A Benchmark for Best View Selection of 3D Objects

Helin Dutagaci, Chun Pan Cheung, Afzal Godil
National Institute of Standards and Technology
{helin, bcheung, godil}@nist.gov

ABSTRACT

The best view selection corresponds to the task of automatically selecting the most representative view of a 3D model. In this paper, we describe a benchmark for evaluation of best view selection algorithms. The benchmark consists of the preferred views of 68 3D models provided by 26 human subjects. The data was collected using a web-based subjective experiment where the users were asked to select the most informative view of a 3D model. We provided a quantitative evaluation measure based on this ground truth data, and compared the performances of seven best view selection algorithms.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information System-evaluation/methodology; I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding-3D/stereo scene analysis, shape; I.4.8 [**Computer Graphics**]: Computational Geometry and Object Modeling

General Terms

Algorithms, Measurement, Performance, Experimentation, Standardization

Keywords

Best view selection, 3D shape analysis

1. INTRODUCTION

The best view selection problem corresponds to automatically selecting the most representative view of a 3D model. Automatic thumbnail generation of large 3D databases is one important application of this problem. With the expansion of 3D collections in CAD, molecular biology, medicine, and entertainment, fast visual browsing of 3D models is possible through automatic thumbnail generation. The thumbnails should be pleasant and enable fast recognition of the objects by the user. Some other applications of best view selection are automatic camera replacement, 3D scene generation, surgery planning, and view-based 3D object recognition.

In the last decade, a number of algorithms have been proposed to solve the best view selection problem. However, there exists no quantitative measure for objective comparison of the performances of these algorithms. In most cases, the authors display the images of the best views selected by their algorithms and base their evaluation on subjective comments.

In this paper, we present a benchmark which provides a method for quantitative evaluation of best view selection algorithms. We designed a web-based subjective experiment where the users are asked to rotate a model in 3D so that the most informative view is displayed. 26 people participated in these experiments; hence for each of the 68 3D models 26 views were obtained (Figure 1). We suggested a quantitative performance measure based on these subjective view preferences.

The paper is organized as follows: In Section 2, we summarize recent work on best view selection. In Section 3, we give brief descriptions of the seven best view selection algorithms we evaluated for comparison in this paper. In Section 4, we describe our benchmark; the database, the web interface, the construction of the ground truth, and a quantitative evaluation measure. In Section 5, we provide the results, and finally, we conclude in Section 6.

2. PREVIOUS WORK

Analysis of humans' preferred views of 3D objects has been a subject of research for psychological studies [1, 13, 3]. These preferred views are referred to as "canonical views" or "best views". The criteria that determine the best view of a model, such as the first imagined view, the most aesthetic view, and the most familiar view, are various and interconnected. The previous psychophysical experiments point to the result that for certain types of familiar objects the preferred views are consistent among the human subjects.

Polonsky et al. [11] analyzed a number of best view selection algorithms which associate a goodness measure to a number of candidate views. This goodness measure is a function of some objective related to the geometrical or statistical properties of the object itself or the projected views. Examples of such objectives are visible projected area [10], viewpoint entropy [16], recognition performance [12], curvature entropy [9], and silhouette entropy [9]. Polonsky et al. [11] compared these algorithms via presentation of the images of the best views of a few objects.

Yamauchi et al. [17] used mesh saliency as a measure of

3DOR'10, October 25, 2010, Firenze, Italy.

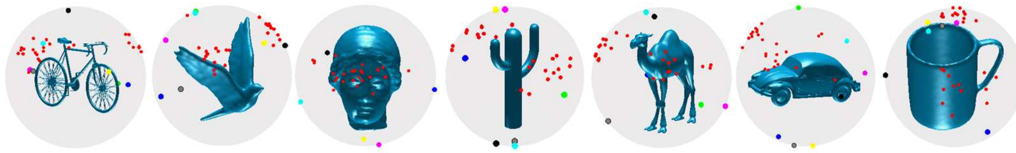


Figure 1: Sample models rendered from one view. The red dots are the viewpoints selected by human subjects and the other colors correspond to viewpoints selected by different algorithms.

the "goodness" of a view. The candidate views are selected through a partitioning of the view-sphere. The partitions are based on the similarity of the projected views and the centroids of the partitions are considered as representative views. Then these views are ranked based on the mesh saliency measure.

