ABSTRACT
The rapid growth of 3D models has urged the problem of searching in large collections of models. Although there are a large number of tools for 3D object retrieval, these tools do not take advantage of recent technologies. The results are presented as thumbnails, which hinders the interpretation of results, and allow little or no manipulation at all. With the spreading of stereoscopic viewing and last generation interaction devices outside lab environment, richer results could be presented instead of a list of thumbnails.

This thesis describes a new approach for the visualization of query results for 3D object retrieval, which takes advantage of virtual reality immersive environment. We devised a system where users can navigate through the results of a query on an immersive virtual reality environment by taking advantage of a post-WIMP interface. The user can explore the results, navigate in the three-dimensional space and manipulate the scattered objects. The combined use of virtual environments and six degree-of-freedom devices, provides a more complete visualization of models and makes interaction more natural, with direct manipulation. To validate our prototype we performed a evaluation with users.

The goal and objectives of this thesis were achieved, proving it to be a very promising work. With the validation of our approach we opened the way to research and solve new challenges in the context of 3DOR using immersive environments.

Keywords
Multimedia Information Retrieval, 3D Object Retrieval, Immersive Virtual Environment
Instead of presenting the query results as thumbnails, we present them as 3D objects scattered in a 3D space, where the user can literally navigate and manipulate the results. The objects are distributed according to their geometric similarity, using matching algorithms for each axis of the coordinate axes. Furthermore, with the appearance of new off-the-shelf devices, it is required that our approach also aims to be modular and scalable, in order to add new devices for both visualization and interaction.

Taking into consideration the requirements described above, we merged the benefits of 3D object retrieval with virtual reality environments, building a prototype for 3D model search where query results are presented in a 3D virtual space. Our system provides a post-WIMP interface, using new off-the-shelf devices, for both the visualization and interaction. Using a modular architecture, we were also able to combine different sets of devices, as well as, make it easy to add new devices.

In order to validate our approach, an user test evaluation was conducted, in which our prototype was compared to a traditional 3D object retrieval system. Our approach proved to provide a better interpretation of results, resulting in less errors and number of steps required in order to find a specific object.

2. RELATED WORK

We built an immersive VR system for 3D object retrieval (Im-O-Ret) which will extend benefits of VR environments and post-WIMP approach, to the 3D object retrieval. As such, throughout this section we discuss related work that is relevant to our proposal, from each of these areas.

2.1 3DOR Search Engines

Last decades, several 3D shape search engines have been presented. In 2001, the Princeton 3D model search engine is introduced by Thomas Funkhouser et al. [16]. This work provides as query specification four options: text based; by example; by 2D sketch; and by 3D sketch. The results of these queries are presented as an array of model thumbnails, has depicted in Figure 1. After a search, it is also possible to choose a result as query-by-example for a new search.

Additionally to queries by example and sketch-based queries, Ansary et al. proposes the FOX-MIIRE search engine[1], which introduces the query by photo. However, and similarly to previous engines, the results are displayed as a thumbnail list.

Outside the research field, Google 3D Warehouse [17] offers a text-based search engine. This online repository contains a very large number of different models, from monuments to cars and furniture, humans and spaceships. However, searching for models in this collection is limited by textual queries or, when models represent real objects, by its geographic reference. On the other hand, the results are displayed by model images in a list, with the opportunity to manipulate a 3D view of a selected model.

Generally, the query specification and visualization of results of other commercial tools for 3D object retrieval, usually associated with 3D model online selling sites, did not differ much from those presented above. The query is specified through keywords and results are presented as a list of model thumbnails.

Despite the growing of current hardware and software, these approaches to query specification and result visualization do not take advantage of latest advances of neither computer graphics or interaction paradigms. Indeed, to the extent of our knowledge, there has not been presented any research or new solution that take advantage of immersive virtual environments for information retrieval since Nakazato’s 3D MARS [27].

2.2 3D Mars

A decade ago, Nakazato proposed 3D MARS [27], an immersive virtual reality environment for content-based image retrieval. The user selects one image from a list and the system retrieves and displays the most similar images from the image database in a 3D virtual space, as shown in Figure 2. The image location on this space is determined by its distance to the query image, where more similar images are closer to the origin of the space. The distance in each coordinate axis depend on a pre-defined set of features. The X-axis, Y-axis and Z-axis represent color, texture and structure of images respectively. Nakazato focused his work on query result visualization. Thus 3D MARS supports only query-by-example mechanism to specify the search.
The 3D MARS system demonstrates that the use of 3D visualization in multimedia retrieval has two main benefits. First, more content can be displayed at the same time without occluding one another. Second, by assigning different meanings to each axis, the user can determine which features are important as well as examine the query result with respect to three different criteria at the same time.

