VIEWPOINT QUALITY AND GLOBAL SCENE EXPLORATION STRATEGIES

Dmitry Sokolov, Dimitri Plemenos and Karim Tamine

XLIM Laboratory UMR CNRS 6172, University of Limoges
83 rue d'Isle, 87000 Limoges, France

Keywords: Scene understanding, automatic virtual camera, good point of view.

Abstract: Virtual worlds exploration techniques are used in a wide variety of domains — from graph drawing to robot motion. This paper is dedicated to virtual world exploration techniques which have to help a human being to understand a 3D scene. An improved method of a viewpoint quality estimation is presented in the paper, together with a new method for an automatic 3D scene exploration, based on a virtual camera. The automatic exploration method is related to off-line exploration and is made in two steps. In the first step, a set of "good" viewpoints is computed. The second step uses this set of points of view to compute a camera path around the scene.

1 INTRODUCTION

One of the reasons for rapid development of computer science was introduction of human-friendly interfaces, which have made computers easy to use and learn. The increasing exposure of the general public to technology means that their expectations of display techniques have changed. The increasing spread of the internet has changed expectations of how and when people are to access information. As a consequence, a lot of problems raised. One of them is automatic exploration of a virtual world. During last years, people pay essentially more attention to this problem. They realized the necessity of fast and accurate techniques for better exploration and clear understanding of various virtual worlds. A lot of projects use results of intelligent camera placement researches, from the "virtual cinematographer" (He et al., 1996) to motion strategies (Marchand and Courty, 2000).

Quality of a view direction is a rather intuitive term and, due to its inaccuracy, it is not easy to precise for a selected scene, which viewpoints are "good" and which are not. Over the last decades, many methods were proposed to evaluate qualities of view directions for a given scene and to choose the best one. All of them are based on the fact that the best viewpoint gives to the user maximum amount of information about a scene. And again, an imprecise term "information" is met.

The proposed viewpoint evaluation methods could be divided into two groups, characterized by the nature of input information:

1. *Low-level methods*, which consider only quantity of visible surfaces;
2. *Middle-level methods*, which take into account geometry of visible surfaces.

In this paper, a new group of *high-level methods*, operating with the visibility of scene *objects* is introduced.

The rest of the paper is organized as follows: section 2 gives a brief description of previous works. A new method of viewpoint quality estimation is described in section 3. A comparison of methods is given in section 4. A new method of global scene exploration is presented in section 5. Examples of the new technique application are given in section 6. Finally, section 7 concludes our work and points out directions of future work.

2 PREVIOUS WORK

2.1 Low-level Methods of Viewpoint Quality Evaluation

This group of strategies is the most numerous one. Among these methods, one can point at the Kamada and Kawai (Kamada and Kawai, 1988) approach. They have proposed to minimize the angle between

a direction of view and the normal of the considered plane for a single face, or to minimize the maximum angle deviation for all the faces of a complex scene.

In (Plemenos and Benayada, 1996) Plemenos and Benayada have proposed a heuristic that extends the definition given by Kamada and Kawai. The heuristic considers a viewpoint to be good if it minimizes the maximum angle deviation between a direction of view and normals to the faces and gives a high amount of details. The viewpoint quality according to (Plemenos and Benayada, 1996) can be computed by the following formula:

$$C(p) = \frac{\sum_{i=1}^{n} \left[\frac{P_i(p)}{P_i(p)+1} \right]}{n} + \frac{\sum_{i=1}^{n} P_i(p)}{r}, \quad (1)$$

where:

1. $C(p)$ is the viewpoint quality for the given viewpoint p,

2. $P_i(p)$ is the number of pixels corresponding to the polygon number i in the image obtained from the viewpoint p,

3. r is the total number of pixels of the image (resolution of the image),

4. n is the total number of polygons in the scene,

5. $[a]$ means the ceiling function, i.e the smallest integer number $a_c \in \mathbb{N} : a_c \geq a$.

