Eurographics Workshop on 3D Object Retrieval (2012)
M. Spagnuolo, M. Bronstein, A. Bronstein, and A. Ferreira (Editors)

Towards an Immersive Interface for 3D Object Retrieval

Pedro Pascoal and Alfredo Ferreira and Joaquim Jorge

INESC-ID/IST/Technical University of Lisbon
Lisboa, Portugal
http://vimmi.inesc-id.pt

Abstract

In this paper we propose the use of immersive virtual reality to visualize and explore query results in 3D object retrieval. These results are shown in a three-dimensional virtual space as 3D objects, organized according to their similarity to the query. Therefore, instead of browsing through a list of thumbnails, the user can explore the query results by navigating among 3D objects in a virtual space and even manipulate these objects for closer a inspection. We performed a preliminary user study and concluded that the proposed solution provides better performance than traditional approaches.

Categories and Subject Descriptors (according to ACM CCS): H.3.3 [Information Systems]: Information Search and Retrieval—Search process H.5.2 [Information Systems]: User Interfaces—Interaction styles H.5.2 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality

1. Introduction

The number of 3D objects available in digital format has been growing. This rapid growth raises the issue of searching in large collections. To approach this challenge, different approaches to a 3D object search engine have been proposed [FMK*03, AVD07]. Still, most of the existing solutions exhibit major drawbacks and challenges that need to be tackled. Among those identified in Datta's survey [DJLW08] and directly applicable to 3D object retrieval (3DOR), we highlight two. First, queries rely mostly on meta-information, often keyword-based. This means that, searches can be reduced to text information retrieval of multimedia objects. Second, results are presented as a list of items on a screen. These items are usually thumbnails, often combined with filenames or metadata. However, thumbnails may not provide the best view of the model, as Dutagaci *et al.* concluded in their study [DCG10].

In this paper, we address the second issue by proposing a novel approach for query result visualization in 3D object retrieval (3DOR). Instead of using a list of model thumbnails, our system displays three-dimensional representations of retrieved models distributed in a 3D virtual space. Following a query, retrieved object are shown in a virtual reality (VR) environment, organized according to their degree of similarity. Then, the user is able to navigate through the search results and explore them in detail.

To that end, we introduce the Immersive 3D Object Retrieval (**Im-O-Ret**) prototype. In Im-O-Ret, the user can explore the results of a query to a collection of 3D objects in an easy and natural fashion. In order to support our claims, the proposed solution was evaluated against a traditional thumbnail-based approach for result visualization.

2. Related Work

As a response to the necessity for search by content on growing 3D model collections, special attention has been given to algorithms for 3D retrieval. However, the user interfaces seem to have been out of researchers focus. Namely with respect to most recent interaction paradigms.

2.1. 3D Object Retrieval

In 2001, Thomas Funkhouser *et al.* [FMK*03] introduced the Princeton 3D model search engine, providing content-based retrieval of 3D models. Queries can be specified using text, by example, 2D sketch, or 3D sketch. The results are presented as an array of model thumbnails. After a search it is also possible to choose a result model as query-by-example to initiate a search.

More recently, Ansary *et al.*() introduced FOX-MIIRE search engine [AVD07], providing query by photo. This tool

DOI: 10.2312/3DOR/3DOR12/051-054

retrieves a 3D model similar to an object in a photo. Additionally, this system offers interfaces both for desktop and mobile devices. However, and similarly to previous solutions, results are displayed as a list of thumbnails.

Outside the research field, Google 3D Warehouse offers a search engine for 3D models. However, while this repository contains a very large number of different 3D models, searching for models in this collection is limited by textual queries or by its geo-reference. As in other 3DOR systems, query results are displayed in a thumbnail list. Nevertheless, in this search engine the user can manipulate a 3D view of a single selected model. We consider this feature a relevant, yet shy, step towards powerful interfaces for result visualization in 3D search engines.