The interaction with the query results was done on the NCSA CAVE [10], which provided fully immersive experience, as depicted in Figure 3, and later on desktop VR systems. By wearing shutter glasses, the user can see a stereoscopic view of the world. In such solution, visualizing query results goes far beyond scrolling on a list of thumbnails.

3. EXPLORING 3D ENVIRONMENTS

Immersive 3D environments are often associated with virtual reality environments. Virtual Reality (VR) is defined by Aukstakalnis Blatner [4] as “a way for people to visualize, manipulate and interact with computers and extremely complex data”. Based on this definition and the fact that 3D models are multimedia data that is very complex to handle, we believe that the use of virtual reality environments is the ideal for our work.

3.1 Navigation

There is a large set of parameters to take into account and that obviously depend on the metaphor used for navigation. Bowman et al. [6] proposes the division into two types of navigation based on their movement as: wayfinding and travel.

The wayfinding navigation must have prior knowledge of the space, in order to create a cognitive map of the environment. The movement is done based on a coordinated that specifies the position where the camera is moved to. Mackinlay et al. [24] proposed the navigation by points of interest. A point of interest is defined as a desired location on a particular object of the scene. Chosen a point of interest it is calculated the trajectory so that there is an approximation of it. This method is not only used for calculating the translation point of view, but also for calculating the correct orientation using the information for that of the normal vector point of interest chosen.

Compared with the technique of point of interest where the user simply sets the target, this new technique offers the possibility to control the distance to the target as well as direction for the visualization of it. In the first face, the user is able to position the focus area by directly moving the virtual ray in the scene. If the ray intersects an object during the dragging movement the half-sphere is placed at the intersected position, else the sphere sticks with the ray within the distance of the last intersection. The second face relates to the positioning of the camera. In this step the user can define the radius of the semi-sphere that defines the distance to objects, as well as moving the cursor over the hemisphere in order to define the point of view camera.

Devices like the Head Mounted Displays, are another example of the use travel of navigation. Typically, these devices have the guidance of the viewport being controlled by “head tracking”. A pioneer in this type of interaction was Sutherland et al. [32], with motion sensors that could detect the correct rotation of the head, which is used in camera orientation. This system only considers techniques of viewport orientation.
Later Fisher et al. [13] extended this approach, by using devices such as gloves with which was possible to detect predefined movements that allowed the navigation and interaction of the scene. An example of these type of devices we have the 5DT Data Glove used in our prototype. The 5DT Data Glove is a glove with flexible fibre optic sensors that capture gestures and hand movements, as depicted in Figure 5a. It has six sensors that detects finger bending, as well as rotation and closing the hand.

### 3.2 Visualization

The visualization of virtual reality environments can be broadly divided into two types: non-immersive and immersive. The immersive is based on the use of helmet or projection rooms (eg. HMDs [31], CAVEs [10]) while the non-immersive virtual reality based on the use of monitors. The notion of immersion, is based on the idea of the user being inside the environment.

HMDs have two LCD screens in a relative position of each eye, which draws the virtual world depending on the orientation of the user's head via a tracking system. The Figure 6 illustrates an example of an HMD. HMD's can display the same picture in both eyes or be stereoscopic by combining separate images from two offset sources. Both of the offset images are then combined in the brain to give the perception of 3D depth.

Virtual Environments (VE) as a large space enables the display of more content at the same. Also the use of stereoscopic view of the world, gives a visualization of the world that goes far beyond 2D view of the screens. Additionally, with the growingly and common use of new human-computer interaction (HCI) paradigms and devices brought new possibilities for multi-modal systems.

![Figure 6: Head-Mounted Display: Z800 3DVisor](image)

### 3.3 Post-WIMP Interaction

The recent dissemination among common users of new HCI paradigms and devices (e.g. Nintendo Wiimote[23] or Microsoft Kinect[25]) brought new possibilities for multi-modal systems. For decades, the WIMP interaction style prevailed outside the research field, while post-WIMP interfaces were being devised and explored [35], but without major impact in everyday use of computer systems.

Particularly, the use of gestures to interact with system has been part of the interface scene since the very early days. In 1980, Bolt wrote "Put-that-there"[5], a pioneering multi-modal application where the use of natural language helps direct manipulation to describe arguments to select actions. In “Put-that-there”, the user commands simple shapes on a large-screen graphics display surface. This approach combined gestures and voice commands to interact with the system.

