

Enhanced Visual-Attention Model for Perceptually Improved 3D Object Modeling in Virtual Environments

Maude Chagnon-Forget · Ghazal Rouhafzay · Ana-Maria Cretu · Stéphane Bouchard

Received: 25 August 2016 / Revised: 4 October 2016 / Accepted: 6 October 2016 / Published online: 17 October 2016
© 3D Research Center, Kwangwoon University and Springer-Verlag Berlin Heidelberg 2016

Abstract Three-dimensional object modeling and interactive virtual environment applications require accurate, but compact object models that ensure real-time rendering capabilities. In this context, the paper proposes a 3D modeling framework employing visual attention characteristics in order to obtain compact models that are more adapted to human visual capabilities. An enhanced computational visual attention model with additional saliency channels, such as curvature, symmetry, contrast and entropy, is initially employed to detect points of interest over the surface of a 3D object. The impact of the use of these supplementary channels is experimentally evaluated. The regions identified as salient by the visual attention model are preserved in a selectively-simplified model obtained using an adapted version of the QSlim algorithm. The resulting model is characterized by a higher density of points in the salient regions, therefore ensuring a higher perceived quality, while at the same time ensuring a less complex and more compact representation for the object. The quality of the resulting models is compared with the performance of

other interest point detectors incorporated in a similar manner in the simplification algorithm. The proposed solution results overall in higher quality models, especially at lower resolutions. As an example of application, the selectively-densified models are included in a continuous multiple level of detail (LOD) modeling framework, in which an original neural-network solution selects the appropriate size and resolution of an object.

Keywords Visual attention · Saliency · 3D modeling · Interest points · Mesh simplification · Level of detail

1 Introduction

One of the most common representations of 3D models nowadays is the polygonal representation. Points in 3D (i.e. vertices), are connected by line segments to form faces that together constitute a polygonal mesh. In most of the cases the faces are triangles, because graphics cards can render them quickly. However, because the faces are planar, in order to represent complex objects, especially objects made out of curved surfaces, an accurate representation requires the use of a very large number of polygons. On the other hand, in interactive applications, the user expects to interact with a virtual world composed of many objects in real-time. From here stems the interest into developing means to simplify object meshes in order to represent them in a compact manner that allows a real-time interaction, while

M. Chagnon-Forget · G. Rouhafzay · A.-M. Cretu (✉)
Department of Computer Science and Engineering,
Université du Québec en Outaouais, Gatineau, QC,
Canada
e-mail: ana-maria.cretu@uqo.ca

S. Bouchard
Department of Psychoeducation and Psychology,
Université du Québec en Outaouais, Gatineau, QC,
Canada

still containing enough details and important information such as not to hinder the visual quality. The degree of simplification can be controlled by managing the level of detail (LOD) of objects. The main idea is to gradually adjust the complexity of a 3D object model by removing unimportant details when the object is small or when it is situated far from the user in the 3D environment. In this way, the farther the object is situated within a 3D environment with respect to the user, a less detailed representation is used to ensure a higher rendering speed. However, this reduction of quality of an object mesh should remain invisible to the user. Even at the lowest resolution, the shape and the most salient details of the object should still be distinguishable. Luebke et al. [1] classify the LOD managing methods as discrete, continuous and view-dependent. Discrete LOD methods use multiples copies of different resolutions of an object, with details uniformly and gradually reduced according to the distance to the viewer. These copies are created offline and one of them is chosen at run-time according to the distance from the viewer to the object. Continuous LOD methods use specific data structures storing a continuous spectrum of details from which the desired level is extracted at run-time. View-dependent methods are variations of continuous LOD techniques that select dynamically the most appropriate level of detail for the current view of an object. All these methods provide in most of the cases appealing results. However, they generally perform poorly at very low level of detail because the object geometry is simplified uniformly, without considering that some areas or features that characterize the object and that can be sometimes small with respect to the object size (e.g. the ears and tail of a cat), could be perceptually more salient than others. One solution to this problem is to employ human users to provide inputs on the desired quality of a model, by selecting areas over the surface of an object where a local improvement of detail is desirable [2–5]. Alternatively, quality adjustments can be made automatically, for example, based on the object's surface properties [6].

This paper contributes to the advancement of 3D modeling techniques by making them more adapted to human visual system capabilities without requiring the human input. We aim at building compact selectively-densified object models, in which the mesh is denser only in areas considered important by the human visual attention. This results in a reduction of the number of faces in the model and hence of the rendering time, while preserving the perceived quality of the object, even at low resolution. Classical models of visual

attention employ only a limited set of human visual system characteristics (i.e. intensity, color and orientation). According to Frintrop et al. [7], considering more features results in more accurate and biologically plausible models. The proposed model therefore incorporates other biologically and psychologically inspired features, namely curvature, symmetry, contrast and entropy. While these characteristics were already tested separately in various computational visual attention models in the literature, to our knowledge no other author used them together and in the context of interest point detection for 3D object modeling. In fact, very few researchers have used inspiration from the biological human visual attention for the detection of regions of interest in 3D [6, 8]. The main contributions of this paper can be summarized as it follows: (1) the adaptation of a visual attention model for the detection of points of interest over the surface of 3D objects, (2) the incorporation of identified regions of interest in compact selectively-densified object models in which the mesh is denser in salient regions, (3) the inclusion of selectively-densified models in continuous LOD modeling applications, including a novel strategy to automatically select the appropriate mesh size according to the distance with respect to the user, and (4) an experimental study of the impact on the quality of models as a result of the incorporation of several characteristics of the human visual perception.

2 Literature Review

The human visual system performs two stages of visual processing: a pre-attentive parallel stage during which the entire visual field is processed at once, and a slow serial attentive processing stage, where regions of interest are selected by attention for further analysis. The role of visual attention is to break down the problem of understanding a scene into a rapid series of computationally less demanding, localized visual analysis problems [9]. It also decides the order in which a scene is investigated, or the order of fixations [9], in which the fovea is positioned on specific regions of the object, maximizing the focus on identified regions, making the central areas clearer (i.e. center-surround mechanism of the human visual receptive field). Most computational implementations of visual attention are based on bottom-up features that can capture attention during free viewing conditions. A measure that has been shown to be particularly relevant is the local image

saliency, which corresponds to the degree of conspicuity between that location and its surroundings. In other words, the responsible feature that guides the deployment of attention needs to be sufficiently discriminative with respect to its surroundings. In spite on the fact that opinions on features that guide human visual attention are still controversial, Wolfe and Horowitz's study on the deployment of attention in visual search tasks [10] led to a relatively complete description of attributes, including undoubted (color, motion, orientation and size), probable (flicker, luminance polarity, offset, stereoscopic depth and tilt, pictorial depth cues, shape, line termination, closure and curvature) and possible attributes (lighting direction, glossiness, aspect ratio). Aside these, psychological studies showed the influence of other less understood properties of the visual attention, including the influence of symmetry of the object shape on attracting visual attention [11].

Computational models of visual attention have been shown to significantly improve the speed of scene understanding [12], by attending only the regions of interest and distributing the resources where they are required. It was proven that attention systems are especially well suited to detect discriminative features and that the repeatability of salient regions is higher than the repeatability of non-salient regions provided by classical feature descriptors such as corners or SIFT keypoints [13, 14]. Only few authors use inspiration from the bottom-up human visual attention for the detection of interest points. Lee et al. [6], starting from the observation that changes in curvature are correlated to regions that attract visual attention, identify points of interest in the curvature map, a mapping from each vertex of the mesh to its mean curvature. In Castellani et al. [8], the identification of interest points is based on difference-of-gaussian filtering (i.e. the center-surround mechanism).

The detection of points of interest attracted the preoccupation of the research community for various 3D applications such as mesh and shape retrieval [15, 16], matching of objects [8], or mesh simplification guidance [5, 6]. Song et al. [5] compute mesh saliency using the geodesic measure to identify the neighborhood of a point and further incorporate multi-scale information in a conditional random field framework to impose consistency constraints between neighboring points. Yang et al. [15] calculate vertex saliency based on the distance of a vertex with respect to its neighborhood. The approach for interest point detection in [16] uses a voxel grid and capitalizes on the SIFT algorithm. A 3D version of Harris corner for interest point detection is proposed by Sipiran

and Bustos [17]. Also exploiting the corner detection idea is the work of Novatnack and Nishino [18]. In Sun et al. [19], points of interest are identified as local maxima of the Heat Kernel Signature computed over a triangular mesh. Some researchers focused towards the guidance of mesh simplification process using points of interest. A survey of polygonal simplification methods is available for the interested reader in [1, 20]. Most of the work from the literature uses QSlim [21] as the algorithm of choice, due to its best balance between speed, fidelity and robustness among similar algorithms [14]. A few research articles proposed user-guided versions of QSlim to improve locally the quality of a mesh. Kho and Garland [3] adapt QSlim to preserve features situated in regions labelled by a user as being important. The authors of [4] allow the user to improve unsatisfactory regions by first weighting and then applying local refinements to the desired region. Song et al. [5], bias the simplification process by amplifying the saliency values in regions of interest, while Lee et al. [6] control the order of simplification contractions of the QSlim algorithm by weighting stronger important regions in a mesh saliency map. In the work of Pojar and Schmalstieg [2], the user controls the simplification of a mesh by painting the desired regions in a Maya plug-in. Howlett et al. [22] propose the idea of saliency-guided simplification, where saliency is captured in form of eye fixations. In the current work the simplification process is based on visual attention and biased by constraining the maximum resolution in the regions considered perceptually salient.

