3D content-based search using sketches

K. Moustakas[1,2], G. Nikolakis[1], D. Tzovaras[1], S. Carbini[3], O. Bernier[3] and J.E. Viallet[3]
[1] Informatics & Telematics Institute
1st Km Thermi-Panorama Road, PO Box 361,
GR-57001 Thermi-Thessaloniki, Greece
{tzovaras, moustak, gniko}@iti.gr
[2] Electrical and Computer Engineering Department
Aristotle University of Thessaloniki
GR-54006 Thessaloniki, Greece
[3] France Telecom R&D, Lannion, France
{sebastien.carbine, olivier.bernier,
jeanemmanuel.viallet}@rd.francetelecom.com

Abstract. The present work presents a novel framework for 3D content-based search and retrieval. On contrary to most state-of-the-art approaches, the query model can be not only an existing object from a database but also a model manually generated by the user using sketches. In the context of the proposed framework, three interfaces to the sketch-based 3D search application were tested and comparative results were extracted according to usability and efficiency criteria.

1. Introduction

Search and retrieval of 3D objects is nowadays a very challenging research topic and has application branches in numerous areas like recognition in computer vision and mechanical engineering, content-based search in e-commerce and edutainment applications etc. [1]. These application fields will expand in the near future, since the 3D model databases grow rapidly due to the improved scanning hardware and modeling software that have been recently developed.

The difficulties of expressing multimedia and especially 3D content via text-based descriptors, reduces the performance of the text-based search engines to retrieve the desired multimedia content. To resolve this problem, 3D content-based search and retrieval (S&R) has drawn a lot of attention in the recent years.

However, the visualization and processing of 3D models are much more complicated than those of simple multimedia data [2]. The major difference lies in

Please use the following format when citing this chapter:
Moustakas, Konstantinos, Nikolakis, Georgios, Tzovaras, Dimitrios, Carbini, Sebastien, Bernier, Olivier, Viallet, Jean Emmanuel, 2006, in IFIP International Federation for Information Processing, Volume 204, Artificial Intelligence Applications and Innovations, eds. Maglogiannis, I., Karpouzis, K., Bramer, M., (Boston: Springer), pp. 361–368

the fact that 3D models can have arbitrary topologies and cannot be easily parameterized using a standard template, which is the case for images. Moreover, there can be many different models of representing them, i.e. indexed facets, voxel models etc. Finally, processing 3D data is much more computationally intensive, than processing media of lower dimension, and often requires very large amounts of memory.

Many researchers worldwide are currently developing 3D model recognition schemes. A number of approaches exist in which 3D models are compared by means of measures of similarity of their 2D views [3]. More direct 3D model search methods focus on registration, recognition, and pairwise matching of surface meshes [4]. However, these methods require a computational costly search to find pairwise correspondences during matching. Significant work has also been done in matching 3D models using geometric characteristics, where initial configurations are derived from conceptual knowledge about the setup of the acquisition of the 3D scene [5] or found automatically by extracting features such as curvature or edges [6].

A typical S&R system, like the aforementioned ones, evaluates the similarities between query and target objects according to low-level geometric features. However, the requirement of a query model to search by example often reduces the applicability of an S&R platform, since in many cases the user knows what kind of object he wants to retrieve but does not have a 3D model to use as query.

Imagine the following use case: The user of a virtual assembly application is trying to assemble an engine of its spare parts. He inserts some rigid parts into the virtual scene and places them in the correct position. At one point he needs to find a piston and assemble it to the engine. In this case, he has to manually search in the database to find the piston. It would be faster and much more easier if the user had the capability of sketching the outline of the piston using specific gestures combined with speech in order to perform the search. In the context of this project the integration of speech and gestures for the generation of the query model is addressed. Speech commands are used for performing specific actions, while gesture recognition is used to draw a sketch of the object and to manipulate the scene objects in the 3D space. The system is also capable to assemble the built objects so as to generate complex structures. The sketch-based 3D search engine has been tested using three interfaces. Comparative results on the usability and efficiency of the interfaces are presented in the experimental results section.

2. 3D content-based search

For each 3D model, rotation invariant geometrical descriptors are extracted. In particular, the object is initially normalized in terms of translation and scaling, i.e. it is translated to the center of the coordinate system, and is scaled uniformly so that the coordinates of all its vertices lie in the interval [0,1]. Next, N concentric spheres are built centered at the origin of the coordinate system. Each sphere is built using tessellation of a normal icosahedron so that the vertices over its surface are uniformly distributed. In the experiments 20 concentric spheres of 16002 vertices are used. For each sphere the discrete 3D signal $F(r_S, \theta_i, \phi_i)$ is assumed, where i is the

index of the sphere vertices. The values of function $F(r_S,\theta_i,\phi_i)$ are calculated using the Spherical Trace Transform (STT) [7].