In 1994, Koons et al. introduces the ICONIC [22]. With the ICONIC, they present a new form of interaction based on the detection of hand gestures. This aims to better represent some operations allowed to the user. This way facilitates some commands that, so far, were unnatural. A few years later, Sharma et al. [29], describes a system with bi-modal interface, voice and gestures, integrated into a 3D visualization system, where users model and manipulate 3D objects, in this case, models of biomolecules.

Recently such interaction paradigms have been introduced in off-the-shelf commodity products. Now, with limited resources, novel and more natural HCI can be developed and explored. For instance, Lee [23] used a Wiimote and took advantage of its high resolution infra-red camera to implement multipoint interactive whiteboard, finger tracking and head tracking for desktop virtual reality displays.

In the context of 3DOR, Holz and Wilson [19], present a system which allows users to describe spatial object through gestures. The system capture gestures with a Kinect camera and then finds the most closely matching object in a database of physical objects. This system represents a good example of the use of new interaction paradigms in 3DOR, which help give the user a more natural way of interaction.

Another device which provides a very accurate targeting precision and greater control of movement is the SpacePoint Fusion [9] developed by PNI Sensor Corporation. With the use of a magnetometer, a gyroscope and an accelerometer, all of three axes it self-calibrates and maintains pinpoint accuracy, thus allowing a better immersion experience.

Also, by combining these new devices with stereoscopy, it is possible to make a multi-modal immersive interaction with direct and natural manipulation of objects shapes within virtual environments. In our current prototype we used as Wii Remote and the Spacepoint Fusion, which were tested through various metaphors for navigation and interaction.
4. IMMERSIVE 3D OBJECT RETRIEVAL

In this Section we describe our immersive VR system for 3D object retrieval, (Im-O-Ret), in order to overcome the limitations of the current 3DOR search engines. We aimed to make use of display devices for viewing and interaction of our results in immersive virtual reality environment, and existing 3DOR techniques to implement our retrieval system, not to create or enhance new methods for retrieval of 3D objects. For the end result we also aim to make our prototype accessible to anyone. Therefore, we used commercial devices for both viewing and for interaction.

4.1 Overview

With our work, we wanted a system that would allow the user to browse, navigate and manipulate, query results of a search in an immersive VR environment. Based on this propose, we conceptualized our architecture to fulfil the set of technical requirements. As such, our system was divided into two major modules, the Three-Dimensional (3D) module and the Object Retrieval (OR) module.

The OR module, would be responsible for indexing and retrieving 3D models. This should ultimately receive a model, and return a list of similar models ordered by the degree of similarity between these and the query model. Also, similar to what is done in the 3D MARS [27], were for each coordinate axis is pre-defined set of features, we intend to assign an different 3DOR algorithm for each axis. For the propose of studying and comparing different retrieval methods, it should be easy to integrate or replace these algorithms.

In the 3D module, we would create the VE were we display the results of a search. Since we aimed to make our prototype accessible to anyone, thus using of a set of different commercial devices for both the visualization and the interaction, it is required that our system is both modular and scalable, in order for an easy integration of new devices. As such, we divided this module in three sub-modules: Core, View, and Controller.

4.2 Architecture

For the implementation of our system we required a framework that would provide easy usage of multiple input devices. As such, we used the OpenIVI framework [3], which relies on both OpenSG and OpenTracker technologies to provide an integrated solution with a powerful 3D scenegraph and support for tracking devices. Thanks to flexible XML based configuration files, it enables customized prototyping and fast integration mechanisms with a complete separation of inputs, interaction and functionality. As such, it enables the usage of customized visualization devices such as Head Mounted Displays, Tiled display systems or traditional Desktop monitors. The architecture of our prototype integrated with the OpenIVI is depicted in Figure 7.

4.2.1 Object Retrieval Module

The OR module, responsible for indexing and retrieving 3D models. For the models we used the Princeton Shape Benchmark [30] model database. This benchmark contains a collection of 1,814 polygonal models, all collected from various sources in the World Wide Web. Each model, has a JPEG image file with a thumbnail view of the model, which we used for visualization in our traditional system. This traditional system was created for the purpose of compare our proposal with current traditional systems.

For matching algorithms we used the Light-Field Descriptors [8] (LFD), the Cord and Angle Histogram [28] (CAH), the Spherical Harmonics Descriptor [20] (SHA). The executables of the LFD and CAH, we implemented both in C++, based on the description provided by their respective authors. For the LFD, we also used OpenGL, for the drawing of the 2D silhouettes. For the executable of the SHA, was used a created by one of the authors\(^1\).

We used these shape matching algorithms for three reasons. First, because each targets a different set of features [34].

