Towards Stable and Salient Multi-View Representation of 3D Shapes

Hitoshi Yamauchi Waqar Saleem Shin Yoshizawa Zachi Karni Alexander Belyaev Hans-Peter Seidel
MPI Informatik
Saabrücken, Germany
Email:{hitoshi|wsaleem|shin|karni|belyaev|hpseidel}@mpi-inf.mpg.de

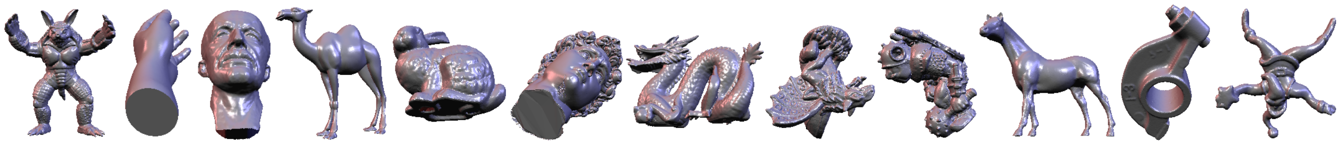


Figure 1. Best views generated by our approach.

Abstract

An approach to automatically select stable and salient representative views of a given 3D object is proposed. Initially, a set of viewpoints are uniformly sampled along the surface of a bounding sphere. The sampled viewpoints are connected to their closest points to form a spherical graph in which each edge is weighted by a similarity measure between the two views from its incident vertices. Partitions of similar views are obtained using a graph partitioning procedure and their "centroids" are considered to be their representative views. Finally, the views are ranked based on a saliency measure to form the object's representative views. This leads to a compact, human-oriented 2D description of a 3D object, and as such, is useful both for traditional applications like presentation and analysis of 3D shapes, and for emerging ones like indexing and retrieval in large shape repositories.

1 Introduction

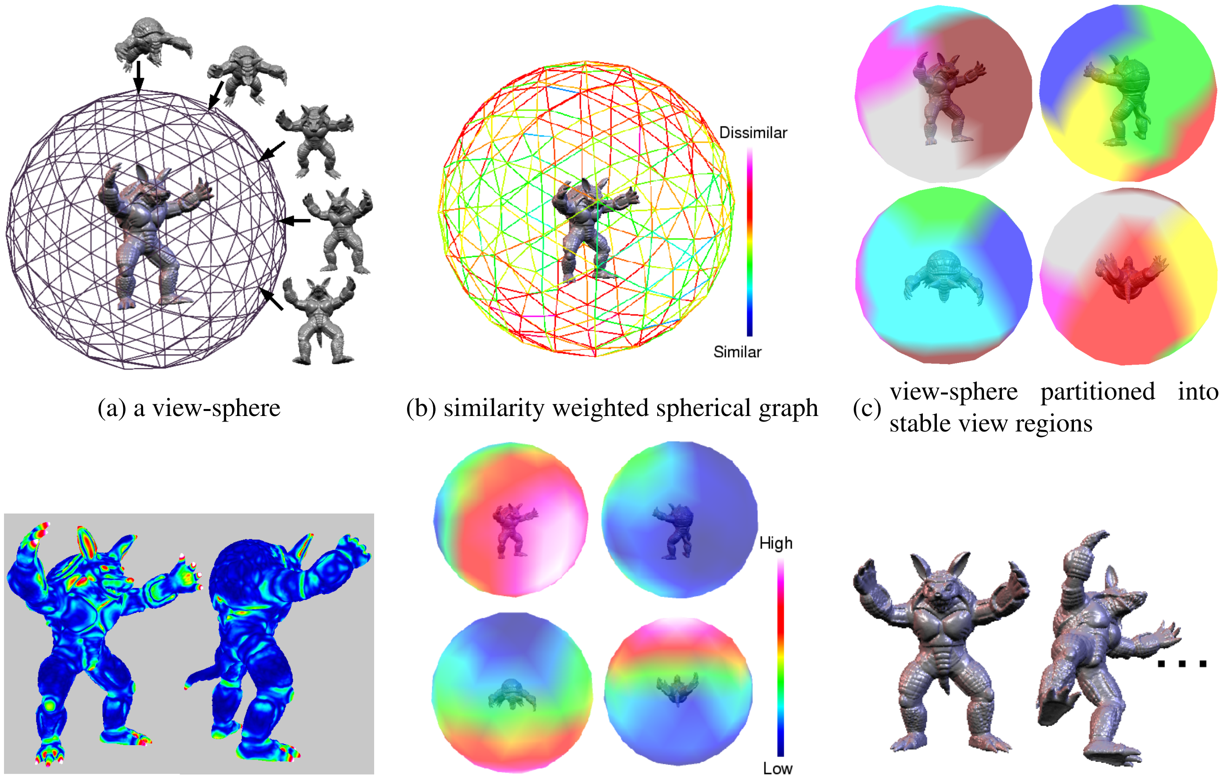
The nature of 3D shape perception has intrigued humans for many centuries, and continues even today to be a subject of intensive research in psychology, neuroscience, psychophysics, and computer science [1,19,20] (see also references therein).

