A Unified Information-Theoretic Framework for Viewpoint Selection and Mesh Saliency

MIQUEL FEIXAS, MATEU SBERT and FRANCISCO GONZÁLEZ
University of Girona

Viewpoint selection is an emerging area in computer graphics with applications in fields such as scene exploration, image-based modeling, and volume visualization. In particular, best view selection algorithms are used to obtain the minimum number of views (or images) in order to understand or model an object or scene better. In this paper, we present a unified framework for viewpoint selection and mesh saliency based on the definition of an information channel between a set of viewpoints (input) and the set of polygons of an object (output). The mutual information of this channel is shown to be a powerful tool to deal with viewpoint selection, viewpoint stability, object exploration and viewpoint-based saliency. In addition, viewpoint mutual information is extended using saliency as an importance factor, showing how perceptual criteria can be incorporated to our method. Although we use a sphere of viewpoints around an object, our framework is also valid for any set of viewpoints in a closed scene. A number of experiments demonstrate the robustness of our approach and the good behavior of the proposed measures.

Categories and Subject Descriptors: I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism

General Terms: Algorithms, Human Factors, Experimentation

Additional Key Words and Phrases: Viewpoint selection, mesh saliency, visual perception, information theory

1. INTRODUCTION

In computer graphics, several *viewpoint quality measures* have been applied in areas such as scene understanding [Plemenos and Benayada 1996; Vázquez et al. 2001; Polonsky et al. 2005], scene exploration [Andújar et al. 2004; Sokolov et al. 2006], image-based modeling [Vázquez et al. 2003], and volume visualization [Bordoloi and Shen 2005; Takahashi et al. 2005; Viola et al. 2006]. In other areas, such as object recognition and mobile robotics, best view selection is also a fundamental task. Many works have demonstrated that the recognition process is view-dependent [Palmer et al. 1981; Bülthoff et al. 1995; Tarr et al. 1997; Blanz et al. 1999]. In [Tarr et al. 1997], the authors found that "visual recognition may be

This project has been funded in part with grant numbers TIN2004-07451-C03-01 of the Spanish Government and IST-2-004363 (GameTools: Advanced Tools for Developing Highly Realistic Computer Games) from the VIth European Framework.
Authors' address: Miquel Feixas, Mateu Sbert, and Francisco González, Universitat de Girona, Campus Montilivi, P4, 17071-Girona, Spain; e-mail: {feixas,mateu,gonzalez@ima.udg.edu}.

explained by a view-based theory in which viewpoint-specific representations encode both quantitative and qualitative features". In robotics, the Simultaneous Localization And Mapping (SLAM) problem requires that the robot decides on its own the necessary motions to construct the most accurate map possible. In [González-Baños and Latombe 2002], an algorithm is proposed to guide the robot through a series of good positions, where 'good' refers to the expected amount and quality of the information that will be revealed at each new location.

The basic question underlying the viewpoint selection study and application is "what is a 'good' scene viewpoint?" Obviously, this question does not have a simple answer. Depending on our objective, the best viewpoint can be, for instance, the most *representative* one or the most *unstable* one, i.e., the one that maximally changes when it is moved within its close neighborhood [Bordoloi and Shen 2005]. Palmer et al. [1981] and Blanz et al. [1999] have presented different experiments demonstrating that observers prefer views (called *canonical views*) that avoid occlusions and that are off-axis (such as a three-quarter viewpoint), salient (the most significant characteristics of an object are visible), stable and with a large number of visible surfaces.

Extending the work initiated in [Vázquez et al. 2001; Sbert et al. 2005], we present here a unified and robust framework to deal with viewpoint selection and mesh saliency. Given a set of viewpoints surrounding an object, we define an *information channel* between the viewpoints and the polygons of the object. From this channel, the *viewpoint mutual information* is used to obtain the best views of an object, to calculate the stability of a viewpoint, and to guide the object exploration. Then, we invert the channel and we compute both the information and the saliency associated with each polygon. Finally, this polygonal saliency is used to calculate how salient a viewpoint is and it is incorporated into viewpoint mutual information to drive the viewpoint selection. Our framework is also applicable to any set of viewpoints in a closed scene and, although only the geometric properties of an object have been considered, other aspects such as lighting could be incorporated.

The main contributions of this paper can be summarized as follows. First, an information channel between the set of viewpoints and the polygons of the object is defined (Section 3). Second, a new viewpoint quality measure based on mutual information is introduced (Section 3) and some of its fundamental properties are used to deal with viewpoint similarity and stability (Section 4). Third, a new best view selection algorithm, also used for viewpoint clustering and object exploration, is presented (Section 5). Fourth, the information and the saliency associated with each polygon are defined from the inverted viewpoint channel. From the polygonal saliency, the viewpoint saliency is also computed (Section 6). Fifth, the viewpoint quality measure is extended by incorporating the saliency as an importance factor (Section 7).

2. BACKGROUND

In this section we review some basic concepts of information theory (see [Cover and Thomas 1991]) and related work.

2.1 Information-Theoretic Concepts

Let $\mathcal{X}$ be a finite set, let X be a random variable taking values x in $\mathcal{X}$ with distribution $p(x) = Pr[X = x]$. Likewise, let Y be a random variable taking values y in $\mathcal{Y}$. An information channel between two random variables (input X and output Y) is characterized by a *probability transition matrix* (composed of conditional probabilities) which determines the output distribution given the input.

The *Shannon entropy* $H(X)$ of a random variable X is defined by

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x). \tag{1}$$

It is also denoted by $H(p)$ and measures the average uncertainty of a random variable X. All logarithms are base 2 and entropy is expressed in bits. The convention that $0\log 0 = 0$ is used. The *conditional entropy* is defined by

$$H(Y|X) = -\sum_{x\in\mathcal{X}} p(x)\sum_{y\in\mathcal{Y}} p(y|x)\log p(y|x), \tag{2}$$

where $p(y|x) = Pr[Y = y|X = x]$ is the conditional probability. The conditional entropy $H(Y|X)$ measures the average uncertainty associated with Y if we know the outcome of X. In general, $H(Y|X) \neq H(X|Y)$, and $H(X) \geq H(X|Y) \geq 0$.

The *Mutual Information* (MI) between X and Y is defined by

$$I(X,Y) = H(X) - H(X|Y) = \sum_{x\in\mathcal{X}} p(x)\sum_{y\in\mathcal{Y}} p(y|x)\log\frac{p(y|x)}{p(y)}. \tag{3}$$

It is a measure of the shared information between X and Y. It can be seen that $I(X,Y) = I(Y,X) \geq 0$. A fundamental property of MI is given by the *data processing inequality* which can be expressed in the following way: if $X \to Y \to Z$ is a Markov chain, i.e., $p(x,y,z) = p(x)p(y|x)p(z|y)$, then

$$I(X,Y) \geq I(X,Z). \tag{4}$$

This result demonstrates that no processing of Y, deterministic or random, can increase the information that Y contains about X.

The *relative entropy* or *Kullback-Leibler distance* between two probability distributions $p = \{p(x)\}$ and $q = \{q(x)\}$ defined over $\mathcal{X}$ is given by

$$KL(p|q) = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)}, \tag{5}$$

where, from continuity, we use the convention that $0\log 0 = 0$, $p(x)\log\frac{p(x)}{0} = \infty$ if $p(x) > 0$, and $0\log\frac{0}{0} = 0$. The relative entropy $KL(p|q)$ is a divergence measure between the *true* probability distribution p and the *target* probability distribution q. It can be proved that $KL(p|q) \geq 0$.

A convex function f on the interval $[a,b]$ fulfils the Jensen inequality: $\sum_{i=1}^{n}\lambda_i f(x_i) - f\left(\sum_{i=1}^{n}\lambda_i x_i\right) \geq 0$, where $0 \leq \lambda \leq 1$, $\sum_{i=1}^{n}\lambda_i = 1$, and $x_i \in [a,b]$. For a concave function, the inequality is reversed. If f is substituted by the Shannon entropy, which is a concave function, we obtain the *Jensen-Shannon*

inequality [Burbea and Rao 1982]:

$$JS(\pi_1, \pi_2, \ldots, \pi_N; p_1, p_2, \ldots, p_N) \equiv H\left(\sum_{i=1}^{N} \pi_i p_i\right) - \sum_{i=1}^{N} \pi_i H(p_i) \geq 0, \qquad (6)$$

where $JS(\pi_1, \pi_2, \ldots, \pi_N; p_1, p_2, \ldots, p_N)$ is the *Jensen-Shannon divergence* of probability distributions $p_1, p_2, \ldots, p_N$ with prior probabilities or weights $\pi_1, \pi_2, \ldots, \pi_N$, fulfilling $\sum_{i=1}^{N} \pi_i = 1$. The JS-divergence measures how 'far' are the probabilities p_i from their likely joint source $\sum_{i=1}^{N} \pi_i p_i$ and equals zero if and only if all the p_i are equal. It is important to note that the JS-divergence is identical to $I(X, Y)$ when $\pi_i = p(x_i)$ and $p_i = p(Y|x_i)$ for each $x_i \in \mathcal{X}$, where $p(X) = \{p(x_i)\}$ is the input distribution, $p(Y|x_i) = \{p(y_1|x_i), p(y_2|x_i), \ldots, p(y_M|x_i)\}$, $N = |\mathcal{X}|$, and $M = |\mathcal{Y}|$ [Burbea and Rao 1982; Slonim and Tishby 2000].