Also, by using shape descriptors from different categories we can increase the diversity of query results. With the combined use of different matching methods, we aimed to obtain a best overall performance.

Using each of the three shape descriptors, we extracted the feature vectors from all models. The extracted feature vectors extracted from each shape descriptor, were then indexed into three separated NB-Trees[14]. Each indexed feature vector, would have a corresponding identifier, in order to match to the corresponding model. When querying, we extract the feature vectors of the query model, for each of shape descriptors. Then using the query feature vectors, in the corresponding NB-Tree, we receive the identifiers corresponding to the model which are more similar to the query model. For querying the NB-Tree we used its k-Nearest Neighbours query.

4.2.2 Three-Dimension Module

The 3D module, will use the OpenIVI framework in order to create the VE for the display of the query results. Since OpenIVI already enables the usage of customized visualization devices, we used it’s visualization module for the management of the visualization device. For the interaction devices, we implemented our modules, using the tracking module only for the management of the NaturalPoint TrackIR cameras and the interaction module for the process of input events of the keyboard and mouse.

The module Core, which manages all the system logic will handle all the communication with the OR module for retrieval tasks. In the event of a search the Core module will communicate with the 3D module, receiving as result, a list of objects similar to the query. Then the Core module will update his list of present objects. For new objects, it will add the object to the display. If the object already exists on the list, it will issue an order to the View module, to move the object to a new position in the 3D space. For object that are displayed, but are not present in the new result list, it is removed from the display.

The purpose of the scale and relocation of models being done gradually and periodically, until reaching the goal, is to ease the load time required for the addition of all the objects in the list of result. By adding each, model separately, and adding animation to each added model, we lessen the system load and captures the user attention.

The management of input devices will be handled by the Controller module. Each sub-module will interact with existing modules and implement a predefined behaviour or functionality. This implementation, allows easy to extension and creation of new sub-module with specific interactive behaviour. With minimal effort is possible to have the system working in a context using other input devices, such as the Kinect.

In the occurrence of an event in a sub-module, it is published by the Controller module in the event sink. The published event will contain the action to be performed, and which module should handle the corresponding action. For example, the voice command used to start a new search, will be addressed to the Core module, which communicates with the OR module, while the grabbing and direct manipulation of an object will be handled by the View module. The View module will, not only, handle the presentation of the data handled by Model, but also the transformations to the objects position, scale and rotation.

4.3 Im-O-Ret

Taking into consideration the requirements described above, we built a prototype for 3D model search where query results are presented in a 3D virtual space. In our prototype, (Im-O-Ret), the results of a query are displayed and distributed in the three-dimensional space as 3D objects, instead of the traditional list of thumbnails. As such, the user can explore the results, navigate and manipulate the scattered objects in the three-dimensional space.

4.4 Query Specification

Since query specification was outside the scope of this thesis, little work was done regarding it. As such, for query specification, only a simple query-by-example was implemented. When starting a search, the user is presented with 36 models, randomly selected, that are placed in the plane xOz, equally distant from one another, as depicted in Figure 8.

To perform a search the user selects one of the initial presented models, for a search for similar objects. Then, the user can either perform a new search, in which the system will present a new set of 36 models or another search using one of the query result models as query.

4.5 Spacial Distribution of Results

The query results are distributed in the virtual 3D space according to their similarity. To each axis it can be assigned 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 first, the objects are distributed in the space with
equal distance from each other. This allows to view the objects individually without occlusion. In this mode, although the more similar are closer to the origin of the 3D space, there is no information of the different of similarity between two objects. In the second mode, the objects position is the ground-truth value retrieved from each algorithm. This cause the creation of clusters, when the retrieved values of different models are very similar, which provides a better study and analyse of algorithms and query-results.

4.6 Exploration of Query Results
In order to view the scattered models of a search result, the user can explore the results, by navigating in that space and directly manipulating the objects. For the navigation we considered the division proposed by Bowman et al. [6], were he divides the navigation into two types based on their movement. As such we created both a travel and a wayfinding.

For the travel it consists in exploratory movements were the viewpoint is moved from one location to another, by using either the arrows in the keyboard, the wiimote nunchuck, or tracking tools, similar to the process of walking. The wayfinding was described by using points of interest, where the movement is done by selecting an object that specifies the position where the camera is moved to. For the wayfinding, we used a combination of pointing and voice commands.

In Figure 10 is illustrated the use of tracking tools for the exploration of the 3D world. This interaction does not require neither much teaching or previous knowledge in order to use. Also, the use of shutter glasses, provides the user with a stereoscopic view of the world, which enables a more complete immersive experience. As seen, the use of stereoscopic view and multi-modal ways of interaction, provide a visualization of query results that goes far beyond the traditional scrolling on a list of thumbnails. In this interaction scenarios, the user literally navigates among the results in the three-dimensional space.