2.2. Immersive Environments for Retrieval

Indeed, despite the growing capabilities of current hardware and software, the approaches referred above, as well as others, do not take advantage of recent advances in computer graphics or interaction paradigms to improve result visualization. Even in the broader domain of multimedia information retrieval, to the extent of our knowledge, the single approach that uses immersive virtual environments is Nakazato's 3D MARS [NH01].

3D MARS is a content-based retrieval system for images, but clearly demonstrates that the use of immersive environments in multimedia retrieval can offer many benefits, which served as a catalyst for the approach proposed in this paper: explore immersive visualization systems for 3D object retrieval. To that end we depend on more recent and cheap VR technology. Indeed, devices like Nintendo Wiimote or Microsoft Kinect, togheter with 3DTV sets or HMD glasses provide low-cost approaches that brought immersive experiences from the labs into our homes.

In the context of 3DOR, Holz and Wilson [HW11] proposed a system that allows users to describe 3D objects through gestures. Their system captures gestures with a Kinect camera and then finds the most closely matching object in a database of physical objects. This work represents a good combination of new interaction paradigms for 3D object retrieval. However, it addresses the issue of query specification, while the present work focus on query result visualization. Nevertheless, both approaches could be combined to create a complete and fully featured search engine.

3. Immersive 3D Object Retrieval

Taking advantage of new paradigms in HCI, we propose an immersive virtual reality system for 3D object retrieval (**Im-O-Ret**). In our system, query results are displayed as 3D models distributed in a three-dimensional space, instead of the traditional list of thumbnails. The user can explore the results by navigate in the three-dimensional space through the 3D models and inspect them closely. The combined use of virtual environments and control devices with six degrees of freedom (6DoF), affords a complete visualization of models while providing natural interaction, thanks to direct object manipulation.

3.1. Query Specification

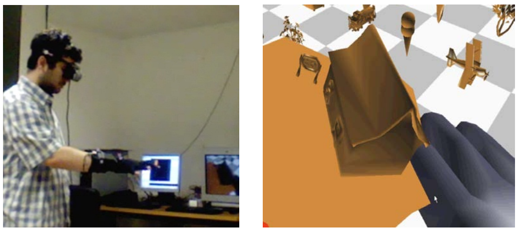
The work presented in this paper is focused on the visualization, navigation and manipulation of search results. As such, for query specification, only a simple query-by-example was implemented. From the query results, users can then perform another search for similar to one of the results, or start a new search.

3.2. Spatial Distribution of Results

The query results are distributed in the virtual 3D space according to their similarity. Each axis is assigned to a different shape matching algorithm. The coordinate value is determined by the similarity to the query given by the corresponding algorithm. As such, when performing a search, the query model is used to find similar models, using each algorithm. The results are then merged, giving a 3D position for each similar model retrieved. The query mechanism can be adapted to specific domains, producing more precise results.

There are two modes for the distribution the 3D objects. In the **equidistant** mode objects are distributed in the space with equal distance from each other. This mode minimizes the occlusion effect, maintaining the order. In this mode, although the more similar results are closer to the origin of the 3D space that represents the query, there is no implicit information regarding similarity between objects. In the **exact** mode, the objects position is the given directly by the value retrieved from each matching algorithm. This originates clusters, when retrieved models are very similar, providing a more precise perception of similarity between query results, but might led to undesirable occlusions.

Thanks to architecture modularity, the query mechanism can be adapted to specific domains, producing more precise results. This means that different shape descriptors can be assigned to each axis, as well as different distance measurements. In the current version of **Im-O-Ret**, Euclidean distance are used to measure the similarity between the three chosen descriptors. In this particular case, we used the lightfield descriptors [CTtSO03] on the X-axis, the cord and angle histogram [PR97] for the Y-axis, and the spherical harmonics descriptor [KFR03] for the Z-axis. One of the reasons behind this choice of descriptors is that each of them targets a different set of features [TV08]. The other, more practical, was the availability of the corresponding descriptors.