3 Modeling Framework for Perceptually Improved 3D Object Modeling Based on Visual Attention

The overall approach for creating perceptually improved 3D object models in the context of modeling at multiple LOD using regions of interest derived from visual attention is illustrated in Fig. 1.

An enhanced computational visual attention model is applied on images captured from multiple viewpoints of a 3D object, in order to identify regions that attract the attention of a human viewer. This process aims at capturing the discriminant details that characterize the shape and the identity of the object. Within these regions, points of interest are identified as centroids, and projected back in 3D to obtain the points of interest over the entire surface of the object. Multiple copies of the same points

identified in different viewpoints are eliminated. Given the appropriate number of faces for each sample of an object within a LOD hierarchy, for which a novel neural-network solution is proposed in this paper, the QSlim algorithm is adapted to simplify only those faces of the objects that do not contain as vertices the identified interest points and their immediate neighbours. Due to the computational cost associated to the computation of the enhanced visual attention model, on which details are provided in the experimental section of the paper, the detection of interest points takes place offline, while once the interest points are identified, the simplified models with region of interest preservation at various resolutions are constructed online.

3.1 Enhanced 3D Visual Attention Model

Computational attention models are designed to work on images. Because the proposed solution is expected to work in 3D, they are two possible solutions to deal with

this issue: one is to expand the visual attention model to 3D, the other one is to capture multiple images, IM_v , from different viewpoints of a 3D object in order to ensure a relatively complete description of the regions of interest over its whole surface. Because certain features that characterize the visual attention system are less understood in 3D, in this paper it was chosen to follow the second option. In order to capture multiple images from different viewpoints of an object, a virtual camera model is employed. As only meshes of object are available in our dataset, the objects are rendered with a smooth material of neutral, grey color. The headlight, a source of light situated in the front of the object at an infinite distance, is the only light used in the scene. To avoid that attention is captured mainly due to the contrast around the contour of the object, a simple black background is used for testing. Once these images obtained, an enhanced version of a classical computational visual attention model that uses additional features is applied on each collected

Fig. 1 Perceptually-improved 3D object modeling framework

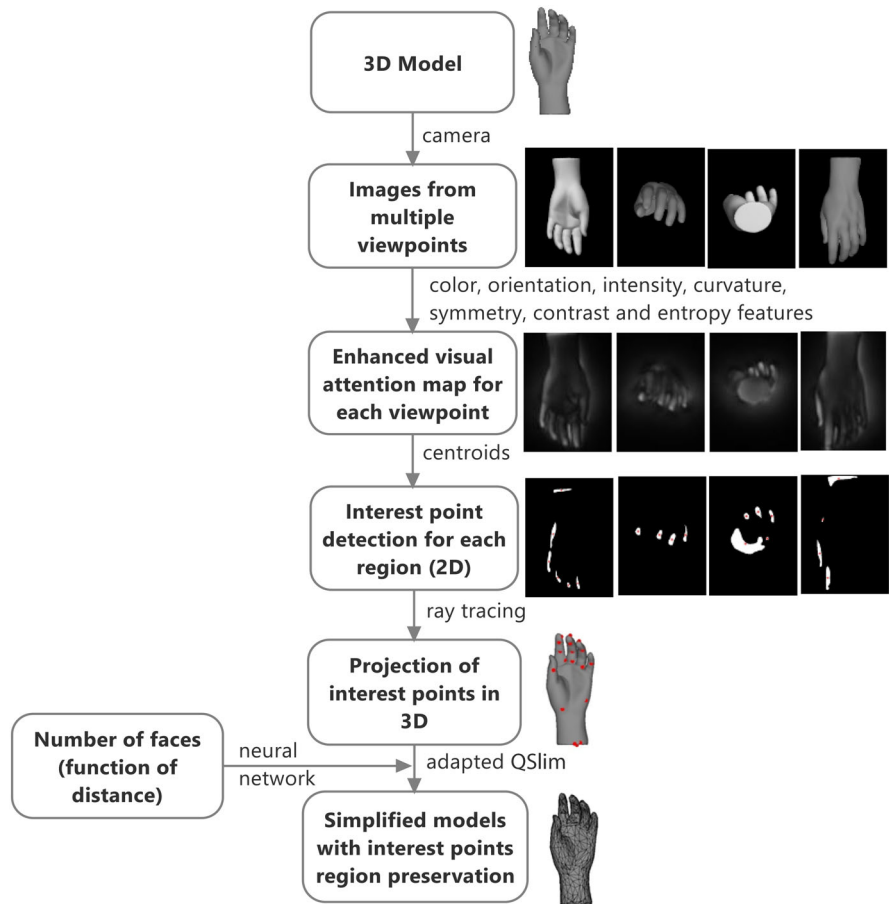


image from the multiple viewpoints to build the saliency map.

The model of Itti et al. [23], that employs intensity, colour, and orientation, is used as a base model. It uses nine spatial scales, created from each image using dyadic Gaussian pyramids. Each feature is calculated as a series of center-surround operations similar to human visual receptive field. Typical visual neurons are more sensitive in a small region of the visual space, namely its center, while stimuli in a broader and weaker region concentric with the center inhibit neural responses. The center-surround mechanism is implemented as a difference between fine and coarse scales, where the center is a pixel at scale $c \in \{2, 3, 4\}$ and the surround pixel corresponds to a scale $s = c + \delta$, where $\delta \in \{3, 4\}$. Given r, g and b , the red, blue and green channels of an initial image, IM_v , the intensity map \mathcal{I} is obtained as $\mathcal{I} = (r + g + b)/3$ and the corresponding conspicuity map is computed as

$$\bar{C}_{\mathcal{I}} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} \mathcal{N}(\mathcal{I}(c) \ominus \mathcal{I}(s)) \tag{1}$$

with \ominus representing an across-scale difference operation, \oplus across-scale addition, involving a reduction of scales to 4 and a point by point addition, and $\mathcal{N}(\cdot)$ is a normalization operation by $(M - \bar{m})^2$ that promotes globally the maps with a small number of strong saliency peaks and inhibits maps with many similar peaks. M is the global maximum of the map and \bar{m} the average of all local maxima. The information on local orientation is obtained from \mathcal{I} using oriented Gabor pyramids $O(\sigma, \theta)$ where $\sigma \in [0..8]$ represents the scale and $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$, the preferred orientations. The orientation conspicuity map is given by:

$$\bar{C}_O = \sum_{\theta} \mathcal{N}(\bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} \mathcal{N}(|O(c, \theta) \ominus O(s, \theta)|)) \tag{2}$$

To compute the color conspicuity map, four broadly tuned color channels are initially created as: $R = r - (g + b)/2$, $G = g - (r + b)/2$, $B = b - (r + g)/2$, and $Y = (r + g)/2 - |r - g|/2 - b$ and two maps quantifying the red/green and blue/yellow opponency are computed as:

$$RG(c, s) = |(R(c) - G(c)) \ominus (G(s) - R(s))| \tag{3}$$

$$BY(c, s) = |(B(c) - Y(c)) \ominus (Y(s) - B(s))| \tag{4}$$

The color conspicuity map is then calculated as:

$$\bar{C}_{RG} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} [\mathcal{N}(RG(c, s)) + \mathcal{N}(BY(c, s))] \tag{5}$$

An additional color feature conspicuity map is introduced in the model following the Derrington-Krauskopf-Lennie (DKL) color space [24] that refers to the color opposition model in the early visual processing [25]. According to this model, color vision starts by the extraction of different signals transmitted by cones and is then processed by three post-receptor mechanisms, one for luminance and two for red/green and blue/yellow opponency, denoted R_{Lum} , R_{L-M} and R_{S-Lum} . A look-up table extracted from [26], is used to convert the r, g and b channels of an image to R_{Lum} , R_{L-M} and R_{S-Lum} components and the conspicuity map becomes:

$$\bar{C}_{DKL} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} [\mathcal{N}(R_{Lum}(c, s)) + \mathcal{N}(R_{L-M}(c, s)) + \mathcal{N}(R_{S-Lum}(c, s))] \tag{6}$$

In spite of the fact that most of our objects are grey, the use of color channels improves the precision of the detection of regions of interest, as demonstrated in Sect. 4.