Instead of relying on local geometric features to identify the information content of a view, Mortara and Spagnuolo [8] suggested using mesh segmentation to identify meaningful salient parts of the object. Then the viewpoint that maximizes the visibility of the salient parts is selected as the best view. The authors tested their algorithm on a large database of 400 models and showed the resulting best views to 20 human subjects to be classified as "good", "acceptable", and "rejected" views. However, the evaluation is not based on a "ground truth" that is obtained before the experimentation, and quantitative comparison with other algorithms is not possible.

Jagadeesan et al. [5] conducted a subjective experiment with Amazon's mTurk Crowdsourcing system to investigate the best views of CAD models. In the experiment, the subjects were shown 20 views of the models and asked to select the best one. However, the assessment is based on a fixed number of viewpoints, as opposed to our system where users are free to rotate the object in 3D space.

Theetten et al. [14] proposed the use of equilibrium planes to select the characteristic views of a model and used these views for 3D object recognition and retrieval. They defined the equilibrium plane as a plane on which the object can stand statically. Laga [6] formulated the best view selection problem in terms of a classification problem, hence employed a learning strategy to identify the views that lead to highest classification accuracy. This approach is based on the assumption that objects belonging to the same semantic categories share similar best views.

The performance of the view selection algorithms is mostly assessed via visual demonstration of the selected views, and the analysis lacks a quantitative measure to compare the performance of various algorithms. In this paper, we propose a methodology to evaluate the performance of a view selection algorithm based on ground truth and a quantitative measure.

3. BEST VIEW SELECTION ALGORITHMS

We implemented a number of view selection algorithms and evaluated them using the ground truth supplied by the human subjects. These algorithms select the best view based on the view descriptors which are assumed to measure the geometric complexity of the visible surface as viewed from a point in the view sphere. The viewpoint maximizing the geometric complexity is assumed to provide the most informative view of the object. The algorithms differ with respect to the descriptor they use to assess the goodness of a view. We selected seven descriptors in order to provide baseline performance figures for future algorithms and for further research. These descriptors are listed as follows:

1) *View area*: The area of the projection of the object as seen from a particular viewpoint. Maximizing the area descriptor results in the view with the largest silhouette.

2) *Ratio of visible area* [11]: Ratio of the visible surface area to the total surface area of the object. The difference from the previous view descriptor is that, here we sum up the actual areas of the visible triangles instead of their projected areas.

3) *Surface area entropy* [15]: In this method, the ratio of the projected area of a triangle to the total projected area of the object is assigned to be the "probability" of the triangle with respect to a particular viewpoint. The entropy over this probability distribution is the surface area entropy-based view descriptor.

4) *Silhouette length* [11]: Length of the outer contour of the silhouette of the object as seen from a particular viewpoint.

5) *Silhouette entropy* [9]: Entropy over the curvature distribution of the outer contour of the silhouette.

6) *Curvature entropy* [9]: Entropy of the curvature distribution over the visible surface of the object. By maximizing this quantity, the view with the most diverse curvature values is selected as the best view. We used the mean curvature at the visible vertices to calculate the curvature entropy.

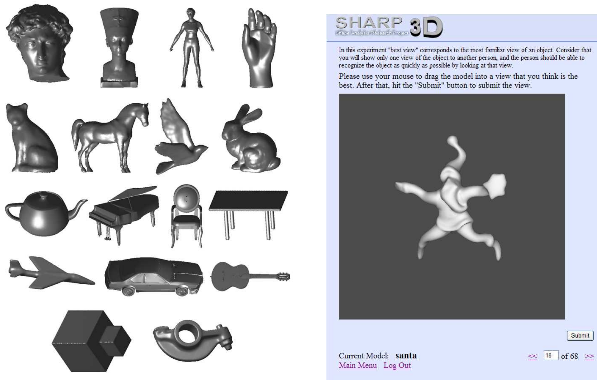
7) *Mesh saliency* [7]: Mesh saliency is also based on the local curvature over the surface. The mean curvature at each vertex is weighted by two Gaussian filters one with scale twice the other. The absolute difference between the weighted curvatures at two scales corresponds to the mesh saliency at that scale pair. Then, the total mesh saliency at a vertex is calculated as the sum of mesh saliency values at successive scale pairs. The best view is selected as the one which maximizes the sum of saliency values at the visible vertices.