In (Vazquez et al., 2001) Vazquez et al. have provided an information theory-based method, that can be classified as low-level method. To select a good viewpoint they propose to maximize the following function they have called a "viewpoint entropy":

$$I(p) = \sum_{i=0}^{N_f} \frac{A_i}{A_t} \cdot \log_2 \frac{A_t}{A_i}, \quad (2)$$

where:

1. p is the viewpoint,

2. N_f is the number of faces of the scene,

3. A_i is the projected area of the face number i,

4. A_0 is the projected area of background in open scenes,

5. A_t is the total area of the projection.

2.2 Middle-level methods

Directly or implicitly, all the methods from the low-level group use only two global parameters as input:

1. Quantity of visible surfaces (projected area, amount of voxels, angle between direction of sight and normal to a face),

2. Number of visible faces.

In other words, all of them consider a viewpoint quality as a sum of qualities of separate faces, but don't take into account *how* a polygon is connected to the adjacent ones.

The number of visible faces is a quite weak criterion for viewpoint quality estimation. For example, if we consider a very simple scene that consists of one square (figure 1(a)), then equation 2 gives us $I(p) = 0$ for a viewpoint p lying on the perpendicular to the square's center. If we subdivide the square (figure 1(b)), topology of the scene does not change, but $I(p)$ grows.
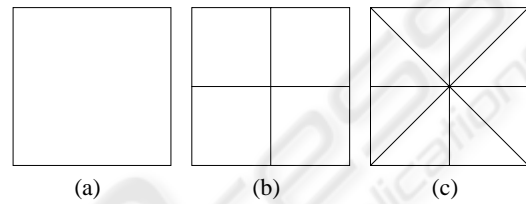

(a)   (b)   (c)

Figure 1: Three scenes represent the same square **(a)**, but subdivided into 4 parts **(b)** and 8 parts **(c)**. Equation 2 gives us $I(p) = 0$ for **(a)**, $I(p) = \log 4$ for **(b)** and $I(p) = \log 8$ for **(c)**.

Thus, the methods using a number of faces to evaluate a viewpoint quality, depend on initial scene subdivision. Using the projected area of a face as a criterion of quality, the dependence appears also if we don't use an additive metric, i.e., the sum of areas.

Recently Sokolov and Plemenos (Sokolov and Plemenos, 2005) have proposed to consider the total curvature of visible surfaces as an amount of information appropriate to a viewpoint:

$$I(p) = \sum_{v \in V(p)} \left| 2\pi - \sum_{\phi_i \in \phi(v)} \phi_i \right| \cdot \sum_{f \in F(p)} P(f), \quad (3)$$

where:

1. $F(p)$ is the set of faces visible from p,

2. $P(f)$ is the projected area of the face f,

3. $V(p)$ is the set of visible vertices of the scene,

4. $\phi(v)$ is the set of angles adjacent to the vertex v.

The proposed heuristic is invariant to any subdivision of a scene keeping the same topology. Indeed, if we subdivide a flat face to several ones, then all the edges and vertices inside the face are to be discarded due to zero angles.

An important property of this viewpoint quality definition is the possibility to extend it, using the total integral curvature $\int_{\Omega} |K| dA$, into the class of continuous surfaces, such as NURBS etc., more and more usable nowadays.

3 A NEW HIGH-LEVEL METHOD

Now let us suppose that, having a complex scene, there exists some proper (in human perception) division of a scene into a set of objects. Figure 2 shows us an example of such a scene. These objects are: the computer case, the display, the mouse, the mouse pad, two cables, the keyboard and several groups of keys.
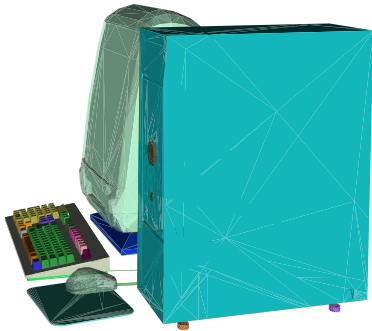

Figure 2: The scene is subdivided into a set of objects. The display is almost completely hidden by the case, but we could clearly recognize it.