Within the pattern-recognition and computer-vision communities, the problem of defining representative 2D views for recognition and representation of 3D objects has recently received significant attention [6, 8, 12, 13]. The main idea consists of studying similarity and stability relationships between different 2D views of the same 3D object [4, 5]. However, these techniques do not take into account human perception factors. For example, when a 3D object is viewed, some of its features are more salient than others. Therefore, a preferred view of the object should make these parts visible.

A different approach for optimal positioning of viewpoints and lighting is gaining popularity among graphics researchers [2,7,11,17,18,21]. These papers introduce image goodness/saliency measures and then select views that maximize these measures. A comparison of various measures for 3D triangulated meshes is presented in [15]. The main conclusion derived in [15] is that all descriptors tested in the paper are reasonably good but each of them fails to produce satisfactory results for certain types of 3D shapes.

In this paper, we suggest a method that combines similarity and goodness/saliency approaches. Such a combination inherits the strengths of each approach while compensating for their individual disadvantages. Our method is based on similarities between different views. However, instead of comparing fine details of a view, we consider only 2D silhouettes, that result in computing similarities between binary images. The similarity measures guide a clustering process that groups together a set of viewpoints, all sharing a similar view of an object. From each cluster, one representative view is selected by using the cluster's "centroid". To decide which of the views is most *important*, we use a similar approach to [11]. The saliency measure orders the representative views from the most salient to the least.

Our interest in representative views is mainly in context of shape repositories, where information on the shape of stored models has to be presented to users in the form of a few representative views. Two applications of our method are to automatically compute such views for a newly added model and 2D-based matching for retrieval of 3D objects.



(a) a view-sphere (b) similarity weighted spherical graph (c) view-sphere partitioned into stable view regions

(d) mesh saliency (e) view saliency (f) computed views

Figure 2. Overview of our stable and salient view selection method. Top row: Partitioning the view-sphere into stable view regions. (a) view-sphere and observed object. (b) similarity weighted spherical graph. (c) colored stable view regions. Bottom row: using view saliency to select the final views. (d) mesh saliency [11]. (e) visualization of view saliency. (f) selected representative views.

2 Stable and Salient View Selection

Figure 2 presents an overview of our method. Our method is divided into two main parts: finding stable views and defining their saliency. We consider the most salient among all the stable views as the best view of the given 3D object.

A common approach to most view selection methods is the definition of a criterion for comparison between different viewpoints [2, 11, 15, 18, 21] (e.g. view entropy). The criterion quantifies the level of importance of features in the object and hence, defines a best view as the one that maximizes this quantity. Polonsky et al. [15] suggested and compared several criteria.

Researchers within the object recognition area claim that a single view is insufficient and suggest to use multiple views instead [12, 13]. Their strategy is to examine a large set of views and to filter out those that look similar, since they do not add any new information. The remaining views are the selected multi-views of the object. Although this method suffices for object recognition purposes, it considers all views to be equal. Hence, a view that includes many details about the object has no priority over the others.

Our method combines the above general ideas. It first selects a small set of stable views and then sorts them according to the value of information they carry.

Throughout the paper, we use the following notations. Given a 3D object in world coordinates system, we translate the object such that its center of gravity aligns with the origin of the coordinate system. A view of the object is taken by a virtual camera that is positioned at a *viewpoint* and aimed (*lookat*) at the origin. The camera's *upvector* defines the rotation angle around the viewpoint-origin axis, and the camera's *field-of-view* parameter is always set to 45^o. When a view is taken, the result is always a 2D image. Therefore, whenever we use the term *view* or *a view of the model*, we refer to the resulting 2D image.