We use the vertices of the geodesic sphere to sample viewpoints on the view sphere of a model. We measure the geometric complexity of each view according to a selected criterion, for example, the area or silhouette entropy of the view. The view that maximizes the criterion is assigned as the best view selected by the particular algorithm. We denote the associated vertex as v_m^a, i.e., the best viewpoint of model m, yielded by the algorithm a.

4. BENCHMARK DESIGN

4.1 3D model dataset

The 3D object dataset used in our experiments consists of 68 triangular meshes. Some of the models are standard mod-



(a) (b)

Figure 2: (a) Samples from the model set. (b) User interface for subjective view selection experiments.

els that are widely used in 3D shape research; and they have been used as test objects by researchers working on the best view selection problem. Examples are Armadillo, David's head, Utah teapot, Bunny, etc. We chose some of the models from The Stanford 3D Scanning Repository [18] and some others from the SHREC2007 watertight model database [19]. Figure 2-a shows samples from the model set. Almost all of the models are common objects highly familiar to humans. Only the two models that are shown in the last column of Figure 2-a are rather unfamiliar objects.

4.2 User interface for collecting ground truth

We created a web-page where users can login using an alias and participate in the experiments [20]. The user is shown the 68 3D models one at a time. Each model is initially rendered with a random pose. The user is asked to rotate the model via dragging the mouse into a view that he/she thinks is the best, and then to click on the submit button. We give the user an explanation for the best view of a model that ensures that the user does not try to select an aesthetic view but rather an informative view that will maximize the rate of recognition. Figure 2-b shows a snapshot of the user interface.

Up to now, 26 participants have submitted their preferred best views for the 68 models. We have constructed the ground truth and evaluated view selection algorithms based on the judgments of these 26 subjects. The 3D model dataset and the ground truth data are available at our website [21].

4.3 Evaluation measures

4.3.1 Ground Truth

An object can be seen from an infinite number of points of the view sphere. Our subjective experiment provides close to continuous viewpoints on the view sphere. However, computer-based view selection algorithms operate on a finite set of views of the model, therefore the view sphere should be sampled. Starting from an octahedron, we iteratively subdivide the triangular mesh in order to get vertices on the sphere. Figure 3 shows the process of obtaining a geodesic sphere. We select the geodesic sphere with 258 vertices to sample the view sphere. For each model, we have a

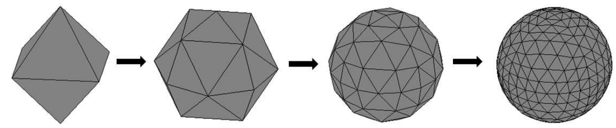


Figure 3: Construction of geodesic sphere.

set, V of 258 viewpoints. The viewpoints are defined by the vertices of the unit geodesic sphere, $v \in V$.

A model can have symmetries that lead to very similar views from different viewpoints. For a viewpoint v of a model m, we detect similar views as seen from other vertices and associate a symmetry set, $\Sigma_v = \{w \mid d\left(I_w, I_v\right) \leq t,\, w \in V\}$. Here, I_v corresponds to the depth image as seen from the viewpoint v, and $d\left(I_w, I_v\right)$ is a rotation invariant dissimilarity measure between image I_v and I_w. In this work, we have used Fourier-Mellin based image matching [2], however, other matching schemes can also be used. We set a low threshold t, since we prefer to detect almost-identical views among the viewpoints of a symmetric model.