Only 20% of the display surface is visible, but it does not embarrass its recognition. Thus, we could conclude that if there exists a proper subdivision of a scene into a set of objects, the visibility of the objects could bring more information than just the visibility of the faces, and this leads us to the third group of methods — *the high-level class*.

The requirement of a scene division into objects does not limit the area of the method application. There are many ways to get it. First of all, complex scenes often consist of non-adjacent simple primitives, and this leads to the first disjunction of a scene. Otherwise, if a scene (or some parts of a scene) is represented by a huge mesh, it could be decomposed. The domain of the surface decomposition is well-studied and there are a lot of methods giving excellent results. One can point at results of Zuckerberger et al. (Zuckerberger et al., 2002) and Chazelle et al. (Chazelle et al., 1995).

The method could be used also in declarative modelling. In this case, the decomposition could be provided by a modeler directly, or it can be combined with the information extracted from a scene geometry.

An accurate definition of the new heuristic is given further. Let us suppose that for each object ω of a scene Ω importance $q(\omega) : \Omega \to \mathbb{R}^+$ is specified.

We would like to generalize the method and do not want to be limited by a strict definition of the importance function, because it could be done in different ways, especially, if some additional knowledge about a scene is supplied. For example, if the method is incorporated into some dedicated declarative modeler, the importance of an object could be assigned in dependence on its functionality. Moreover, after the first exploration the importances could be rearranged in a different manner to see various parts of a scene more precisely than during the previous exploration.

If no additional information is provided and the user takes into account scene geometry only, then the size of object bounding box could be considered as the importance function:

$$q(\omega) = \max_{u,v \in V_\omega} |u_x - v_x| + \max_{u,v \in V_\omega} |u_y - v_y| + \\ + \max_{u,v \in V_\omega} |u_z - v_z|,$$

where V_ω is the set of vertices of the object ω.

It is also necessary to introduce a parameter characterizing the predictability of an object: $\rho_\omega : \Omega \to \mathbb{R}^+$. In other words, the parameter determines the quantity of object surface to be observed in order to well understand what the object looks like. If an object is well-predictable, then the user can recognize it even if he sees its small part. The bad predictability forces the user to observe attentively all the surface.

We propose to consider the function $f(t) = \frac{\rho_\omega + 1}{\rho_\omega + t} t$ as the measure of observation quality for an object, where $0 \leq t \leq 1$ is the explored fraction of the object (for example, the area of the observed surface divided by the total area of the object surface). Refer to figure 3 for an illustration. If the percentage t for the object ω is equal to zero (the user has not seen the object at all), then $f(t)$ is zero (the user cannot recognize the object). If all the surface of the object ω is observed, then $f(t)$ is 1, the observation is complete.

If nothing is known about a scene except its geometrical representation, then in order to observe it, the parameter ρ could be taken as rather small value, for example, $\rho_\omega \equiv 0.1 \forall \omega \in \Omega$. In such a case even exploration of a part of an object gives a good comprehension.

Now let us suppose that there exists some additional knowledge, for example, a virtual museum is considered. Then for all the paintings the parameter could be taken equal to 1000 and, for all the walls, chairs, doors equal to 0.1. Now, in order to get a good comprehension of a painting, one should observe all its surface, but only a small part of door or wall is necessary to recognize them.

Let us consider a viewpoint p. For scene objects it gives a set of values $\Theta(p) = \{0 \leq \theta_{p,\omega} \leq 1, \omega \in \Omega\}$, where $\theta_{p,\omega}$ is the fraction of visible area for the object ω from the viewpoint p. $\theta_{p,\omega} = 0$ if the object is not
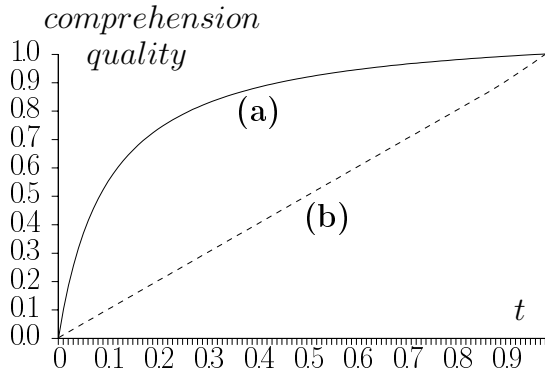
Figure 3: The behavior of the function $f(t) = \frac{\rho+1}{\rho+t} \cdot t$ for two values of the parameter ρ. **(a)** $\rho = 0.1$, even a part of an object provides a good knowledge. **(b)** $\rho = 1000$, the user should see all the object to get a good knowledge.

