3DSSE – A 3D Scene Search Engine
Exploring 3D Scenes Using Keywords

Anestis Koutsoudis  Konstantinos Stavroglou  George Pavlidis  Christodoulos Chamzas
akoutsou@ipet.gr  kstavrog@ipet.gr  gpavlid@ipet.gr  chamzas@ipet.gr

Cultural and Educational Technology Institute/‘Athena’ Research and Innovation Centre in Information, Communication and Knowledge Technologies, 58 Tsimiski street, Xanthi 67100, Greece

Abstract

The evolution of real time 3D graphics technologies in combination with high bandwidth Internet connections and modern Web browsers enable users to explore complex 3D scenes. As a rule, a virtual visitor has to manually explore the geometrically complex 3D model in order to discover points of interest. This manual exploration is a time consuming process that, in some cases, can be assisted by sets of predefined points of interest. In this paper, we propose the annotation of 3D scenes in order to equip the user with a text based 3D scene search engine. The search engine provides a query mechanism that unburdens the user from the time consuming process of manually exploring vast 3D scenes. It responds to queries by exploiting the metadata of each 3D model and returns textual and visual information along with a group of links that correspond to relative points of interest within the 3D scene. The search engine allows the virtual visitor to automatically be transferred to a specific point of interest. We have built a Web accessible prototype system that is able to handle queries related to historical data, topological relationships and architectural properties of buildings. A number of 3D reconstructions covering urban areas of cultural importance located in Northern Greece have been annotated and used in the search engine as case studies. The prototype system is based on open source technologies and on a hybrid metadata schema that is derived from the MIDAS Heritage and MACE schemas.

Keywords: 3D reconstructions, annotation, search engine, viewpoint, virtual reality, metadata, native XML, database

1. Introduction

Nowadays, high accuracy 3D digitisation of cultural heritage monuments is considered a common practice [1]-[19]. Modern real time 3D graphics accelerators in combination with viewpoint aware geometry occlusion algorithms allow the virtual exploration of such high geometric complexity 3D scenes. Some of the produced 3D models have been published on the Internet and thus can be accessed using a Web browser [1][3][7]. Unlike a 3D digital replica of an artefact, a 3D model of a monument or one of an archaeological site can be considered as a collection of 3D objects. This collection can also be considered as a static database. A manual virtual exploration of such a collection is not always efficient. This is due to the geometrical-morphological complexity of the scene and due to the imponderable factor of the virtual visitor’s special interests. In cases where a 3D scene is annotated with metadata that describe thematic aspects and properties of collection’s entities (e.g. artefacts, statues, buildings or architectural) then a keyword-based search engine can provide the basis of an information retrieval mechanism that enhances the user’s virtual exploration experience and study of the monument.

In this paper, we discuss the idea of facilitating virtual visits in complex 3D scenes by employing a keyword-based search engine. We propose the annotation of 3D scenes with information that encloses archaeological details, historical facts, architectural properties and spatial relationships of entities. The proposed methodology provides a tool that can be applied to shift the virtual visitor from the time consuming manual exploration procedure of
identifying areas of interest into an efficient and user-oriented approach for exploring the
details of complex 3D scenes. In order to support the previous statement, we have developed
a prototype search engine using open-source technologies. The functionality of the search
engine enables the user to perform keyword-based queries, spatial-topological queries and
predefined conditional queries that are related with architectural and historical properties. We
have selected three 3D reconstructions of urban areas of Northern Greece as case studies
[3][7][9]. These were built using photogrammetry, 3D terrestrial range scanner data,
topographical maps and manual measurements. The VRML 2.0 standard file format has been
used for bringing the 3D scenes to the Web. The 3D scenes were annotated and the metadata
that have been created follow a hybrid XML schema that is based on the MIDAS Heritage
[20] schema enhanced with custom elements and elements derived from the MACE [21]
schema.

The rest of this paper is organised as follows. In Section 2, we describe some of the related
works and we emphasis on the novelty of this paper. In Section 3, we describe the point-of-
interest annotation approach, the proposed hybrid schema, the architecture and the
functionality of the prototype search engine. We conclude in Section 4 by proposing some
ideas on the future development of the prototype.

