

Personalised 3D navigation and understanding of Geo-referenced Scenes

Nikolaos Doulamis

National Technical University of Athens
9 Heroon Polytechniou Str. 15773, Zografou, Greece
Email: {ndoulam@cs.ntua.gr}

Christos Yiakoumettis, Georgios Miaoulis

Technological Education Institute of Athens
Ag. Spyridonos St., 122 10 Egaleo, Greece.
Email: {chgiak@teiath.gr, gmiaoul@hol.gr}

Abstract— Google Maps and Google Earth were milestones in the evolution of Geographic Information Systems (GIS). A new era of 3D virtual geo-referenced worlds began. The need of exploration and creation personalized virtual tours is imperative. This work implements a best viewing algorithm for presenting a 3D model by considering both semantic and geometric features of the 3D model. The camera trajectory is estimated by respecting the user's preferences using a personalized entropy-based measurement. Moreover the camera speed is adjusted properly based on the complexity of the projected view of the model.

Keywords— *personalized path planning, 3D scene exploration, virtual tour, scene understanding, knowledge extraction, camera speed estimation*

I. Introduction

GIS services have recently come in the forefront of people's interest. Existing GIS have spatial reference on a real or on a virtual world. It is imperative the use of a system that can evaluate a number of parameters, design a path and navigate on a three dimensional environment presenting all the interesting points and important information of a 3D model based on user's preferences. Applications of such systems can be found on any large scale scenes based on real geographic data or on simulators, or even on imagine virtual worlds. In addition the above navigation techniques can be applied on any three dimensional space, such as presentations of 3D models of house or contractions.

In this paper, it is presented an automatic approach of virtual tour around a 3D model based on geometric and semantic metadata by respecting the user's preferences. Furthermore focus will take place on the camera speed estimation, considering the projected image complexity of the 3D models.

II. State of the Art

This paper is dealing with personalized virtual tour/presentation of 3D objects. According to the literature review there are two different approaches on camera's trajectory estimation for providing an optimal solution on scene exploration. There are approaches that are based on geometric features of the scene and some others are based on semantic ones. These methodologies are known as best

viewing algorithms. Still there are no best viewing algorithms that combine camera speed estimation with geometric and semantic features of the scene

A. Best Viewing Algorithms

The exploration of a virtual environment on behalf of the users can be classified in three distinct categories: (a) navigation, presenting data from the surface of the world, (b) presentation of an object or a path and (c) presentation of an object or item from specific views in order to reveal as much information as possible [9].

In the area of best viewing algorithms for individual objects or scenes, several approaches have been implemented in the literature. An efficient approach for the calculation of the best view of a 3D scene has been presented by Sokolov and Plemenos [5][6][7][8]. It is based on the coverage of the surface of the scene. Each projected view is independently evaluated based on its complexity, measured as the total number of faces projected onto the 2D plane. In particular, [5], [14] optimizes the area of the visible surface over the sum of the area of the surface of the whole scene. Similar methodologies have been implemented by Tab and Vazquez [9], [10]. Instead Vazquez introduces entropy metric that measures the projection complexity; views that maximize this information are selected as the best ones.

Recent work of ours, has implement a best viewing algorithm that combines both geometric and semantic features of the scene. According to our approach each viewing angle is evaluated and ranked [13].

Finally, relative methodologies of moving the camera on a virtual sphere around an object have been published. In these approaches the object can be watched closer, either by user's intervention or by the system automatically [9].

B. Image Complexity

Image complexity is relevant to the time that an 'observer' requires in order to understand the content of the image. The more complex an image is, the more time is required. Image complexity measure can be defined upon the semantic and the syntax of the image [3].

According to Prkio and Hyvarinen a natural measure of image complexity can be defined by the entropy of the discrete probability distribution of Shannon [3],[4].

A well-known approach of measuring image complexity is based on edge level of image [1], [2]. According to literature review the complexity of an image can be measured with a fuzzy based approach and it can be classified in discrete values such as ‘Little Complex’, ‘More or Less Complex’ and ‘Very Complex’ [1], [2].

C. Related Architectures

The current scene understanding and exploration approaches of 3D scenes are using either geometric features of scene [5], or semantic ones [10]. Previous work of ours has implemented a new approach that combines both geometric and semantic features in a single architecture [13].