visible and $\theta_{p,\omega} = 1$ if one can see all its surface from the viewpoint p.

The fraction $\theta_{p,\omega} = 0$ may be computed in various ways. The simplest one is to divide the area of the visible surface by the total area of an object. A bit more complicated way is inherited from the middle-level method (see equation 3). If we divide the curvature of the visible surface by the total curvature of an object, we obtain the fraction equal to 0 if an object is not visible at all and equal to 1 if we could see all its surface.

Thus, we propose to evaluate a viewpoint quality as a sum of scene object importances with respect to their visibility:

$$Q(p) = \sum_{\omega \in \Omega} q(\omega) \cdot \frac{\rho_\omega + 1}{\rho_\omega + \theta_{p,\omega}} \theta_{p,\omega}. \qquad (4)$$

4 AN EXAMPLE OF APPLYING THE NEW TECHNIQUE

In this section, the computer model, that we have met before, is considered more precisely. Figures 4 and 5 show the results of applying two techniques from the middle and the high level classes. For the new method no additional information is provided, so, the bounding box sizes are taken as the importance function $q(\omega)$ and $\rho_\omega \equiv 0.1 \forall \omega \in \Omega$.

The best viewpoints, chosen with the two methods, are quite close (the picture is shown at figure 6), but there are significant differences in the estimation of other view directions.

Compare figure 2, showing the scene from the viewpoint number 10, and figure 6, showing it from the best viewpoint. It is clear that the viewpoint 10 is less attractive, but it still gives a good representation
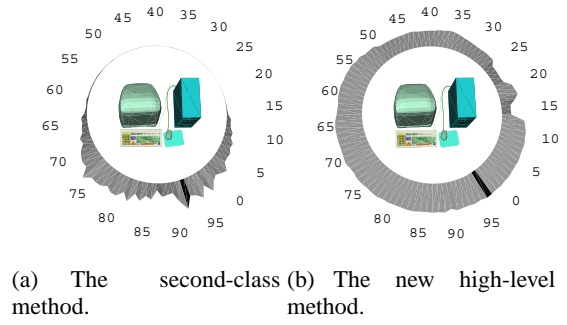

(a) The second-class method. (b) The new high-level method.

Figure 4: The qualities for 100 viewpoints equally distanced from the center of the model. The best viewpoint is shown by the black sector.

(a) The second-class method. (b) The new high-level method.
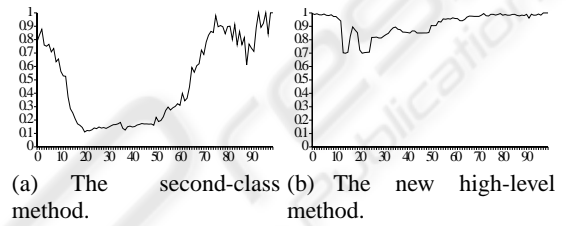
Figure 5: The plain graphs of the qualities for 100 viewpoints around the scene (see figure 4).
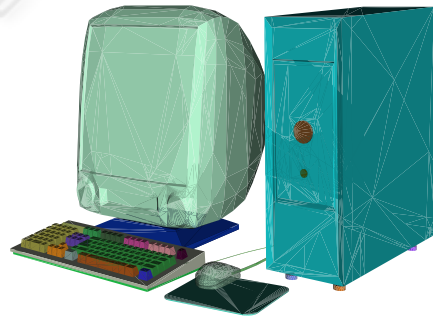

Figure 6: The best viewpoint for the computer model.

of the scene. The function on figure 5(b) decreases smoothly in this region, while figure 5(a) shows a drastic fall. At the viewpoint 17 the function from figure 5(b) grows, because a back side of the display and a part of the keyboard are visible simultaneously. Then it decreases again because the case covers the keyboard. The new method also shows a better quality than the old one from the back side of the scene. From each viewpoint some parts of the mouse or of

the keyboard are visible, so the estimation should not be so small as at figure 4(a).