2. Related Work

Nowadays, 3D content has become very popular not only due to the current technological
trends (e.g. 3D films and television) but also because it provides a mean for better
comprehension and appreciation of the visual content. Currently, research projects like the
CARARE, work on brining architectural and archaeological 3D content to the public through
the Europeana search engine [22]. Several research groups have proposed methods of
annotating 3D scenes in order to exploit their content in various applications. Gobetti et al.
[23] proposed the exploitation of the anchor link feature covered by the VRML 2.0 standard
[24] in order to present historic and tourist information. Singh et al. [25] described a method
to visualise 3D scenes by combining non-linear projections and by performing data mining
from manual explorations. Bilasco et al. [26] proposed a generic semantic annotation model
for low complexity 3D scenes that aims in the reuse of geometric primitives. Tenmoku et al.
[27] proposed a view management method which emphasizes on annotating 3D replicas of
artefacts. In [28], Maass et al. presented a technique that models annotations as separate three-
dimensional scene elements that are automatically positioned and oriented according to the
shape of the referenced object. Kadobayashi et al. [29] proposed the idea of allowing the user
to create, edit and store annotation metadata of 3D models.

The combination of annotation and augmented reality has also been exploited by Feiner et al.
[30]. They proposed a prototype 3D mobile augmented reality system for exploring real urban
environments while presenting the metadata using a head unit display. Okuma et al. [31] has
also proposed a similar wearable 3D model annotation system based on object tracking. Jung
et al. [32] presented the post-it note metaphor as a mean to allow special interest groups to
annotate 3D models in order to exchange notifications in a collaborative 3D environment.
Kleinermann et al. [33] described a system to annotate the content of 3D virtual environments
in order to support virtual touring. Chittaro et al. [34][35] proposed the adoption of guided
tours in virtual environments and the use of a radar metaphor as an effective aid for
navigation. The lack of semantic description in virtual environments has also been noticed by
Pittarello et al. [36]. They described an approach to associate semantic information with 3D
scenes based on the integration of the X3D [37] and Semantic Web [38] standards. The
annotation of 3D models has also been researched by bioengineers. Balling proposed a system
that allows the annotation of volume data provided by computer aided tomography (CAT)
[39].

The novelty in our work arises from the fact that it departs from the idea of visualising
information within the 3D environment. It describes an information retrieval mechanism that
integrates a common graphical user interface of low learning overhead that allows efficient retrieval of information related to the content found in 3D scenes. Assuming, that a user is searching for the Roman Colosseum and this monument exists within a 3D scene that covers a large portion of the ancient Rome. It is important for the virtual visitor to be able to efficiently explore this point of interest. Based on this example, the proposed system focuses on the idea of unburdening users from a time-consuming manual exploration process and allows them to identify parts of 3D scenes that fall in their particular interests, whether these are scientific (e.g. architectural-archaeological features and properties, historical information) or not (e.g. cultural tourism, scene exploration). Such a system may find applications in areas such as cultural resources management and heritage interpretation. Cultural resource managers might use such a system in order to annotate 3D replicas of physical assets and scarce elements with information and details in order to enhance their resource management procedures. Museums may improve essential functions such as artefact position inspection and identification or virtual collection management. It may be deployed in the significant sector of cultural tourism as it allows visitors to identify points of interest prior-to-the-visit or retrieve in situ information (heritage interpretation) when wireless network technologies allows it. Another possible application may be the integration of the proposed search engine in an e-tourism service that exploits Web 2.0 technology (commenting, tagging, etc) in order for the visitor to adapt his/her cultural experiences according to other’s reviews, comments and ratings.

Furthermore, we consider a 3D scene as a static database and thus we suggest the use of a text-based search engine as a tool to identify entities in such a database. This is achieved by following a 3D scene annotation approach that is based on points-of-interest (POI). We present a prototype that allows users to perform keyword-based queries related to the content of the 3D scenes. We focus on 3D reconstructions that depict urban areas of cultural importance and thus we have used a hybrid schema that is able to handle historical, archaeological, architectural and topological information.

3. A 3D Scene Search Engine Prototype

In this Section, we describe the proposed annotation approach of 3D scenes, which includes the procedure of defining points of interest, the hybrid schema that has been used for storing the metadata and the architecture and functionality of the prototype search engine.

3.1 3D Scene Annotation Based on Points-of-Interest

Annotation is the procedure of building relationships between entities that belong in a single or multiple information domains. In this work, we establish relationships between points-of-interest (POIs) in 3D scenes and sets of textual information that add value to the POIs’ content by describing historical, archaeological, architectural and topological-spatial aspects. According to the proposed annotation approach, entities such as a building, a statue, a monument or any part of them are considered as POIs as long as they exhibit important aspects in the given archaeological-architectural context domain. Hence, the procedure of identifying POIs is absolutely interlinked with the content of the 3D scene, the scope of its creation and the information context that can be extracted from it.