When a user s selects the best view of a model m, we denote the corresponding viewpoint as p_m^s. We map the point p_m^s to the closest vertex of the geodesic sphere: $v_m^s = \arg\min_{v \in V} GD\left(p_m^s, v\right)$, where $G\left(a,b\right)$ defines the geodesic distance between the two points, a and b, on the unit sphere. In order to take into account the symmetries of the model, we determine the symmetry set $\Sigma_{v_m^s}$, which is associated by the vertex v_m^s. As a result, for a particular subject, we have as many best views as the number of elements in the symmetry set $\Sigma_{v_m^s}$. The ground truth for a model m is then defined as the symmetry set $\Sigma_{v_m^s}$.

4.3.2 Evaluation Procedure

Having obtained the ground truth and the result of a view selection algorithm, we attempt to measure the "success" of the algorithm. As already mentioned, both the subjects' choices and the best view given by an algorithm are mapped onto the vertices of the geodesic sphere.

Consider a model m and its associated ground truth $\Sigma_{v_m^s}$ supplied by subject s. Consider the best viewpoint v_m^a calculated by an algorithm a. Then the "goodness" of v_m^a is related to the following error from the human subject's choice of a viewpoint:

$$W_s\left(v_m^a\right) = \min_{v} \frac{1}{\pi} GD\left(v, v_m^a\right), \quad v \in \Sigma_{v_m^s} \qquad (1)$$

where $GD\left(a,b\right)$ defines the geodesic distance between the two points, a and b, on the unit sphere. This measure corresponds to a particular model and the ground truth supplied by a particular subject. $W_s\left(v_m^a\right)$ is a number between 0 and 1, since the maximum geodesic distance on the unit sphere is π. A value close to zero means that the algorithm gave a close view to that chosen by the subject. We average this measure over the choices of all human subjects and obtain a single measure, which we call as "View Selection Error" (VSE):

$$VSE = \frac{1}{S} \sum_{s} W_s\left(v_m^a\right) \qquad (2)$$

where S corresponds to the number of subjects.

Table 1: Inconsistency values for some of the models

Models with low inconsistency		Models with high inconsistency	
bunny	0.107	camel	0.319
teddy	0.113	ant	0.329
cat	0.114	dog	0.330
cow	0.126	dragon	0.342
David's head	0.127	helicopter	0.391
vase	0.140	cup	0.392
human	0.149	octopus	0.395
Utah teapot	0.198	rockerarm	0.444
armadillo	0.212	hand	0.456

5. RESULTS

5.1 Consistency

The consistency of the selected views among the subjects is an important issue. We measure the inconsistency of the choices of the human subjects over a model in a similar way we calculate the performance of a view selection algorithm. We measure the distance of one subject's preference of view to another subject's preference. Put in a more formal way, we calculate $W_p\left(v_m^r\right)$, the "error" of the selection of subject r with respect to the selection of subject p. Then we average $W_p\left(v_m^r\right)$ over all subjects:

$$W_p\left(v_m^r\right) = \min_v \frac{1}{\pi} GD\left(v, v_m^r\right), \quad v \in \Sigma_{v_m^p} \tag{3}$$

$$Inconsistency_m = \frac{1}{S^2} \sum_{p,r} W_p\left(v_m^r\right) \tag{4}$$

Notice that this measure takes into account the symmetries of a model. The higher this inconsistency value is, the more diverse are the subjects' preferences. Table 1 gives the inconsistency measure of the subjects' view selection for some of the models in our database. Our results indicate that human subjects consistently prefer some views for most of the objects. In Figure 4, we marked the viewpoints preferred by the human subjects as red dots. Each model is rendered from side, front and top. We can observe that for most of the models the red dots are concentrated in certain regions.

For the "vase" model, although the red dots are at diverse locations on the unit sphere, the view inconsistency score is low (indicating the users' preferences are consistent). That is due to the highly symmetric nature of the vase model. For the "bunny" and "David's head" models all users preferred viewpoints directly looking to the face.

The "rockerarm", as the least familiar model in our set, yields a high inconsistency score. Although not displayed here, for the "hand" model, half of the users preferred the back of the hand while the other half selected a view towards the palm. For highly articulated animals, such as "ant" and "octopus", we get high inconsistency values; however most of the users chose a top view of the "ant".

