



Point cloud saliency detection via local sparse coding

Esmeide Leal^a, German Sanchez-Torres^b & John William Branch-Bedoya^c

^a Facultad de Ingenierías, Universidad Autónoma de Colombia, Barranquilla, Colombia. esleal@uac.edu.co

^b Facultad de Ingenierías, Universidad del Magdalena, Santa Marta, Colombia. gsanchez@unimagdalena.edu.co

^c Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia. jwbranch@unal.edu.co

Received: November 4th, 2018. Received in revised form: March 19th, 2019. Accepted: April 22th, 2019.

Abstract

The human visual system (HVS) can process large quantities of visual information instantly. Visual saliency perception is the process of locating and identifying regions with a high degree of saliency from a visual standpoint. Mesh saliency detection has been studied extensively in recent years, but few studies have focused on 3D point cloud saliency detection. The estimation of visual saliency is important for computer graphics tasks such as simplification, segmentation, shape matching and resizing. In this paper, we present a method for the direct detection of saliency on unorganized point clouds. First, our method computes a set of overlapping neighborhoods and estimates a descriptor vector for each point inside it. Then, the descriptor vectors are used as a natural dictionary in order to apply a sparse coding process. Finally, we estimate a saliency map of the point neighborhoods based on the Minimum Description Length (MDL) principle. Experiment results show that the proposed method achieves similar results to those from the literature review and in some cases even improves on them. It captures the geometry of the point clouds without using any topological information and achieves an acceptable performance. The effectiveness and robustness of our approach are shown by comparing it to previous studies in the literature review.

Keywords: point clouds; sparse coding; saliency; minimum description length.

Detección de prominencia de nubes de puntos por medio de codificación dispersa local

Resumen

El sistema visual humano (SVH) puede procesar grandes cantidades de información visual al instante. La percepción de la salencia visual es el proceso de localizar e identificar regiones con un alto grado de prominencia desde un punto de vista visual. La detección de la salencia de la malla se ha estudiado ampliamente en los últimos años, pero pocos trabajos se han centrado en la detección desde las nubes de puntos 3D. La estimación de la salencia visual es importante para tareas de gráficos por computadora tales como simplificación, segmentación, ajuste de formas y cambio de tamaño. En este artículo presentamos un método para la detección directa de la salencia en nubes de puntos no organizadas. Primero, nuestro método calcula un conjunto de vecindarios superpuestos y estima un vector descriptor para cada punto dentro de él. Luego, los vectores descriptores se utilizan como un diccionario natural para aplicar un proceso de codificación disperso. Finalmente, estimamos un mapa de prominencia de los vecindarios de puntos según el principio de Longitud de descripción mínima (MDL). Los resultados experimentales muestran que el método propuesto alcanza los resultados del estado del arte y algunos casos los logra superar. Captura la geometría de las nubes de puntos sin utilizar ninguna información topológica y logra un rendimiento aceptable. La efectividad y la solidez de nuestro enfoque se muestran en comparaciones con trabajos anteriores en el estado del arte.

Palabras clave: nubes de puntos, código disperso, salencia, longitud de descripción mínima.

1. Introduction

The human visual system (HVS) can process large amounts of visual information instantly and is able to locate objects of interest and distinguish them from complex background scenes. Research shows that the HVS pays more

attention to infrequent features and suppresses repetitive ones [1-2]. Visual saliency plays an important role in the process in which HSV identifies scenes and detects objects. It is also concerned with the way the biological system perceives the environment. For example, every time we look at a specific

place, we pay more attention to particular regions which are distinct from the surrounding area.

Visual saliency is an active research field in areas such as psychology [2], neuroscience [3], computer vision [4-5] and computer graphics [6-7]. There are many computational methods for simulating the HVS and visual saliency, but it nonetheless remains an unexplored field because of the difficulty of designing algorithms to simulate this process [6]. As for computer graphics, while the concept of visual saliency has been widely explored for mesh saliency [6], [8-18], few studies have explored visual saliency in point clouds [19-22]. Visual saliency is an important topic for 3D surface study and has important applications in 3D geometry processing such as resizing [1], simplification [23-24], smoothing [25], segmentation [26], shape matching and retrieval [27-28], 3D printing [29], and so forth.

Due to the development of 3D data scanning technology, current scanners generate thousands of points for every scanned object. Therefore, for rendering, point clouds have become an alternative to triangular meshes. A common way to process a point cloud is to reconstruct the surface using methods such as triangular mesh, NURBS representation and Radial Basis Functions. However, due to a large number of points, different sampling densities and the inherent noise produced by the scanning process, reconstruction is an expensive and challenging computational task. For these reasons, it is necessary to develop geometry processing algorithms that operate directly on the point sets. Applying existing saliency detection techniques to point clouds is not a trivial task. This is due to the absence of topological information, which is not a problem for mesh-based methods. The method we propose here is inspired by Li *et al.* [30] and we made it fit for general use supported by the Minimum Description Length (MDL) principle. In this way, MDL is used as a criterion to distinguish regions in the point clouds.

1.1. Contribution

The main contributions of this paper are: (1) a method for saliency detection without topological information using only the raw point sets, (2) the use of the MDL principle for defining saliency measurements, extended to sparse coding representation to obtain saliency maps of point clouds.

Experimental results show that the proposed algorithm performs favorably in the capture of the geometry of the point clouds without using any topological information and achieves an acceptable performance when compared to previous approaches.

1.2. Organization

This paper is organized as follows. A review of related work on saliency detection is provided in Section 2. The theoretical basis for Sparse Coding and MDL for the proposed Sparse Coding Saliency method is presented in Section 3. Section 4 lays out the results and discussion. Finally, we present the conclusions and future studies in Section 5.

2. Related work

Visual saliency detection has its origin in the area of computer vision, specifically in 2D images. Inspired by this research, visual saliency detection has been applied successfully to 3D meshes and point clouds. In recent years, much research has tried to develop methods for visual saliency on 3D surfaces [10-12,17,19-21,32].

Early advances in the field of Mesh Saliency used 2D projections for 3D applications. While such works often ignored (or were unable to exploit) the importance of depth in human perception, they set a basis for further research by highlighting the importance of human perception in image analysis [18]. Some of the first researchers to exploit saliency for 3D mesh processing were Lee *et al.* [32]. The authors introduced the concept of mesh saliency and computed it using a Gaussian-weighted center-surround mechanism. Results were then exploited for the implementation of mesh simplification and best point-of-view selection algorithms. Wu *et al.* [13] presented an approach based on the principles of local contrast and global rarity. This method has applications for mesh smoothing, simplification and sampling. Leifman *et al.* [33] propose a method for detecting a region of interest on surfaces where they capture the distinctness of the vertex descriptor, characterizing the local geometry around it. The method is applied to viewpoint selection and shape similarity. Tao *et al.* [11] put forward a mesh saliency detection approach which reached state-of-the-art performance even when handling noisy models. Firstly, the manifold ranking was employed in a descriptor space to imitate human attention, and then descriptor vectors were built for each over-segmented patch on the mesh, using Zernike coefficients and center-surround operators. Afterward, background patches were selected as queries to transfer saliency, helping the method to handle noise better. An approach for mesh saliency detection based on a Markov Chain is proposed by Liu [10]; the input mesh was partitioned into segments using Neuts algorithms and then over-segmented into patches using Zernike coefficients. Instead of employing center-surround operators, background patches were selected by determining feature variance to separate the insignificant regions. Limper *et al.* [8] applied the concept of Shannon entropy, which is defined as the expected information value within a message, for 3D mesh processing. In this case, the curvature is established to be the primary source of information in the mesh. Song *et al.* [18] argued for the benefits of using not only local but also global criteria for mesh saliency detection. Their method consisted of computing local saliency features while considering the later computation of global saliency using a statistic Laplacian-based algorithm which captures salient features at multiple scales.