2.2 Related Work

We review now some viewpoint quality measures for polygonal models. In [Plemenos and Benayada 1996], the quality of a viewpoint v of a scene is computed using the *Heuristic Measure* (HM) given by

$$C(v) = \frac{\sum_{i=1}^{n} \left\lceil \frac{P_i(v)}{P_i(v)+1} \right\rceil}{n} + \frac{\sum_{i=1}^{n} P_i(v)}{r} \ , \qquad (7)$$

where $P_i(v)$ is the number of pixels corresponding to the polygon i in the image obtained from the viewpoint v, r is the total number of pixels of the image (resolution of the image), and n is the total number of polygons of the scene. In this formula, $\lceil x \rceil$ denotes the smallest integer, greater than or equal to x. The first term in (7) gives the fraction of visible surfaces with respect to the total number of surfaces, while the second term is the ratio between the projected area of the scene (or object) and the screen area (thus, its value is 1 for a closed scene).

From (1), the *Viewpoint Entropy* (VE) [Vázquez et al. 2001] has been defined from the relative area of the projected polygons over the sphere of directions centered at viewpoint v. Thus, the viewpoint entropy was defined by

$$H_v = -\sum_{i=0}^{N_f} \frac{a_i}{a_t} \log \frac{a_i}{a_t}, \qquad (8)$$

where N_f is the number of polygons of the scene, a_i is the projected area of polygon i over the sphere, a_0 represents the projected area of background in open scenes, and $a_t = \sum_{i=0}^{N_f} a_i$ is the total area of the sphere. The maximum entropy is obtained when a certain viewpoint can see all the polygons with the same projected area. The best viewpoint is defined as the one that has maximum entropy. In molecular visualization, both maximum and minimum entropy views show relevant characteristics of a molecule [Vázquez et al. 2006].

From (5), a new viewpoint quality measure, called *Viewpoint Kullback-Leibler distance* (VKL) [Sbert et al. 2005], has been defined by

$$KL_v = \sum_{i=1}^{N_f} \frac{a_i}{a_t} \log \frac{\frac{a_i}{a_t}}{\frac{A_i}{A_T}}, \qquad (9)$$

where a_i is the projected area of polygon i, $a_t = \sum_{i=1}^{N_f} a_i$, A_i is the actual area of polygon i and $A_T = \sum_{i=1}^{N_f} A_i$ is the total area of the scene or object. The VKL measure is interpreted as the distance between the normalized distribution of projected areas and the 'ideal' projection, given by the normalized distribution of the actual areas. In this case, the background can not be taken into account. The minimum value 0 is obtained when the normalized distribution of projected areas is equal to the normalized distribution of actual areas. Thus, to select views of high quality means to minimize KL_v.

Apart from the previous references on viewpoint quality measures, Polonsky et al. [2005] describe a number of different ways to measure the goodness of a view of an object. After analyzing different view descriptors, they conclude that no single descriptor does a perfect job and possibly a combination of them would amplify the advantage that each one has. Given a sphere of viewpoints, Yamauchi et al. [2006] compute the similarity between each two disjoint views using Zernike moments analysis and obtain a similarity weighted spherical graph. A view is considered to be stable if all edges incident on its viewpoint in the spherical graph have high similarity weights. Andújar et al. [2004] and Sokolov et al. [2006] present two different exploration algorithms guided by viewpoint entropy and the total curvature of a visible surface, respectively. In the volume rendering field, Bordoloi and Shen [2005] and Takahashi et al. [2005] use an extended version of viewpoint entropy and Viola et al. [2006] introduce the viewpoint mutual information. Castelló et al. [2007] use viewpoint entropy as a perceptual measure for mesh simplification.

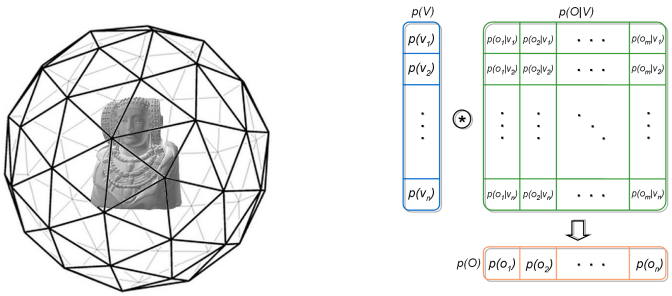
Based on the investigation on canonical views, Gooch et al. [2001] present a new method for constructing images, where the viewpoint is chosen to be off-axis, and Lu et al. [2006] obtain the viewing direction from the combination of factors such as saliency, occlusion, stability and familiarity. Lee et al. [2005] have introduced the saliency as a measure for regional importance for graphics meshes and Kim and Varshney [2006] presented a visual-saliency-based operator to enhance selected regions of a volume. Gal and Cohen-Or [2006] introduced a method for partial matching of surfaces by using the abstraction of salient geometric features and a method to construct them.

3. VIEWPOINT CHANNEL

In this section, we introduce an information channel between a set of viewpoints and the set of polygons of an object to deal with viewpoint selection. Then we define the viewpoint mutual information to select the most representative views of an object. At the end of this section we compare the behavior of that measure with the ones reviewed in Section 2.

3.1 Viewpoint Mutual Information

Our viewpoint selection framework is constructed from an information channel $V \rightarrow O$ between the random variables V (input) and O (output), which represent, respectively, a set of viewpoints and the set of polygons of an object (see Figure 1(a)). This channel, which we call *viewpoint channel*, is defined by a conditional probability matrix obtained from the projected areas of polygons at each viewpoint. Viewpoints will be indexed by v and polygons by o. Throughout this



(a) Viewpoint sphere. (b) Probability distributions of channel $V \rightarrow O$.

Fig. 1. Viewpoint information channel.

paper, the capital letters V and O as arguments of $p()$ will be used to denote probability distributions. For instance, while $p(v)$ will denote the probability of a single viewpoint v, $p(V)$ will represent the input distribution of the set of viewpoints.

The viewpoint channel can be interpreted as an *observation channel* where the conditional probabilities represent the probability of *seeing* a determined polygon from a given viewpoint (see Figure 1(b)). The three basic elements of this channel are:

—Conditional probability matrix $p(O|V)$, where each element $p(o|v) = \frac{a_o}{a_t}$ is defined by the normalized projected area of polygon o over the sphere of directions centered at viewpoint v (a_o is the projected area of polygon o and a_t is the total area of the sphere of directions). Conditional probabilities fulfil $\sum_{o \in \mathcal{O}} p(o|v) = 1$. In our approach, the background is not taken into account but it could be considered as another polygon.

—Input distribution $p(V)$, which represents the probability of selecting each viewpoint. In our experiments, $p(V)$ will be obtained from the normalization of the projected area of the object at each viewpoint. This can be interpreted as the probability that a random ray originated at v hits (or *sees*) the object. This assignation is consistent with the objective of selecting the viewpoints which *see* more projected area. Let us remember that this is a characteristic of a canonical view (see Section 1). The input distribution can also be interpreted as the *importance* assigned to each viewpoint v. For instance, the input distribution could be defined by $p(v) = \frac{1}{N_v}$, where N_v is the number of viewpoints and, in this case, the same importance would be assigned to each viewpoint.

—Output distribution $p(O)$, given by

$$p(o) = \sum_{v \in \mathcal{V}} p(v)p(o|v), \tag{10}$$

which represents the average projected area of polygon o, i.e., the probability of polygon o to be hit (or *seen*) by a random ray cast from the viewpoint sphere.

From the previous definitions, the *conditional entropy* (2) is given by the average

of all viewpoint entropies:

$$H(O|V) = -\sum_{v\in\mathcal{V}} p(v) \sum_{o\in\mathcal{O}} p(o|v) \log p(o|v) = \sum_{v\in\mathcal{V}} p(v) H(O|v), \tag{11}$$

where $H(O|v) = -\sum_{o\in\mathcal{O}} p(o|v) \log p(o|v)$ is the *viewpoint entropy* H_v (8) and measures the degree of uniformity of the projected area distribution at viewpoint v. Let us observe that H_v has been now rewritten in a different form. Both entropies $H(O|v)$ and $H(O|V)$ tend to infinity when polygons are infinitely refined. This makes these measures very sensitive to the discretisation of the object and, in general, not appropriate to evaluate the quality of a viewpoint.

We now devote our attention to the *mutual information* (3) between V and O, that expresses the degree of *dependence* or *correlation* between the set of viewpoints and the object. From (3), mutual information is given by

$$I(V,O) = \sum_{v\in\mathcal{V}} p(v) \sum_{o\in\mathcal{O}} p(o|v) \log \frac{p(o|v)}{p(o)} = \sum_{v\in\mathcal{V}} p(v) I(v,O), \tag{12}$$

where we define

$$I(v,O) = \sum_{o\in\mathcal{O}} p(o|v) \log \frac{p(o|v)}{p(o)} \tag{13}$$

as the *Viewpoint Mutual Information* (VMI), which gives us the degree of dependence between the viewpoint v and the set of polygons, and it is interpreted as a measure of the *quality* of viewpoint v. Consequently, mutual information $I(V,O)$ gives us the average quality of the set of viewpoints. *Quality* is considered here equivalent to *representativeness*.

In our framework, the best viewpoint is defined as the one that has *minimum* VMI. High values of the measure mean a high dependence between viewpoint v and the object, indicating a highly *coupled* view (for instance, between the viewpoint and a small number of polygons with low average visibility). On the other hand, the lowest values correspond to the most *representative* or *relevant* views, showing the maximum possible number of polygons in a balanced way.

3.2 Discussion

It is important to observe that $I(v,O) = KL(p(O|v)|p(O))$, where $p(O|v)$ is the conditional probability distribution between v and the object and $p(O)$ is the marginal probability distribution of O, which in our case corresponds to the distribution of the average of projected areas. It is worth observing that $p(O)$ plays the role of the *target* distribution in the KL distance and also the role of the *optimal* distribution since our objective is that $p(O|v)$ becomes similar to $p(O)$ to obtain the best views. On the other hand, this role agrees with intuition since $p(O)$ is the average visibility of polygon o over all viewpoints, i.e., the *mixed distribution* of all views, and we can think of $p(O)$ as representing, with a single distribution, the knowledge about the scene. Note the difference between VMI (13) and VKL (9), due to the fact that in the last case the distance is taken with respect to the actual areas.