2.1 Generating Stable Multi-Views

Our method for selection of stable views utilizes similarity measures [12, 13]. We define a sphere that is aligned,

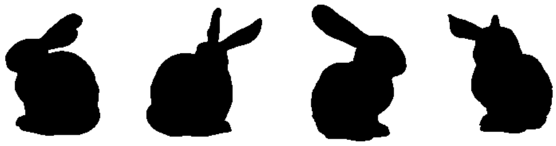
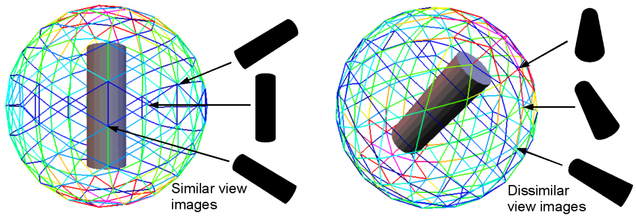
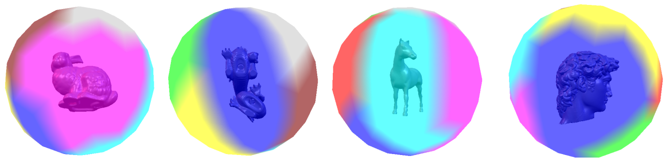


Figure 3. Sample of binary (silhouette) views.

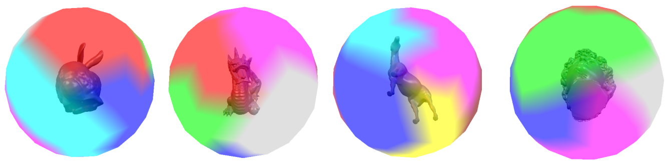


(a) similar views (b) dissimilar views

Figure 4. Similarity weighted spherical graph of a cylinder.



(a) each stable view region is represented by one of its views.



(b) unstable views lie at the intersection of stable view regions.

Figure 5. Stable view regions.

together with the 3D model, at the coordinate system's origin, and set its radius such that it always bounds the model. The sphere's surface is used as a ground for the camera's viewpoints from which all necessary views of the object are acquired. Given the set of views, we measure the similarity between each two disjoint views and group together those that are similar. However, to avoid the high computational effort that is required to check each pair of views, we suggest the following alternative.

We construct a *view-sphere* using two iterations of the Loop subdivision scheme on an initial icosahedron mesh and refer to the resulting mesh structure as a *spherical graph*. Only views acquired from the spherical graph's vertices are considered and similarities between adjacent views (two views sharing an edge) are computed. The similarity measure is then assigned to the edge as a weight. Figure 2a shows a view-sphere bounding the Armadillo model together with several sampled views. Figures 3 shows examples of binary (silhouette) images that we used as an input to the view similarity measurements.

Similarity between views is computed using Zernike moments analysis [14]. The method employs frequency analysis in polar coordinates, therefore, the comparison is rotation invariant. It is also simple to achieve scale and translation invariance. The similarity is computed as the L^2 norm of the difference between the views' Zernike moments. We implement the "direct method" of Zernike moments computation [3] and use moments up to order 15. The property of rotational invariance is crucial for view selection since any view has the degree of freedom of rotating around its *viewpoint-lookat* axis. Thus, our viewpoint selection method considers such views to be one and the same.

Figures 2b and 4 show the spherical graph with edges colored by similarity weights. A blue edge represents high similarity between its incident views while red edges represent dissimilar views. We consider a view to be stable if all edges incident on its viewpoint in the spherical graph have high similarity weights. *A stable view region* is an area on the view-sphere that groups several stable viewpoints together. Figure 4a shows that views that are rotated version of themselves are considered similar (blue edges).

To find stable view regions, we partition the similarity weighted spherical graph based on its edge weights. One possible way to achieve this is to find an edge cut that segments the graph into the requested number of partitions while minimizing the total weights of edges in the cut. This way we prevent region with high stability to be partitioned into two disjoint parts. MeTiS [10] is a graph partitioning application that partitions a graph into the requested number of sub-graphs. Given a graph with weighted edges, MeTiS finds an edge cut which minimizes total weights and generates sub-graphs with balanced number of vertices. Although our application does not require the balancing property, our experiments show that it does not create any bias. Figure 2c shows an example of this partition method (each partition has a different color). The partitioning quality is also influenced by the sampling density of the view-sphere. Sparse sampling will result in a small graph which will be difficult to partition. In the field of object recognition, around 50 uniform distributed samples are used [13, 16]. In our implementation, 162 samples are used, same as in [18], and from our experience this number is sufficient.