(a) (b)

Figure 1: Immersive object retrieval: (a) exploring the 3D space with HMD and data gloves; (b) user's view.

(a) (b)

Figure 2: Immersive object retrieval: (a) TV Screen and Wii-Remote; (b) Display Wall and SpacePoint Fusion.

3.3. Exploring Query Results

Once a query is submitted, the system returns a set of similar objects that are organized in a 3D virtual space according with the distribution mode (*equidistant* or *exact*) and the similarities returned by the shape descriptors. Independent of object distribution, a major contribution of our work is the possibility the user has to explore the results. This exploration is is accomplished through navigation in the virtual space and direct manipulation of 3D objects.

For the navigation techniques we considered the division proposed by Bowman [BDHB99], into two types based on their movement. As such, we created both a *travel* mode and a *wayfinding* mode. The *travel* mode consists in exploratory movements were the viewpoint is moved from one location to another, by using either the arrows in the keyboard, the Wiimote, or the SpacePoint sensor, similar to the process of flying. The *wayfinding* mode consists on using points of interest, where the movement is done by selecting an object that specifies the position where the camera is to be moved. This can be accomplished by pointing with a mouse, the Wiimote$^{\text{TM}}$, the data glove or the SpacePoint Fusion$^{\text{TM}}$ device.

Object inspection can be more or less natural according to the device used. The mouse-based interaction is far from natural and requires some training. However, with gloves or other gadgets the users intuitively know how to manipulate the object. Also, the use of shutter glasses or head mounted displays, provides the user with a stereoscopic view of the world, thus enabling a more complete immersive experience. Figure 1 illustrates a scenario where a user is exploring the result space using a HMD display and a data glove. Here the user can easily navigate through the results and manipulate any of them for inspection by literally holding it. The combination of depth perception with direct manipulation offer a simple and natural way to explore a set of results.

3.4. Modularity and Scalability

Our system can be configurable to work with a wide range of different interaction devices and displays. We can combine a different set of visualization and interaction devices, and create multiple interaction paradigms. Also, with minimal implementation effort, it is possible to prepare the system for different input and output devices. Indeed, the **Im-O-Ret** system can be used in wide range of scenarios: from an expensive large screen display with tracking to a TV and Wiimote commodity setup, depicted in Figure 2.

4. Evaluation

In order to validate our approach, we compared it with the thumbnail approach for result display used by traditional search engines. To avoid biasing the study, or adding additional complexity to the test scenario, we developed a simple system where users could perform 3D object retrieval. In the THumbnail Object Retrieval (**THOR**), the query results are presented as an ordered list of thumbnails.

4.1. Test Description

The twelve participants involved in the tests performed three search tasks of increasing difficulty in two different test environments, as depicted in Figure 3. The first environment consisted on a computer screen with a mouse as pointing device. For the second environment we asked the user to wear a pair of data gloves and a set of HMD glasses with tracking.

For evaluation purposes, we measured the number of steps, errors, and time required to find a specific object. This object was given to the user at the beginning of each task. The task was considered complete when the user positively identified the object in the returned results.

To simplify query specification and to make the starting point for each task in both systems equal, we used a static group of 36 models as example queries. To perform the search task, the user should select from these model the one he considers most appropriate and submit it as a query to the system. Then he must find in the returned results the object that was presented to him in the beginning of the task.

(a) (b)

Figure 3: Immersive object retrieval test environment: (a) **THOR** (b) **Im-O-Ret**

4.2. Results

This evaluation allowed us to compare our approach with a traditional 3DOR system both quantitatively and qualitatively. For the quantitative evaluation we measured the number of steps, errors and time spent. The steps correspond to the actions performed by the user, the time corresponds to the time took to complete the task, and an error is when an user either rolled back to a previous search result, or restarted the search because did not found the object he was looking for.

From the number of steps measurement we found no meaningful difference between the two approaches. A similar observation can be made for the number of errors measurement, although in this case, a slight difference between the two approaches can be detected. However, for the more complex second and third search tasks, our approach are less error-prone than traditional thumbnail visualization. Although without statistical significance, these results allow us to conclude that our approach might provide better visualization and interpretation of the query results.