3.1.1 Curvature

Wolfe and Horowitz's study [10] identifies the curvature as a probable attribute that guides the deployment of visual attention. However, in spite of their visual importance, small high-curvature details over relatively large and uniform regions will be likely ignored by most simplification methods, because simplifying them introduces minimal error [8]. This justifies the interest into associating more importance to high-curvature regions in order to improve the detection of salient regions [6, 27]. To compute the curvature map, we use inspiration from the approaches in [26, 28, 29]. The result is a 3D curvature model, M_c , similar to a saliency map, in which lighter areas are characterized by a higher curvature. To compute the conspicuity map, the 3D curvature model is projected using the camera model in 2D for each given point of view v . The resulting image IM_{cv} is filtered to simulate the center-surround mechanism, and the curvature conspicuity map becomes: $\bar{C}_{Curv} = \mathcal{N}(IM_{cv})$. Alternatively, the interest points can be extracted directly from each view of the curvature model M_c (see Sect. 3.2) and merged with the visual attention derived interest points to enable comparison.

3.1.2 Symmetry

Research published in the literature also suggests that symmetry of visual shapes has an impact on visual attention. Locher and Nodine [11] demonstrated that if an object exhibits a symmetry of shape, the eye fixations follow the symmetry axis, therefore sustaining the theory that symmetry is an attribute guiding visual attention deployment. Koostra et al. [30] compare the use of isotropic, radial and color symmetry operators and merge them in a symmetry saliency map using multiple-scale computations. Their work demonstrates that while there is no significant difference between the results, all of these operators offer better results (validated by human eye fixation data) [26] than the model of Itti and that the radial symmetry operator seems to provide slightly better performance. This is the reason why we have chosen to include it in our model. Moreover, bilateral symmetry is more readily detectable by humans than other types of symmetry [31], justifying the interest of including this type of symmetry in our model as well. The approach in [32] is adapted to compute 3D bilateral and radial symmetric points over the surface of an object. In order to incorporate the symmetry in the form of a saliency map, saliency maps *Sym* are created from different viewpoints, in which points of interest and their immediate neighbors are shown in white and the background in black. Center-surround operations are applied on the resulting map, and the conspicuity map of symmetry becomes $\bar{C}_{Sym} = \mathcal{N}(Sym)$. Similar to the curvature model, we also consider separately the interest points derived from the various types of symmetry, for comparison purposes.

3.1.3 Contrast and Entropy

When looking at an image, people are attracted to regions of strong contrast, while weaker contrast regions tend to be ignored. Zhang et al. [33] use the luminance, texture and colour contrast as the three components of their attention model, while in [34] a histogram-based contrast method is proposed to improve salient region detection. In this paper, the grayscale contrast map *Con* is calculated using the luminance variance in a local neighbourhood of 80×80 pixels [35], and the contrast conspicuity map is built as: $\bar{C}_{Con} = \mathcal{N}(Con)$.

Kadir and Brady [36] propose the idea of using entropy as a measure of local signal complexity or

unpredictability in an image. It is expected that using entropy in the computation of visual attention will yield better results because small areas that are uniquely salient because of lighting (e.g. a local light spot) or color uniqueness are not necessarily salient in general [37]. In this work, the input image is pre-processed with a median filter and then the entropy is encoded as local entropy value of a 9×9 neighbourhood around the corresponding pixel in the filtered image [38]:

$$Ent = - \sum_{i=0}^{L-1} p(\mathcal{I}_i) \log_2 p(\mathcal{I}_i) \quad (7)$$

where $p(\mathcal{I}_i)$ is the histogram of the intensity levels in the region i and L the number of possible intensity levels (e.g. $L = 256$ for experimentation). The entropy conspicuity map is computed as: $\bar{C}_{Ent} = \mathcal{N}(Ent)$.

The final saliency map is calculated as an average of independently calculated conspicuity maps:

$$S_{avg} = \sum_i \bar{C}_i / |i| \quad (8)$$

where $i = \{I, O, RG, DKL, Sym, Con, Ent\}$ and $|i|$ denotes the cardinality of the set i . The grayscale saliency map is then thresholded to retain 30 % of highest saliency values and therefore to identify the most interesting regions from the visual attention perspective.

3.2 Interest Point Identification

Because the identified regions of interest in the saliency map are too large to constrain all their points in the simplification, the interest points are identified as the centroids of each identified region. The resulting 2D points are computed in each image taken from various viewpoints of the object in order to obtain a relatively complete coverage and projected back onto the 3D model using the virtual camera model. To achieve this, we have used four principal points of view (i.e. where the camera is located on positive z axis, negative z axis, positive y axis and negative y axis, respectively, all targeting the origin) to identify salient points based on the visual attention model. Only these four viewpoints are used in the current implementation, because our research revealed that these viewpoints led to the lowest error rates for the modelling of objects in our dataset (see details in Sect. 4). The largest length of the object along the z axis in pixels,

computed from the image, divided by the real dimension in world units, results in the number of pixels per world unit, such that for each point, two elements of the coordinate can be readily obtained. In order to find the third coordinate, we have adopted the ray/triangle intersection model introduced by Moller and Trumbore [39]. This algorithm is a fast solution to find all intersections of the ray passing from each point in parallel with the third axis and thus the closest intersection with the object surface is considered as the third coordinate of the visual attention-based salient point. Only these salient points and their immediate neighbours will be preserved at full resolution in the simplification process.

3.3 3D Object Simplification and Multiresolution Modeling

To allow the simplification to only affect faces whose defining edge points are not among the identified points of interest or their immediate neighbours, an adaptation of QSlim algorithm is proposed. The complete description of the algorithm is available in [21]. Starting from a triangular mesh given by a set of vertices and a set of faces, the algorithm simplifies it by repeated edge collapses using an error metric (i.e. the quadric error, representing the sum of squared distances from the vertex to the planes of neighbouring triangles). If its value is large, the corresponding vertex could represent a distinctive feature or detail on the mesh, and therefore will be removed later from the mesh. Otherwise, it will be removed earlier. This metric is used to compute the cost associated with a contraction as well as the optimal position for the new unified vertex. All the edges from a mesh are extracted along with their associated cost and are stored in an ordered list of costs. At each step, the edge with the least cost is removed from the mesh, its neighbourhood is updated and the costs of edges connected to the unified vertex recomputed. Most solutions in the literature propose means to weigh stronger the regions of interest [4–6], mainly by adjusting their cost in order to delay their simplification. In the current work, the QSlim algorithm is adapted such as the faces of the mesh that contain points of interest and their immediate neighbours are eliminated from the list of faces to be affected by the simplification process [40]. The experiments performed (see Sect. 4) led to the conclusion that a 3-neighborhood around the points of interest provides the best simplification results over our dataset.

As it is difficult for a user to judge the number of faces required for a certain object at a given distance, in the current approach a novel solution is proposed based on neural networks. In particular, a series of two-layer feed-forward architectures is used, one network being associated with each version of visual attention model consisting of various combinations of feature channels. The role of each is to learn the number of faces that should be used in a simplified model based on the object characteristics. The interest of using neural networks for this purpose stems from their capability to provide estimates for data that was not part of the training set, meaning in the current context that the number of faces can be predicted for objects whose characteristics are similar, but not identical to the ones used during training. In the current implementation, the object characteristics are based on the tolerated error, the distance with respect to the user, and the object complexity, the latter described by initial size of the object mesh and the identified number of salient points. The justification for using these characteristics is as follows: in a first place we want to give the user the choice to control the desired accuracy of the model; a value of 0.05 is selected by default if the user doesn't want to intervene in the process. The distance with respect to the user plays an important role in LOD modelling. As stated in the introduction, the farther the object is situated within a 3D environment with respect to the user, a less detailed representation should be used in order to ensure a higher rendering speed. The object complexity has also an important impact on the quality of results. A more complex model might require more faces in order to preserve a good representation of the object (see details in the experimental section). In order to take into account all these factors, each proposed neural network, with an empirically determined size of 30 neurons in the hidden layer, has 4 inputs, namely the tolerated error, the distance with respect to the user, the initial size of the object mesh and the number of salient points and one output, namely the number of faces.

Each network (corresponding to a combination of feature channels) has to be trained to learn the mapping between the number of faces at output and the input variables. In order to train the network, it is necessary to provide values for each input and output variable. The series of error measures are computed as detailed in Sect. 3.4 within a certain range of resolutions, namely

from 1500 to the number of faces in the initial mesh, for the various combinations of visual attention feature channels. For each of combination, the size of initial mesh and the number of interest points are also stored. The distance values are determined in VRML by gradually moving the object further from the user and marking the distance values when important features seem to disappear. A change in resolution is expected to occur at these milestones. Once all this data is available, each network is trained using gradient descent back-propagation with a constant momentum value of 0.95 and an adaptive learning rate. A null sum-squared-error is targeted over 1000 training epochs. Once the network

trained, it will provide an estimate of the number of faces for each input variable combination. The final number of faces is computed as an average over the results provided by each of the networks in the series. The simplification algorithm with regions of interest preservation is applied to constrain the selectively-densified mesh to the calculated number of faces. If desired, the algorithm can be included in a continuous LOD scheme that monitors the distance in the environment and creates the appropriate model according to it.