5 EXPLORING A SCENE

The viewpoint quality estimation is only the first step in the domain of the scene understanding. In order to help a user to get a good knowledge of a scene, methods, allowing to choose a best viewpoint (or a set of viewpoints), should be proposed. Dynamic exploration methods could be very helpful too, since a set of static images is often not sufficient for understanding of complex scenes.

There are two classes of methods for virtual world exploration. The first one is the global exploration class, where the camera remains outside the explored world (see figure 10). The second class is the local exploration. In this class the camera moves inside a scene and becomes a part of the scene. Local exploration may be useful and even necessary in some cases, but only global exploration could give the user a general knowledge on a scene. In this section we are mainly concerned with global exploration of fixed unchanging virtual worlds. But it should be said that interesting results have been obtained with local exploration techniques in some works.

There are few works dedicated to the problem of virtual world exploration. Based on the definition of good viewpoint (Plemenos and Benayada, 1996), Barral et al. in (Barral et al., 2000b) present an incremental method for automatic exploration of objects or scenes. The technique does a global exploration of a scene, i.e. it creates a "movie" with a camera, whose trajectory lies on a sphere, surrounding the scene.

Marchand and Courty in (Marchand and Courty, 2000) have presented the general framework that allows an automatic control of a camera in dynamic environment. The method is based on image-based control approach.

Vázquez et al. in (Vazquez et al., 2001) present a measure, the viewpoint entropy, based on Shannon's entropy. Then they propose the extension of the method given by Barral et al. in (Barral et al., 2000a).

In this section a non-incremental method of global scene exploration is presented. Since we would like to explore the exterior of a scene, it is reasonable to restrict the space of feasible viewpoints to a surrounding sphere. Moreover, a viewpoint quality is quite smooth function, so the sphere could be easily discretized. Thus, the scene is placed in the center of the sphere, whose discrete surface represents all possible points of view.

Having the viewpoint quality criterion and a rapid algorithm for visibility computations (refer to (Sokolov and Plemenos, 2005)), we are ready to choose good views. The main idea of the method is to find a set of viewpoints, giving a good representation of a scene, and then to connect the viewpoints by curves in order to get a simple path on the surface of the sphere — trajectory of the camera. The views should be as good as possible and the number of views should not be too great. These criteria are satisfied with a greedy search scheme. Let us give a more strict formulation.

Let us suppose that two sets are given for a scene: a set of faces $F = \{f_i, 1 \le i \le n_f\}$ and a set of vertices $V = \{v_j, 1 \le j \le n_v\}$. The scene disjunction into a set of objects is supplied: $\Omega = \{\omega_k, 1 \le k \le n_\omega\}, V = \bigcup_{k=1}^{n_\omega} \omega_i, k \ne l \Rightarrow \omega_l \bigcap \omega_k = \emptyset$. For each viewpoint s of the discrete sphere S the set of visible vertices $V(s) \subseteq V$ is given.

Let us denote the curvature in a vertex $v \in V$ as $C(v)$ and the total curvature of a mesh $V_1 \subseteq V$ as $C(V_1) = \sum_{v \in V_1} C(v)$. We suppose that all objects in Ω have non-zero curvatures.

In addition to equation 4, let us introduce the quality of a set of viewpoints:

$$Q(S_1 \subseteq S) = \sum_{\omega \in \Omega} q(\omega) \cdot \frac{\rho_\omega + 1}{\rho_\omega + \theta_{S_1, \omega}} \theta_{S_1, \omega},$$

where $\theta_{S_1, \omega} = \frac{C(V(S_1) \bigcap \omega)}{C(\omega)}, V(S_1) = \bigcup_{s \in S_1} V(s)$. Since the camera remains outside the scene and always points to the center of the sphere, there is no need to define the view angle.