The work of Sokolov and Plemenos focuses only on best view algorithm by optimizing the visible area of the 3D scene [5]. According to this approach, the semantic features of the scene are not considered in the camera’s trajectory estimation. Moreover, it is successfully provides an optimal camera trajectory based on geometry of the scene, but the generated trajectory doesn’t consider the user’s preferences

On the other hand the work of Vazquez [10] considers only semantic features of the scene and it doesn’t consider the geometric features of the scene and the user’s preferences.

The main differences of the aforementioned research approaches with the one presented are:

- 1) the use of a 3D navigation that enables users’ personalized 3D experiences,
- 2) a best viewing algorithm that considers both semantic and geometric features of the scene,
- 3) the genetic optimization algorithm that respects the user’s preferences on the camera’s trajectory estimation,
- 4) the estimation of the camera speed based on the complexity of the projected face of the viewing angle of the 3D model of the scene,
- 5) the integration of the aforementioned innovative aspects in a single platform in order to offer the aforementioned facilities in real-world cases.

III. Contribution.

In this paper, a 3D personalized scene exploration and understanding architecture is proposed. The architecture automatically provides the optimal personalized camera trajectory around a 3D model to the user that fully respects user’s preferences. To accomplish this, initially the 3D scene has been enriched with semantic features (metadata). An optimization process is adopted either to estimate the best preferred routes based on a combined maximization of several cost functions that reflect geometric, thematic and user-centric criteria or to recommend a set of 2D views from the selected 3D objects within the itinerary that simultaneously optimize geometric constraints and user’s preferences [13]. Finally the camera speed of the calculated trajectory is estimated based on the projected complexity of the viewing angle of the scene.

In contrast with the current approaches dealing with personalized 3D scene understanding and exploration approach, our architecture incorporates a personalized best viewing algorithm that considers both semantic and geometric features of the scene and estimates the camera speed based on the complexity of projected image of the scene.

The main contributions of this work are summarized in the following:

1. Optimal 3D scene exploration is accomplished based on complex cost functions that measures the quality of the projected viewing angle of a 3D scene.
2. The complexity of the projected viewing angles is considering in order estimating the required time that a user needs to have in order to understand and extract the information of the content of the presented 3D model. The camera speed is adjusted on the calculated path considering the projected complexity of the texture of the image, by giving to the user the required time to observe the scene.
3. A hierarchical framework is adopted for route planning estimation. The framework comprises all the above components in a single algorithm which is entitled as “micro-path” in this paper.

IV. System Overview and Basic Requirements

In this section, we introduce a new best viewing algorithm, used in our proposed architecture. This approach is based on previous work of ours [11],[12],[13]. An existing algorithm has been adopted and modified in a way to consider semantic metadata of 3D objects for navigation in 3D scene. Moreover the texture image complexity of the 3D objects is considered in order to estimate efficiently the camera speed of the 3D trajectory. This approach is able to present a 3D object in a unique way, based on user’s preferences and on special properties of the 3D object.

This implies that we need to implement a methodology for selecting the best view of a 3D model (building). An entropy-based oriented scheme has been adopted in this work to estimate the best view selection. This process is called ‘Micro Path’ layer and it can be divided in parts (see Fig. 1)

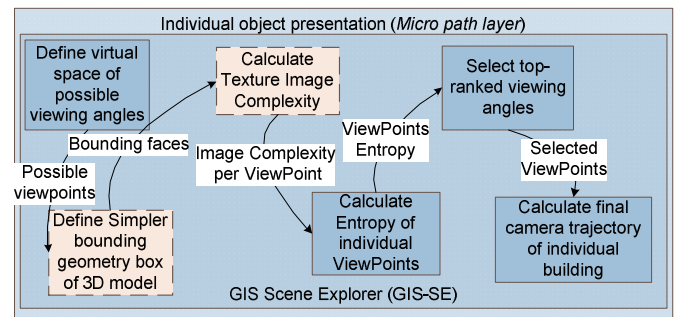


Fig. 1: An overview of the proposed personalized route planning architecture

Best Viewing for Micro Path Creation: The focus of this paper is on the creation of the micro-path, which is a

methodology for best viewing of the selected 3D objects. Our approach consists of six individual processes. A virtual space is defined of all possible viewing angles of the 3D model. Afterwards a simpler 3D representative of the object is created.