Recently, point saliency detection-based methods have been introduced. Shtrom *et al.* [21] propose a multi-level approach to find distinct points, basing their approach on the context-aware method for image processing. Tasse *et al.* [20] propose a cluster-based approach; the method decomposed the point cloud into the small clusters using an adaptive fuzzy clustering algorithm and is applied in the detection of key-points.

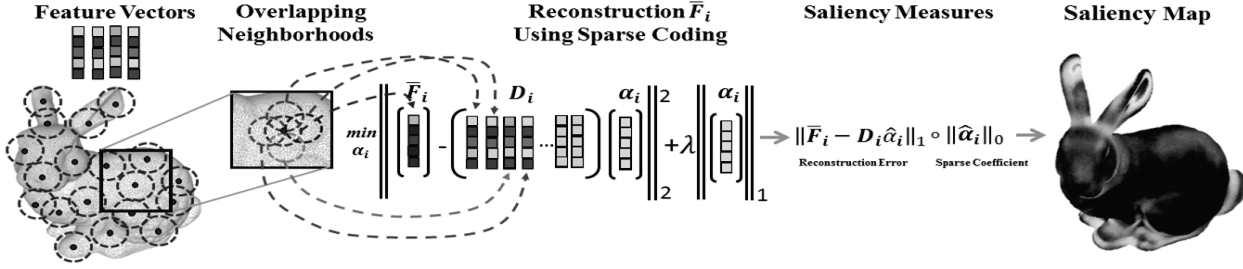


Figure 1. Processing pipeline of our sparse coding-based saliency detection method.
Source: The Authors.

Finally, Guo *et al.* [19] propose a saliency detection method based on a covariance descriptor to capture the local geometry information, using a sigma set descriptor to transform the covariance descriptor from a Riemannian space to a Euclidian space to facilitate the application of Principal Component Analysis for the inner structure analysis to classify whether a point is salient or not.

3. Proposed method

The saliency detection framework based on sparse coding is shown in Fig.1. The input point cloud is represented in sparse form and the saliency is estimated by a center-surround hypothesis [34-35], which states that a region is salient if it is distinct from its surrounding regions. Taking a point cloud as input, we estimate a neighborhood with radius r for each point in the cloud, then select some of its boundary points. We estimate a neighborhood with radius r around each selected point. Overlapping is allowed between neighborhoods in order to capture the structure of the cloud. Then, for each point in the neighborhood, we compute a set of basic features to build a descriptor vector and finally, we estimate the mean of the feature vectors to build only one descriptor vector per neighborhood. Similarly, we do the same with the neighborhoods surrounding the central one. Afterward, we construct a dictionary D , using the feature vectors from the surrounding neighborhoods, and then a sparse coding is carried out using the feature vector of the central neighborhood and by searching for a sparse representation for it with the basis in the dictionary D . Finally, we compute the neighborhood saliency using the MDL principle based on the sparse representation measure of vector x , and the residual of the sparse reconstruction measure. The final saliency map is computed using a fusion of both measures.

3.1. Sparse coding

The purpose of sparse coding is to approximate a feature input vector as the linear combination of basis vectors, called atoms, which are selected from a dictionary which has been created from the data. In other words, sparse coding provides a low-dimensional approximation of a given signal in a given set of a basis [36].

Formally, let x be a signal of dimension n , the sparse coding aims to find a dictionary $D = \{d_1, d_2, \dots, d_N\}$, such that x can be approximated by a linear combination of the

atoms $\{d_i\}_{i=1}^N$, this is $x \approx D\alpha = \sum_{j=1}^N \alpha_j d_j$, (if the dictionary is overcomplete then $N > n$) where most of the coefficients α_j are zeros or close to zero [37]. We propose that the sparse coding problem can typically be formulated as an optimization problem, as in (1):

$$\hat{\alpha} = \arg \min_{\alpha} \|x - D\alpha\|_2^2 \text{ s.t. } \|\alpha\|_0 \leq L \quad (1)$$

In this formulation, the dictionary D is given and L controls the sparsity of x in D . The term $\|\alpha\|_0$ measures the dispersion of the decomposition and can be understood as the number of non-zero coefficients in α , or, sparse coefficients, in order to approximate the signal x in as sparse a way as possible. Or, alternatively, it can be formulated as (2):

$$\hat{\alpha} = \arg \min_{\alpha} \|x - D\alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (2)$$

This is an optimization problem where the norm $l_0(\|\cdot\|_0)$ is changed by the norm $l_1(\|\cdot\|_1)$, where λ is the regularization parameter). The solution to equation (1) with the norm l_0 is an NP-hard problem; fortunately, under certain conditions, it is possible to relax the problem using the norm l_1 and find an approximated solution using equation (2) with the norm l_1 .

3.2. Minimum description length

The Minimum Description Length (MDL) principle [4][31][38] states that, given a model class \mathcal{M} of candidate models, its parameter M , and its data sample x , the MDL will provide a generic solution to the model selection that minimally represents the data x . Formally, given a set of candidate models \mathcal{M} , and a data vector x , MDL searches for the best model $\hat{M} \in \mathcal{M}$, that can be used to describe x in its entirety with the shortest length.

$$\hat{M} = \arg \min_{M \in \mathcal{M}} \mathcal{L}(x, M) \quad (3)$$

$\mathcal{L}(x, M)$ is a coding assignment function which gives the code length required to describe (x, M) uniquely.

In [39], a method based on information theory was introduced for image saliency detection. This method measures the saliency concerning the likelihood of a patch given the patches that surround it. The method measures the self-information using the negative log-likelihood. Defining

x as an image and $p(x)$ as the probability of occurrence of the patch x , given its surrounding neighborhood patches, the saliency measure of the patch is defined as $-\log p(x)$, that is, the self-information characterizes the raw likelihood of the n -dimensional vector values given by x .