In accordance with the previous studies on humans' view preferences, this study shows that humans tend to choose three-quarter views for most of the objects (Figure 4-a, c, d). However, for some types of objects, such as human faces and bodies, some users tend to select a frontal view (Figure 4-b, e, f).

Another trivial observation is that, whenever an object is supposed to stand on the ground, none of the users chose a viewpoint seeing the bottom of the object. This observation confirms that automatic detection of the object base will greatly reduce the search space for best view as suggested in [14] and [4].

Table 2: Average View Selection Error of the view selection algorithms

View Selection Algorithm	VSE	n
View area	0.517	9
Ratio of visible area	0.473	7
Surface area entropy	0.396	15
Silhouette length	0.446	12
Silhouette entropy	0.484	7
Curvature entropy	0.474	9
Mesh saliency	0.430	9

5.2 Performance results

Table 2 gives the View Selection Error, averaged over the 68 models, of the seven automatic view selection algorithms we have described in Section 3. In the table, n gives the number of models for which an algorithm gave the lowest VSE (best view) among all the seven algorithms. In Figure 4, we marked the selected viewpoints of these algorithms with dots of different colors on the view sphere, and displayed the selected views enclosed by boxes of the corresponding colors. Under the views we have printed the VSE yielded by the corresponding algorithm. Observing Table 2 and Figure 4, we can list our comments as follows:

- On average, surface area entropy yielded the best performance (the lowest average VSE), followed by mesh saliency. Surface area entropy is the most successful approach for 15 of the objects, but does not perform consistently better for all the objects.

- None of the methods is consistently the best (or worst) over all the objects; they are rather complementary. This result is in agreement with the conclusion in [11]. Fusion of these methods, accompanied by learning schemes that adjust the relative weight of each method, will be beneficial for increasing the best view selection performance.

- Maximizing geometric complexity, independent of the context of the object, does not always lead to a best view. Examples are the "bunny" in Figure 4-a and "chair" in Figure 4-d. The tail, the ears pointing backwards and the holes at the base of the "bunny" deceive most of the algorithms. Similarly, the legs of the "chair" cause the mesh saliency algorithm to choose the wrong view. Therefore, the best view selection algorithms will benefit from learning schemes as pointed out in [6].

- The "goodness" of the views displayed in little square frames in Figure 4 is in accordance with the VSE values printed under them, i.e. the lower the value, the better the view. This measure can be utilized for optimization procedure of algorithms, if accompanied by training objects with ground truth.

6. CONCLUSION

In this paper we attempted to provide a benchmark that involves a methodology and a quantitative measure to evaluate the performance of view-selection algorithms. We designed web-based subjective experiments to analyze humans' view preferences and to construct a ground-truth. We assessed the performance of a view-selection algorithm with respect to the geodesic distance between the viewpoint determined by the algorithm and the viewpoint preferred by a human subject. The viewpoints are limited to the number of the vertices of the unit geodesic sphere. We also took into account the symmetry properties of the models while calculating the performance measure.

Notice that, in this work, we did not attempt to cluster similar views among the view sphere and choose candidate views, which is the approach carried out in [11, 17, 6]. Instead, we directly calculated the complexity measures described in Section 3 on each of the 258 viewpoints and chose the one which maximizes the measure as the best view. Prior view clustering is beneficial not only to reduce the search space, but also to guarantee the view stability. However, our objective in this work is suggesting a quantitative evaluation method, rather than maximizing the performances of individual methods.

We intend to make a couple of improvements on the evaluation procedure: One is to weight the geodesic distances between vertices with the similarity of the views as seen from these vertices and reformulate the VSE accordingly. A second improvement can be the elimination of outliers among the subjects' choices. The subjects' choices can be clustered, and another evalutaion measure can be defined based on the distance to the cluster centers.

Another important issue is the correction of the in-plane orientation after the viewpoint is determined. This step is essential in automatic thumbnail generation applications. The methods we evaluated in this work do not directly address this problem. However, this problem was mentioned in [17] and [8] with suggestions on automatic upright orientation. In [4], the authors have concentrated on the upright orientation of man-made objects through detection of static equilibrium and learning discriminative attributes of views. Although not discussed in this paper, our ground truth data include the subjects' preferences of in-plane orientation. As a future work, we are planning to develop a quantitative measure to evaluate the success of the algorithms aiming to perform this final step.