Then the complexity of the texture of 3D model is considered. Based on the complexity of the texture, an indicator is defined. The complexity indicator represents the time that requires a 'visitor' to understand the conceptual content of the texture. The complexity of the image measure is independent of the human perception criteria [1],[2]. An indicator that represents the image complexity is used. As the image complexity measure increases, more time is required by a 'visitor' to understand the conceptual content of the object. Then, as the time of understanding increases, the speed of the camera's trajectory decreases.

We assume that the real 3D model and the bounding box are the same. Moreover, the complexity of the texture of the image is calculated once during the initial 3D object modelling process. Generalizing the process, the sequence of the processes has to be as presented in Fig. 1. Afterwards, an entropy based optimization methodology is adopted so that the angle that maximizes the projection entropy is selected for the navigation around an object.

A. Definition of Semantic 3D Model Specifications

Each 3D model of a building was characterized by a set of descriptors. These features will help us to answer questions such as, what type of buildings they are, or, what they are used for, or what the purposes are for their existence or for visiting them. Moreover, descriptors that describe the outside features of the buildings were added [13].

Due to the complexity of the 3D model of the of a real building, it was assumed that each 3D model, no matter the individual shape and size, could be considered as a rectangular cube. In this way each 3D model is contained in a bounding box. There was an assumption that all the 3D models were touching the ground with one of their faces. A bounding box with 6 faces was used to represent the geometry of the 3D model. The features were applied in each individual bounding box side. Each 3D model has five bounding sides, four of them are around the building and the last of them is on the roof of it (see Fig. 2).

All the figures of the model, have been assigned to the 5 faces (ground face has been excluded) of this simpler geometrical model. Moreover, an indicator of image texture complexity of each face has been assigned to each face of the bounding box, based on real textures of the 3D building model. The estimation of the image texture complexity is calculated based on classification of image complexity introduced by Chanon and Alma [1],[2].

B. Definition of Virtual Space of Possible Viewing Angles

A geometrical figure of semi sphere was used in order to define the virtual space of all possible viewing points. The virtual semi sphere is adjusted every time, according to 3D model geometrical requirements. For simplifying our

experiments we assume that every 3D model is far enough of each other, in a way not to exclude points of the virtual semi sphere.

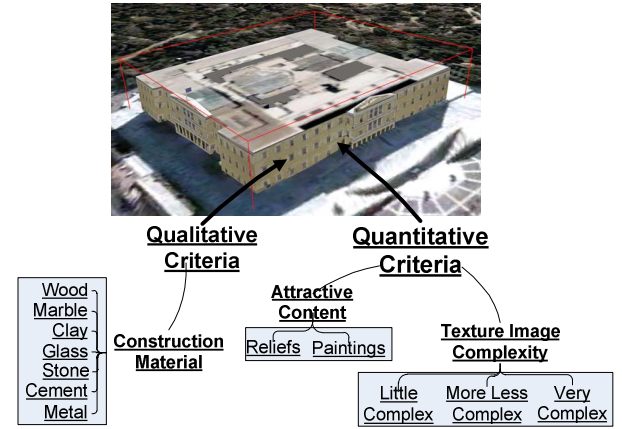


Fig. 2: Bounding side features

C. Entropy-Based View Selection.

According to Shannon's theory [4], the entropy of a discrete random variable X with values in the set $\{a_1, a_2, \dots, a_n\}$ is given by

$$H(X) = -\sum_{i=1}^n p_i \log_2 p_i \quad (1)$$

where $p_i = \Pr(X = a_i)$ is the probability of the random variable X to take the value of a_i and the $\log_2 p_i$ the 2-base logarithm of p_i . For continuity reasons the $p_i \log_2 p_i = 0$ in case of $p_i = 0$.

$$S(u) = \sum w_u^j f_u^j \quad (2)$$

In equation (2), f_u^j denotes the j -th feature element for object u , while w_u^j the respective weight. Equation (2) means that the most important feature elements with respect to current user's information needs are characterized by greater weight values in contrast to feature elements of low significance to which low weight values are assigned.