2.2 Sorting of Partitions' Importance

Once the spherical graph is partitioned into stable view regions, the stable partitions have to be sorted according to their importance and the representative viewpoint inside each partition needs to be found.

To sort the stable partitions, we first need to value their importance. Each partition is assessed by the average of its viewpoints' value, where viewpoints are assigned val-

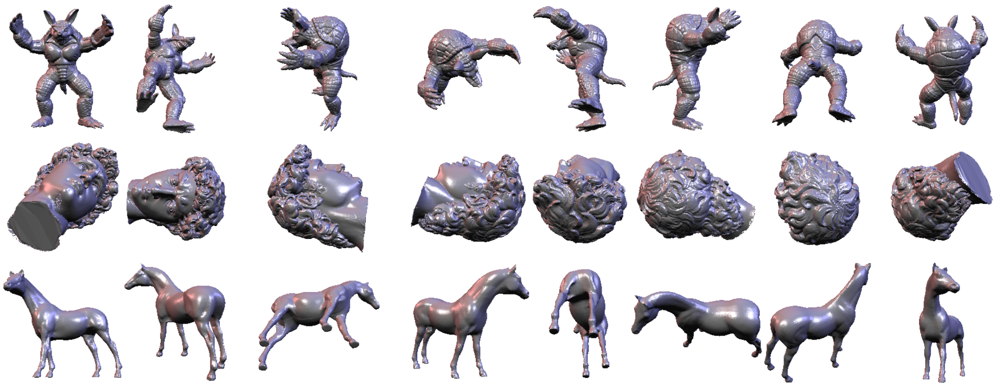


Figure 6. Selection of stable and salient views using binary (silhouette) views.

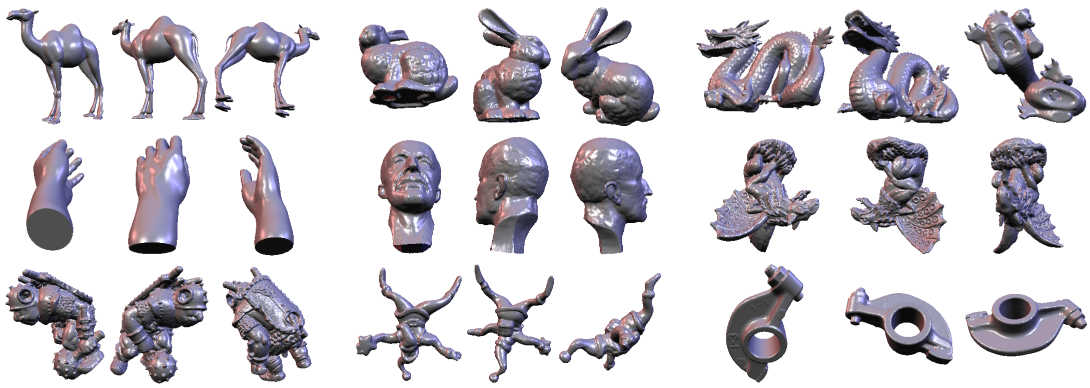


Figure 7. Best three views generated by our approach.

ues using a mesh saliency [11] method, which is based on a vision image saliency technique [9]. This method evaluates saliency by computing mean curvature of a 3D mesh model and employing multi-scale mesh smoothing to remove small details, leaving behind the salient features.

A representative viewpoint $\mathbf{R}_i$ of a partition P_i is given by the following view saliency weighted centroid,

$$\mathbf{R}_i = \frac{\sum_{j \in P_i} s_j \cdot \mathbf{p}_j}{\sum_{j \in P_i} s_j},$$

where $\mathbf{p}_j$ is the view-sphere vertex and s_j is its corresponding view saliency.

2.3 A Suggestion for Model Orientation

One of the major challenges in view selection is to find the proper orientation of the 3D model. For example, when viewing a model of a four-legged animal, we expect the view-selection method to orient the view such that the animal will be on its feet and not on any other body-part. To the best of our knowledge, this problem is not yet to be addressed successfully, and in former view-selections publications (e.g., [15, 18]), the final orientation was set by hand.

A solution to this problem is outside the scope of this paper, nor do we claim to have one. However, we would like to suggest a hypothesis, which claims that for an object to be properly oriented, its least important part should be facing down. The justification is that usually the lower part of an object is hidden from the viewer and the viewer would thus choose the least important part to be the lower one.