The extraction of the final descriptor vectors, which will be used for the matching algorithm, is achieved by applying the spherical functionals "T", as described in [7], to the initial features $F(r_S,\theta_i,\phi_i)$ generated from the STT. The spherical functionals for each concentric sphere "ρ" are summarized below:

$$T_1(F)=\max\{F(r_S,\theta_i,\phi_i)\} \quad (1)$$

$$T_2(F)=\sum_{j=1}^{N_S}\left|F'(r_S,\theta_i,\phi_i)\right| \quad (2)$$

$$T_3(F)=\sum_{j=1}^{N_S}F(r_S,\theta_i,\phi_i) \quad (3)$$

$$T_4(F)=\max\{F(r_S,\theta_i,\phi_i)\}-\min\{F(r_S,\theta_i,\phi_i)\} \quad (4)$$

$$T_l(F)=A_l^2=\sum_m a_{lm} \quad (5)$$

where N_S is the total number of sampled points $(\eta_j, j=1,\ldots,N_S)$ at each concentric sphere, $l=0,\ldots,L$ and $-l<m<l$. The values of a_{lm} are the expansion coefficients of the Spherical Fourier Transform [8]:

$$a_{lm}=\sum_{i=1}^{N_S}F(r_S,\theta_i,\phi_i)\cdot Y_{lm}(\eta_i)\frac{4\pi}{N_S} \quad (6)$$

where $Y_{lm}(\eta_i)$ corresponds to the spherical harmonic function, which is defined through:

$$Y_{lm}(\theta,\phi)=k_{l,m}P_l^m(\cos\theta)e^{jm\phi} \quad (7)$$

where P_l^m is the associated Legendre polynomial of degree l and order m, $k_{l,m}$ a normalization constant and j the imaginary unit.

The quantities A_l^2 are invariant to any rotation of the 3D model. Choosing a sufficiently large number of L coefficients of the Spherical Fourier Transform, a total number of $L+4$ spherical functionals is used for each concentric sphere.

Finally, the descriptor vectors $D(l)$ are created, where $l=0,\ldots,(L+4)N_C$ is the total number of descriptors and N_C is the number of concentric spheres. In the experiments described in the sequel, $L=26$ and $N_C=20$ were chosen.

Now, let A, B, be two 3D part models, and D_A, D_B, their descriptor vectors respectively. The two parts are compared in terms of similarity according to the following formula:

$$D_{dissimilarity}=\sqrt{\sum_{l=1}^{(L+4)N_C}\left|D_A(l)-D_B(l)\right|} \quad (8)$$

Fig. 1 depicts the retrieved objects using as input the first model of each column.

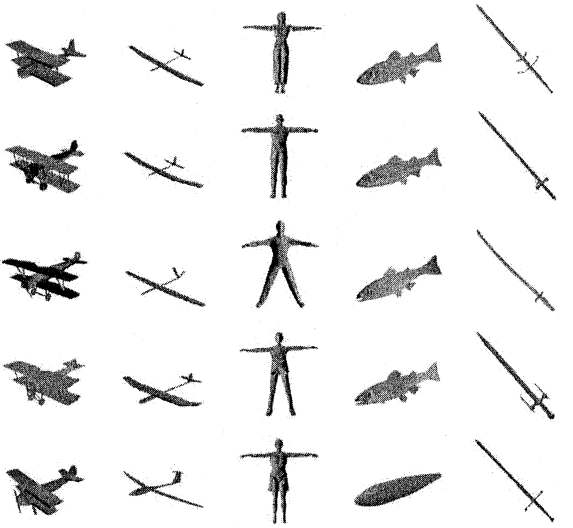


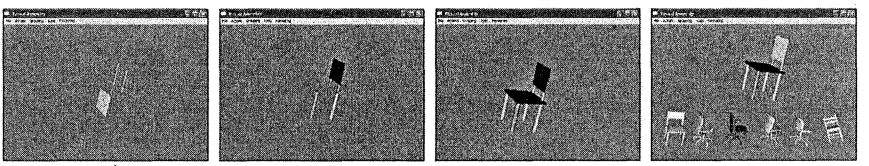
Fig. 1. 3D search results using as query the models of the first row.