A set of viewpoints, giving a good scene representation, could be obtained by a greedy search. The greediness means choosing the best viewpoint at each step of the algorithm. More strictly: having given a threshold $0 \le \tau \le 1$, one should find a set of viewpoints $M_k \subseteq S$ such as $\frac{Q(M_k)}{Q(S)} \ge \tau$. At the beginning $M_0 = \emptyset$, each step i of the algorithm adds to the set the best viewpoint s_i: $Q(M_{i-1} \bigcup \{s_i\}) = \max_{s \in S} Q(M_{i-1} \bigcup \{s\}), M_i = M_{i-1} \bigcup \{s_i\}$.

Figure 7 shows the amount of acquired information in dependence on the number of algorithm steps. It is easy to see that often only few viewpoints are necessary to get a good knowledge of a scene.

The next question is: if the camera has to move from one viewpoint to another, what path on the sphere is to be chosen? A naive answer is to connect the viewpoints with a geodesic line, the shortest one. This preserves the camera from brusque changes of trajectory during traversal from one point to another and gives the shortest solution, but acute angles still could appear in control points of trajectory. Such connection does not guarantee that the path consists of good viewpoints. This drawback is serious, and, in
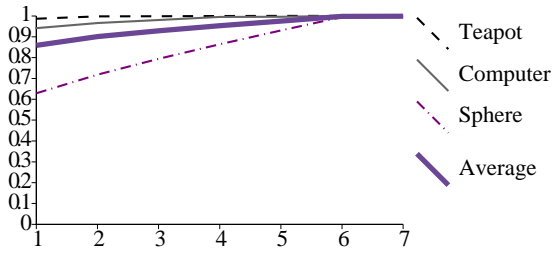
Figure 7: The amount of acquired information in dependence of number of viewpoints selected by greedy search for the Utah teapot model (figure 11), the sphere with imposed objects (figure 12) and the computer model (figure 6).

order to avoid it, we have to introduce additional costs and discounts.

The main idea is to make the distances vary inversely to the viewpoint qualities. It means the augmentation of path's length if it contains bad viewpoints and the reduction of the length otherwise. For example, it can be done in the following way: if two vertices s_1 and s_2 are adjacent in a sphere tessellation, then the new distance between s_1 and s_2 is calculated with the formula:

$$\hat{d}_{s_1,s_2} = ||\vec{s_1} - \vec{s_2}|| \cdot c_q(s_1,s_2), \qquad (5)$$

where $||\vec{s_1} - \vec{s_2}||$ is the Euclidean distance between points s_1 and s_2 and $c_q(s_1,s_2) = 1 - \frac{Q(s_1)+Q(s_2)}{2\max\limits_{s \in S} Q(s)}$ is the discount that forces the camera to pass via "good" viewpoints.
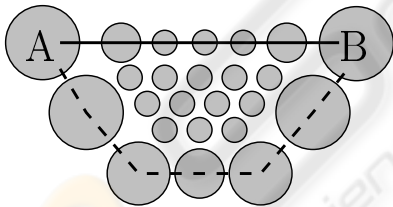

Figure 8: The reason to change the metric. The circles represent viewpoints: larger circles denote better viewpoints. The solid line shows the geodesic line between viewpoints A and B, the dashed line shows the shortest path according to the new metric.

This empiric formula augments distances near "bad" viewpoints and reduces near "good" ones. The reason is shown at figure 8. Figure 9 gives an example. It is easy to see that the camera trajectory presented at figure 9(b) brings to a user more information than the shortest one.