Figure 8 shows several examples of model orientation selection based on our hypothesis. Although only 3 out of the 12 models that we tested were oriented correctly, we still believe that this idea has potential and plan to continue research on it. For all other view selection results in this paper, the up vector was set to $(0, 1, 0)$. View orientation in the figures has not been adjusted in any other way.

Figure 8. Best views generated by our approach together with our suggestion for model orientation.

Figure 9. The best views selected by view saliency [11] only.

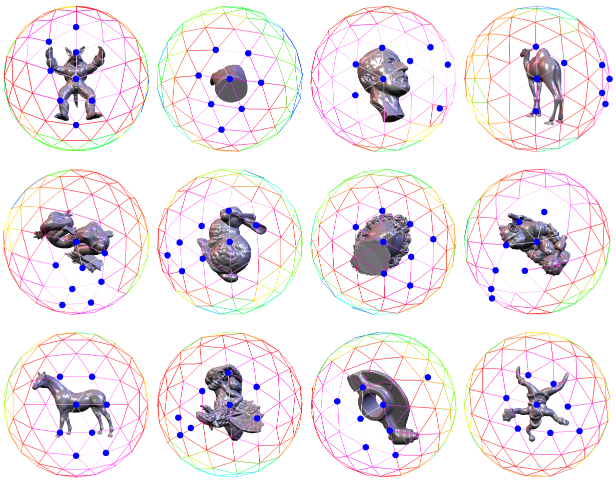


Figure 10. Top 8 most salient viewpoints. (Blue points) The view-sphere's edge color reflects saliency value of the incident viewpoints. Notice the back of view-sphere is culled for visualization.

Figure 11. Top 3 salient view examples of Figure 10.

3 Results and Discussion

Figure 5 shows some results obtained by partitioning the similarity weighted spherical graph using MeTiS. Figure 5a shows a representative stable view for each model while Figure 5b shows an unstable view. Note, that the unstable view lies at the intersection of several stable view regions. We can observe that the stable regions for the dragon and horse models are stretched in the sphere's longitudinal direction. This implies that a movement of the viewpoint along a longitudinal line will result in minimal change in view compared to movement in a latitudinal direction.

We fixed the number of partitions (also the number of final views) to 8. From our experiments, we have noticed that this number is usually sufficient to cover all interesting parts of the model. A larger number of partitions may not reveal any new information, while a smaller number might not suffice.

Figures 1, 6, and 7 show several results of our experiments using our automatic multi-view selection method. We also examined illuminance images (grayscale images) using the Gouraud shading model besides binary (silhouette) images. However, the shaded images are sensitive to environmental conditions, such as lighting, and hence tend to bias the results. Due to this, we conclude the use of binary images is more appropriate for our purposes.

Figure 10 shows a multi-view selection that is based on saliency alone. It is easy to see that all high saliency viewpoints are concentrated in a small region on the view-sphere. This is because small deviations in viewpoint do not affect the saliency value much. By taking into account the stability of the view, we force the views to be spread all over the view-sphere, resulting in a better distribution.

Figure 9 shows the best views selected by the mesh saliency method [11]. Comparing Figure 9 with Figures 6 and 7, we can say that our results are comparable or, in some cases, better. For example, the bottom of the dragon model is the most salient (Figure 9). However, when stability is taken into account, the best view changes to the side of the dragon, which is a much better suggestion (Figure 7), although this is not the most salient view.

Similarly, the neck of David's head is also salient due to its high and consistent curvature. However, this view is unstable (Figure 5). Therefore, our method avoids the uninteresting view of the neck and recommends better views that cover the front and top of the head (Figure 6). Note that our method does not ignore the saliency recommendation in the neck region, but combines it with a stable view. These examples show the benefits of our method achieved by combining two human perception elements, stability and saliency.

Computation time of constructing the similarity weighted spherical graph is about 40 minutes. This does not depend on the model since the computations are on the rendered views and the number of views is fixed (number of vertices of the view-sphere). The view resolution is set to 256×256. Graph partitioning by MeTiS takes less than 0.01 seconds. Readers can find the mesh saliency computation time in [11]. The view saliency computation takes less than a minute, depending on the rendering time of the model. Currently, no optimization has been done throughout our implementation. All timings were measured on a 2.8GHz Pentium 4 PC with an ATI Radeon R300.