We assume that all the weight values have been properly rectified by an ontological schema [13]. The weight element w_u^j expresses the degree of importance for the j -th feature element of the u -th 3D object. Since the j -th feature element refers to one of the multiple faces that 3D object is decomposed to, we can derive a personalized entropy using the values of w_u^j . In particular, let us denote as $\psi(u, f)$ an operator that returns the weight indices which are related with those metadata that describe the f -th face of the u -th object. For example, let us assume for a 3D object that the material of one face consists of gypsum and the material of the other face of marble. Then, operator $\psi(\cdot)$ returns the weight indices

referring to the marble feature element only in case that the specific face has this property. Therefore, the personalized entropy for a particular object u is defined as by the equation (3).

$$H(\xi, u) = - \sum_{i=0}^k \frac{NP_i}{NP_{total}} \cdot \frac{\sum_{j=1}^n w_{u,i}^j}{\sum_{j=1}^n w_{u,i}^j} \log \left(\frac{NP_i}{NP_{total}} \cdot \frac{\sum_{j=1}^n w_{u,i}^j}{\sum_{j=1}^n w_{u,i}^j} \right) \quad (3)$$

Using equation (3), we can estimate the best object view that maximizes the entropy $H(\xi, u)$. The main problem for performing the aforementioned maximization is that the number of projections ξ is infinity. To overcome this difficulty, we evaluate $H(\xi, u)$ over discrete (limited) viewpoints on the spherical coordinate system that encloses the 3D object. The spherical coordinate system and the discrete points used to estimate best object views is depicted in Fig. 3 around a 3D object of a church.

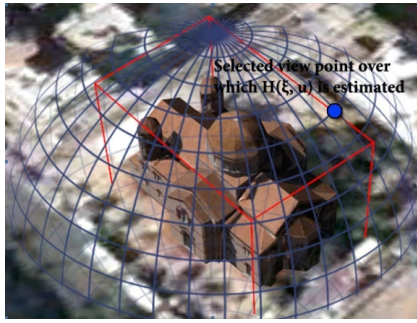


Fig. 3: The spherical coordination system used to estimate the best micro-path view through the maximization of the entropy $H(\xi, u)$

Table 1 shows the main algorithm of the viewpoint entropy estimation.

ViewPoints Entropy Estimation
1. Set azimuth angle $\Phi \leftarrow 0$ and inclination angle $\Theta \leftarrow 0$
2. Set $Best_View \leftarrow Minimum_Value$
3. For $\Theta=0$ to 90 , $\Theta+=Step$ do
a. For $\Phi=0$ to 360 , $\Phi+=Step$ do
i. $Entropy \leftarrow Equation(4)$
ii. If ($Entropy > Best_View$)
1. $Best_View \leftarrow Entropy$
$Best_Phi \leftarrow \Phi$, $Best_Theta \leftarrow \Theta$

Table 1: The main steps of the Micro Path Estimation Algorithm (ViewPoint Entropy)

D. Estimation of the Camera Speed

Besides the entropy we introduce an indicator for texture image complexity. Each face of a bounding box has been characterized by a discrete value 'Little Complex', 'Less or More Complex' or 'Very Complex', represented by numerical

values from zero to two. It is clear that the more complex a scene is the lower the camera speed is; thus complexity of the scene and camera speed is inverse proportional. Equation (4) expresses the measure regarding image texture complexity, as a function of entropy. The feature of image complexity It is calculated during the creation or reconstruction of the 3D model, for each face of it. It is calculated based on

$$C(\xi, u) = - \sum_{i=0}^k \frac{I_i}{4\pi} \log \frac{I_i}{4\pi} \quad (4)$$

where I_i is the image complexity of face i , while ξ denotes the view (2D plane) over which the 3D object is projected onto. Table 2 shows the main algorithm of the camera speed estimation based on the texture image complexity.

Camera Speed Estimation based on Texture Image Complexity
4. $Camera_Speed \leftarrow Normal_Speed$
5. For each Best_Phi as Visiting_Point
a. $Image_Complexity \leftarrow Equation(5)$
b. Switch (Image_Complexity)
i. Case $\geq Very_Complex$
6. $Camera_Speed \leftarrow Extremely_Low_Speed$
a. Case $\geq More_Less_Complex$
i. $Camera_Speed \leftarrow Low_Speed$
b. Case $\geq Little_Complex$
$Camera_Speed \leftarrow Normal_Speed$

Table 2: The main steps of Camera Speed estimation

E. Calculation of Final Camera's Trajectory

This is the final step of the Micro path component. All the data are gathered and the camera trajectory is calculated from the top ranked viewpoints. The final path respects the requirements of the "Macro Path" by defining the starting and ending point of the camera trajectory in order to create a smooth overall path of all visiting points (scenes) on the whole 3D GIS platform.