Technically, POIs are defined as vectors that describe the properties of a perspective virtual camera model. These properties are the spatial coordinates of the centre of projection, the orientation of the projection plane against each Cartesian axis and the field of view (Figure 1.i). All coefficients are defined within the local coordinate system of a 3D scene and are used to describe a viewpoint that focuses on an entity’s façade or detail.

Figure 1. Point of Interest (POI) identification example

In order to extract the coefficients of each POI vector, we have followed a process of manual exploration and identification of entities in all 3D scenes. The POIs were defined by moving
the virtual camera in positions where its projection plane is able to capture the important aspects of each entity (Figure 1.ii). The 3D reconstructions that have been used contain mainly buildings and thus all POIs where selected by attempting to minimise the geometric declinations between the entities’ façades and the virtual camera’s projection plane.

3.2 3D Scene content documentation and XML schema encoding

After the definition of the POIs, the documentation of all selected entities took place. This was done in cooperation with archaeologists of CETI’s cultural heritage department [40]. The MIDAS Heritage XML schema [20] was selected as the basis for encoding the documentation information of each entity. The specific schema carries elements that allow to describe heritage assets (archaeological monuments, landscapes areas, buildings, etc), related activities (management, designation, events, etc) and information sources (bibliography, archives, narratives, etc). For this work, we have encoded information related to the entity’s history, current and previous owners, architectural properties, usage, preservation state, possible interventions, designations (e.g. listed as a scheduled building), spatial and topological data by using a hybrid metadata schema that is based on MIDAS Heritage and MACE schemas [21].

More specifically, a number of extensions to the MIDAS schema were implemented in order to achieve the search engine’s required functionality. The MIDAS meta element is the primary node (root) for storing a set of basic information related to a heritage dataset [41]. We extended this node by introducing the Scene Details (Figure 2) and Generation Details (Figure 3) elements. The Scene Details element defines the primary properties of a 3D scene. These are a unique identification number, a uniform resource identifier (URI) to a VRML/X3D model and a set of environmental details such as the sky colour vector and the dimensions of the avatar. The Generation Details element is used to provide fundamental information about the methodologies that have been followed to create the 3D scene. The creation purpose sub-element describes the scope of use of the 3D model. The geometrical accuracy and fidelity of a 3D model is one of the dominant factors that determine the scope of its use. For example, a low density 3D model created using photogrammetry is adequate for the dissemination of the monument or for online virtual touring but not sufficient for scientific analysis. Moreover, the raw data types sub-element describes, using controlled vocabularies, the data types being provided by the 3D digitisation systems that were used during the digitisation phase of the scene.

Furthermore, two extensions are also proposed to the monument element of the MIDAS schema [42]. These are the Scene Object and the Asset Details elements. The Scene Object element (Figure 4) is used to describe an entity’s POI that holds a unique identification number. The inheritance sub-element provides an alternative approach to define parent-to-child and child-to-parent relationships between the entities of a 3D scene. It provides a method of relating entities of a 3D scene under any given knowledge domain. Although, the spatial appellation element of the MIDAS schema provides efficient means for extracting spatial and topological relationships between entities based on geographic coordinates of a given datum, the inheritance sub-element operates as a supplemental method to relate entities in different knowledge domains. The definition of the knowledge domain in which the entities relationships occur, depends on the metadata author. In our case study, we have used the inheritance sub-element in order to relate entities details that cannot be identified by geographic coordinates. Such examples are the decorative details of a building or a wall engraving. The Image Digital Resources sub-element
holds URIs to digital photographs of an entity while the viewpoint vector sub-element holds the virtual camera’s coefficients.

The asset details element (Figure 5) focuses on the entity’s architectural classification. The preservation sub-element provides construction and restoration dates along with current preservation state information. The observations sub-node may provide brief descriptions about the entity’s condition. The architectural details sub-element is contains attributes defined by the MACE architectural classification schema [21]. We have selected an efficient basis of elements in order to describe the entity’s functionality, basic form, architectural style, materials being used, etc.

Figure 4. Graphical representation of the Scene Object element

Figure 5. Graphical representation of the Asset Details element

In our implementation, the metadata structure follows the tree metaphor. Each 3D scene is composed by a root node that provides the 3D scene’s basic information. This information is described in the Scene Details and Generation Details elements (Figure 2 and 3). The root node is followed by a number of sub-nodes that represent the scene’s entities. A complete 3D scene metadata example in XML can be found at http://www.ceti.gr/3dsse/SceneExample.xml.