In [Viola et al. 2006], it has been shown that the main advantage of VMI over VE is its robustness to deal with any type of discretisation or resolution of the

volumetric dataset. The same advantage can be observed for polygonal data. Thus, while a highly refined mesh will attract the attention of VE, VMI will be almost insensitive to changes in the mesh resolution. This behavior of both measures with respect to the discretisation can be deduced from the mathematical analysis of VE and VMI. For instance, let us assume that a regular polygon o of the object is subdivided into two equal parts o_1 and o_2 such that $p(o_1|v) = p(o_2|v)$, $p(o_1) = p(o_2)$, $p(o|v) = p(o_1|v) + p(o_2|v)$ and $p(o) = p(o_1) + p(o_2)$. Assuming that only the term referred to polygon o can change in the formulas of VE (8) and VMI (13), we analyze their variation after the subdivision of o. The variation of VE is given by

$$\delta H(O|v) = -p(o_1|v)\log p(o_1|v) - p(o_2|v)\log p(o_2|v) - (-p(o|v)\log p(o|v)) = p(o|v).$$

Therefore, VE increases with a value $p(o|v)$ after the subdivision. On the other hand, the variation of VMI is given by

$$\delta I(v,O) = p(o_1|v)\log\frac{p(o_1|v)}{p(o_1)} + p(o_2|v)\log\frac{p(o_2|v)}{p(o_2)} - p(o|v)\log\frac{p(o|v)}{p(o)} = 0. \quad (14)$$

Thus, VMI remains invariant to the proposed subdivision. In general, if we compare both measures for finer and finer discretisations, VMI will converge to an upper bound and VE will increase to infinite [Feixas 2002]. Note that HM is also highly dependent on the discretisation, since the first term in (7) is given by the quotient between the number of visible polygons and the total number of polygons. The behavior of all these measures with respect to the discretisation will be experimentally shown in the next section.

3.3 Results

In this section, the behavior of VMI (13) is compared with the one of HM (7), VE (8), and VKL (9). To compute these viewpoint quality measures, we need a *preprocess step* to estimate the projected area of the visible polygons of the object at each viewpoint. Before projection, a different color is assigned to each polygon. The number of pixels with a given color divided by the total number of pixels projected by the object gives us the relative area of the polygon represented by this color (conditional probability $p(o|v)$). In this paper, all measures have been computed without taking into account the background and using a projection resolution of 640×480.

In our experiments, all the objects are centered in a sphere of 642 viewpoints built from the recursive discretisation of an icosahedron and the camera is looking at the center of this sphere. Our framework could be extended to any other placement of viewpoints but the choice of a sphere of viewpoints permits us to analyze an object in an isotropic manner. Note that all the measures analyzed here are sensitive to the relative size of the viewpoint sphere with respect to the object. In this paper, the viewpoint sphere is built in the following way: first, the smallest bounding sphere of the model is obtained and, then, the viewpoint sphere adopts the same center as the bounding sphere and a radius three times the radius of the bounding sphere.

In Table I we show the number of polygons of the models used in this section and the cost of the preprocess step, i.e., the cost of computing the probability distributions $p(V)$, $p(O|V)$ and $p(O)$. Even though a large number of viewpoints

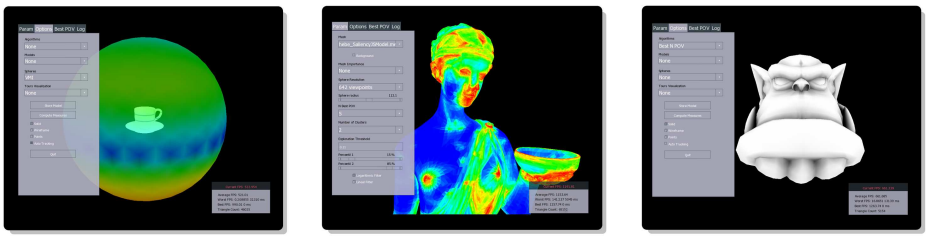


Fig. 2. The figure shows the interface of our viewpoint software.

	Cow	Coffee cup	Ship	Lady of Elche
Number of triangles	9593	43935	47365	51978
Computational cost	41 sec	81 sec	62 sec	80 sec

Table I. Number of triangles of the models used and computational cost of the preprocess step for each model.

(a) (b) (c)

Fig. 3. Cow, ship and lady of Elche wireframe models.

have been used, an acceptable quality could be achieved with less viewpoints and the consequent reduction of timings. To show the behavior of the measures, the sphere of viewpoints is represented by a color map, where red and blue colors correspond respectively to the best and worst views. Note that a good viewpoint corresponds to a high value for both HM (7) and VE (8), and to a low value for both VKL (9) and VMI (13). Figure 2 shows the interface of our viewpoint software created using the 3D-rendering engine Ogre3D (`http://www.ogre3d.org`). Our tests were run on a 3GHz machine with 2 GB RAM and an Nvidia GeForce 8800 GTX with 768 MB.

To evaluate the performance of the four viewpoint quality measures presented, five models have been used: a cow (Figure 3(a)), two coffee-cup-and-dish with two different discretisations of the dish (Figures 5(i.a) and 5(ii.a)), a ship (Figure 3(b)), and the lady of Elche (Figure 3(c)). Figure 4 has been organized as follows. Rows (i), (ii) and (iii) show, respectively, the behavior of HM, VE and VMI measures. Columns (a) and (b) show, respectively, the best and worst views, and columns (c) and (d) show two different projections of the viewpoint spheres. Figure 4 illustrates how VMI selects better views than both HM and VE. Observe how VE chooses to *see* the most highly discretised parts of the cow. The same occurs with HM,

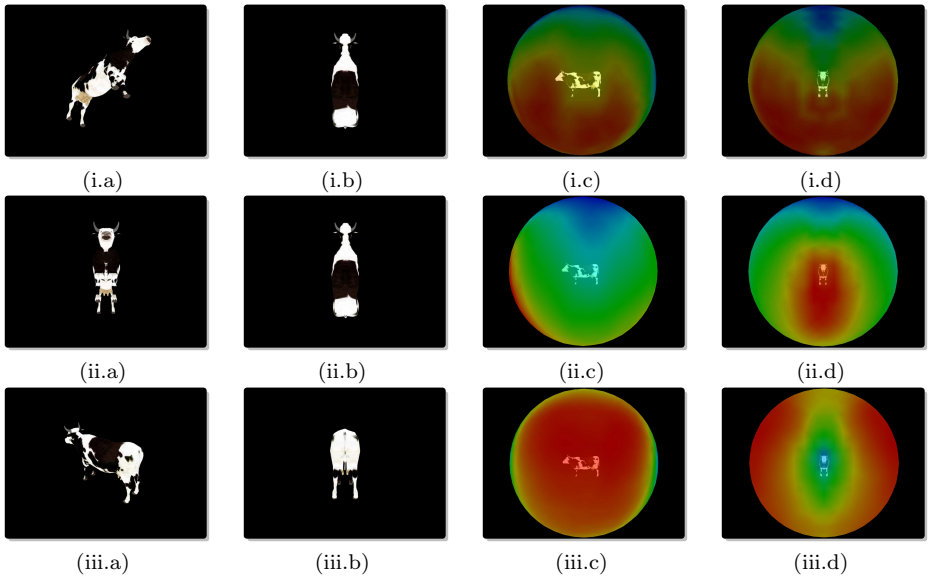


Fig. 4. (a) The most representative and (b) the most restricted views, and (c-d) the viewpoint spheres obtained respectively from the (i) HM, (ii) VE and (iii) VMI measures. Red colors on the sphere represent the highest quality views and blue colors represent the lowest quality views.

although this one also searches for a view with higher projected area. While the worst views for the HM and VE measures correspond to the ones that see the less discretised parts, in the VMI case a true restricted view is obtained.

Figure 5 shows the behavior of the HM, VE and VMI measures when the discretisation of the object varies outstandingly. Rows (i) and (ii) show the viewpoint spheres computed respectively for the coffee-cup-and-dish model of Figure 5(i.a) and for the same model with a more refined dish (Figure 5(ii.a)). We can clearly observe how the spheres obtained from HM and VE change according to the discretisation variation, whereas VMI spheres are almost insensitive to this variation.

The different behavior between VKL and VMI is shown in Figure 6. Remember that the main difference between VMI and VKL is that while the former computes the distance between the projected areas of the polygons and the average area *seen* by the set of viewpoints, the later calculates the distance with respect to the *actual* areas of polygons. Due to this fact, the reliability of VKL is outstandingly affected by the existence of many non visible or poorly visible polygons, as in the case of the ship and lady of Elche models.