Based on [4,31,38], we use the MDL principle to propose a method for saliency detection in point clouds based on sparse coding. The coding assignment function $\mathcal{L}(x, M)$ can be defined in terms of probability assignment $P(x, M)$, based on the Ideal Shannon Codelength Assignment [28], that is, $\mathcal{L}(x, M) = -\log P(x, M)$. Using Bayes theorem, we can establish $P(x, M)$ as $P(x, M) = P(x|M)P(M)$ and then by applying maximum a posteriori (MAP), the penalized likelihood form of the coding model is formulated thus:

$$\hat{M} = \arg \min_{M \in \mathcal{M}} -\log P(x|M) - \log P(M) \quad (4)$$

where $-\log P(x|M)$ is the term that describes how well the model adjusts the data x and $-\log P(M)$ describes the model complexity, model cost or prior term.

Once \hat{M} is estimated, the terms in equation (4) can be interpreted as follows:

- $P(\hat{M})$ is the description length or codelength of the model; and

- $P(x|\hat{M})$ is the description length or Codelength of the data encoded using the model.

3.3. Relating the MDL principle to sparse coding

The standard Gaussian is assumed to be coding assignment function L , and assuming that x can be represented in a sparse way given a basis dictionary D and a sparse vector α , the probability distribution of the reconstruction error follows a Gaussian distribution $P(x|D\alpha) \sim N(D\alpha, \sigma^2)$, with known variance σ^2 . The term $-\log P(x|M) - \log P(M)$ in (4) becomes $-\log P(x|D\alpha) = \frac{1}{2\sigma^2} \|x - D\alpha\|_2^2$, and the term $-\log P(M)$ in (4), assuming sparsity constraint, becomes $-\log P(M) = \|\alpha\|_0$. The aforementioned calculations adhere to the sparse coding model described in Section 3.3.

The sparsity condition and the sparse reconstruction error conform to the MDL principle if we set the estimation parameter $\hat{M} = \hat{\alpha}$ and evaluate it with the prior term $P(\hat{\alpha})$, and with the likelihood term $P(x|\hat{\alpha})$, and so $\|\hat{\alpha}\|_0$ and $\|x - D\hat{\alpha}\|_2$ are obtained respectively. We conclude that the sparsity of the vector $\hat{\alpha}$ is the codelength of the model and the residual of the sparse reconstruction error is the codelength of the data given the model.

The MDL principle selects the best model that produces the shortest description of the data. The more regularity that is presented in the data, the shorter the description the model will produce. If a neighborhood is equal or slightly different with respect to its surroundings in terms of information, it means these signals (neighborhoods) are redundant and can be represented sparsely with a suitable basis dictionary, meaning that its description length (the sparsity of vector $\hat{\alpha}$ i.e., $\|\hat{\alpha}\|_0$) will be short and we can conclude that this neighborhood is not salient; on the other hand, if the

neighborhood is very different from its surroundings, its description length will be longer. In other words, it cannot be represented sparsely by the basis dictionary and we can conclude that the neighborhood is salient.

If the sparse reconstruction error (i.e., $\|\hat{\alpha}\|_0$) produces a high residual, it means that its description length will be longer and this implies that the neighborhood is dissimilar from its surroundings and therefore more salient. On the other hand, if the sparse reconstruction error produces a low residual, its description length will be short and this implies that the neighborhood is similar with respect to its surroundings and, consequently, less salient. Based on these saliency measurements, our point cloud saliency detection method is described below.

3.4. Sparse coding saliency detection

Unlike some saliency detection methods for point clouds compiled in the literature review, our method does not estimate saliency using individual points directly, rather, a neighborhood is used to determine its saliency.

3.4.1. Feature vector

Given the point set $P = \{p_i \in R^3\}, i = 1, 2, \dots, N$, a neighborhood $N(p_i)$ is estimated for every point $p_i \in P$ with radius $r > 0$, i.e., $N(p_i) = \{p_j | \|p_i - p_j\| \leq r\}$ is the number of neighbor points p_j inside a sphere with radius r , and center p_i . See Fig.1. In order to establish whether a neighborhood is different, first it is important to characterize it with a descriptor, of which there are several which describe low level features for each p_i , for example, normal, curvatures, shape index, etc. In this method the normal and Gaussian curvatures are selected; these features are rotationally invariant. A third feature, d_i , is also selected. With the features defined, a five dimensional feature vector F_i is formed for every point p_i of $N(p_i)$, i.e., $F_i = (Nx_i, Ny_i, Nz_i, k_i, d_i)$, where Nx_i, Ny_i, Nz_i are the three components of the normal vector N_i ; k_i is the Gaussian curvature and d_i is the fifth component coordinate of F_i , which will be defined below. It is necessary to have a single descriptor vector for each neighborhood, so the mean of the characteristic vectors belonging to the neighborhood are estimated as follows:

$$\bar{F}_i = \frac{1}{|N(p_i)|} \sum_{j=1}^{n_i} F_j \quad (5)$$

where $n_i = |N(p_i)|$ is the cardinality of $N(p_i)$.

d_i is a global measure, which establishes the difference between the feature vector of each neighborhood F_i' and the global mean \bar{F}_g (6) of all the feature vectors of the point cloud, that is:

$$\bar{F}_g = \frac{1}{N} \sum_{j=1}^N F_j \quad (6)$$

$$d_i = \|\bar{F}_i - \bar{F}_g\| \quad (7)$$

3.4.2. Surrounding neighborhoods

Once the neighborhood descriptor is established, the next step is to find the surrounding neighborhoods to each $N(p_i)$. To find these neighborhoods, first, the 3x3 covariance matrix (8) of $N(p_i)$ is computed as:

$$Cm_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (p_i - p') (p_i - p')^T \quad (8)$$

where \bar{p} is the mean of $N(p_i)$ and $n_i = |N(p_i)|$.

After the covariance matrix Cm_i has been estimated, we will use its two largest eigen-values with its corresponding eigen-vectors that follow the x and y axes, which expand the tangent plane T_i to $N(p_i)$ at the point p_i . Next, the points within $N(p_i)$ are projected onto the 2D plane T_i as shown in Fig. 2(a). Before projecting the points within $N(p_i)$, we establish the area of the surrounding neighborhoods as the n -ring, as can be seen in Fig. 2(b), that is, the 1-ring corresponds to radius r , the 2-ring corresponds to $2r$ and so on. In these experiments we used the 1-ring. Then, we proceed to trace a set of radii, separated by an θ angle, from the center of the projected neighborhood in Fig. 2(b); the length of radii is n -ring. In each of the radii, we mark points (green dots) depending on if we have used the 1-ring, 2-ring or n -ring, as shown Fig. 2(b). Then, we find the nearest point to each marked point within the projected neighborhood and find its corresponding point within the 3D neighborhood; then we estimate a neighborhood for each of the 3D points with radius r , as seen in Fig. 2(c). Finally, a feature vector is estimated for each surrounding neighborhood, in the same way as in the previous section.