The pseudo-algorithm for our approach can be summarized as it follows:

Step 1:

```
// interest points detection for an object O represented by a mesh M
collect a predefined number n of images of the object for multiple viewpoints
for each image do
    apply visual attention model to build the saliency map: SMi, i=1,...,n
    recuperate regions of interest Rj by thresholding the saliency map
    for each region of interest do
        determine the point of interest as the centroid of the region
    end
    recuperate 3D coordinates of the point of interest using ray intersection model
    project point of interest on the initial object model and, if not already there, add it to the list of interest points P
end
```

Step 2:

```
// simplification for a mesh M to obtain selectively-densified mesh Ms
Calculate PN = the n-nearest neighborhood of points P in M, i.e. n=3
Calculate f = desired number of faces in the simplification using neural network
Recuperate edges on the mesh that are not points of interest or neighbors of points of interest:
ED = {e(u,v) | e(u,v) ∈ M, e(u,v) ∉ PN}
Initialize Ms=ED
// apply QSlim
Compute quadric error at each vertex of Ms
Determine contraction cost at each edge e(u,v) in Ms
Create an ordered list of edges based on cost
while f not reached do
    Remove the edge e(u,v) with lowest cost
    Use quadric error to choose the optimal contraction target
    Contract u and v and recalculate costs for the adjacent edges in Ms
```

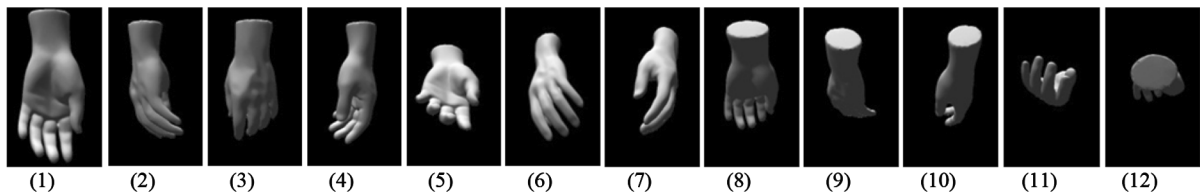



Fig. 2 Viewpoints for visual attention calculation: (1) initial pose and rotations of: (2) 90° along z , (3) 180° along z , (4) 270° along z , (5) 45° around x , (6) 120° around z and of 45° around x ,

(7) 240° around z and of 45° around x , (8) -45° around x , (9) 120° around z and of -45° around x , (10) 240° around z and of -45° around x , (11) 90° around z , and (12) 180° around x

3.4 Mesh Quality Evaluation

The quality of the resulting simplified models is evaluated from quantitative and qualitative points of view. Metro [41] allows comparing two meshes (e.g. the original, full-resolution mesh of an object and its simplified version) based on the computation of a point-surface distance (i.e. Hausdorff distance) and returns the maximum and mean distance as well as the variance (RMS). The lower this error is, the better is the quality of the simplified object. Since our interest is into improving the perceptual quality of models, other three measures of perceptual error are employed. The first one is based on the structural similarity metric (SSIM), proposed based on the observation that the human visual perception is highly adapted to extract structural information in a scene [42]. In particular, the inverse of this metric is employed as an error measure, as a lower similarity between the simplified mesh and the initial mesh implies a higher error. The second category of errors are Laplacian pyramid-based image quality assessment errors [43], two image quality metrics based on early vision transformations, namely local luminance subtraction and contrast gain control. The authors suggest that representing the image in a nonlinear multi-scale decomposition can result in a better account of human perceptual quality judgments. The two forms reported are the predicted distance in Laplacian domain and in normalized Laplacian domain. Because these errors are meant to be used on images, in order to apply them, images are captured over the simplified models of objects from the same viewpoints from which the visual attention model is computed and are compared with the images of the initial, not simplified object from the same viewpoints. The error measures for each object are reported as an average over the viewpoints and overall results are reported as an average over all the objects in the dataset. A qualitative evaluation of the results is

obtained using Cloud Compare [44] that allows visualizing in an intuitive, color-coded manner the regions most affected by error in the simplified object with respect to its original version.

4 Experimental Results

In order to evaluate the proposed framework, it was tested on the objects from the benchmark for 3D interest points [45]. The choice of this dataset is justified because it contains the interest points obtained by several detectors from the literature, therefore allowing for a direct comparison with the proposed solution. In order to identify the most promising viewpoints to use and their number, experiments were performed for an increasing number of viewpoints starting from 4 to 12 [46], each viewpoint resulting from various rotations along x and z axis respectively, as illustrated in Fig. 2.

The object is initially situated at the origin of the system of axis. The Metro error measures, namely the maximum error, the average error and the variance (RMS) described in Sect. 3.4 are then computed as an average over the objects in the dataset, for various viewpoint combinations. Figure 3a illustrates a few tested combinations and shows that, in general, the error increases with the number of viewpoints. This is justified by the fact that a larger number of viewpoints leads to an increased number of the interest points. Due to the fact that our simplification algorithm preserves the interest points and their neighborhoods, only a reduced number of faces is left to represent the rest of the surface of the object, leading to larger errors. As Fig. 3a shows, the combination of four viewpoints, denoted Four (1, 2, 11 and 12) in the figure, leads globally to the smallest error measures and therefore only these viewpoints were used in this article.

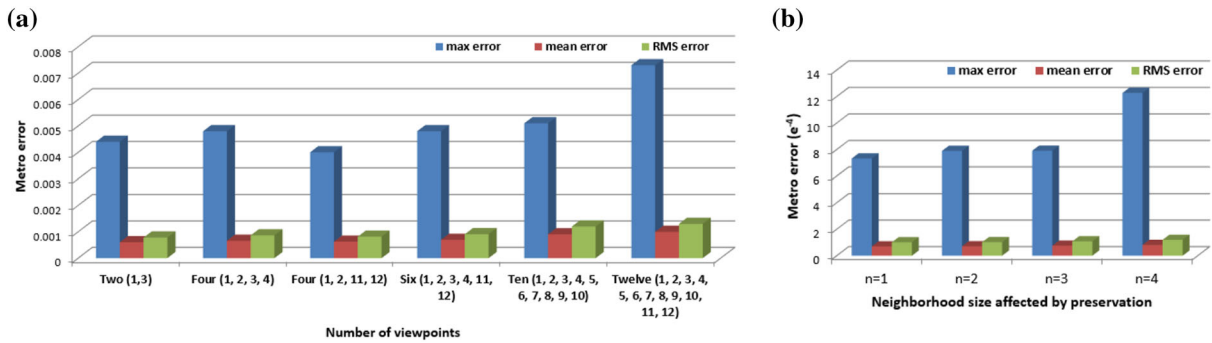
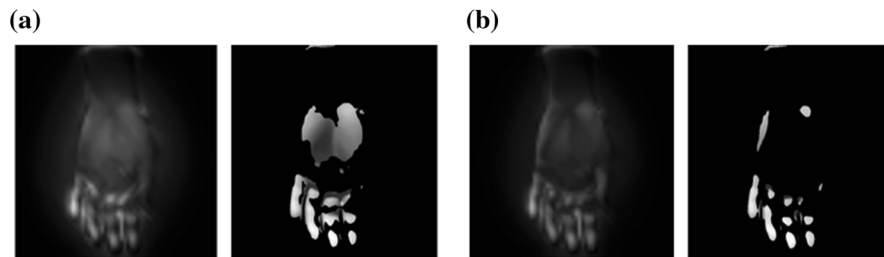


Fig. 3 Error measures illustrating **a** the influence of viewpoints, and **b** the influence of neighborhood size, n

Fig. 4 Saliency map and regions of interest: **a** without and **b** with the use of color (RGB and DKL) features



Another parameter that has also an influence on the quality of the simplification is the number of neighbors to be preserved around an interest point. As in the case of multiple viewpoints, the Metro error measures tend to increase with an increase in the size of the neighborhood n , as illustrated for the $n = 1$ to $n = 4$ and for 15,000 faces in Fig. 3b. This number of faces is chosen because it ensures that for all objects in the database the shape is preserved for $n = 4$. Due to the fact that certain methods return many interest points, the remaining number of faces when constraining $n = 4$ neighbors of each interest point is not sufficient to preserve the general shape of the object. As already stated in the case of viewpoints, this occurs because the interest points and their neighbors constitute a large part of the total number of points of the object and therefore leaving too few points to represent the rest of the surface of the object. As a consequence, for coarser meshes it is desirable to only affect smaller neighborhoods ($n = 1$ or 2), while for higher detailed meshes larger neighborhoods ($n = 3$ or 4) can be used. In experimentation with different meshes made of up to 35,000 faces, generally a 3-neighborhood gave good results for meshes larger than 5500 faces, a 2-neighborhood for meshes between 3500 and 5500 and 1-neighborhood for those below 3500. Because the

objects in the dataset we used for testing have over 5500 faces, we have used a value of $n = 3$ for the remainder of tests performed.

A series of tests aimed at studying the impact of various feature channels over the quality of the simplified mesh. In terms of color channels, experiments revealed that in spite of the use of a dull grey material, their use allows to better identify the interest regions, as shown in Fig. 4.

Another series of experiments dedicated to the identification of an appropriate background color, showed that a black background is more appropriate for the identification of interest points. The average error over all the objects in the dataset (for 1500 faces in the simplified model) calculated for a black background is 0.001, followed by gray (0.002), and by white (0.006).