Now, having defined the metric and having found the set of viewpoints, we would like to determine a trajectory of the camera. It is not hard to construct a complete graph of distances $G = (M_k, E)$, where the
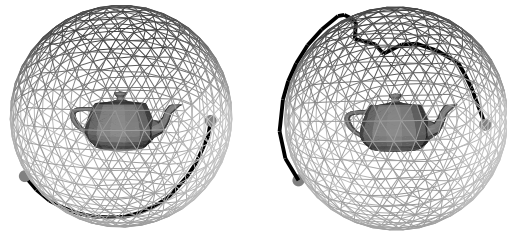
(a) Shortest line connecting two viewpoints. (b) Shortest line with respect to the viewpoint qualities.

Figure 9: The trajectories between two selected points on the surface of the surrounding sphere.

weight of an arc $(v_1, v_2) \in E$ is equal to the metric between the viewpoints v_1 and v_2 (equation 5).

Now the trajectory could be computed as the shortest Hamiltonian path (or circuit, if we would like to return the camera to initial point). The problem is also known as the *travelling salesman problem* (TSP). Unfortunately, the TSP problem is NP-complete even if we require that the cost function satisfies the triangle inequality. But there exist good approximation algorithms to solve the problem. Moreover, often $|M_k|$ is rather small (see figure 7), and the problem in such a case could be solved even by the brute-force algorithm in real-time.

6 EXPLORATION EXAMPLES

Figures 10 and 11 show camera trajectories for the Utah teapot model. The first one is obtained by applying the incremental technique with the viewpoint entropy as the quality heuristic, and the second one is obtained by our method.

Both of them show 100% of the surface of the teapot model. The new method could give brusque changes of the trajectory, and the old one is free of this disadvantage. A simple way to smooth the trajectory is to construct a NURBS curve. Control points for the curve are to be taken from the approximation of the minimal set of viewpoints, and its order is to be defined by solving the TSP task. The new technique gives significantly shorter trajectories, and this advantage is very important.

One more example of the new method application is shown at figure 12. This model is very good for exploration technique tests, it represents six objects imposed into the sphere with holes, and the explorer should not miss them.
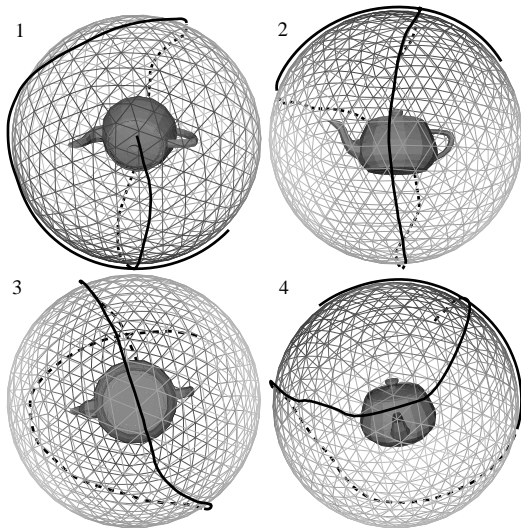
Figure 10: The exploration trajectory for the Utah teapot model. The trajectory is computed by the incremental method using the viewpoint entropy as the quality heuristic. Images are taken consequently from the "movie", first one is the best viewpoint.
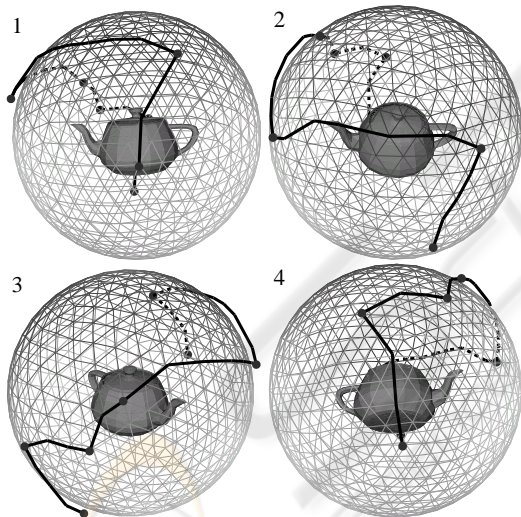

Figure 11: The exploration trajectory for the Utah teapot model, obtained with the new technique. Images are taken consequently from the "movie". Black knots are the control points of the trajectory, i.e. an approximation of the minimal set of viewpoints sufficient to see all the surface of the teapot model.