3.4.3. Dictionary construction and sparse coding model

It can be observed that the surrounding feature vectors \bar{F}_j are a natural over complete basis dictionary D , i.e., $D = \{\bar{F}_1, \bar{F}_2, \dots, \bar{F}_N\}$ and the central feature vector \bar{F}_i acts as the sparse linear combination of these basis or atoms, i.e., $\bar{F}_i = D\alpha = \sum_{j=1}^N \alpha_j \bar{F}_j$. Now we can write the sparse model as so:

$$\hat{\alpha} = \arg \min_{\alpha} \|\bar{F} - D\alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (9)$$

Equation (9) is a linear regression problem optimization for estimating $\hat{\alpha}$, known as Lasso. The LARS algorithm gives the solution and the SPAMS library is used to carry out the sparse coding solution.

3.4.4. Saliency detection

Previously, in Section 3.4.1, d_i was mentioned as being the fifth component of the feature vectors F_i , and it was estimated using (7). This component was added because often, a local neighborhood has similar surroundings but the local and surroundings are globally distinct over the entire point cloud. Using only normal vectors and Gaussian curvature can produce areas in the cloud which have saliency but the center of these areas can be empty. Adding d_i solves this difficulty.

When a sparse solution to (9) is achieved, the codelength for the neighborhoods is established as was laid out in Section 3.3. Our model is based on the MDL principle, therefore, the saliency of each neighborhood $N(p_i)$ is proportional to $\|\hat{\alpha}\|_0$ and $\|\bar{F} - D\hat{\alpha}\|_1$. We replaced the L2-norm with the L1-norm in the residual error, because it is more discriminative and robust to outliers. Next, following the instructions in Borji and Itti [40], we rewrote the equations $\|\bar{F} - D\hat{\alpha}\|_1 = S_i(N(p_i))$ and $\|\hat{\alpha}\|_0 = S_M(N(p_i))$, both saliency measurements are then normalized and combined (10).

$$S_T(N(p_i)) = \mathcal{N}(S_M) \circ \mathcal{N}(S_i) \quad (10)$$

where $S_M(\cdot)$ is the saliency produced by the sparse reconstruction error, $S_i(\cdot)$ is the saliency produced by the sparse coefficient and S_T is the combination of both saliency measurements; $\mathcal{N}(\cdot)$ normalizes the saliency measurements for a better fusion. The symbol \circ in (10), is an integration scheme $\{*, +, \max, \min\}$ [40], see Fig. 3.

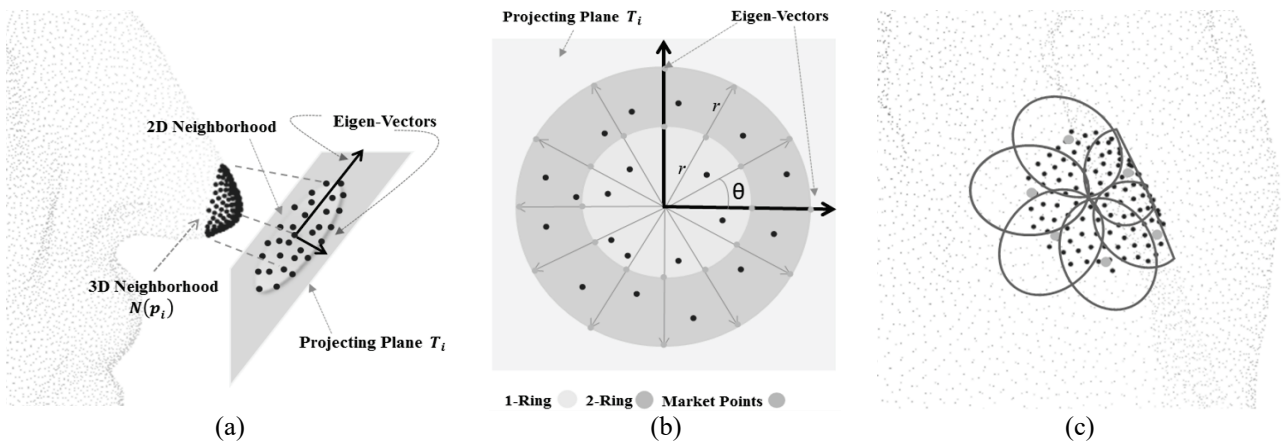


Figure 2. Surrounding neighborhoods selection. (a) Projecting the 3D points to 2D, (b) mark points and n-ring area, (c) selecting surrounding neighborhoods. Source: The Authors.

The best results in all experiments were produced using the * operator. Since overlapping is allowed between neighborhoods, the total saliency map is obtained by accumulating the saliency by neighborhood. Our method differs from that proposed in [30], since it only uses the dispersion vector (i.e., $\|\hat{\alpha}\|_0$) as a measure of saliency, while our method also takes into account the reconstruction error or residual (i.e., $\|x - D\hat{\alpha}\|_1$) as an additional measure of saliency; finally when these measurements are combined, we obtain a better saliency map. Our method can be seen as a generalization of the framework presented in Li *et al.* [30], improving the final result of the saliency map. This can be observed in Fig. 4. This is a result of the residuals performing as a weighting factor with respect to the measure of rarity given by the dispersion vector.

4. Results and discussion

In this section, we evaluate the proposed method on a set of objects and compare it against 3D model saliency detection approaches outline in the literature review, taking into account methods based on point clouds and meshes. The object shapes were obtained from the Watertight Models of SHREC 2007 and the Stanford 3D Scanning repository. Our method is compared to three-point cloud-based methods from the literature review: Tasse *et al.* [20], Shtrom *et al.* [21] and Guo *et al.* [19], as well as six mesh-based methods from the literature review: Lee *et al.* [32], Wu *et al.* [13], Leifman *et al.* [33], Tao *et al.* [11], Song *et al.* [18] and Liu *et al.* [10]. It is also compared against the pseudo-ground truth [41].

All the experiments were run on a PC with an Intel Core i7-2670QM CPU@2.20 GHz and 8GB RAM. As an example, for the model of a girl in Fig. 8, with 15.5 K vertices, the saliency computation cost of our method was 81.7s, being slower than Tasse [20] and Guo [19] but it has some optimizations and if a language like C++ is used, performance will improve.

4.1. Parameter selection

The only parameter in our method is λ ; this is the regularization parameter. It produces smoother results as value increases in the range (0, 1); its visual effect is appreciated in Fig. 5. Experimentally we found that, on average, a value of lambda of near to 0.9 generates the best qualitative as well as quantitative results, therefore in all our experiments, we fix lambda at $\lambda = 0.9$.

The size of the neighborhood is calculated according to the local characteristics of points like density and curvature, and its size is increased or decreased. To achieve this, a method—such as that proposed in [42]— is used to estimate the size of the neighborhood, taking into account the local characteristics named.