4. VIEWPOINT SIMILARITY AND STABILITY

As we have mentioned in Section 1, a basic property of a canonical view is its *stability* [Blanz et al. 1999]. That is, observers prefer a view which minimally changes when it is moved within its nearest neighborhood. In this section, viewpoint stability is defined from the notion of dissimilarity between two viewpoints, which is

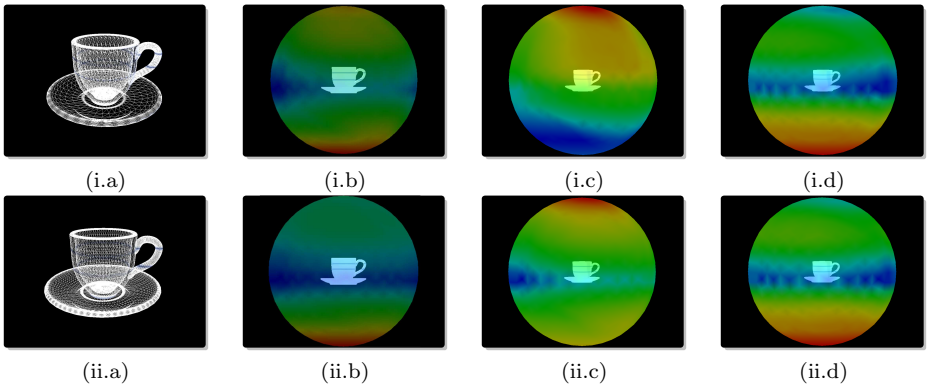


Fig. 5. Column (a) shows the models used to compute the viewpoint spheres, where the dish in (ii.a) is more refined than the one in (i.a). The viewpoint spheres are obtained respectively from the (b) HM, (c) VE and (d) VMI measures.

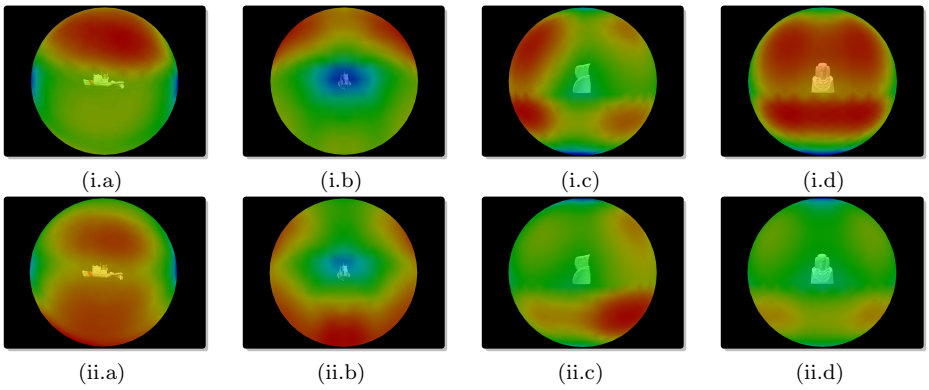


Fig. 6. Viewpoint spheres obtained respectively from the (i) VKL and (ii) VMI measures.

given by the Jensen-Shannon divergence between their respective distributions. The use of Jensen-Shannon as a measure of *view similarity* has been previously proposed by Bordoloi and Shen [2005] in the volume rendering field. In our approach, this measure appears naturally from the variation of mutual information.

If we apply the data processing inequality (4) to the channel $V \rightarrow O$, we find that any clustering over V or O, respectively denoted by $\widehat{V}$ and $\widehat{O}$, will reduce $I(V,O)$. Therefore, if neighbor viewpoints (or polygons) are clustered, then $I(\widehat{V},O) \leq I(V,O)$ (or $I(V,\widehat{O}) \leq I(V,O)$). The result of clustering (or merging) two viewpoints v_i and v_j is defined as a 'virtual' viewpoint $\widehat{v} \equiv v_i \oplus v_j$ such that

$$p(\widehat{v}) = p(v_i \oplus v_j) = p(v_i) + p(v_j) \tag{15}$$

and

$$p(o|\widehat{v}) = p(o|v_i \oplus v_j) = \frac{p(v_i)p(o|v_i) + p(v_j)p(o|v_j)}{p(\widehat{v})}. \tag{16}$$

The reduction of MI when two viewpoints v_i and v_j are merged is given by

$$\begin{aligned}\delta I &= I(V,O) - I(\widehat{V},O) \\ &= (p(v_i)I(v_i,O) + p(v_j)I(v_j,O)) - p(\widehat{v})I(\widehat{v},O) \\ &= p(\widehat{v})\left(\frac{p(v_i)}{p(\widehat{v})}I(v_i,O) + \frac{p(v_j)}{p(\widehat{v})}I(v_j,O) - I(\widehat{v},O)\right) \\ &= p(\widehat{v})D(v_i,v_j), \end{aligned} \tag{17}$$

where we define

$$D(v_i,v_j) = \frac{p(v_i)}{p(\widehat{v})}I(v_i,O) + \frac{p(v_j)}{p(\widehat{v})}I(v_j,O) - I(\widehat{v},O) \tag{18}$$

as the *viewpoint dissimilarity* between v_i and v_j. That is, the loss of information when two viewpoints are merged is interpreted from the dissimilarity between them. It can be seen that the dissimilarity will be null when the two viewpoints capture the same distribution of projected areas: if $p(O|v_i) = p(O|v_j)$, then $\delta I = 0$.

From the definition of the Jensen-Shannon divergence, it can be shown that the viewpoint dissimilarity can also be written as

$$D(v_i,v_j) = JS\left(\frac{p(v_i)}{p(\widehat{v})}, \frac{p(v_j)}{p(\widehat{v})}; p(O|v_i), p(O|v_j)\right), \tag{19}$$

where the second term is the Jensen-Shannon divergence (6) between the distributions $p(O|v_i)$ and $p(O|v_j)$ captured by v_i and v_j with weights $\frac{p(v_i)}{p(\widehat{v})}$ and $\frac{p(v_j)}{p(\widehat{v})}$, respectively (see [Slonim and Tishby 1999]) . If two views are very similar, i.e., the JS-divergence between them is small, the channel could be simplified by substituting these two viewpoints by their merging, without a significant loss of information.

Two interesting properties follow:

—It can be seen that the clustering $\widehat{V}$ of *all* viewpoints would give $\delta I = I(V,O)$ and, thus, $I(\widehat{V},O) = 0$.

—$H(O) = H(O|V) + I(V,O) = H(O|\widehat{V}) + I(\widehat{V},O)$, where $H(O)$ is the entropy of $p(O)$. Note that if two viewpoints are clustered the decrease of $I(V,O)$ is equal to the increase of $H(O|V)$ since $H(O)$ remains constant (the discretisation of the object has not been changed).

View unstability was defined by Bordoloi and Shen [2005] as the maximum change in view that occur when the camera position is shifted within a small neighborhood. Thus, a small change corresponds to a stable viewpoint and a large change to an unstable one. We now define the unstability of a viewpoint v as the average variation of dissimilarity between v and its neighbor viewpoints. That is, v_i is a stable viewpoint if $p(O|v_i)$ is close to the probability distributions $p(O|v_j)$ of its neighbors, where v_j stands for a neighbor of v_i. Thus, the *viewpoint unstability* of v_i is defined by

$$U(v_i) = \frac{1}{N_n}\sum_{j=1}^{N_n} D(v_i,v_j), \tag{20}$$

where v_j is a neighbor of v_i and N_n is the number of neighbors of v_i.

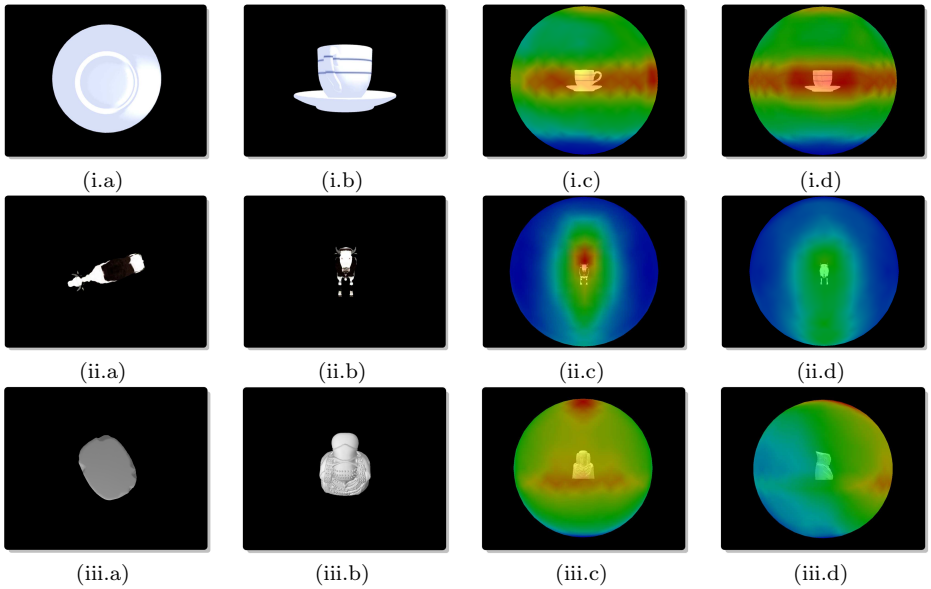


Fig. 7. The (a) most stable and (b) most unstable viewpoints, and (c-d) the unstability spheres obtained for the (i) coffee-cup-and-dish, (ii) cow and (iii) lady of Elche models. Red colors on the sphere represent high unstability values, blue colors represent low unstability values.

Figure 7 shows the behavior of the viewpoint unstability measure for the coffee-cup-and-dish, cow and lady of Elche models. Observe how the results obtained agree with intuition.

5. BEST VIEW SELECTION AND OBJECT EXPLORATION

In order to understand or model an object, we are interested in selecting a set of representative views which provides a complete representation of the object. In this section, new algorithms based on the concepts introduced in Sections 3 and 4 are applied to both the selection of the N best representative views and object exploration.

5.1 Selection of N Best Views

With the goal of obtaining the best representation of the object using the minimum number of views, a new viewpoint selection algorithm based on VMI is presented. If we look for a good set of views within the set of viewpoints, we will obtain the most representative set by selecting the views such that their mixing (merging) minimizes the distance to the target distribution $p(O)$. We consider that this mixing provide us with a *balanced* view of the object.

