, T. S. (2000). Attention is fast but volition is slow. Nature, 406, 691.

Wolfe, J. M., Cave, K. R. & Franzel, S. L. (1989). Guided research: An alternative to the feature integration model for visual research.. Journal of Experimental Psychology: Human Perception & Performance, 15(3), 419-433.

Wolfe, J. M., Palmer, E. M., & Horowitz, T. S. (2010). Reaction time distributions constrain models of visual research. Vision Research, 50(14), 1304–1311. <http://doi.org/10.1016/j.visres.2009.11.002>

Wu, J., Shen, X., Zhu, W. & Liu, L. (2013). Mesh saliency with global rarity.. Graphical Models, 75, 255-264

Yarbus, A. L. (2002). *Eye movements and vision*. Plenum press.

Yee, Y. H., Pattanaik, S. N. & Greenberg, D. P. (2001). Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments.. ACM Trans. Graph., 20, 39-65.

Zhao, Y., Liu, Y., Song, R. & Zhang, M. (2012). Extended non-local means filter for surface saliency detection.. ICIP (p./pp. 633-636), : IEEE. ISBN: 978-1-4673-2534-9

Zhao, Y. & Liu, Y. (2012). Patch based saliency detection method for 3D surface simplification.. ICPR (p./pp. 845-848), : IEEE Computer Society. ISBN: 978-1-4673-2216-4

Zhao, Y., Liu, Y. & Zeng, Z. (2013). Using Region-Based Saliency for 3D Interest Points Detection.. In R. C. Wilson, E. R. Hancock, A. G. Bors & W. A. P. Smith (eds.), CAIP (2) (p./pp. 108-116), : Springer. ISBN: 978-3-642-40245-6

Zhu, Q., Zhao, J., Du, Z. & Zhang, Y. (2010). Quantitative analysis of discrete 3D geometrical detail levels based on perceptual metric.. Computers & Graphics, 34, 55-65.

KEY TERMS AND DEFINITIONS

Geometric saliency:saliency detected from the geometry of the 3D non colored mesh.

Colorimetric saliency:saliency detected from colors of the 3D colored mesh.

HVS: Human visual system.