3. Sketching the query model

The sketch-based query model generation algorithm aims to provide the sufficient means for the easy and fast design of an approximation of the target model using primitive objects. It consists of the following steps:

1. Sketching the 2D contour of the desired primitive object.
2. Choosing among the corresponding 3D shapes using speech commands [9] (e.g. for a circle choose between sphere, cylinder and cone) and define its height, which cannot be drawn in 2D.
3. If a new primitive is desired go to Step 1, otherwise proceed to Step 4.
4. Assemble the primitives to form the final shape

The user initially sketches, using one of the sketching interfaces that will be described in the sequel, the 2D contour of the primitive to be inserted, e.g. circle for a sphere, a cylinder or a cone, rectangle for parallelepipeds, cubes and triangles for a pyramid or a prisma. These shapes are recognized using least squares minimization with the Levenberg-Marquardt algorithm [10][11] and a sample primitive is automatically inserted in the scene. Next, the degrees of freedom that cannot be defined just from the 2D sketch are defined and the primitive is manipulated. In other words the user defines the height of the object and translates, scales, rotates it until it reaches its target position. After inserting all the primitives they are assembled to the final target query model that is used as input to the 3D content based search procedure described in Section 2. An example of the sketching procedure is illustrated in Fig. 2.



a) b) c) d)

Fig. 2. a,b,c) Sketching procedure and d) 3D search results

4. Query interfaces

The query interface to the 3D search engine is a multimodal gesture-speech interface. The following table describes, the actions that are controlled with gestures and with speech.

Table 1. Speech-Gesture controlled actions

Speech controlled actions	**Actions performed using gestures or automatically by the system**
No speech	The 3D pointer follows the motion of the user's hand
Selection	Point at the object to be selected
Translation	Move the hand until the object reaches the target 3D position
Rotation	Rotate the hands like grabbing and rotating a sphere
Scaling	Increase decrease the distance between the hands
Sketching	Freehand sketching
Search	Use the selected object as query and search for similar content
Select group	Initiate grouping the primitives and call the selection command for each primitive
Retrieve	Retrieve the objects from the database starting with the most similar
Next	Retrieve next object
Delete	Delete selected object
Clone	Clone selected object
Stop action	Stop currently performed action

Speech recognition is performed as described in [9][10] and pointing gestures are extracted using one of the following interfaces:

4.1. Unobtrusive interface

The first interface is totally unobtrusive. The head and hands of the user are captured using a stereo camera and are efficiently tracked [10] using a statistical model

composed of a color histogram and a 3D spatial Gaussian function [12], while the user sketches or performs specific actions.

4.2. Virtual reality haptic interface

The backbone of this interface is a haptic glove that is used as input to the application, as it is capable of identifying hand gestures, and as output since it provides tactile or force feedback. It handles both human-hand movement input and haptic force-feedback for the fingers using Immersion's CyberGlove® (Fig. 3a) and CyberGrasp™ (Fig. 3b) haptic devices [13]. CyberGlove® is a widely used human-hand motion-tracking device of proven quality. CyberGrasp™ is currently one of the very few force-feedback devices that are offered commercially, providing high quality of construction, operation and performance. The 350g CyberGrasp™ exoskeleton is capable of applying a maximum of 12N per finger force-feedback at interactive rates and with precise control. The direction of the force feedback is approximately perpendicular to the fingertips.

(a) (b) (c)

Fig. 3. a) CyberGlove, b) CyberGrasp, c) Motionstar wireless tracker

Additionally to the haptic devices a position tracker device for providing information on the accurate position of the hand is used. Based on the requirements of the proposed application, the MotionStar Wireless Tracker of Ascension Technologies Inc. has been selected as the appropriate device, mainly due to its wireless nature (Fig. 3c). Combining CyberGrasp with the motion tracker can create a workspace of six meters diameter hemisphere where the user can move and interact with the virtual model, in contrary with the usual systems that limit the user workspace to be less than half a meter (just in the front of a personal computer).

4.3. Air-mouse interface

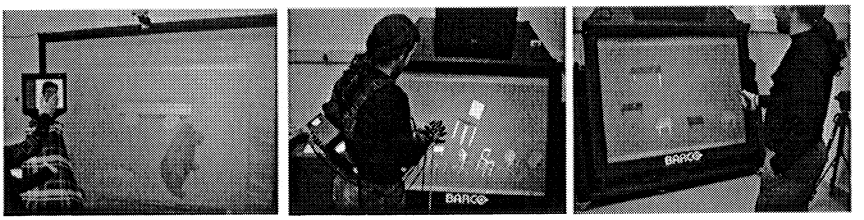
The third interface consists of a wireless air-mouse [14] that has the exact functionalities of a typical 2D mouse and can additionally by operated in the air since it utilizes a gyroscope sensor to identify changes in its orientation. Notice that, despite the fact that it can be operated in the 3D space, it is not a 3D mouse.