7 CONCLUSIONS AND FUTURE WORKS

In this paper a new criterion of viewpoint quality evaluation is presented. The criterion introduces a new level of a scene comprehension methods, so-called
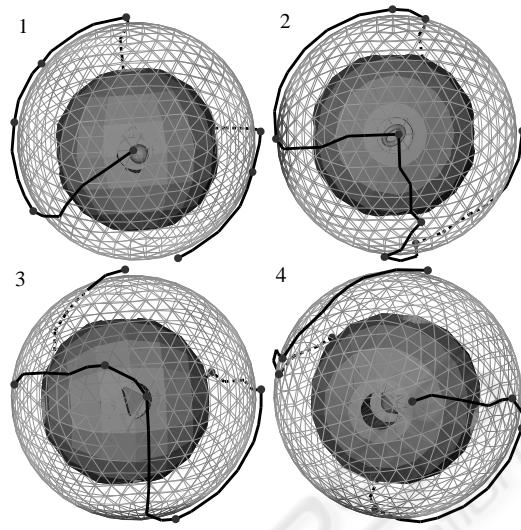

Figure 12: The exploration trajectory for the sphere with several embedded objects. Images are taken consequently from the "movie", black knots are the control points of the trajectory.

high-level methods.

A new non-incremental method of global scene exploration is also presented.

In the future we shall investigate, what kinds of additional knowledge can be efficiently used in scene exploration. For example, if we have a room with paintings, it is more reasonable to pay more attention to the objects of art than to walls and chairs.

It would be interesting to develop automatic methods of scene exploration allowing interaction with the user, where the user can point at parts of a scene he would like to explore in details.

ACKNOWLEDGMENTS

REFERENCES

Barral, P., Dorme, G., and Plemenos, D. (2000a). Intelligent scene exploration with a camera. In *International Conference 3IA'2000*, Limoges (France).

Barral, P., Dorme, G., and Plemenos, D. (2000b). Scene understanding techniques using a virtual camera. In *Eurographics'2000*, Interlagen (Switzerland).

Chazelle, B., Dobkin, D. P., Shouraboura, N., and Tal, A. (1995). Strategies for polyhedral surface decomposition: an experimental study. In *SCG '95: Proceedings*

of the eleventh annual symposium on Computational geometry, pages 297–305, New York, NY, USA. ACM Press.

He, L.-W., Cohen, M. F., and Salesin, D. H. (1996). The virtual cinematographer: a paradigm for automatic real-time camera control and directing. In *Proceedings of SIGGRAPH'96, Annual Conference Series*, pages 217–224.

Kamada, T. and Kawai, S. (1988). A simple method for computing general position in displaying three-dimensional objects. *Comput. Vision Graph. Image Process.*, 41(1):43–56.

Marchand, E. and Courty, N. (2000). Image-based virtual camera motion strategies. In Fels, S. and Poulin, P., editors, *Graphics Interface Conference, GI'00*, pages 69–76, Montreal, Quebec. Morgan Kaufmann.

Plemenos, D. and Benayada, M. (1996). Intelligent display in scene modelling. new techniques to automatically compute good views. In *GraphiCon'96*, Saint Petersburg (Russia).

Sokolov, D. and Plemenos, D. (2005). Viewpoint quality and scene understanding. In Mudge, M., Ryan, N., and Scopigno, R., editors, *VAST 2005: Eurographics Symposium Proceedings.*, pages 67–73, ISTI-CNR Pisa, Italy. Eurographics Association.

Vazquez, P. P., Feixas, M., Sbert, M., and Heidrich, W. (2001). Viewpoint selection using viewpoint entropy. In *VMV '01: Proceedings of the Vision Modeling and Visualization Conference 2001*, pages 273–280. Aka GmbH.

Zuckerberger, E., Tal, A., and Shlafman, S. (2002). Polyhedral surface decomposition with applications. *Computers & Graphics*, 26(5):733–743.