4.2. Qualitative evaluation

Our method is compared with methods from the literature review, both mesh-based and point cloud-based. In Fig. 6, we compare the point-based methods, including Shtrom *et al.* [21], Tasse *et al.* [20] and Guo *et al.* [19]. The method in

Shtrom *et al.* can obtain a reasonable saliency result and highlights a relevant saliency region on the Max Planck model, but there are also extensive areas with noise around the principal saliency features. In the method proposed by Tasse *et al.*, less noise is perceived, but there are large regions around the features with greater saliency. In the result of the Guo *et al.* method we observe a clean saliency map concentrated on the most representative saliency features, however, our result, in addition to achieving the same, highlights areas such as eyes, the lower part of the nose, ears and lips with greater saliency. Regarding the dragon model, there was a similar result as with the Max Planck model. The proposed method produces a clean saliency map compared to Shtrom *et al.* and Tasse *et al.* The saliency map produced by Guo *et al.* is very similar to ours, which only highlights some of the finer details such as eyes and ridges.

We also compare our results with mesh-based methods, shown in Fig. 7, with those of Lee *et al.* [32], Wu *et al.* [13], Leifman *et al.* [33], and Tao *et al.* [11]. It can be observed that the local changes in the curvature had less influence when using the proposed method, as seen in Lee *et al.* [32].

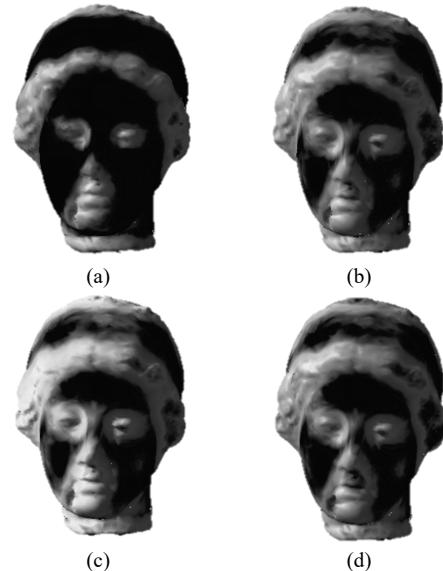


Figure 3. Saliency maps generated by different operator functions, using equation (11). (a) Using * operator, (b) Using + operator, (c) Using min operator and (d) Using max operator. Source: The Authors.

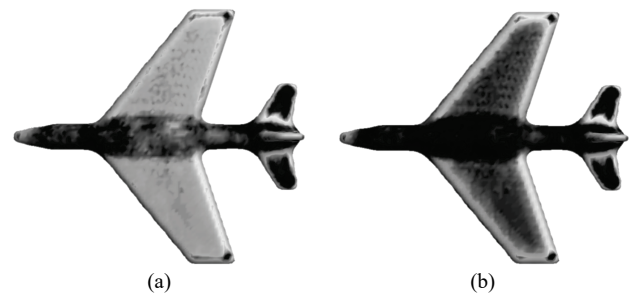


Figure 4. Saliency maps results, (a) Using only the dispersion vector $\|\hat{\alpha}\|_0$ as saliency measurement (b) Using the dispersion vector and residual error $\|x - D\hat{\alpha}\|_1 * \|\hat{\alpha}\|_0$ as saliency measurement (b). Source: The Authors.

In the bunny and dragon models, the saliency map is not correct due to variation of the level of saliency in different areas, the dragon, for example, has lost some fine details like crests and ridges.

The method of Wu *et al.* produces a better saliency map than that of Lee *et al.*, but some salient areas are missing as in the case of the dragon. The method proposed by Leifman *et al.* produces a reliable saliency map for the dragon model, in the case of bunny, the feet are missing in the final saliency map. For the Tao *et al.* method, it can be seen that in the dragon model, some areas are shown to be salient where they are not. In the visual comparisons, it should be pointed out that our method generally achieves better results than the other four methods.

Fig. 8 compares our results with [10,11,18], and shows that the way in which our method detects saliency is more coherent, as it detects small salient regions, such as the ears and the little tie, and the facial regions like the eyes and mouth of the bust of the girl. Furthermore, the hair (bun and braids) show up better with the pseudo-ground truth [41], in comparison to the other four methods. We observe, in reference to the bird model, how our method detects the saliency points in the wings, tail, beak and finally the light area of saliency between the wings compared to the pseudo-ground truth bird model.



Figure 5. Saliency map produced by our method with different values of λ . From left to right: $\lambda = 0.1, 0.2, 0.4, 0.6, 0.8, 0.9$. Source: The Authors.

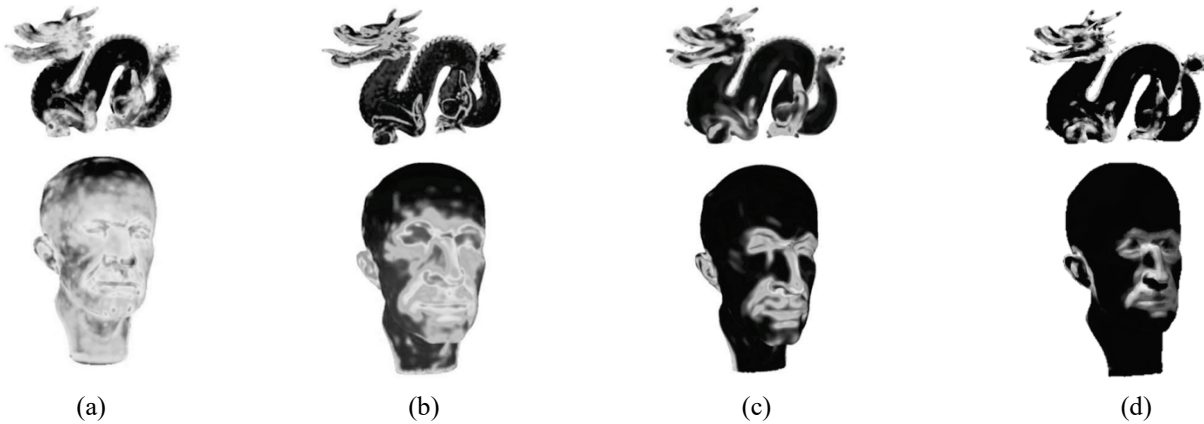


Figure 6. Point-based methods saliency comparison. (a) Shtrom et al. [23]. (b) Tasse et al. [3]. (c) Guo et al. [24], and (d) our proposal. Source: [24,3,23].