To study the influence of the curvature, simplified models when only the classical visual attention model is used (e.g. colour, intensity and orientation, denoted VisAtt), are compared with the case when only the curvature information is used (Curv), when the curvature points are added to the visual attention interest points, and when the curvature is incorporated into the visual attention map. The results, illustrated in Fig. 5a for 1500 faces in the simplified model, show

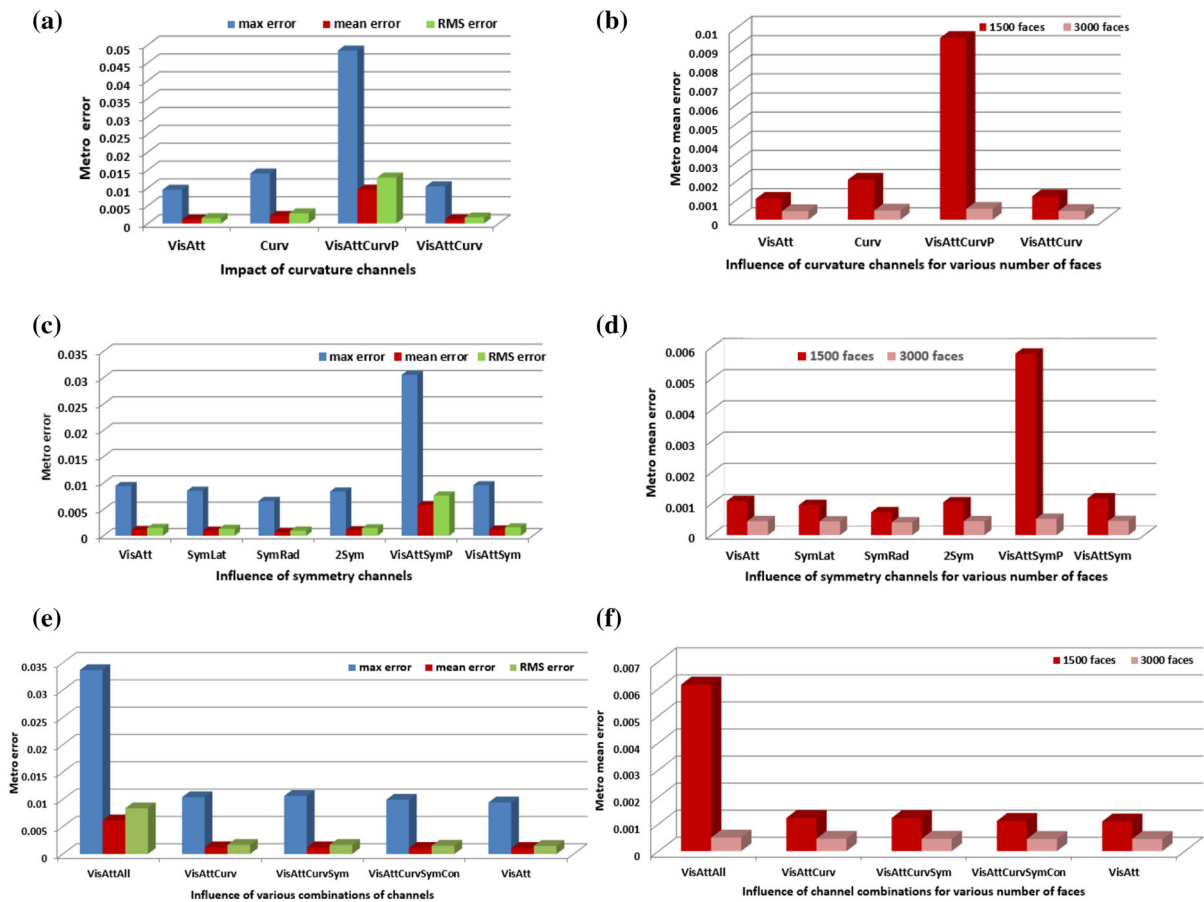


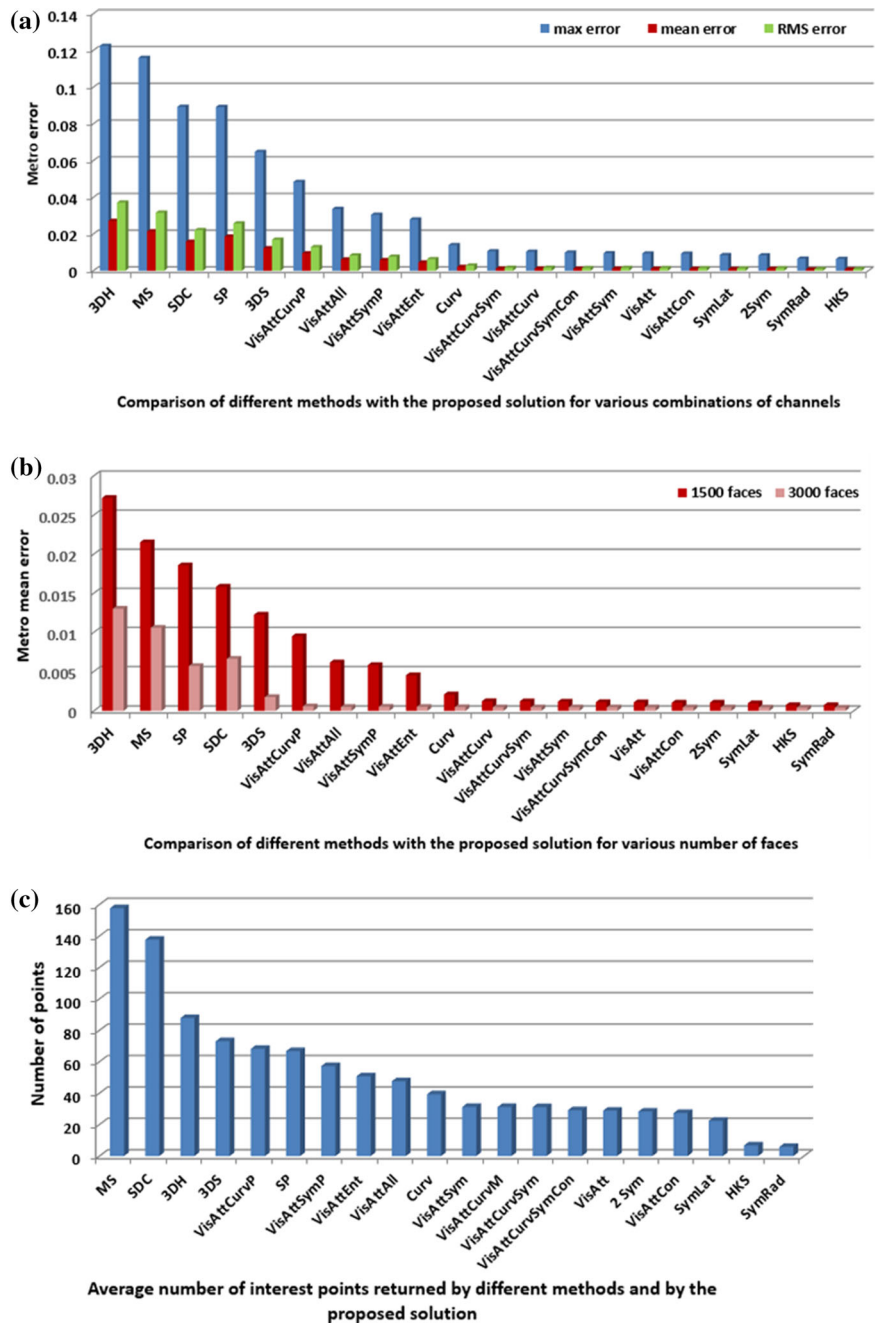
Fig. 5 The influence of visual attention channels and the impact of the number of faces in the simplification on the error measures when: **a, b** curvature, **c, d** symmetry, and **e, f** various combinations of channels are used

that the highest errors are associated with the visual attention with curvature points (VisAttCurvP), while the classical model is close to the error obtained when merging the curvature conspicuity map in the saliency map (VisAttCurv), with the latter obtaining a slightly better performance according to perceptual errors (Fig. 7). As expected and confirmed in Fig. 5b and Fig. 6b, that compare the error measures for 1500 faces (in red) and for 3000 faces (in pink), the errors decrease with an increased number of faces in the simplified model. The difference in errors is also more visible at lower resolutions than at higher resolutions. Similar results can be noticed for the symmetry in Figs. 5c, d, 6a, and 7.