In order to understand the results of the personalized scene understanding algorithm, several use case scenarios will be presented. In our implementation, each projected view of the 3D model is calculated on discrete points around the semi spherical coordination system (see Fig. 3 for more details). Then, for the given set of selected discrete view points, the ones that maximize the overall personalized entropy metric for all the 5 faces are selected as the most appropriate for the user. In other words, the personalized entropy metric includes not only geometric properties of the projected view but, in addition, the importance of a face with respect to user's preferences.

We assume that the user desires to visit Athens for sightseeing. Moreover, it is assumed that a user has already created a profile using the active learning component of GIS-SE. Macro path component provide the list of selected visiting points (sightseeing).

Among the selected buildings, Parthenon was chosen. As it is observed, Fig. 4(a) indicates the projection that maximizes the total number of projected points of the view. Instead, Fig. 4(b), (c) and (d) indicate the selected view where additional criteria (apart from the total number of projected points on the 3D image plane) that capture user's preferences are incorporated in the best view selection.

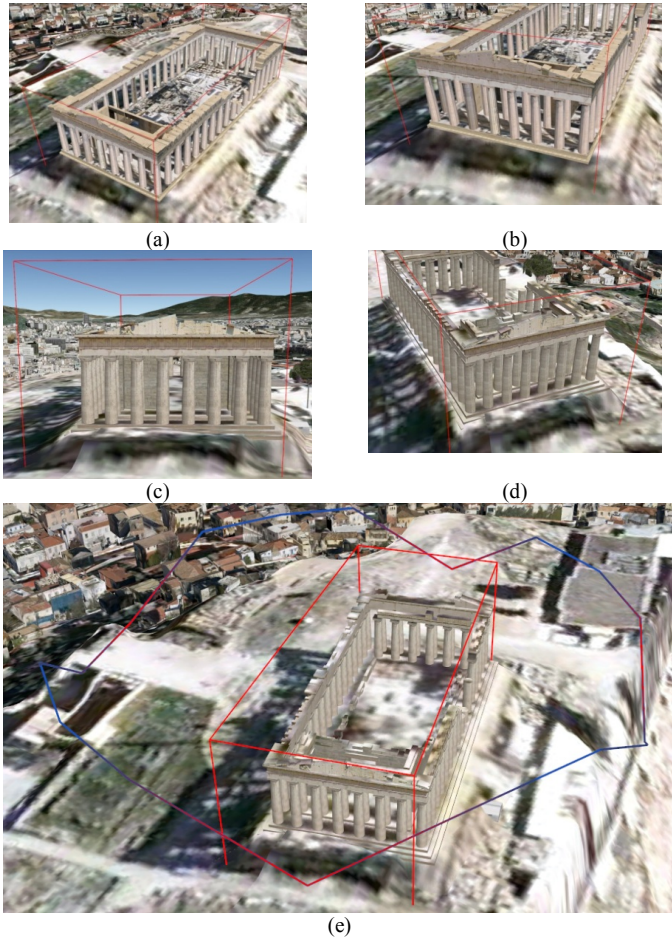


Fig. 4: (a) Best view selected using only geometrical properties. (b)(c)(d) Best view selected using geometrical and human factors. (e) Indicating camera trajectory and the speed of it for the selected viewpoints

Parthenon is included in a bounding box, each of the four surrounded faces of the bounding box have the same vector values ("Made from Marble", "Provides Historical Education", "Provides Cultural Education", "Is Accient", "Popularity"), but the upper face of the bounding box doesn't have "Made from Marble" (there is no roof). Besides the specifications of the 3D model, the two faces that contain the entrance of Parthenon has been defined as "very complex" image texture complexity and the rest rounded faces as "little complex". According to user's profile, "Made from Marble" vector is preferred by the user. The aforementioned case leads the best viewing "Micro Path" algorithm to rank higher viewing angles such as Fig. 4(b), (c) and (d) rather Fig. 4(a). Moreover the camera speed with be slower among viewing points of Fig. 4(b) to Fig. 4(c) and then to Fig. 4 (d) rather than any other viewing point of the 3D model.