5. Experimental results

The developed sketch-based 3D search platform was evaluated in many scenarios where the user had to sketch the query object in order to search for similar content. The aim of the evaluation was to test and compare the three different interfaces with respect to several parameters, which are:

- User immersion
- Usability
- 3D manipulation efficiency
- Mobility
- Robustness
- Computational efficiency
- Device intrusiveness
- Cost

Fig. 4 illustrates three snapshots, while using the sketch-based 3D search platform, while Table 2 presents the comparative results of their evaluation.



a) b) c)

Fig. 4. a) Unobtrusive interface, b) Haptic interface, c) Air-mouse interface

Table 2. Comparison of the interfaces

	Unobtrusive	**Haptic VR**	**Air-mouse**
User Immersion	Very high	High	Very low
Usability	Very high	Very high	Moderate
3D manipulation efficiency	Very high	Very high	Very low
Mobility	Very low	Very low	Very high
Robustness	High	Very high	Very high
Computational efficiency	Moderate	High	Very high
Cost	High	Very high	Very Low
Device intrusiveness	Very low	High	Low

6. Conclusions

In the present paper a sketch-based 3D search system was presented. The user is capable of creating the query object using speech and gesture instead of using an existing model to search for similar 3D content. Three different interfaces for human computer interaction were tested and comparative results were extracted that indicate that each interface has its advantages and disadvantages. Which one to use? It depends absolutely on the context of the application to be developed.

Acknowledgment. This work has been supported by the EU funded SIMILAR Network of Excellence.

References

1. S. Berchtold and H.P. Kriegel, "S3: Similarity Search in CAD Database Systems", Proc. of SIGMOD, J. Peckham, Ed. ACM, pp. 564-567, 1997.
2. E. Paquet and M. Rioux, "Nefertiti: A Tool for 3-D Shape Databases Management", SAE Transactions: Journal of Aerospace, vol. 108, pp. 387-393, 2000.
3. J. Loffler, "Content-based Retrieval of 3D models in Distributed Web Databases by Visual Shape Information", Proc. of Int. Conf. on Information Visualisation (IV2000), 2000.
4. A. E. Johnson and M. Hebert, "Using Spin-images for Efficient Multiple Model Recognition in Cluttered 3-D Scenes", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 21, no. 5, pp. 433-449, 1999.
5. G. Blais and M. Levine, "Registering Multiview Range Data to Create 3D Computer Objects", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 17, no.8, pp. 820-824.
6. C. S. Chua and R. Jarvis, "3D Free-Form Surface Registration and Object Recognition", Proc. of Int. Journal of Computer Vision}, Kluwer Academic Publishers.
7. P.Daras, D.Zarpalas, D.Tzovaras and M.G.Strintzis, "3D Model Search and Retrieval based on the Spherical Trace Transform", IEEE Intl Workshop on Multimedia Signal Processing, Sienna, Italy, 2004.
8. D.V. Vranic and D. Saupe, "Description of 3D-shape using a complex function on the sphere", Proc. IEEE International Conference on Multimedia and Expo, pp. 177-180, 2002.
9. R. Schwartz and Y.L. Chow, "The N-Best Algorithm: an Efficient and Exact Procedure for Finding the N Most Likely Sentence Hypothesis", ICASSP 1990, pp. 81-84, 1990.
10. K. Moustakas, D. Tzovaras, S. Carbini, O. Bernier, J.E. Viallet, S. Raidt, M. Mancas, M. Dimiccoli, E. Yagci, S. Balci and E.I. Leon, "MASTER-PIECE: A Multimodal (Gesture+Speech) Interface for 3D Model Search and Retrieval Integrated in a Virtual Assembly Application", Proceedings of the eNTERFACE 2005, pp. 62-75, August 2005.
11. D.W. Marquardt, "An Algorithm for the Least-Squares Estimation of Nonlinear Parameters", SIAM Journal of Applied Mathematics, vol. 11, no. 2, pp. 431-441, 1963.
12. S. Carbini, J. E. Viallet and L. Delphin-Poulat, "Context dependent interpretation of multimodal speech-pointing gesture interface", International Conference on Multimodal Interfaces, Trento, Italy, 2005.
13. Immersion Technologies Inc., "Virtual Hand Suite 2000: User & Programmer Guides", http://www.immersion.com/3d/products/virtualhand_sdk.php.
14. Gyration Inc., http://www.gyration.com/go24airmouse.htm.