Considered separately, the symmetry channels, whether lateral (SymLat), radial (SymRad) or both (2Sym) obtain roughly the same errors that are slightly lower than the classical visual attention model as

highlighted by the perceptual error in Fig. 7. The model that includes the symmetry conspicuity map (VisAttSym) obtains a slightly higher perceptual error than the classical model (but within a 0.1 difference). Figure 5e, f show the error measures when various combinations of supplementary features are considered in the computation of the saliency map. It can be noticed that the various combinations of features, except for the case when all channels are considered, obtain roughly the same error as the classical model (within a difference of 0.0002), which implies that the addition of channels brings information that is not already available in the classical model. However, when all channels are considered the error is slightly higher. This is mainly due to the addition of entropy information (VisAttEnt) and is believed to come from the fact that the tested objects do not have texture, while entropy can be indirectly considered a measure

Fig. 6 Comparison with other salient point detectors: **a** Metro errors, **b** influence of number of faces over the mean Metro error and **c** number of interest points

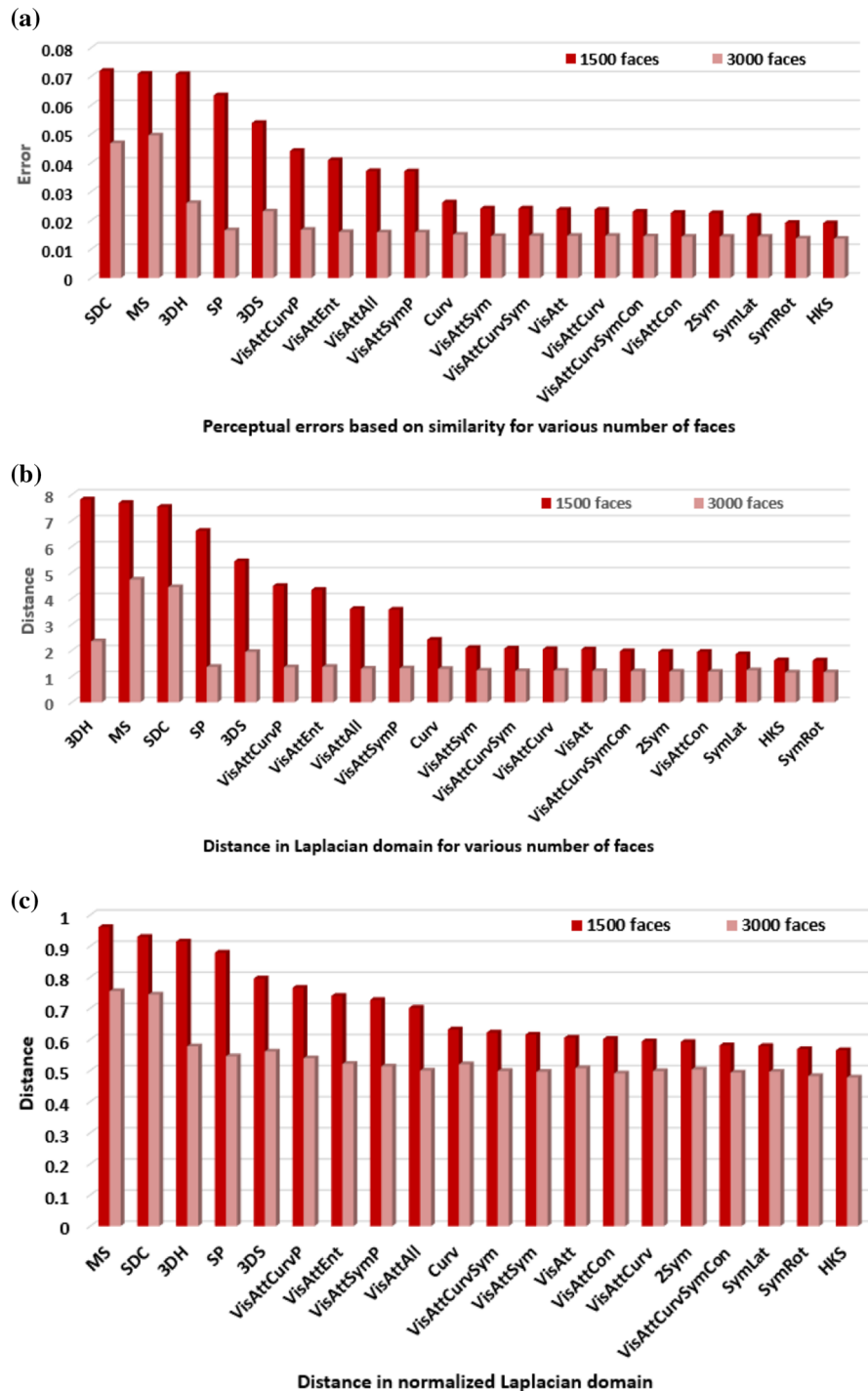


of texture change. The entropy information is more relevant for textured objects, as it will be further demonstrated.

Overall, the model containing curvature, symmetry and contrast (VisAttCurvSymCon) and the one adding contrast only (VisAttCon) lead to the best quality. The

proposed solution is also compared, both in its classic version and with the proposed additional feature channels, with a series of interest point detectors proposed in the literature, including mesh saliency (MS) [6], salient points (SP) [8], 3D-SIFT (3DS) [16], 3D Harris (3DH) [45], scale-dependent corners (SDC)

Fig. 7 Perceptual errors based on: **a** similarity (1-SSIM), **b** distance in Laplacian domain, and **c** distance in normalized Laplacian domain



[18] and Heat Kernel Signature (HKS)[19], embedded in a similar manner in the simplification algorithm (Step 2).

Comparing the error measures (computed as average over all objects) in Fig. 5a–e, it can be noticed that

all proposed solutions based on visual attention lead in general to a better performance for selectively-densified simplification, except HKS approach. A certain correlation exists between the associated number of points of interest in Fig. 6c and the error measures. A

higher number of interest points seems to be, in general, associated with larger error measures. This is due to the fact that our simplification algorithm preserves the regions around the interest points, and therefore only a limited number of faces are impacted by the simplification process. These faces get redistributed to cover the remaining surface of the object, outside the regions of interest.

However a drastic reduction in the number of interest points does not necessarily lead to a drastic decrease in the errors (e.g. for the HKS method or SymRad). It is interesting to notice that the visual attention with additional curvature points (VisAtt-CurvP) obtains smaller errors than the SP method in spite of an almost equal number of interest points. This is due to a better distribution of points of interest ensured by the proposed visual attention approach illustrated in Fig. 8c versus Fig. 8f. It is also worth mentioning that beyond being associated with larger errors, methods that obtain a large number of salient points, such as SDC or 3DH can lead at low resolution to distortion after the simplification. A final remark is related to the fact that more points of interest do not necessarily lead to better results as illustrated in Fig. 9.

In this figure, the selectively-densified simplification results are compared between a method that obtains many points, e.g. SDC (Fig. 9a), and the proposed method with curvature, symmetry and contrast (VisAttCurvSymCon) (Fig. 9b). Too many points lead to the creation of clusters of dense triangles on the mesh, as those in Fig. 6a. On the other hand fewer points of interest, as obtained by HKS (Fig. 6c),

lead to a model closer to uniform simplification, with less well defined characteristics.

Most of the studies performed in this paper are performed on a uniform greyscale dataset because no color information is available for the object meshes in the benchmark dedicated to interest point detection [45] that is used for testing. The reason for choosing this benchmark is that we wanted to be able to directly compare with similar methods for point of interest detection. However, the proposed method works without adjustments, on colored and textured meshes as well. Figure 10 shows an example of texture mesh (Fig. 10a) and presents the results obtained when comparing the error rates without and with the use of texture for the same object.

As it can be observed in Fig. 10b, c, in general, there is no significant error difference (max. 0.01) between the case when the texture is used or not. The perceptual error (Fig. 10c) is computed in this case as the normalized mean error based on SSIM and the distance in the normalized Laplacian domain. Slight differences are expected and visible in Fig. 10b, c in the case of the symmetry and the entropy channels. In case of the symmetry, the texture has an impact on the computation of this channel (i.e. the current texture is not symmetrical) and this leads to slightly higher errors when the symmetry channel is used. In the case of the entropy, the resulted errors are lower when the texture is used, showing that the use of this channel is beneficial for textured objects.

In terms of computation time, the whole procedure, starting from capturing images to the visualization of

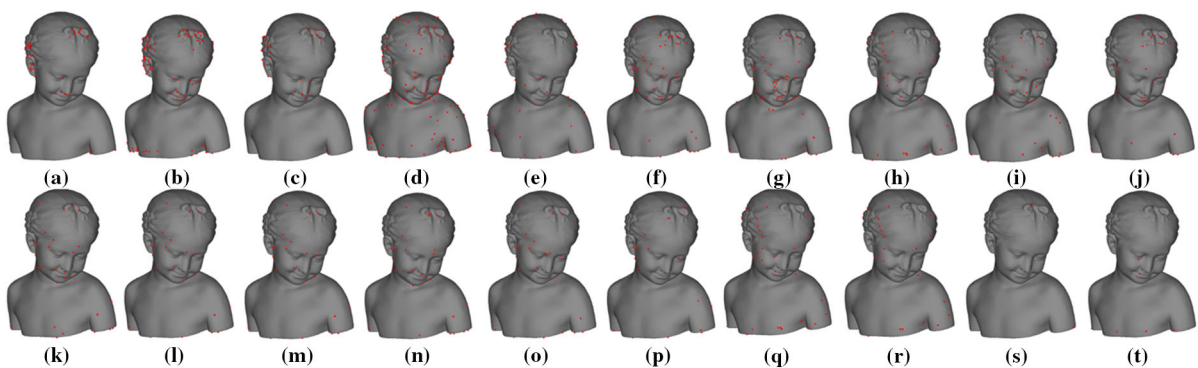


Fig. 8 Comparison of various interest point detectors: **a** 3DH, **b** MS, **c** SP, **d** SDC, **e** 3DS, **f** VisAttCurvP, **g** VisAttAll, **h** VisAttSymP, **i** VisAttEnt, **j** Curv, **k** VisAttCurv,

l VisAttCurvSym, **m** VisAttSym, **n** VisAttCurvSymCon, **o** VisAtt, **p** VisAttCon, **q** 2Sym, **r** SymLat, **s** HKS and **t** SymRad