Thus, our selection algorithm should select the N viewpoints so that their merging $\widehat{v}$ minimizes the viewpoint mutual information $I(\widehat{v}, O)$. Due to the fact that this optimization algorithm is NP-complete, we adopt a greedy strategy by selecting successive viewpoints that minimize $I(\widehat{v}, O)$. That is, at each merging step we aim to maximize the JS-divergence between the set of previously merged viewpoints

	Coffee cup		Armadillo		Lady of Elche	
Best view	$I(\widehat{v},O)$	**Ratio**	$I(\widehat{v},O)$	**Ratio**	$I(\widehat{v},O)$	**Ratio**
a	1.471	0.730	1.791	0.850	1.355	0.703
b	0.692	0.343	0.837	0.397	0.644	0.334
c	0.346	0.172	0.616	0.292	0.458	0.237
d	0.262	0.130	0.416	0.197	0.275	0.143
e	0.207	0.103	0.310	0.147	0.219	0.113
f	0.190	0.095	0.238	0.113	0.153	0.079
Cost	36 sec		77 sec		38 sec	

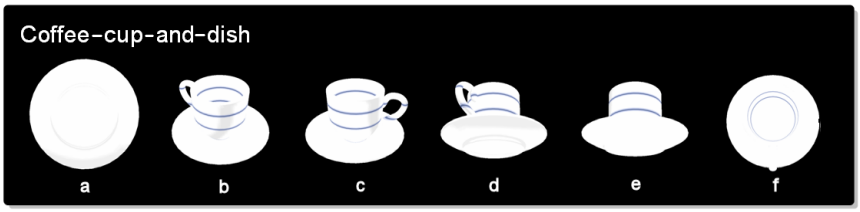
Table II. For the coffee cup, armadillo, and lady of Elche models, we show the $I(\widehat{v},O)$ values from the merging of the six best selected viewpoints (see Figure 8), the corresponding VMI-ratio, and the computation cost of selecting the six best views.

and the new viewpoint to be selected. This algorithm permits us to find in an automated and efficient way the minimal set of views which represent the object or scene.

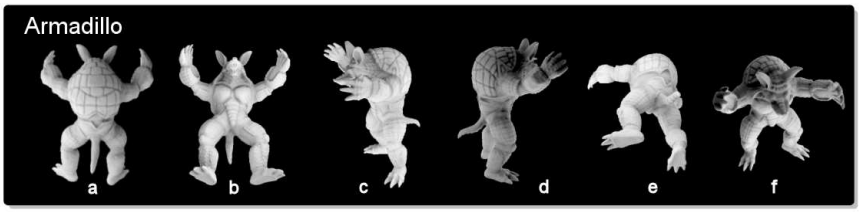
The algorithm proceeds as follows. First, we select the best viewpoint v_1 with distribution $p(O|v_1)$ corresponding to the minimum $I(v,O)$. Next, we select v_2 such that the mixed distribution $\frac{p(v_1)}{p(\widehat{v})}p(O|v_1)+\frac{p(v_2)}{p(\widehat{v})}p(O|v_2)$ will minimize $I(\widehat{v},O)$, where $\widehat{v}$ represents the clustering of v_1 and v_2 and $p(\widehat{v})=p(v_1)+p(v_2)$. At each step, a new mixed distribution $\frac{p(v_1)}{p(\widehat{v})}p(O|v_1)+\frac{p(v_2)}{p(\widehat{v})}p(O|v_2)+\ldots+\frac{p(v_n)}{p(\widehat{v})}p(O|v_n)$, where $p(\widehat{v})=p(v_1)+p(v_2)+\ldots+p(v_n)$, is produced until the VMI-ratio given by $\frac{I(\widehat{v},O)}{I(V,O)}$ is lower than a given threshold or a fixed number of views is achieved. This ratio can be interpreted as a measure of the goodness or representativeness of the selected viewpoints.

Figure 8 show the six best views obtained with our VMI-based selection algorithm for three different models. In Table II, for each new viewpoint selected we show the VMI of the clustering of selected viewpoints ($I(\widehat{v},O)$) and the corresponding VMI-ratio. For instance, to achieve a degree of representativeness given by a VMI-ratio lower than 0.15, four views are needed for the coffee-cup-and-dish and lady of Elche models, and five for the armadillo model. Table II also shows the computation cost of selecting the six best views. The behavior of our algorithm is also analyzed in Figure 9, where we observe how the $I(\widehat{v},O)$ values obtained from the successive mixed distributions converge asymptotically to zero. It is important to note that the best views for the selected models (Figure 8(a)) are not the ones our intuition would expect as more representative. This is due to the fact that, from a purely geometric approach, the best views of Figure 8 correspond to the viewpoints that their projected area distribution is more *similar* (in the Kullback-Leibler sense) to the average projected area distribution (target distribution). This problem will be tackled in the next sections, introducing perceptual criteria to select the best views.

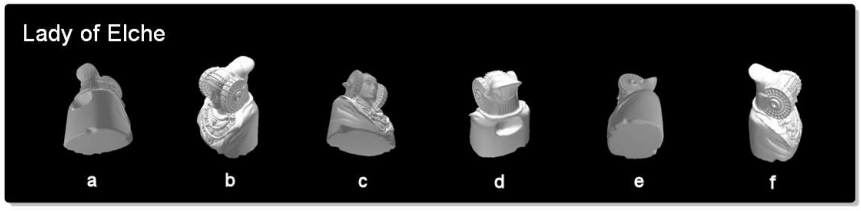
From the N best representative viewpoints, a simple greedy clustering algorithm is proposed in order to partition the sphere of viewpoints. The two main steps of this algorithm are the following. First, we select the N best viewpoints from a given VMI-ratio. These viewpoints will play the role of centroids in our algorithm. Second, each viewpoint is assigned or clustered with the 'nearest' centroid, where



(i)



(ii)



(iii)

Fig. 8. From (a) to (f), the six most representative views selected by the VMI-based algorithm for the (i) coffee-cup-and-dish, (ii) armadillo, and (iii) lady of Elche models.

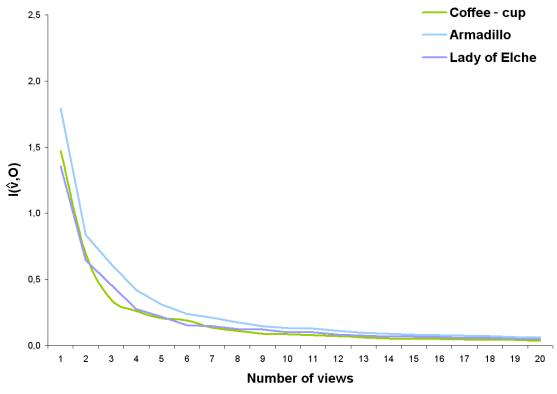
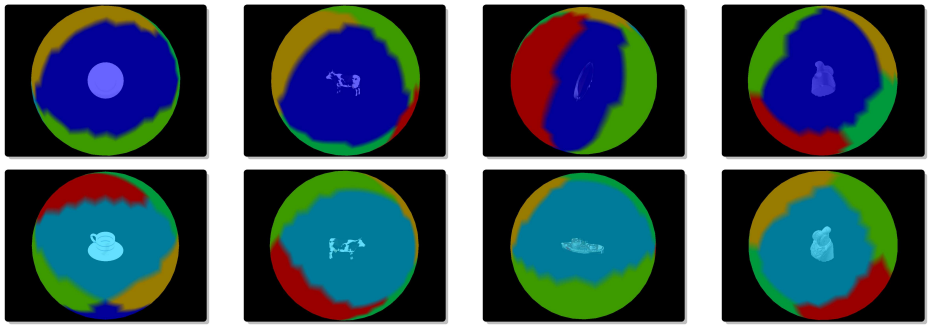


Fig. 9. For the models used in Figure 8, $I(\widehat{v}, O)$ values (vertical axis) are shown from the successive mixed distributions corresponding to the viewpoints (horizontal axis) obtained by our selection algorithm.



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

Fig. 10. Viewpoint clustering spheres with six clusters for the (a) coffee-cup-and-dish, (b) cow, (c) ship and (d) lady of Elche models.

the distance is given by the Jensen-Shannon divergence between two viewpoints. Finally, the set of viewpoints will be grouped around the N best viewpoints. In Figure 10, the behavior of this clustering algorithm is shown for the (a) coffee cup, (b) cow, (c) ship and (d) lady of Elche models.

5.2 Object Exploration

In this section, two greedy algorithms are presented to explore the object. In both cases, the best viewpoint (minimum VMI) is the starting point of the object exploration. In the first algorithm (*guided tour*), the path visits a set of N preselected best views which ensure a good exploration of the object. In the second algorithm (*exploratory tour*), the successive viewpoints are selected using the maximum *novelty* criterion with respect to the parts seen of the object. In `http://www.gametools.org/viewpoint/index.html`, several videos show the performance of these methods.

Guided tour. First, we obtain the list of the N best viewpoints. Then, the algorithm starts at the best viewpoint and visits all the other best viewpoints as follows. From the best viewpoint, we find the *nearest* (minimum JS-divergence) best viewpoint in the list. This is now the *target* viewpoint. Thus, from the best viewpoint, successive neighbor viewpoints will be selected so that, without any viewpoint repetition, their *distance* to the target viewpoint is minimum. The distance between two viewpoints is always calculated from the JS-divergence. When the first target viewpoint is achieved, we select a new target one among the rest of best viewpoints in the list. Then we proceed in the same way until the last best view is reached or the cycle is completed arriving at the initial best viewpoint. Our algorithm, being a greedy one, is fast but it can cause a small detour over the minimum path. Figure 11(i) shows the exploration of the coffee-cup-and-dish and the lady of Elche models from the six best views obtained in each case (the blue, green and red light points correspond to the starting, intermediate and ending viewpoints, respectively). Two different projections of the sphere are shown to see the path better.