The same cannot be concluded for the time required to perform the search tasks, which showed that the participants wasted more time in the tasks performed using the **Im-O-Ret**. Nevertheless, this was caused by most users inexperience with immersive virtual reality environments, wasting more time exploring the virtual space, navigating and viewing the query results, than focusing on the task and performing the supposed search.

5. Conclusions and Future Work

We believe that recent advances in low-cost post-WIMP enabler technology can be seen as an opportunity to overcome some drawbacks of current multimedia information retrieval solutions. Following this belief, we present in this paper a new approach to search interfaces for 3D Object Retrieval, namely regarding exploration of query results. Although it does not describe a complete, final and closed prototype, it lays the ground for future work by demonstrating the validity of the concept while identifying challenges to be tackled in future research.

From the work presented in this paper, we identified one major challenge to be tackled in a near future: the organization of results in a three dimensional space. While in traditional thumbnail listing the results are presented in a 2D space as ordered and equally distributed images, in our approach we implemented a simple solution that distributes the results in the three-dimensional space.

Moreover, although only focused on visualization, our solution can be greatly improved by integrating gesture-based query specification, following the concepts proposed by Holz and Wilson [HW11]. Having such we might say that a complete and powerful search engine is ready, supporting the whole search process and taking advantage of recent visualization and interaction technology.

References

[AVD07] Ansary T. F., Vandeborre J.-P., Daoudi M.: 3d-model search engine from photos. In *Proc. of the 6th ACM int. conf. on Image and video retrieval* (2007), CIVR '07, ACM. 1

[BDHB99] Bowman D. A., Davis E. T., Hodges L. F., Badre A. N.: Maintaining spatial orientation during travel in an immersive virtual environment. In *Presence: Teleoperators and Virtual Environments* (1999), pp. 618–631. 3

[CTtSO03] Chen D.-Y., Tian X.-P., te Shen Y., Ouhyoung M.: On visual similarity based 3d model retrieval. *Computer Graphics Forum 22*, 3 (2003), 223–232. 2

[DCG10] Dutagaci H., Cheung C. P., Godil A.: A benchmark for best view selection of 3d objects. In *Proc. of the ACM workshop on 3D object retrieval* (2010), 3DOR '10, ACM, pp. 45–50. 1

[DJLW08] Datta R., Joshi D., Li J., Wang J. Z.: Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv. 40* (2008), 5:1–5:60. 1

[FMK*03] Funkhouser T., Min P., Kazhdan M., Chen J., Halderman A., Dobkin D., Jacobs D.: A search engine for 3d models. *ACM Trans. Graph. 22* (2003), 83–105. 1

[HW11] Holz C., Wilson A.: Data miming: inferring spatial object descriptions from human gesture. In *Proc. of the 2011 annual conference on Human factors in computing systems* (2011), CHI '11, pp. 811–820. 2, 4

[KFR03] Kazhdan M., Funkhouser T., Rusinkiewicz S.: Rotation invariant spherical harmonic representation of 3d shape descriptors. In *Proc. of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing* (2003), SGP '03, Eurographics Association, pp. 156–164. 2

[NH01] Nakazato M., Huang T. S.: 3d mars: Immersive virtual reality for content-based image retrieval. In *Proc. of 2001 IEEE int. Conference on Multimedia and Expo (ICME2001)* (2001). 2

[PR97] Paquet E., Rioux M.: Nefertiti: a query by content software for three-dimensional models databases management. In *Proc. of the int. Conf. on Recent Advances in 3-D Digital Imaging and Modeling* (1997), NRC '97, IEEE Computer Society, pp. 345–. 2

[TV08] Tangelder J. W., Veltkamp R. C.: A survey of content based 3d shape retrieval methods. *Multimedia Tools Appl. 39* (2008), 441–471. 2