Fig. 9 Simplification results when using: **a** a large number of interest points (SDC), **b** an intermediate number of interest points (VisAttCurvSymCon), and **c** a small number of interest points (HKS)

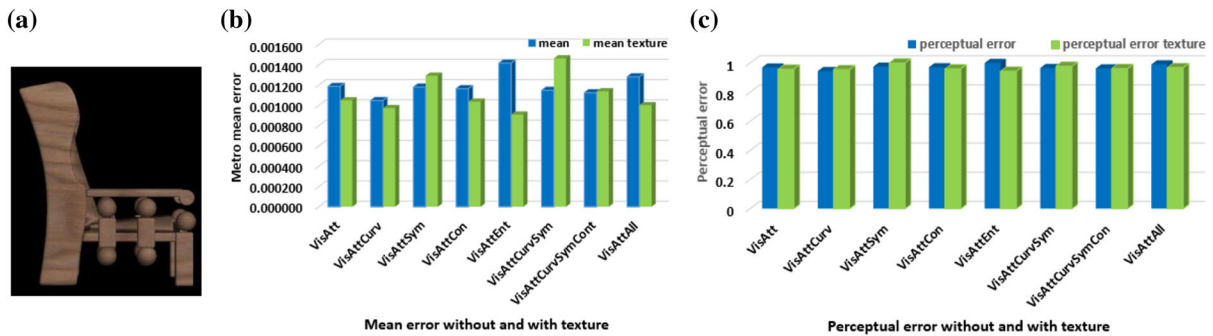
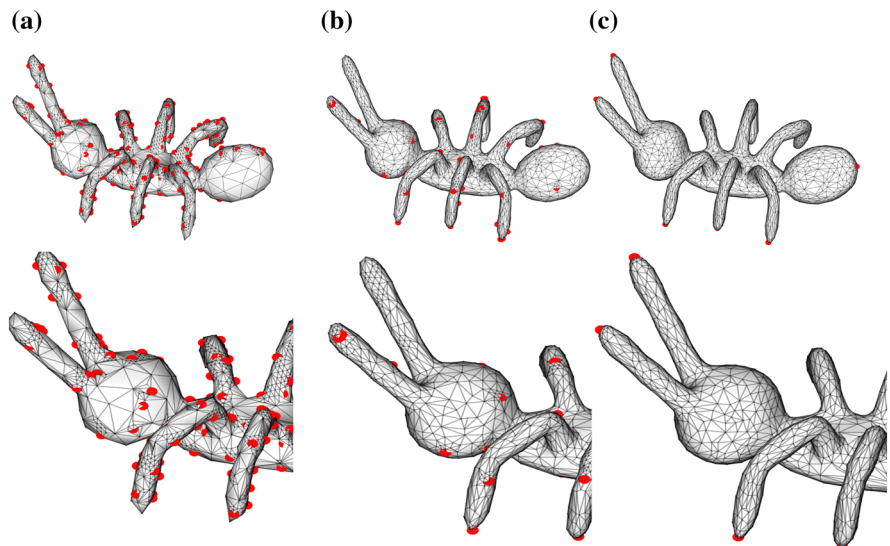


Fig. 10 **a** Textured object mesh, and error rate comparison when the texture is used or not, **b** metro mean error and **c** perceptual error

the simplified mesh, it takes on average 120.4 s per object or roughly 69.1 s per 10,000 faces. Overall, about 30 % is dedicated to the capture of snapshots and 70 % to the identification of interest points. The simplification takes maximum 0.2 s per object. It is also important to state that these computations take place offline. Once the neural network is trained, it provides estimates of the number of faces in 0.03 s and the simplification time is maximum 0.2 s, leading to a maximum of 0.23 s and therefore a real-time performance for the generation of meshes at various resolutions. The experiments are performed using Matlab code, on an Intel Core i7 2.2 GHz machine with 8 GB of memory.

As well, tests were performed to illustrate the degree of compactness that can be achieved by employing the proposed method for the simplification

of objects with region of interest preservation. These revealed an average reduction of 91.5 % in the number of faces from the initial mesh when using 1500 faces in the simplified model and of roughly 83 % when using 3000 faces. It is worth to mention that some of the methods returning a large number of points, such as SDC, cannot be employed with less than 4500 faces without provoking a distortion in the mesh, and thus resulting in a maximum reduction of 75 % in the number of faces.

Finally, in order to show the use of the simplified models in LOD modeling applications, Fig. 11 shows an example of different models created automatically by the proposed method. The last row of the Fig. 7 shows the distribution of errors over the surfaces of the object, as visualized in Cloud Compare, with smaller errors in green, medium errors in yellow and large

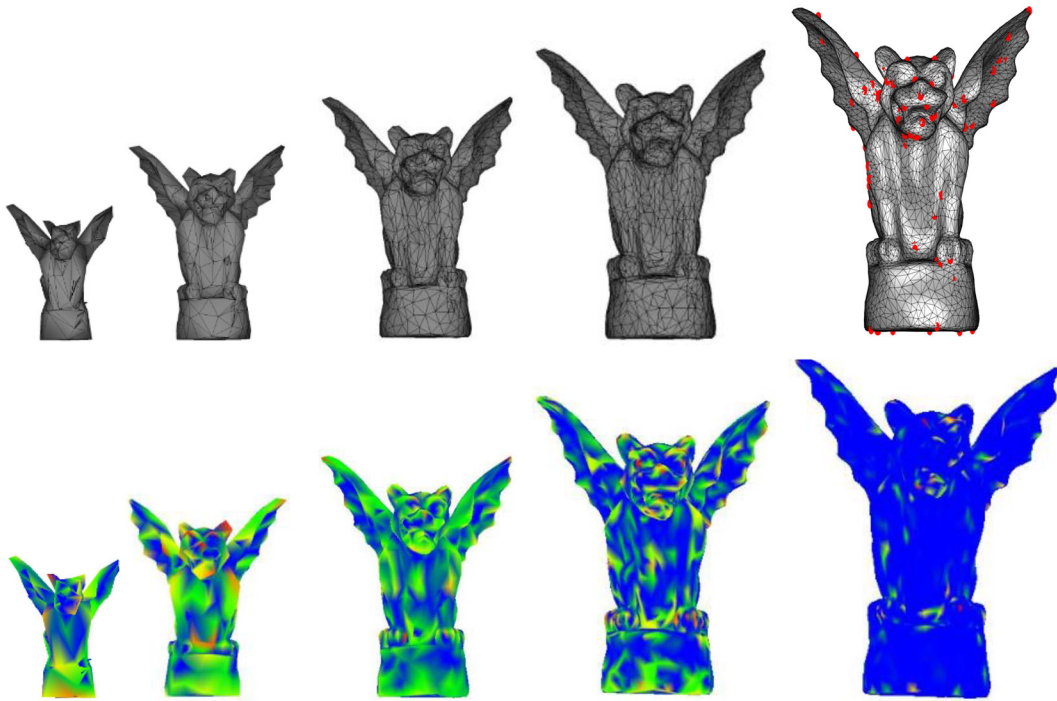


Fig. 11 Object model and color-coded errors at various LOD using visual attention-based interest point identification (VisAttCurvSymCon method)

errors red, while the regions in blue represent a perfect match to the initial mesh. The interest points are shown in red over the mesh. One can notice that even at the lowest resolution (i.e. 950 faces in this case), the fine and perceptually important details of the model (e.g. ears, wings) are preserved.

The tests performed show that overall for the given dataset, the proposed method using the combination with curvature, symmetry and contrast (VisAttCurvSymCon) offers the best trade-off between errors and the quality of the final simplification. However it is important to state, that depending on the nature of the object to be simplified and its characteristics, the same results might not necessarily be the same, in the sense that a certain channel might not have the same importance as in the case of the dataset used for testing. In particular, when a textured object is tested, slight differences are expected in the color (i.e. non-uniform colored texture), symmetry (i.e. non-symmetrical patterns) and entropy channels. Therefore, if a user desires to apply our framework, it is preferable that all the combinations of saliency channels are tested, using ideally the 4 viewpoints identified and a 3-neighborhood around the points of interest to be

preserved from the simplification. Guidelines for the selection of a different size of neighborhood is provided in the experimental section. The information in each saliency channel has to be merged directly in the visual attention map computation as described in Sect. 3.1. The tests showed that it is not worth to consider separately the interest points derived from the various channels, such as symmetry or curvature, and this is expected to be consistent regardless the object's characteristics. Once the various combinations are constructed, the user should proceed to the computation of the error measures for various number of faces and according to the tolerated error for the envisioned application, he or she can retain one or several combinations as possible solutions. In terms of number of faces, it is suggested to start from roughly 15 % of the number of faces in the initial mesh and increase gradually to up to roughly half the number of faces in the initial mesh, according to needs of the user application. If multiple solutions are retained based on error, an in-depth visual comparison using a tool such as Cloud Compare, could allow for the selection of the best solution. If an integration in LOD modelling framework is desired, the user can proceed with the

training of the neural network as detailed in Sect. 3.3. If the best solution is identified, a single neural network would suffice. If multiple solutions are retained based on error, the user can proceed with training of the series of networks. Finally, if the user desires to intervene himself in the selection of the distance, he or she can follow the procedure described for the computation of distance and select himself or herself the desired number of faces. The selectively-densified simplification algorithm can then be applied to obtain the various versions of simplified models.