We have annotated three 3D scenes to be used as case studies. All of them have the speciality to exhibit mixed traditional Greek, European and Oriental architectural features of late 18th century and are located in the region of Eastern Macedonia and Thrace. These are:

i. A part of the old town of Xanthi, which is one of the biggest traditional settlements in Greece (Figure 6.a).

ii. A part of the old town of Kavala, located on the Panagia peninsula (Figure 6.b).

iii. A part of the Saint’s Barbara springs region in city of Drama (Figure 6.c).

A total of twenty entities were identified within the 3D scenes and documented according to the needs of the proposed metadata schema.

Figure 6. Viewpoints from the 3D scenes that were as case studies

3.3 Search engine’s architecture and functionality

The prototype is based on open source technologies and can be found at http://www.ceti.gr/3dsse. Figure 7 illustrates the basic components of the system. The core functionality has been implemented using the PHP script language [43]. A native XML database has been used for metadata storage [44]. For the evaluation of the prototype system, we have performed a number of tests that were focused on ensuring the correct operation of each individual function in the PHP source code followed by operational tests that verified the correct response of the system. Additionally, we have conduct distinct queries based on the metadata content in order to evaluate the system’s responses. As this is an experimental prototype, we relied on the known performance of the open source technologies being used and thus a complete evaluation of the system in terms of performance metrics has not being performed.

Figure 7. Architecture of the prototype search engine

Furthermore, the system handles queries that contain either keywords or natural language text. The latter are predefined sentences that contain variables that allow users to perform specialised queries in order to narrow down the retrieved information. We have implemented
only two predefined queries as the purpose of this work is to demonstrate the functionality of
the prototype. The predefined queries are the following:

i. Built between a start period (year) and an end period (year).

ii. Materials used $1^{st}$ material type, $2^{nd}$ material type, . . . , $n^{th}$ material type

The first query will retrieve all entities that fall into the given chronological limits, while the
second query will retrieve all entities that contain any of the materials being described. The
number of predefined queries can be expanded based on the need of handling more complex
user requests.

Additionally, the query handling script is responsible for parsing the query and composing the
XPATH expression which is forwarded to the database. The database responds to the query
and the results are being parsed by the front end script which composes the retrieved
information. Figure 8 depicts a screenshot of a search engine’s reply. Each root entity
represents a 3D scene that is followed by relevant to the query, entities that exist within
that 3D scene. The root element (Figure 8) is accompanied with a map where the area that is
being covered by the 3D scene is approximated by a rectangular and a short description about
the content of the 3D scene. Additionally, for each one of the entities, a thumbnail image is
provided together with a short textual description related to the history of the entity and its
architectural properties. For each entity, a set of options are available to the user. These are
the following:

i. Retrieve a complete record for the specific entity.

ii. Transfer to the POI within the 3D scene.

iii. Examine the entity’s position onto a world’s map.

iv. Request by the system to indicate what other entities are close to a chosen one by
taking under consideration their geographical coordinates.

More specifically, the first option retrieves the complete metadata set from the database and
organises it in a record form that can be printed. The second option, exploits the textual
format of the 3D models in order to stream to the user’s browser the 3D scene with the
viewpoint vector of the specific POI defined as the default viewpoint. The engine merges a
dynamically created portion of VRML 2.0 source code with the 3D scene and forwards it to
the user’s Web browser.

The third and fourth option take advantage of the spatial metadata. The MIDAS Heritage
schema specifies a set of metadata elements for describing geographic datasets. In our
metadata, we have used the WGS84 geodetic system to describe the entities coordinates.

The third option provides a visual indicator on a world map that points to where the entity is
located. This is done by exploiting the Google Maps service API and the entities spatial
coordinates. On the other hand, the forth option represents a special form of query with which
the system is requested to identify what entities exist within a given area. In our
implementation this area is set to $500 \, m^2$. The system considers the longitude and latitude
coordinates of a selected entity as the centre of this rectangular area and identifies what other
entities fall within the limits. Those entities can be part of any 3D scene within the database.
The system replies by sorting the entities according to their distance. The first entity to be
returned by the system is the query entity which is then followed by the next closest entity. A
higher ranking position indicates a shorter distance from the query entity. The topological-
spatial queries can provide users with information that also could be exploited in a real visit to
a monument.
4. Conclusions

In this work, we have developed an on-line keyword-based search engine for 3D scenes. We have based the development of the prototype system on the assumption that a 3D scene is a collection of 3D objects and thus it can be handled as a static database. We applied the points-of-interest annotation approach in order to create sets of metadata that accompany 3D scenes that cover urban areas of cultural importance in the region of East Macedonia and Thrace, Greece. The 3D scenes make extensive use of texture mapping techniques and thus allow users with low bandwidth connections to experience a 3D virtual tour within reasonable downloading times. The metadata follow a hybrid XML schema based on MIDAS Heritage and MACE. A number of additions to the schema were also proposed in order to achieve the required functionality of the search engine. This functionality provides positioning of the virtual visitor in a point of interest within a 3D scene and the ability to perform queries related to archaeological, historical, architectural and spatial-topological properties.