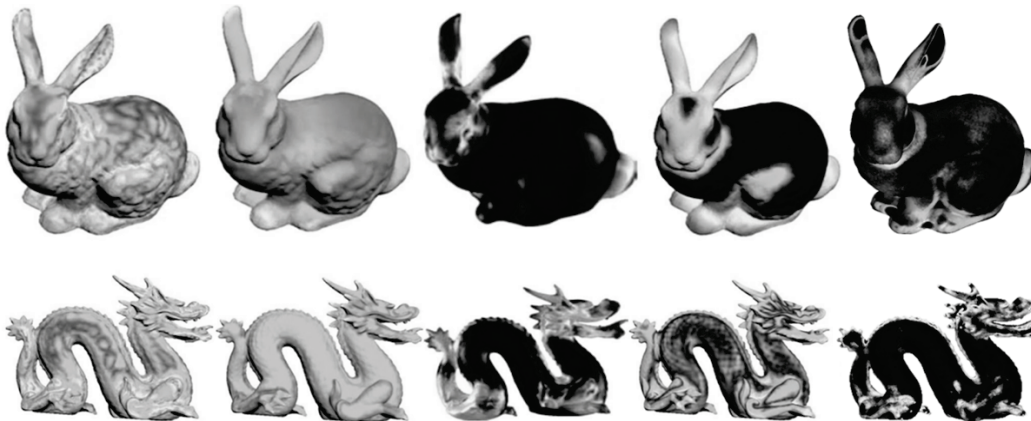


Figure 7. Mesh-based methods comparison. (a) Lee et al. [21]. (b) Wu et al. [10]. (c) Leifman et al. [22], Tao et al. [8] (d), and (e) our proposal. Source: [8,10,21,22].



Figure 8. Mesh saliency results of Tao et al. [11] (a), Liu et al. [7] (b), Song et al. [15] (c), Our method (d), and the pseudo-ground truth. [41]. Source: [7,8,15,41].

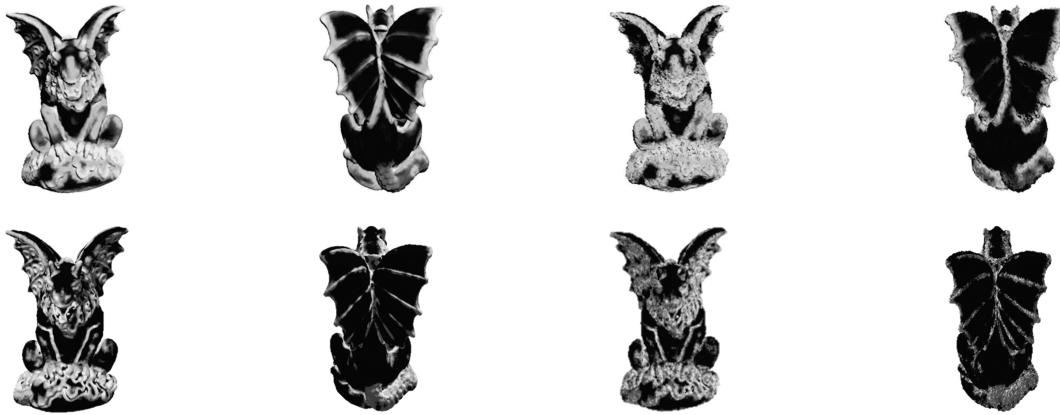


Figure 9. Saliency results on Gargoyle model. Our method (bottom row), and Tao et al. [11] (first row). The first two columns are front and back view of saliency results from Gargoyle model. The second two columns are the same view results but with 30% random noise relative to the average of the nearest distance of each point. Source: [11].

Fig.9 shows the results when the input is a noisy point cloud, we can observe that our method is more robust against noise in comparison to mesh-based methods [11]. We added 30% (relative to the average of the nearest distance of each point) random noise to disturb the points' coordinates, and observe that the method proposed by Tao *et al.* failed to detect some details of the wings in the gargoyle like the little rings between the veins from the front view and the loss of some details in the veins from the back view. We can observe that our method is more consistent in detecting salient features in the gargoyle model, both in the clean and noisy model in comparison to Tao *et al.* [11]. The strength of our method lies in the fact that since the saliency of point clouds is a feature of a sparse nature, modeling based on sparse models is usually a good alternative to adequately represent these characteristics. When it is combined with the MDL principle, this method is the simplest and best choice guarantee among all the possible solutions.

4.3. Quantitative evaluation

We carried out a quantitative evaluation on the 400 watertight models of SHREC 2007, using the distribution of schelling points provided by Chen *et al.* [41] as the ground truth. To evaluate the performance, we used the metrics proposed by Tasse *et al.* [43]; they propose three metrics that adapt 2D image saliency metrics to 3D saliency, these

metrics are: Area Under the ROC curve (AUC), Normalized Scanpath Saliency (NSS) and Linear Correlation Coefficient (LCC). In the AUC metric the ideal saliency model has a score of 1.0, AUC disregards regions with no saliency, and focuses on the ordering of saliency values. The NSS metric measures the saliency values selecting the users as fixation points. The LCC metric has values of between -1 and 1, with values closer to 0 implying weak correlation. This metric compares the saliency map under consideration with the ground-truth saliency map (Schelling distributions). For comparison, we selected three methods from the literature review, two based on point clouds and one based on meshes, these methods are: the clustering method (CS) [20], the point-wise method (PW) [19] and the spectral analysis method (SS) [12]. Also, as a reference score, we incorporated the human performance score (HS), provided by Tasse *et al.* [43].

Table 1 shows the AUC score values of the selected methods; our method competes with the CS and PW methods, as the final average shows. The SS method performs poorly compared to our method and the CS and PW methods. The same is true regarding the HS method which outperforms the SS method. Table 1 shows that for the classes glasses, ant, octopus, bird and bearing, our method obtained the best results, while for the classes hand, fish, spring, mechanic, airplane and vase, our method equals the PW and CS methods. The above shows that the proposed method achieves the same as the most up-to-date methods of

the literature review and, in some cases, even improves on them.

Fig. 10 shows NSS and LCC metrics, and also confirms that the performance of our method is similar to CS and PW, and outperforms SS, but HS outperforms all methods presented.

5. Conclusions and future work

This paper presents a novel and simple method for point cloud saliency detection via sparse coding. Based on the MDL principle, the proposed method uses a sparse coding representation to find the minimum codelength to establish when a neighborhood is salient or not with respect to its surrounding neighborhoods. It is robust against noise since it takes the mean of the feature vectors of the neighborhood as a unique feature vector.

Table 1. AUC performance per shape class in SHREC 2007.

Class	SS	CS	PW	Ours	HS (1 v. all)
human	0.57	0.57	0.59	0.58	0.58
cup	0.59	0.60	0.64	0.62	0.57
glasses	0.53	0.52	0.53	0.55	0.56
airplane	0.62	0.62	0.60	0.62	0.60
ant	0.57	0.60	0.59	0.61	0.60
chair	0.55	0.57	0.59	0.58	0.62
octopus	0.53	0.54	0.56	0.59	0.60
table	0.58	0.63	0.65	0.60	0.61
teddy	0.55	0.56	0.57	0.56	0.60
hand	0.58	0.61	0.61	0.61	0.58
pliers	0.56	0.59	0.59	0.58	0.56
fish	0.63	0.66	0.66	0.66	0.61
bird	0.56	0.60	0.59	0.61	0.57
spring	0.55	0.55	0.55	0.55	0.54
armadillo	0.60	0.65	0.66	0.65	0.62
bust	0.56	0.62	0.63	0.62	0.59
mechanic	0.53	0.68	0.70	0.70	0.64
bearing	0.50	0.60	0.62	0.63	0.55
vase	0.61	0.62	0.63	0.63	0.58
four-legs	0.60	0.60	0.61	0.60	0.59
Average	0.57	0.60	0.61	0.61	0.59

Source: The Authors.