5 Conclusion

The paper evaluates the impact of interest point detection based on human visual attention in the context of 3D object modeling at multiple LOD for virtual environment applications. The influence of various features for the proposed enhanced visual attention computational model is studied experimentally and the superiority in term of quality of the proposed method is experimentally demonstrated by comparison with various point detectors from the literature. This work demonstrates the importance of considering human visual capabilities in 3D multiple LOD modeling having as consequence an improved perception for the users.

References

- Luebke, D., Reddy, M., Cohen, J. D., Varshney, A., Watson, B., & Huebner, R. (2003). *Level of details for 3D graphics*. San Francisco: Morgan Kaufmann.
- Pojar, E., & Schmalstieg, D. (2003). User-controlled creation of multiresolution meshes. In *Proceedings of the ACM symposium on interactive 3D graphics* (pp. 127–130).
- Kho, Y., & Garland, M. (2003). User-guided simplification. In *Proceedings of the ACM symposium on interactive 3D graphics* (pp. 123–126).
- Ho, T.-C., Lin, Y.-C., Chuang, J.-H., Peng, C.-H., & Cheng, Y.-J. (2006). User-assisted mesh simplification. In *Proceedings of the ACM international conference on virtual reality continuum and its applications* (pp. 59–66).
- Song, R., Liu, Y., Zhao, Y., Martin, R. R., & Rosin, P. L. (2012). Conditional random field-based mesh saliency. In *IEEE international conference on image processing* (pp. 637–640).
- Lee, C. H., Varshney, A., & Jacobs, D. W. (2005). Mesh saliency. *ACM Siggraph*, 174(2005), 659–666.
- Frintrop, S., Rome, E., & Christensen, H. I. (2010). Computational visual attention systems and their cognitive foundations: A survey. *ACM Transactions on Applied Perception*, 7(1), 1–39.
- Castellani, U., Cristani, M., Fantoni, S., & Murino, V. (2008). Sparse points matching by combining 3D mesh saliency. *Eurographics*, 27, 643–652.
- Itti, L., & Koch, C. (2001). Feature combination strategies for saliency-based visual attention systems. *J. Electronic Imaging*, 10, 161–169.
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5, 495–501.
- Locher, P. J., & Nodine, C. (1987). *Symmetry catches the eye* (pp. 353–361). Eye Movements: From Physiology to Cognition.
- Walther, D., Itti, L., Riesenhuber, M., Poggio, T., & Koch, C. (2002). Attentional selection for object recognition—A gentle way. In *Biologically motivated computer vision* (pp. 472–479). Lecture Notes in Computer Science, LNCS2525.
- Frintrop, S., & Jensfelt, P. (2008). Attentional landmarks and active gaze control for visual SLAM. *IEEE Transactions on Robotics*, 24(5), 1054–1065.
- Cretu, A.-M., & Payeur, P. (2011). Biologically-inspired visual attention features for a vehicle classification task”. *International Journal on Smart Sensing and Intelligent Systems*, 4(3), 402–423.
- Yang, Y.-B., Lu, T., & Lin, J.-J. (2009). Saliency regions for 3D mesh abstraction. In *Pacific-rim conference on multimedia* (pp. 292–299). Berlin: Springer.
- Godil, A., & Wagan, A. I. (2011). Salient local 3D features for 3D shape retrieval. In *SPIE 3D Image Processing and Application* (p. 78640S-8).
- Sipiran, I., & Bustos, B. (2010). A robust 3D interest points detector based on Harris operator. In *Eurographics workshop on 3D object retrieval* (pp. 7–14).
- Novatnak, J., & Nishino, K. (2007). Scale-dependent 3D geometric features. In *IEEE international conference on computer vision* (pp. 1–8).
- Sun, J., Ovsjanikov, M., & Guibas, L. (2009). A concise and provably informative multi-scale signature based on heat diffusion. *Eurographics Symp. Geom. Proc.*, 28, 1383–1392.
- Luebke, D. (2001). A developer’s survey of polygonal simplification algorithms. *IEEE Computer Graphics and Applications*, 21(3), 24–35.
- Garland, M., & Heckbert, P. S. (1997). Surface simplification using quadric error meshes. In *Proceedings of ACM Siggraph* (pp. 209–216).
- Howlett, S., Hammil, J., & O’Sullivan, C. (2005). An experimental approach to predicting saliency for simplified polygonal models. *ACM Trans. Applied Perception*, 2(3), 1–23.
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), 1254–1259.
- Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of Physiology*, 357, 241–265.

25. Lu, Z.-L., & Doshier, B. (2013). *Visual psychophysics: From laboratory to theory*. Cambridge, MA: MIT Press.
26. Peyre G. (2007) Toolbox graph, 2007. Accessed August 1, 2016, from <http://www.mathworks.com/matlabcentral/fileexchange/5355-toolbox-graph>.
27. Gal, R., & Cohen-Or, D. (2006). Salient geometric features for partial shape matching and similarity". *ACM Trans. Graphics*, 25(1), 130–150.
28. Steiner, D. C., & Morvan, J.-M. (2003). Restricted Delaunay triangulations and normal cycle. In *Proceedings of the nineteenth annual symposium on computational geometry* (pp. 312–321).
29. Alliez, P., Cohen-Steiner, D., Devillers, O., Levy, B., & Desbrun, M. (2003). Anisotropic polygonal remeshing. *ACM Siggraph*, 22(3), 485–493.
30. Kootstra, G., Nederveen, A., & de Boer, B. (2008). Paying attention to symmetry. In *Proceedings of the British machine vision conference* (pp. 1115–1125).
31. Locher, P., & Nodine, C. (1989). The perceptual value of symmetry. *Computers Math. Applic.*, 17(4), 475–484.
32. Loy, G., & Eklundh, J.-O. (2006). Detecting symmetry and symmetric constellations of features. In *IEEE ECCV* (pp. 508–521).
33. Zhang, J., Sun, J. Liu, J., Yang, C., & Yan, H. (2010). Visual attention model based on multi-scale local contrast of low-level features. In *IEEE ICSP*, 2010 (pp. 902–905).
34. Cheng, M.-M., et al. (2001). Global contrast based salient region detection. In *IEEE Conference on computer vision and pattern recognition* (pp. 409–416).
35. Harel, J., Koch, C., & Perona, P. (2006). Graph-based visual saliency. In *Proceedings of the neural information processing systems* (pp. 545–552).
36. Kadir, T., & Brady, M. (2001). Scale, saliency and image description. *International Journal of Computer Vision*, 45(2), 83–105.
37. Holzbach, A., & Cheng, G. (2014). A fast and scalable system for visual attention, object based attention and object recognition for humanoid robots. In *IEEE international conference on humanoid robots* (pp. 316–321).
38. Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2004). *Digital image processing using Matlab*. Upper Saddle River, NJ: Prentice Hall.
39. Moller, T., & Trumbore, B. (1997). Fast, minimum storage ray/triangle intersection. *Journal of Graphics Tools*, 2(1), 21–28.
40. Monette-Thériault, H., Cretu, A.-M., & Payeur, P. (2014). 3D object modeling with neural gas based selective densification of surface meshes. In *IEEE international conference on systems, man, and cybernetics* (pp. 1373–1378).
41. Cignoni, P., Rocchini, C., & Scopigno, R. (1998). Metro: Measuring error on simplified surfaces". *Computer Graphics Forum*, 17(2), 167–174.
42. Wang, Z., Bovik, A., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Im. Proc.*, 13(4), 1–14.
43. Laparra, V., Balle, J., Berardino, A., & Simoncelli, E. P. (2016). *Perceptual image quality assessment using a normalized Laplacian pyramid* (p. 16). Proc: Sym. Elec. Im., HVEI, vol.
44. CloudCompare. (2016). 3D point cloud and mesh processing software. Accessed August 1, 2016, from <http://www.danielgm.net/cc/>.
45. Dutagaci, H., Cheung C.-P., Godil, A. (2016) A benchmark for 3D interest points marked by human subjects. Accessed August 1, 2016, from <http://www.itl.nist.gov/iad/vug/sharp/benchmark/3DInterestPoint>.
46. Chagnon-Forget, M., & Cretu, A.-M. (2015). Visual attention-based 3D multiple LOD modeling for virtual environments. In *IEEE international symposium on haptic audio visual environments and games* (pp. 69–74).