Currently, novel 3D scenes are being annotated in order to be added in the search engine’s database. At the moment we are also working on enhancing the querying system in order to be able to handle more complex queries and on providing a X3DOM-WebGL [45][46] compatible version for eliminating the need of installing a VRML/X3D plug-in on the user’s Web browser.

At present, the search engine’s functionality, except from the 3D content, can be accessed by all smartphone browsers that support Javascript technology by using either GPRS or WiFi networks. The X3DOM-WebGL technology can be considered as a probable 3D content carrier for smartphones [47]. At present this is not possible as there is only a limited number of Android based smartphones that are able to run the Mozilla Fennec [48], that is the only WebGL-enabled Web browser. On the other hand, IOS based devices such the iPhone and iPad are able to access X3D content only through the instantMini [49] standalone application. Thus, a generic solution for accessing such 3D content through a smartphone’s Web browser is still not available.

Acknowledgements

The authors would like to thank the Cultural Heritage Department of the Cultural and Educational Technology Institute – Research Centre ‘Athena’ (http://www.ceiti.gr) for their contribution in the documentation of the 3D scenes content that were used as case studies.

References


[19] P. Salonia, S. Scolastico, A. Marcolongo, T. Leti Messina, Survey and 3D reconstruction of the St. Orso capitals in Aosta, through three-focal photogrammetry,


**Figure Captions**

Figure 1. Point of Interest (POI) identification example

Figure 2. Graphical representation of the *Scene Details* element

Figure 3. Graphical representation of the *Generation Details* element

Figure 4. Graphical representation of the *Scene Object* element

Figure 5. Graphical representation of the *Asset Details* element

Figure 6. Viewpoints from the 3D scenes that were as case studies

Figure 7. Architecture of the prototype search engine

Figure 8. Partial results of a search with the keyword ‘church’
Figures

Figure 1

I. Perspective camera illustration

Part of the 3D scene in wireframe rendering

Field of view

Projection plane orientation

Centre of projection coordinates

ii. Viewpoint content - Virtual Visitor’s perspective

Figure 2

Scene Details

Scene ID

Unique ID of a 3D scene.

Model Digital Resource

Digital resource (URL) of the 3D scene

Ground Colour Vector

R,G,B coefficients of the 3D model ground colour

Sky Colour Vector

R,G,B coefficients of the 3D model sky colour

Avatar Dimensions Vector

Dimensions of the avatar’s bounding box

Camera Position Coordinates

Spatial coordinates of the projection centre

Viewpoint Vector

Orientation of the projection plane against each axis

Camera Target

Field of View

Field of View (f)

Figure 3

Generation Details

Creation Method

Textual description of the method followed to create the 3D model

Creation Purpose

Textual description of the purpose behind the creation of the 3D model

Raw Data Types

Textual description of the 3D raw data types

Note

Additional textual details for the generation process of the 3D model
Figure 4

[Diagram of object ID, scene object, inheritance, parent of, child of, scene's object ID, parent's object ID, thumbnail URL, thumbnail, date, camera position coordinates, spatial coordinates of the projection center, camera target, orientation of the projection plane against each axis, field of view, field of view (f).]

Figure 5

[Diagram of initial construction date, initial construction date of entity, preservation, restoration date, restoration date, state of entity, incomplete, incomplete, fragmentary, observations, draft description of the condition of the entity in natural language, functional typology, description of the form of the entity, form typology, description of the basic form of the entity, trends typology, description of the architecture styles and periods, materials, description of the materials used in the entity, structural elements, description of the structural elements of the entity, form characteristics, topographic and geometric features that influence the form of the entity, number of stories, number of entity's stories.]

Figure 6

- a. View of Kougoumtzoglou mansion, Old city of Xanthi, Greece
- b. View of Mahmet Ali's house, Old city of Kavala, Greece
- c. View of Saint Barbara's springs, Drama, Greece
Figure 7

Figure 8