4 Conclusion

We present a new view selection approach based on two human perception elements, view stability and saliency. Stable view regions on the view-sphere are computed by a graph partition algorithm where the edges are weighted by view similarity. Recommended views are obtained by ranking the regions with view saliency. Compared to other known methods, our generated views look more plausible and perceptually understandable.

In the future, we are planning to investigate and improve performance of our method in certain cases where best view selection fails. We also plan to research the automatic model orientation problem which will have significant influence on automatic view-selection procedures. In addition, computation time optimizations were not considered in this paper. Speed up of the method is also a significant topic. There are several optimizations that can speed up Zernike moments computation (e.g. [3]) and we plan to enhance our approach by several of them.

Acknowledgements The authors would like to acknowledge the Stanford Data Archives, Farfield Technology, Cyberware, CNR/CNUCE, CNR/IMATI/GE, and AIM@SHAPE Shape Repository for 3D models. This research is partially supported by European FP6 NoE grant 506766 (AIM@SHAPE). We are grateful to our AIM@SHAPE partners from Technion and IMATI for fruitful discussions.

References

[1] V. Blanz, M. J. Tarr, H. H. Bülthoff, and T. Vetter. What object attributes determine canonical views? *Perception*, 28(5):575–600, 1999.

[2] U. Bordoloi and H.-W. Shen. View selection for volume rendering. In *IEEE Visualization*, pages 487–494, 2005.

[3] C. Chong, P. Raveendran, and R. Mukundan. A comparative analysis of algorithms for fast computation of Zernike moments. *Pattern Recognition*, 36(3):731–742, March 2003.

[4] F. Cutzu and S. Edelman. Canonical views in object representation and recognition. *Vision Research*, 34(22):3037–3056, 1994.

[5] F. Cutzu and S. Edelman. Faithful representation of similarities among three-dimensional shapes in human vision. *Proc. Natl. Acad. Sci., Psychology*, 93(21):12046–12050, 1996.

[6] T. Denton, M. Demirci, J. Abrahamson, A. Shokoufandeh, and S. Dickinson. Selecting canonical views for view-based 3-D object recognition. In *17th International Conf. on Pattern Recognition (ICPR'04)*, volume 2, pages 273–276, August 2004.

[7] S. Gumhold. Maximum entropy light source placement. In *VIS '02: Proc. of the conference on Visualization '02*, pages 275–282, Washington, DC, USA, 2002. IEEE Computer Society.

[8] P. M. Hall and M. J. Owen. Simple canonical views. In *The British Machine Vision Conf. (BMVC'05)*, volume 1, pages 7–16, 2005.

[9] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, Nov 1998.

[10] G. Karypis and V. Kumar. MeTiS: A software package for partitioning unstructured graphs, partitioning meshes, and computing fill-reducing orderings of sparse matrices. version 4.0. In *http://www-users.cs.umn.edu/˜karypis/metis/*. Univ. of Minnesota, Dept. of Computer Science, 1998.

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

[12] J. Lee, B. Moghaddam, H. Pfister, and R. Machiraju. Finding optimal views for 3D face shape modeling. In *International Conf. on Automatic Face and Gesture Recognition*, pages 31–36, 2004.

[13] F. Mokhtarian and S. Abbasi. Automatic selection of optimal views in multi-view object recognition. In *The British Machine Vision Conf. (BMVC'00)*, pages 272–281, 2000.

[14] R. Mukundan and K. R. Ramakrishnan. *Moment Functions in Image Analysis: Theory and Applications*. World Scientific Publishing Co Pte Ltd., Singapore, September 1998.

[15] O. Polonsky, G. Patané, S. Biasotti, C. Gotsman, and M. Spagnuolo. What's in an image? In *Visual Computer*, pages 840–847, 2005. Proc. of Pacific Graphics 2005.

[16] M. Seibert and A. M. Waxman. Adaptive 3-d object recognition from multiple views. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(2):107–124, February 1992.

[17] R. Shacked and D. Lischinski. Automatic lighting design using a perceptual quality metric. *Comput. Graph. Forum*, 20(3), 2001.

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

[19] M. J. Tarr and D. J. Kriegman. What defines a view? *Vision Research*, 41:1981–2004, 2001.

[20] J. T. Todd. The visual perception of 3D shape. *TRENDS in Cognitive Science*, 8(3):115–121, 2004.

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