In Fig. 4(e) top ranked viewpoints have been marked. The camera trajectory is presented by red-purple-blue line around the 3D model of Parthenon. The entropy is maximized when the camera is placed straight ahead of the bounding faces (of red bounding box). There are no selected viewpoints from the top of Parthenon because this it is not made from marble. The camera speed change has been marked by the change of colour from blue to red .Blue indicated that the camera speed is "Normal", purple means that the camera speed has to be "Low speed" and red means that the camera speed set to be "Extremely Low Speed".

v. Platform Implementation

The platform has been developed using the Google Earth Building maker tools. The buildings are reconstructed as 3D models, each face of which is described using a set of metadata. We assume that all the 3D objects have five faces (the faces that point to the ground are not considered), no matter of their 3D complexity. In particular, we represent all 3D objects as a "bounding-box" of five faces around the object as indicated in Fig. 3 where the 3D model of the national Gallery/Library of Athens is presented. The platform exploits the Google Earth plug-in and the Google Earth virtual environment to visualize the 3D geographical data. We have developed a front end interface to handle user's interactions with the system and a back end module, called GIS Scene Explorer (GIS-SE), which contains the learning component, the Macro-path and the Micro-path processes as described in the aforementioned sections. Since Google Earth does not allow retrieval and management of 3D geographical content, a separate geo-referenced database was used for this purpose in this work. The morphology and the texture of the ground were retrieved by the Google Earth's database to increase the overall scene visualization.

A. Best View Estimation – Micro Path Layer

The micro-path layer focuses on individualized representation of a selected 3D object according to the maximization of the personalized entropy metric. For simplification purposes, the best view has been estimated by exploiting a bounding box around the 3D model (see above for more details). Therefore, each 3D building is modelled using a 5-faces object (the last face corresponding to the ground is not taken into consideration in our algorithm). Fig. 3 presents the bounding box of a 3D building (in this example a church) as well as the spherical coordination system over which the best view is calculated.

Fig. 5 presents the results regarding a best view selection in case that a geometric criterion is used as well as in case that the proposed personalized entropy based metric is adopted. As is observed, Fig. 5(a) indicates the projection that maximizes the total number of the projected points of view. Instead, Fig. 5(b) indicates the selected view where additional criteria (apart from the total number of projected points on the 3D image plane) that capture user's preferences are incorporated in the best view selection.

The best view selection ("micro-path" layer) process was tested for many buildings in the examined 3D environment. In

the following, we evaluate the complexity required for estimating the best view of a 3D building. It is clear that the complexity is straightforwardly proportional to the number of examined discrete points on the semi-sphere bounds the 3D. The more the examined points on the semi-sphere coordination system are, the more the complexity is.

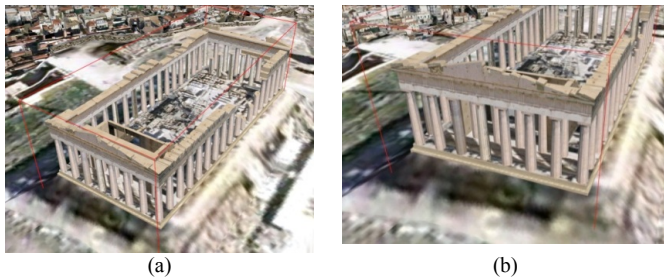


Fig. 5: (a) Best view selected using only geometrical properties. (b) Best view selected using geometrical and human factors

B. Tour Generation

The trajectory of the camera is calculated by estimating several points of the macro and micro path. Sequentially, connecting these points the camera trajectory is created. The camera trajectory is created in KML format and can be viewed using Google Earth, the same way a path is created using path creation tool of Google Earth.

A path in Google Earth is a KML file which contains an array of points. Each point is defined by three coordinates in the three dimensional space of Google Earth. The coordinates for longitude and latitude are expressed in the global coordinate system. The altitude can be expressed as an absolute value relevant to the sea level or to the ground level. According to our implementation requirements, the altitude is set to be always relevant to ground level.

Moving forward, our goal is to generate the camera trajectory. The tour is generated in KML format too. It is composed by a list of points which is called "FlyingList". Each point is referred as "FlyingPoint" and it contains the coordinates of the camera in the 3D space of Google Earth. Moreover, it includes the horizontal and vertical orientation of the viewing angle. Finally it encloses the duration of the "flight" expressed in milliseconds to the next "FlyingPoint". For the smooth creation of the tour the flying mode of the whole tour is set to "smooth". The orientation of the camera is expressed in degrees, 360 degrees for the horizontal orientation and from 90 to -90 for the vertical orientation.