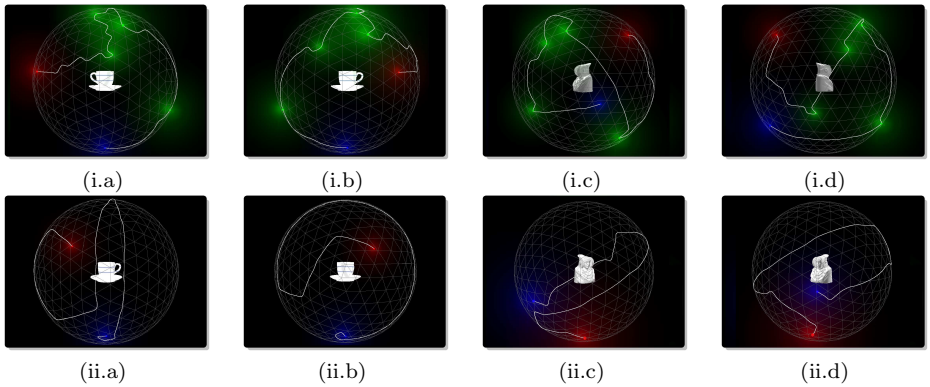


Fig. 11. (i) Guided and (ii) exploratory tours around the coffee-cup-and-dish and lady of Elche models, respectively.

Exploratory tour. From [Itti and Baldi 2006], we know that maximum novelty or surprise attracts the attention of an observer. Following this principle, the algorithm selects the best viewpoint and then successively visits the (non-visited) neighbor viewpoints that minimize the $I(\widehat{v}, O)$ of all visited viewpoints. This means that at each step we select the viewpoint that maximizes its JS-divergence with respect to all visited viewpoints and, consequently, the most dissimilar (*surprising*) viewpoint is selected. This procedure stops if the VMI-ratio is lower than a given threshold. Figure 11(ii) shows the result of the exploration of the coffee-cup-and-dish and the lady of Elche models.

6. VIEW-BASED POLYGONAL INFORMATION AND SALIENCY

As we have seen in Section 3, the information associated with each viewpoint has been obtained from the definition of the channel between the sphere of viewpoints and the polygons of the object. Now, the *information associated with a polygon* will be defined as the contribution of this polygon to the MI of that channel. To illustrate this new approach, the inverted channel $O \rightarrow V$ is considered, so that O is the input and V the output.

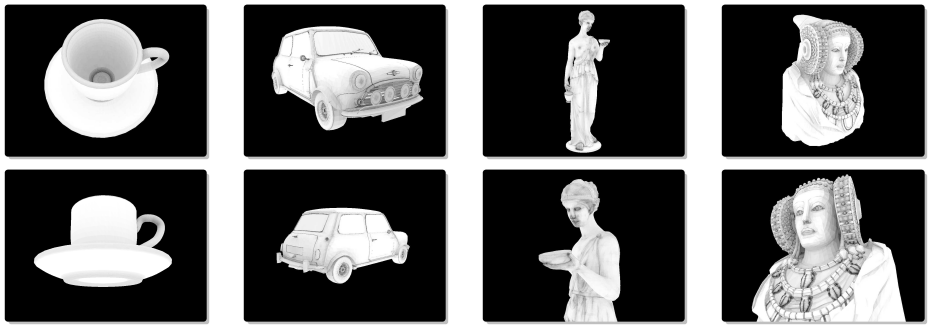
6.1 View-based Polygonal Information

From the Bayes theorem $p(v,o) = p(v)p(o|v) = p(o)p(v|o)$, the mutual information (12) can be rewritten as

$$\begin{aligned} I(O,V) &= \sum_{o\in\mathcal{O}} p(o) \sum_{v\in\mathcal{V}} p(v|o) \log \frac{p(v|o)}{p(v)} \\ &= \sum_{o\in\mathcal{O}} p(o) I(o,V), \end{aligned} \tag{21}$$

where we define

$$I(o,V) = \sum_{v\in\mathcal{V}} p(v|o) \log \frac{p(v|o)}{p(v)} \tag{22}$$



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

Fig. 12. View-based polygonal information for the (a) coffee-cup-and-dish, (b) mini, (c) Hebe and (d) lady of Elche models.

Fig. 13. The figure shows the application of polygonal information as ambient occlusion.

as the *Polygonal Mutual Information* (PMI), which represents the degree of correlation between the polygon o and the set of viewpoints, and can be interpreted as the *information* associated with polygon o. Analogous to VMI, low values of PMI correspond to polygons that 'see' the maximum number of viewpoints in a balanced way, i.e., $p(V|o)$ is close to $p(V)$. The opposite happens for high values. Let us remind that MI is invariant to the inversion of the channel since $I(V,O) = I(O,V)$. As the discrete mutual information converges to the upper bound represented by the continuous mutual information for a finer and finer discretisation (see [Feixas 2002]), the average of PMI values will converge towards the corresponding limiting value. This makes PMI very robust against discretisation. Note that if the model changes with each discretisation level the upper bound property will not be guaranteed.

In Figure 12, we show the information maps of (i) the coffee-cup-and-dish, (ii) mini, (iii) Hebe and (iv) lady of Elche models. To obtain these images, the PMI has been normalized between 0 and 1 and subtracted from 1. Thus, low values of PMI, corresponding to non-occluded or visible (from many viewpoints) polygons, are represented by values near 1 in the grey-map, while high values of PMI, corresponding to occluded polygons, are represented by values near 0 in the grey-map. In Figure 12 we show the polygonal information values computed from the center of each polygon, while in Figure 13 these values have been linearly interpolated at the vertexes of the polygons. Observe that these maps look as an ambient occlusion

or obscurance map (see [Landis 2002; Christensen ; Zhukov et al. 1998; Iones et al. 2003]). In Figure 13 we show one example of the use of polygonal information as ambient occlusion, where this is added to a textured model.

6.2 View-based Mesh Saliency

Itti et al. [1998] maintain that visual attention is saliency-dependent and use a saliency map to represent the conspicuity or saliency at every location in the visual field by a scalar quantity and to guide the selection of attended locations. In [Lee et al. 2005], mesh saliency is captured from surface curvatures and is considered as a perception-inspired measure of *regional importance* and has been used in graphics applications such as mesh simplification and viewpoint selection. We now propose a new definition of mesh saliency based on PMI.

Analogous to the view unstability (Section 4), defined from the dissimilarity between two views, we now define the view-based mesh saliency from the dissimilarity between two polygons, which is given by the variation of mutual information when two polygons are clustered. In this approach, mesh saliency is formulated in terms of how the polygons 'see' the set of viewpoints. Thus, following the same scheme developed in Section 4, the saliency of a polygon is defined as the average dissimilarity between this polygon and its neighbors.

Similarly to (17), the reduction of MI when two polygons o_i and o_j are clustered is given by

$$\begin{aligned}
\delta I &= I(O,V) - I(\widehat{O},V) \\
&= (p(o_i)I(o_i,V) + p(o_j)I(o_j,V)) - p(\widehat{o})I(\widehat{o},V) \\
&= p(\widehat{o})\left(\frac{p(o_i)}{p(\widehat{o})}I(o_i,V) + \frac{p(o_j)}{p(\widehat{o})}I(o_j,V) - I(\widehat{o},V)\right) \\
&= p(\widehat{o})D(o_i,o_j),
\end{aligned} \tag{23}$$

where $\widehat{o} = o_i \oplus o_j$ is the result of clustering o_i and o_j and the *polygonal dissimilarity* between o_i and o_j is defined by

$$D(o_i,o_j) = \frac{p(o_i)}{p(\widehat{o})}I(o_i,V) + \frac{p(o_j)}{p(\widehat{o})}I(o_j,V) - I(\widehat{o},V). \tag{24}$$

This dissimilarity measure can also be written as

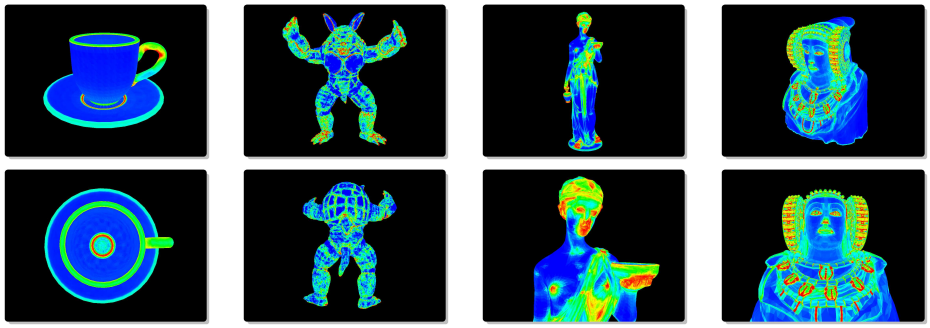
$$D(o_i,o_j) = JS\left(\frac{p(o_i)}{p(\widehat{o})}, \frac{p(o_j)}{p(\widehat{o})}; p(V|o_i), p(V|o_j)\right), \tag{25}$$

where the second term is the Jensen-Shannon divergence (6) between $p(V|o_i)$ and $p(V|o_j)$ with weights $\frac{p(o_i)}{p(\widehat{o})}$ and $\frac{p(o_j)}{p(\widehat{o})}$, respectively. Hence, two polygons are 'similar' when the JS-divergence between them is small.

Some interesting properties follow:

—If two polygons are very 'similar', their clustering involves a small loss of mutual information. If $p(V|o_i) = p(V|o_j)$, then $\delta I = 0$.