We are in the process of increasing the number of subjects via the use of Internet Crowdsourcing. This will also allow us to increase the size of the 3D model dataset so that the results can be generalized to generic objects. An elaborate analysis of these experiments can be inspiring for developing successful view-selection algorithms.

7. ACKNOWLEDGMENTS

We would like to thank Daniela Giorgi and AIM@SHAPE for the models from the Watertight Track of SHREC 2007, and Stanford Computer Graphics Laboratory for the models from The Stanford 3D Scanning Repository.

8. REFERENCES

[1] V. Blanz, T. Vetter, H. Bülthoff, and M. Tarr. What object attributes determine canonical views? *Perception, Suppl.*, 24:199c, 1995.

[2] D. Casasent and D. Psaltis. Position, rotation, and scale invariant optical correlation. *Appl. Opt.*, 15(7):1795–1799, 1976.

[3] F. Cutzu and S. Edelman. Canonical views in object representation and recognition. *Vision Research*, 34:3037–3056, 1994.

[4] H. Fu, D. Cohen-Or, G. Dror, and A. Sheffer. Upright orientation of man-made objects. In *ACM SIGGRAPH'08*, pages 1–7, 2008.

[5] A. P. Jagadeesan, A. Lynn, J. R. Corney, X. T. Yan, J. Wenzel, A. Sherlock, and W. Regli. Geometric reasoning via internet crowdsourcing. In *SPM '09*, pages 313–318, 2009.

[6] H. Laga. Semantics-driven approach for automatic selection of best views of 3D shapes. In *3DOR2010*, 2010.

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

[8] M. Mortara and M. Spagnuolo. Semantics-driven best view of 3D shapes. *Computers & Graphics*, 33(3):280–290, 2009.

[9] D. L. Page, A. Koschan, S. R. Sukumar, B. Roui-Abidi, and M. A. Abidi. Shape analysis algorithm based on information theory. In *ICIP (1)*, pages 229–232, 2003.

[10] D. Plemenos and M. Benayada. Intelligent display in scene modeling. New techniques to automatically compute good views. In *Proc. of GraphiCon*, pages 84–89, 1996.

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

[12] D. Roberts and A. Marshall. Viewpoint selection for complete surface coverage of three dimensional objects. In *Proc.of the Britsh Machine Vision Conference*, pages 740–750, 1998.

[13] M. J. Tarr and D. J. Kriegman. What defines a view? *Vision Res*, 41(15):1981–2004, 2001.

[14] A. Theetten, T. Filali Ansary, and J.-P. Vandeborre. 3D-model view characterization using equilibrium planes. In *3DPVT'08*, 2008.

[15] P.-P. Vázquez, M. Feixas, M. Sbert, and W. Heidrich. Viewpoint selection using viewpoint entropy. In *VMV '01*, pages 273–280. Aka GmbH, 2001.

[16] P.-P. Vázquez, M. Feixas, M. Sbert, and W. Heidrich. Automatic view selection using viewpoint entropy and its applications to image-based modelling. *Comput. Graph. Forum*, 22(4):689–700, 2003.

[17] H. Yamauchi, W. Saleem, S. Yoshizawa, Z. Karni, A. Belyaev, and H.-P. Seidel. Towards stable and salient multi-view representation of 3D shapes. In *SMI '06*, page 40, 2006.

[18] http://www.graphics.stanford.edu/data/3Dscanrep/.

[19] http://watertight.ge.imati.cnr.it/.

[20] http://control.nist.gov/sharp/view/.

[21] http://www.itl.nist.gov/iad/vug/sharp/benchmark/bestview/.

Figure 4: Each model is rendered from side, front and bottom. The red dots are the viewpoints selected by the human subjects. Other points correspond to the viewpoints provided by the algorithms. The views selected by the algorithms are displayed in small images enclosed by the corresponding color. View Selection Error (VSE) is printed under the views.