The exported KML file, according to the aforementioned methodology and calculations is containing the real tour into the three dimensional geo-referenced scene. This file can be viewed and watched by the Google Earth application and plugin.

VI. Acknowledgements

This paper is supported by the project E-Park, "Exploitation of new Technological Trends for payment and handling public parking" approved under the Interreg III Programme, Greek-Cypriot cooperation and funded from European Union and Greek National funds.

VII. Conclusions and Discussions

In this paper, was implemented an improved best viewing algorithm in order to automatically calculate the optimal camera trajectory around an 3D model, by considering both semantic and geometric features of the scene. The final step our aforementioned improved algorithm was to estimate the proper camera speed in a way to make the tour more enjoyable to the 'visitor'. Therefore the complexity of the projected image was considered in order to adjust the camera's speed. Another important aspect in a personalized 3D navigation is to determine the way that the preferred 3D objects should be projected onto 2D screens. In this work, in contrast to previous approaches, we introduce and implement an improved version of a personalized entropy metric best viewing algorithm [13]. All these components have been integrated in a single platform and comparisons have been conducted to demonstrate the outperformance of the proposed 3D scene understanding and exploration

References

- [1] M. Chacon, D. Alma and R. Rodriguez "Image Complexity Measure: a Human Criterion Free Approach," NAFIPS 2005, pp. 241-246, 2005.
- [2] M. Chacon, D. Alma and R. Sandoval "A Fuzzy Approach on Image Complexity Measure," pp.268-284, 2007.
- [3] J. Prkio and A. Hyvarinen "Modelling Image Complexity by Independent Component Analysis, with Application to Content-Based Image Retrieval," ICANN 2009, part II, Springer, pp.704-714, 2009.
- [4] C. E. Shannon "A Mathematical Theory of Communication," Bell System Technical Journal, Vol. 27, pp. 379-423, 623-656, 1948.
- [5] D. Sokolov and D. Plemenos "Viewpoint quality and scene understanding," VAST, Eurographics Symposium Proceedings, Pisa (Italy), pp. 67-73, 2005.
- [6] D. Sokolov and D. Plemenos "High level methods for scene exploration," J. of Virtual Reality and Broadcasting, Vol 3(12), 2006.
- [7] D. Sokolov and D. Plemenos "Intelligent scene display and exploration," International Conference Graphicon 2006, Novosibirsk Akademgorodok, Russia, 2006.
- [8] D. Sokolov and D. Plemenos "Virtual world explorations by using topological and semantic knowledge" The Visual Computer Journal, Vol.23, pp.173-185, 2007.
- [9] D. Tab, G. Robertson and M.Gzewinski, "Exploring 3D Navigation: Combining Speed-coupled Flying with Orbiting," Proc. of the SIGCHI conference on Human factors in computing systems, Vol. 3, pp. 418-425, 2001.
- [10] P. P Vazquez "On the selection of good views and its application to computer graphics," Ph.D thesis, Technical University of Catalonia, Barcelona, Spain, 2003.
- [11] C. Yiakoumettis, G. Bardis, G. Miaoulis, D. Plemenos, and D. Ghazanfarpour "A GIS platform for automatic navigation into Georeferenced Scenes using GIS Scene Explorer (GIS-SE)," Intelligent Computer Graphics 2010, springer, Vol. 321, pp. 105-222, 2010.
- [12] C. Yiakoumettis, G. Bardis, G. Miaoulis, D. Plemenos, and D. Ghazanfarpour "Collaborative Framework for Virtual Globe Based 3D City Modeling," Intelligent Computer Graphics 2010, springer, Vol. 321, pp. 165-184, 2010.
- [13] C. Yiakoumettis, N. Doulamis, G. Miaoulis, and D. Ghazanfarpour "Active Learning of User's Preferences Estimation Towards a Personalized 3D Navigation of Geo-referenced Scenes," Journal of Geoinformatica, in press, 2013.
- [14] J. Zhang, J. Gong, H. Lin, G. Wang, J. Huang, J. Zhu, B. Xu, J. Teng "Design and development of Distributed Virtual Geographic Environment system based on web services," Information Sciences Vol. 177(19), pp. 3968-3980, 2007.