—It can be easily seen that the clustering $\widehat{O}$ of *all* polygons would give $\delta I = I(V,O)$ and, thus, $I(\widehat{O},V) = 0$.



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

Fig. 14. Mesh saliency for the (a) coffee-cup-and-dish, (b) armadillo, (c) Hebe, and (d) lady of Elche models.

—$H(V) = H(V|O) + I(O,V) = H(V|\widehat{O}) + I(\widehat{O},V)$, where $H(V)$ is the entropy of $p(V)$. That is, if two polygons are clustered the reduction of $I(O,V)$ is equal to the increase of $H(V|O)$ since $H(V)$ remains constant (the input distribution of V is not changed).

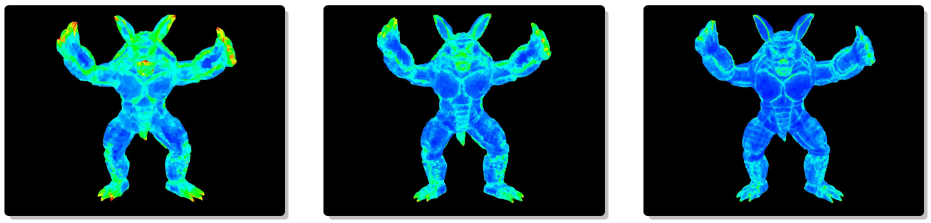
Similarly to the unstability of a viewpoint (20), the *polygonal saliency* of o_i is defined by

$$S(o_i) = \frac{1}{N_o} \sum_{j=1}^{N_o} D(o_i, o_j) \geq 0, \tag{26}$$

where o_j is a neighbor polygon of o_i and N_o is the number of neighbor polygons of o_i. Thus, a polygon o will be salient if the average of JS-divergences between o and its neighbors is high. On the other hand, a polygon at the center of a smooth region will have probably low saliency since the polygons of this region will present small visibility differences with respect to the set of viewpoints. Figure 14 shows the behavior of our saliency measure. The most salient parts are represented in red and the least salient ones in blue. For instance, the handle of the coffee cup and the nose, mouth and eyes of the other models are the most salient surfaces.

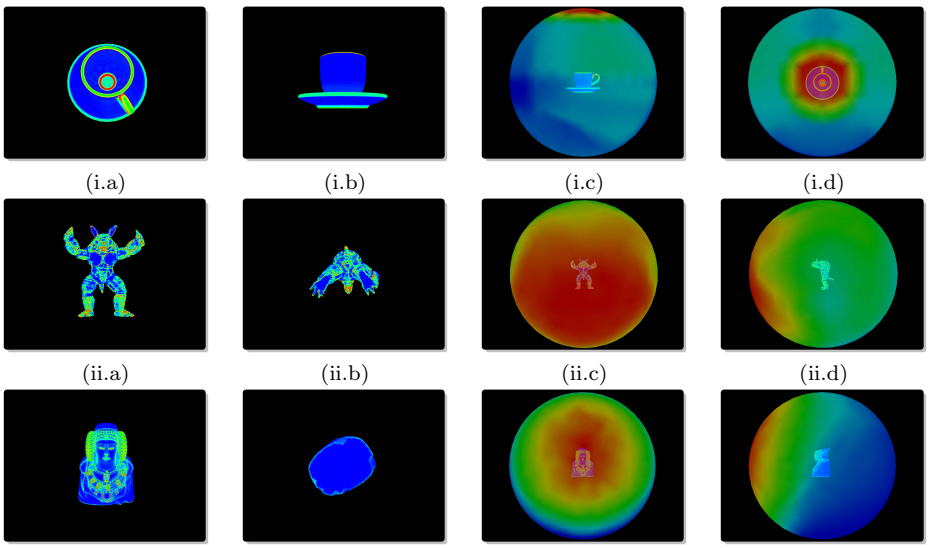
While the average of PMI values converges to a theoretical upper bound for a finer and finer discretisation, polygonal saliency will decrease. This makes saliency dependent on the discretisation. In Figure 15 we illustrate how the polygonal saliency depends on the mesh discretisation level. For comparison purposes, the saliency maps of the armadillo have been represented using the same color range and a projection resolution of 1280×960. The armadillo model has been simplified using the QSlim simplification software (`http://graphics.cs.uiuc.edu/~garland/software/qslim.html`).

Similarly to [Lee et al. 2005], where mesh saliency was used to select the best views, we propose a method to calculate the saliency of a viewpoint. Up to now we have calculated the saliency of a polygon, however we can convey this information to the sphere of viewpoints, using the conditional probabilities of the inverse channel.



(a) (b) (c)

Fig. 15. Mesh saliency maps for different resolutions of the armadillo: (a) 7500, (b) 15000 and (c) 60000 triangles.



(i.a) (i.b) (i.c) (i.d)

(ii.a) (ii.b) (ii.c) (ii.d)

(iii.a) (iii.b) (iii.c) (iii.d)

Fig. 16. The (a) most salient and (b) least salient views, and (c-d) saliency spheres obtained for the (i) coffee-cup-and-dish, (ii) Hebe and (iii) lady of Elche models. Red colors on the sphere represent high saliency values, blue colors represent low saliency values.

Hence, the *viewpoint saliency* is defined by

$$S(v) = \sum_{o \in \mathcal{O}} S(o)p(v|o). \tag{27}$$

Figure 16 shows the viewpoint saliency for the coffee-cup-and-dish, armadillo and lady of Elche models. Columns (a) and (b) illustrate the most salient view and the least one, respectively. Columns (c) and (d) show two different projections of the corresponding saliency spheres. Observe how the most salient views show us the most salient parts of each object.

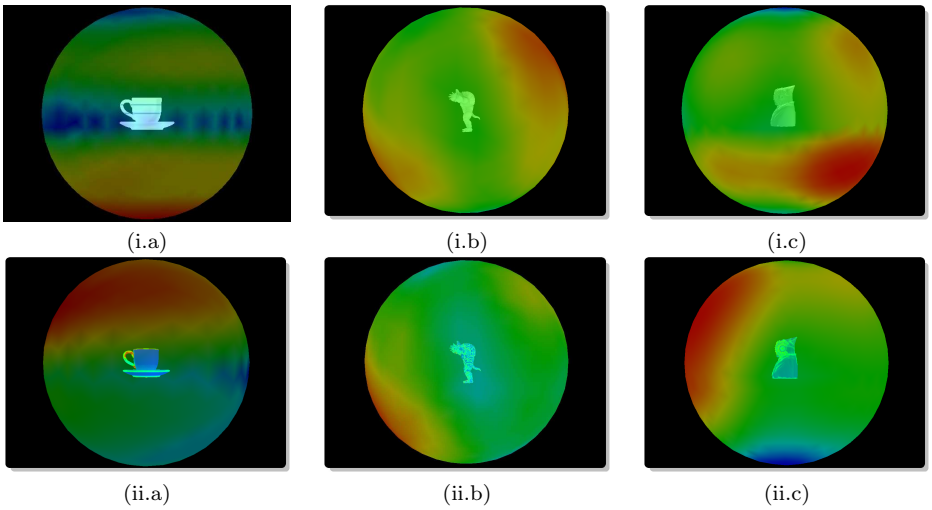


Fig. 17. (i) VMI and (ii) saliency-based EVMI spheres for the (a) coffee-cup-and-dish, (b) armadillo and (c) lady of Elche models.

7. IMPORTANCE-DRIVEN VIEWPOINT SELECTION

As we have mentioned in Section 1, it is desirable that a canonical view of an object shows its most salient parts and also the largest number of visible surfaces [Palmer et al. 1981; Blanz et al. 1999]. However, the viewpoint quality measure VMI only takes into account the geometric relation between the object and the set of viewpoints. Therefore, we can not expect that in general the best VMI-based views fulfill the desired properties for a canonical view. This fact motivates us to investigate how perceptual criteria such as saliency can be introduced into our viewpoint selection quality measure in order to improve the automatic selection of good views.

In the previous section we have presented a method to compute how salient is a viewpoint, but we aim now to incorporate the polygonal saliency to the viewpoint mutual information in order to take into account different factors concerning, respectively, the amount of projected area, the geometric representativeness and the saliency of a polygon. First, we demonstrate how the importance can be introduced into the object space by modifying directly the target distribution $p(O)$. Second, we show the results obtained by the use of the polygonal saliency as an importance factor in the viewpoint mutual information measure.

Due to the fact that VMI represents the distance between the projected visibility distribution $p(O|v)$ at viewpoint v and the target distribution $p(O)$, VMI can be extended by weighting the target distribution with an importance factor. Thus, adding importance to our scheme means simply weighting the original target distribution by an importance factor in order to obtain the new target distribution. The optimal viewpoint would be the one viewing every polygon proportional to its average projected area multiplied by its importance. Hence, the *Extended Viewpoint*

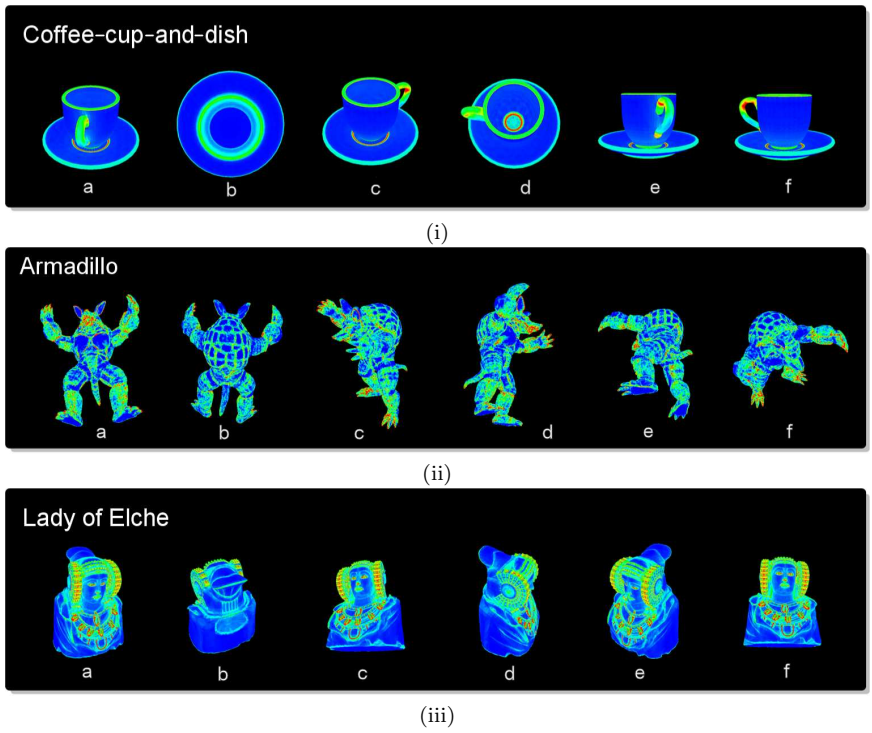


Fig. 18. The six most representative views for the (i) coffee-cup-and-dish, (ii) armadillo and (iii) lady of Elche models using the saliency-based EVMI measure.