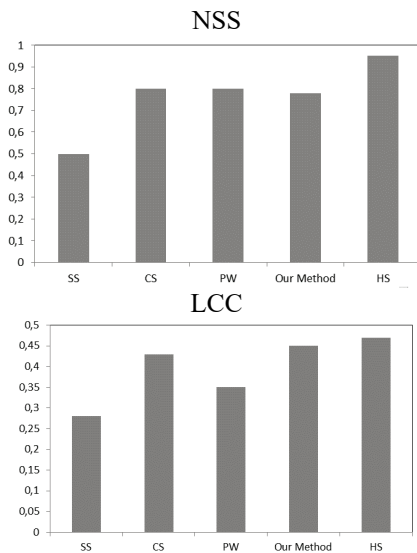


Figure 10. Saliency performance evaluation under the metrics NSS and LCC.

Source: The Authors.

Our approach produces feasible and even faithful results on a variety of models, giving convincing results. We have compared our results to the most recent approaches found in the literature review and we found that the proposed method competes with and in several cases significantly outperforms these approaches, using the pseudo-ground truth provided by Chen *et al.* [41] as a reference. For future studies, we plan to investigate how the incorporation of high-level information in the form of semantic cues into the point cloud saliency detection allows us to identify salience globally on the point cloud.

Acknowledgements

This study is partially supported by Departamento Administrativo de Ciencia, Tecnología e Innovación de Colombia (Colciencias), under the program doctoral scholarships 727.

References

- [1] Jia, S., Zhang, C., Li, X. and Zhou, Y., Mesh resizing based on hierarchical saliency detection, *Graph. Models*, 76(5), pp. 355-362, Sep. 2014. DOI: 10.1016/j.gmod.2014.03.012
- [2] Wolfe, J.M., Guided Search 2.0 A revised model of visual search, *Psychon. Bull. Rev.*, 1(2), pp. 202-238, Jun. 1994. DOI: 10.3758/BF03200774
- [3] Koch, C. and Poggio, T., Predicting the visual world: silence is golden, *Nat. Neurosci.*, 2(1), pp. 9-10, Jan. 1999. DOI: 10.1038/4511
- [4] Somasundaram, G., Cherian, A., Morellas, V. and Papanikolopoulos, N., Action recognition using global spatio-temporal features derived from sparse representations, *Comput. Vis. Image Underst.*, 123(1), pp. 1-13, Jun. 2014. DOI: 10.1016/j.cviu.2014.01.002
- [5] Kalboussi, R., Abdellaoui, M. and Douik, A., A spatiotemporal model for video saliency detection, in *2016 International Image Processing, Applications and Systems (IPAS)*, 2016, pp. 1-6. DOI: 10.1109/IPAS.2016.7880113
- [6] Kim, Y., Varshney, A., Jacobs, D.W. and Guimbretiére, F., Mesh saliency and human eye fixations, *ACM Trans Appl Percept*, 7(2), pp. 1-13, Feb. 2010. DOI: 10.1145/1670671.1670676
- [7] Lau, M., Dev, K., Shi, W., Dorsey, J. and Rushmeier, H., Tactile mesh saliency, *ACM Trans Graph*, 35(4), pp. 1-11, Jul. 2016. DOI: 10.1145/2897824.2925927
- [8] Limper, M., Kuijper, A. and Fellner, D.W., Mesh saliency analysis via local curvature entropy, in: *Proceedings of the 37th Annual Conference of the European Association for Computer Graphics: short papers*, Goslar Germany, Germany, 2016, pp. 13-16. DOI: 10.2312/egsh.20161003
- [9] Wang, S., Li, N., Li, S., Luo, Z., Su, Z. and Qin, H., Multi-scale mesh saliency based on low-rank and sparse analysis in shape feature space, *Comput. Aided Geom. Des.*, 35(36), pp. 206-214, May 2015. DOI: 10.1016/j.cagd.2015.03.003
- [10] Liu, X., Tao, P., Cao, J., Chen, H. and Zou, C., Mesh saliency detection via double absorbing Markov chain in feature space, *Vis. Comput.*, 32(9), pp. 1121-1132, Sep. 2016. DOI: 10.1007/s00371-015-1184-x
- [11] Tao, P., Cao, J., Li, S., Liu, X. and Liu, L., Mesh saliency via ranking unsalient patches in a descriptor space, *Comput. Graph.*, 46, pp. 264-274, Feb. 2015. DOI: 10.1016/j.cag.2014.09.023
- [12] Song, R., Liu, Y., Martin, R. and Rosin, P.L., Mesh saliency via spectral processing, *ACM Trans Graph*, 33(1), pp. 6:1-6:17, Feb. 2014. DOI: 10.1145/2530691
- [13] Wu, J., Shen, X., Zhu, W. and Liu, L., Mesh saliency with global rarity, *Graph. Models*, 75(5), pp. 255-264, Sep. 2013. DOI: 10.1016/j.gmod.2013.05.002