Mutual Information (EVMI) is given by

$$I'(v, O) = \sum_{o \in \mathcal{O}} p(o|v) \log \frac{p(o|v)}{p'(o)}, \tag{28}$$

where

$$p'(o) = \frac{p(o)i(o)}{\sum_{o \in \mathcal{O}} p(o)i(o)} \tag{29}$$

and $i(o)$ is the importance of polygon o. In the experiments of this section, $i(o)$ has been substituted by the polygonal saliency $S(o)$. We follow the convention that if $S(o) = 0$ then polygon o is not taken into account. Other features, such as illumination, could be introduced as importance factors in the EVMI. In [Viola et al. 2006], the object importance has been used to calculate the best views for a volumetric dataset.

The effects of incorporating saliency in our viewpoint selection framework are depicted in Figures 17 and 18, which show for the coffee-cup-and-dish, armadillo and lady of Elche models the saliency-based EVMI spheres and the six most representative views, obtained with the best view selection algorithm (Section 5.1) using saliency-based EVMI. The saliency-based EVMI spheres of Figure 17(ii) show the perceptual improvement obtained with respect to the corresponding VMI spheres

(Figure 17(i)). For instance, whereas the VMI-based best view of the coffee-cup-and-dish shows the bottom of the dish (Figure 8(i.a)), the best view based on EVMI shows a lateral view of the coffee-cup-and-dish (Figure 18(i.a)) which is perceptually much better than the one of Figure 8(i.a). Similarly, the same conclusion can be obtained for the armadillo and the lady of Elche (see the respective best views shown in Figure 8 and Figure 18).

8. CONCLUSIONS AND FUTURE WORK

We have defined a unified framework for viewpoint selection and mesh saliency based on an information channel between a set of viewpoints and the set of polygons of an object. A new viewpoint quality measure, the viewpoint mutual information, has been introduced to quantify the representativeness of a view and has been used to compute viewpoint stability, to select the N best views and to explore the object. From the inversion of the information channel, we have defined both the information and the saliency associated with each polygon, and we have also calculated the saliency of a viewpoint. Finally, the viewpoint mutual information has been extended by incorporating the saliency as an importance factor. Many experiments have demonstrated the robustness of our approach and the good behavior of the proposed measures. There are many research issues that are worth exploring. First, we plan to extend our viewpoint framework to indoor scenes. Second, we wish to analyze how our saliency approach can guide the object exploration. Third, we will investigate the incorporation of importance (for example, obtained from saliency or lighting) to the input distribution. Fourth, we will explore how the best viewpoint selection can be fine-tuned with other perceptual characteristics such as stability. Finally, the mutual information of the viewpoint information channel can also be interpreted as the viewpoint-based shape complexity and could be used for object recognition tasks.

REFERENCES

Andújar, C., Vázquez, P. P., and Fairén, M. 2004. Way-finder: guided tours through complex walkthrough models. *Computer Graphics Forum (Eurographics 2004) 23,* 3, 499–508.

Blanz, V., Tarr, M., and Bülthoff, H. 1999. What object attributes determine canonical views? *Perception 28*, 575–599.

Bordoloi, U. D. and Shen, H.-W. 2005. Viewpoint evaluation for volume rendering. In *IEEE Visualization 2005*. 487–494.

Bülthoff, H., Edelman, S., and Tarr, M. 1995. How are three-dimensional objects represented in the brain? *Cerebral Cortex 5*, 247–260.

Burbea, J. and Rao, C. R. 1982. On the convexity of some divergence measures based on entropy functions. *IEEE Transactions on Information Theory 28,* 3, 489–495.

Castelló, P., Sbert, M., Chover, M., and Feixas, M. 2007. Viewpoint entropy-driven simplification. In *Proceedings of WSCG 2007*. 249–256.

Christensen, P. Ambient occlusion, image-based illumination and global illumination. *Photorealistic RenderMan Application Notes, Note 35*.

Cover, T. M. and Thomas, J. A. 1991. *Elements of Information Theory*. Wiley Series in Telecommunications.

Feixas, M. 2002. An information-theory framework for the study of the complexity of visibility and radiosity in a scene. Ph.D. thesis, Universitat Politècnica de Catalunya, Barcelona, Spain.

Gal, R. and Cohen-Or, D. 2006. Salient geometric features for partial shape matching and similarity. *ACM Transactions on Graphics 25,* 1, 130–150.

González-Baños, H. H. and Latombe, J.-C. 2002. Navigation strategies for exploring indoor environments. *The International Journal of Robotics Research 21,* 10-11, 829–848.

Gooch, B., Reinhard, E., Moulding, C., and Shirley, P. 2001. Artistic composition for image creation. In *Rendering Techniques*. 83–88.

Iones, A., Krupkin, A., Sbert, M., and Zhukov, S. 2003. Fast, realistic lighting for video games. *IEEE Computer Graphics and Applications 23,* 3, 54–64.

Itti, L. and Baldi, P. 2006. Bayesian surprise attracts human attention. In *Advances in Neural Information Processing Systems, Vol. 19 (NIPS 2005)*. 1–8.

Itti, L., Koch, C., and 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.

Kim, Y. and Varshney, A. 2006. Saliency-guided enhancement for volume visualization. *Transactions on Visualization and Computer Graphics 12,* 5, 925–932.

Landis, H. 2002. Renderman in production. In *Course notes of ACM SIGGRAPH*.

Lee, C. H., Varshney, A., and Jacobs, D. W. 2005. Mesh saliency. *ACM Transactions on Graphics 24,* 3, 659–666.

Lu, A., Maciejewski, R., and Ebert, D. S. 2006. Volume composition using eye tracking data. In *Proceedings of EuroVis 2006*. 655–662.

Palmer, S., Rosch, E., and Chase, P. 1981. Canonical perspective and the perception of objects. *Attention and Performance IX*, 135–151.

Plemenos, D. and Benayada, M. 1996. Intelligent display techniques in scene modelling. new techniques to automatically compute good views. In *International Conference GraphiCon'96*.

Polonsky, O., Patanè, G., Biasotti, S., Gotsman, C., and Spagnuolo, M. 2005. What's in an image? *The Visual Computer 21,* 8-10, 840–847.

Sbert, M., Plemenos, D., Feixas, M., and González, F. 2005. Viewpoint quality: Measures and applications. In *Proceedings of 1st Computational Aesthetics in Graphics, Visualization and Imaging*. 185–192.

Slonim, N. and Tishby, N. 1999. Agglomerative information bottleneck. In *NIPS*. 617–623.

Slonim, N. and Tishby, N. 2000. Document clustering using word clusters via the information bottleneck method. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM Press, 208–215.

Sokolov, D., Plemenos, D., and Tamine, K. 2006. Methods and data structures for virtual world exploration. *The Visual Computer 22,* 7, 506–516.

Takahashi, S., Fujishiro, I., Takeshima, Y., and Nishita, T. 2005. A feature-driven approach to locating optimal viewpoints for volume visualization. In *IEEE Visualization 2005*. 495–502.

Tarr, M., Bülthoff, H., Zabinski, M., and Blanz, V. 1997. To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science 8,* 4, 282–289.

Vázquez, P. P., Feixas, M., Sbert, M., and Heidrich, W. 2001. Viewpoint selection using viewpoint entropy. In *Proceedings of Vision, Modeling, and Visualization 2001*. 273–280.

Vázquez, P.-P., Feixas, M., Sbert, M., and Heidrich, W. 2003. Automatic view selection using viewpoint entropy and its applications to image-based modelling. *Computer Graphics Forum 22,* 4, 689–700.

Vázquez, P. P., Feixas, M., Sbert, M., and Llobet, A. 2006. Realtime automatic selection of good molecular views. *Computers & Graphics 30,* 1, 98–110.

Viola, I., Feixas, M., Sbert, M., and Gröller, M. E. 2006. Importance-driven focus of attention. *IEEE Transactions on Visualization and Computer Graphics 12,* 5, 933–940.

Yamauchi, H., Saleem, W., Yoshizawa, S., Karni, Z., Belyaev, A. G., and Seidel, H.-P. 2006. Towards stable and salient multi-view representation of 3d shapes. In *IEEE International Conference on Shape Modeling and Applications*. 265–270.

Zhukov, S., Iones, A., and Kronin, G. 1998. An ambient light illumination model. In *Rendering Techniques*. 45–56.

Received Month Year; revised Month Year; accepted Month Year