- [14] Nouri, A., Charrier, C. and Lézoray, O., Multi-scale mesh saliency with local adaptive patches for viewpoint selection, *Signal Process. Image Commun.*, 38, pp. 151-166, Oct. 2015. DOI: 10.1016/j.image.2015.08.002
- [15] Liu, X., Ma, L. and Liu, L., P2: a robust and rotationally invariant shape descriptor with applications to mesh saliency, *Appl. Math.- J. Chin. Univ.*, 31(1), pp. 53-67, Mar. 2016. DOI: 10.1007/s11766-016-3364-5
- [16] Zhao, Y. et al., Region-based saliency estimation for 3D shape analysis and understanding, *Neurocomputing*, 197, pp. 1-13, Jul. 2016. DOI: 10.1016/j.neucom.2016.01.012
- [17] Jeong, S.W. and Sim, J.Y., Saliency detection for 3D surface geometry using semi-regular meshes, *IEEE Trans. Multimed.*, 19(12), pp. 2692-2705, Dec. 2017. DOI: 10.1109/TMM.2017.2710802
- [18] Song, R., Liu, Y., Martin, R. and Echavarría, K., Local-to-global mesh saliency, *Vis. Comput.*, 34(3), pp. 323-336, Nov. 2016. DOI: 10.1007/s00371-016-1334-9
- [19] Guo, Y., Wang, F. and Xin, J., Point-wise saliency detection on 3D point clouds via covariance descriptors, *Vis. Comput.*, pp. 1-14, Jun. 2017. DOI: 10.1007/s00371-017-1416-3
- [20] Tasse, F.P., Kosinka, J. and Dodgson, N., Cluster-based point set saliency, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 163-171. DOI: 10.1109/ICCV.2015.27
- [21] Shtrom, E., Leifman, G. and Tal, A., Saliency detection in large point sets, in: 2013 IEEE International Conference on Computer Vision, 2013, pp. 3591-3598. DOI: 10.1109/ICCV.2013.446
- [22] Akman, O. and Jonker, P., Computing saliency map from spatial information in point cloud data, in *Advanced Concepts for Intelligent Vision Systems*, 2010, pp. 290-299. DOI: 10.1007/978-3-642-17688-3_28
- [23] Yu, H., Wang, R., Chen, J., Liu, L. and Wan, W., Saliency computation and simplification of point cloud data, 2012, pp. 1350-1353. DOI: 10.1109/ICCSNT.2012.6526171
- [24] An, G., Watanabe, T. and Kakimoto, M., Mesh simplification using hybrid saliency, in: 2016 International Conference on Cyberworlds (CW), 2016, pp. 231-234. DOI: 10.1109/CW.2016.47
- [25] Dutta, S., Banerjee, S., Biswas, P.K. and Bhowmick, P., Mesh denoising using multi-scale curvature-based saliency, in: *Computer Vision - ACCV 2014 Workshops*, 2014, pp. 507-516. DOI: 10.1007/978-3-319-16631-5_37
- [26] Jiao, X., Wu, T. and Qin, X., Mesh segmentation by combining mesh saliency with spectral clustering, *J. Comput. Appl. Math.*, 329(1), pp. 134-146, Feb. 2018. DOI: 10.1016/j.cam.2017.05.007
- [27] Tasse, F.P., Kosinka, J. and Dodgson, N., How well do saliency-based features perform for shape retrieval?, *Comput Graph*, 59(C), pp. 57-67, Oct. 2016. DOI: 10.1016/j.cag.2016.04.003
- [28] Gal, R. and Cohen-Or, D., Salient geometric features for partial shape matching and similarity, *ACM Trans Graph*, 25(1), pp. 130-150, Jan. 2006. DOI: 10.1145/1122501.1122507
- [29] Wang, W. et al., Saliency preserving slicing optimization for effective 3D printing, *Comput. Graph. Forum*, 34(6), pp. 148-160, Sep. 2015. DOI: 10.1111/cgf.12527
- [30] Li, Y., Zhou, Y., Xu, L., Yang, X. and Yang, J., Incremental sparse saliency detection, in: 2009 16th IEEE International Conference on Image Processing (ICIP), 2009, pp. 3093-3096. DOI: 10.1109/ICIP.2009.5414465
- [31] Ramirez, I. and Sapiro, G., An MDL framework for sparse coding and dictionary learning, *IEEE Trans. Signal Process.*, 60(6), pp. 2913-2927, Jun. 2012. DOI: 10.1109/TSP.2012.2187203
- [32] Lee, C.H., Varshney, A. and Jacobs, D.W., Mesh saliency, in: *ACM SIGGRAPH 2005 Papers*, New York, NY, USA, 2005, pp. 659-666. DOI: 10.1145/1186822.1073244
- [33] Leifman, G., Shtrom, E. and Tal, A., Surface regions of interest for viewpoint selection, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 414-421. DOI: 10.1109/CVPR.2012.6247703
- [34] Itti, L., Koch, C. and Niebur, E., A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(11), pp. 1254-1259, Nov. 1998. DOI: 10.1109/34.730558
- [35] Itti, L. and Koch, C., Computational modelling of visual attention, *Nat. Rev. Neurosci.*, 2(3), pp. 194-203, Mar. 2001. DOI: 10.1038/35058500
- [36] Olshausen, and Field, D.J., Sparse coding with an overcomplete basis set: a strategy employed by V1?, *Vision Res.*, 37(23), pp. 3311-3325, Dec. 1997. DOI: 10.1016/S0042-6989(97)00169-7. DOI: 10.1016/S0042-6989(97)00169-7
- [37] Bao, C., Ji, H., Quan, Y. and Shen, Z., Dictionary learning for sparse coding: algorithms and convergence analysis, *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(7), pp. 1356-1369, Jul. 2016. DOI: 10.1109/TPAMI.2015.2487966
- [38] Rissanen, J., Modeling by shortest data description, *Automatica*, 14(5), pp. 465-471, Sep. 1978. DOI: 10.1016/0005-1098(78)90005-5
- [39] Bruce, N.D.B. and Tsotsos, J.K., Saliency based on information maximization, in: *Proceedings of the 18th International Conference on Neural Information Processing Systems*, Cambridge, MA, USA, 2005, pp. 155-162. DOI: 10.1167/9.3.5
- [40] Borji, A. and Itti, L., Exploiting local and global patch rarities for saliency detection, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 478-485. DOI: 10.1109/CVPR.2012.6247711
- [41] Chen, X., Sapiro, A., Pang B. and Funkhouser, T., Schelling points on 3D surface meshes, *ACM Trans Graph*, 31(4), pp. 29:1-29:12, Jul. 2012. DOI: 10.1145/2185520.2185525
- [42] Mitra, N.J. and Nguyen, A., Estimating surface normals in noisy point cloud data, in: *Proceedings of the Nineteenth Annual Symposium on Computational Geometry*, New York, NY, USA, 2003, pp. 322-328. DOI: 10.1145/777792.777840
- [43] Tasse, F.P., Kosinka, J. and Dodgson, N.A., Quantitative analysis of saliency models, in: *SIGGRAPH ASIA 2016 Technical Briefs*, New York, NY, USA, 2016, pp. 19:1-19:4. DOI: 10.1145/3005358.3005380

E. Leal received a BSc. Eng in Systems Engineering in 2000 from Universidad de Antioquia, Colombia, an MSc. in Systems Engineering in 2006, from Universidad Nacional de Colombia in Medellín, Colombia. Currently he is a full professor at the Faculty of Engineering, Universidad Autonoma del Caribe, Barranquilla, Colombia. His research interests include: digital image processing, 3D surface reconstruction, computer vision and computational intelligence techniques.

ORCID: 0000-0003-2468-370X

G. Sanchez-Torres received a BSc. Eng in Systems Engineering in 2005 from Universidad del Magdalena, Colombia, an MSc. in Systems Engineering in 2006, and a PhD degree in Systems Engineering in 2012, both from Universidad Nacional de Colombia in Medellín, Colombia. Currently, he is a full professor at the Faculty of Engineering, Universidad del Magdalena, Santa Marta, Colombia. His research interests include: digital image processing, 3D surface reconstruction, computer vision, computational intelligence techniques and computational simulation and modeling.

ORCID: 0000-0002-9069-0732

J.W. Branch-Bedoya received a BSc. Eng in Mining and Metallurgy Engineering in 1995, an MSc. in Systems Engineering in 1997, and a PhD in Systems Engineering in 2007; all of them from Universidad Nacional de Colombia in Medellín, Colombia. Currently he is a full professor at the Computing and Decision Sciences Department, in the Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia. His research interests include: automation, computer vision, digital image processing and computational intelligence techniques.

ORCID: 0000-0002-0378-028X