4.7 Multimodal Interaction
The use of voice commands, complements the pointing in describing arguments and selecting actions. To effectively use the information from user input through words, direct manipulation helps users know which concept to cover their actions. For instance, to use an object as a query, user can point to it and say ‘search similar to this’, thus triggering a new query. The combined use of virtual environments, devices with six DoF and voice commands provides effective result visualization and makes interaction natural, comprehensible and predictable.

For each action there are some voice commands with few small variations. Most of this variations consist in the order or addition of some words, which gives the user less need of learning the commands.

The proposed modular architecture, makes the system able to be configured to work with a wide range of different interaction devices and displays. Combining different visualization and interaction devices, allows us to create multiple interaction paradigms for our system. With minimal effort is possible to have the system working in a context using other input devices, such as the Kinect. This way we could combine different visualization and interaction devices, and create multiple interaction paradigms for our system.

5. EVALUATION
In order to validate our proposal, it was necessary to conduct a set of evaluation tests. The tests were structured into three stages: a set of preliminary tests, conducted by a small group of users which followed the project development; three search task, using our prototype and a traditional system; and, finally, using a set of different interaction paradigms, followed by a post-questionnaire to rate the prototype’s ease of usage with each of the scenarios.

5.1 Preliminary Evaluation
The first stage of test was composed by informal tests conducted during the development of the prototype. It was carried out using a panel of six users, all of them with knowledge on both retrieval and virtual environments. This test
where mainly focused in two main features: interface and interaction.

Thanks to this preliminary evaluation, it we were able to constantly analyse and refine each of our solution versions.

5.2 Search Efficiency Evaluation

In order to validate our system it was required to compare it with the traditional search engines. As such, we created a web-page where users could perform a traditional 3DOR search, where the query result are presented as a list of thumbnails. We named this system Thumbnail Object Retrieval (THOR).

5.2.1 Test Description

In this test, the participants would perform three searches of increasing difficulty using THOR. The same three searches, would then be perform using Im-O-Ret, as depicted in Figure 13. In order to make the starting point in both systems be the same, we used a static group of 36 models as queries for a new search.

We counted with twelve participants, from which we noted the number of steps, errors, and time required to find a specific object. We considered as an error each time the user either rolled back to a previous search result, or re-started the search. For this test, since we intended to only test the advantages of using the 3D models instead of thumbnails. As such, we used a simple computer screen with the mouse as pointing device. This test was followed by a post-questionnaire to rate the prototype’s usability.

5.2.2 Results and Discussion

This evaluation allowed us to compare our approach with a traditional 3DOR system. In Figure 14 we present a comparison of the number of steps taken from both THOR and Im-O-Ret. In the first search task, there was no meaningful difference in the number of steps. However, the more challenging the search task, the fewer the steps using the Im-O-Ret in comparison with THOR.

The same observation can be seen in the number of error, depicted in Figure 15. The average of both systems had very close values, for the first search task. However, for the more complex second and third search tasks, our approach required less steps and did errors, when comparing with THOR. These results, allow us to conclude that Im-O-Ret provides better visualization and interpretation of the query results.

However, the same cannot be said of the time required to perform the search tasks. The average of time required, depicted in Figure 16, shows that the participants wasted more time in the tasks performed using the Im-O-Ret. During the test we observed two main causes. Since used it used a 3D environment, the users wasted more time navigating and viewing the query results, instead of performing new searches.

The second cause, was the spatial distribution of the 3D objects. Since we presented the 36 query models in a board, most users expected the models to be scattered only in the board, and did not like different height to each object, caused by the shape descriptor assigned to the Y-axis. Since our systems is easily configurable, we removed the shape descriptor of the Y-axis, making the objects being of scattered only using the shape descriptors assigned to the X-axis and Z-axis.
5.3 Interaction Paradigms Evaluation

In this evaluation, we tested the usability of our system, using distinct interaction scenarios with off-the-shelf devices, for both visualization and interaction. This test were followed by a post-questionnaire to rate the prototype’s ease of usage in each of the interaction scenarios. We used four different interaction paradigms.

5.3.1 Test Description

For the first scenario we used a computer screen with the mouse as pointing device. The second scenario, was performed using a commercial TV screen and a Wiimote. As third scenario, we used a large-screen display, the LEMe Wall [2], with a six DoF interaction device, the SpacePoint Fusion. Finally, for the fourth scenario used a HMD, the Z800 3DVisor, the 5DT Data Glove, and the tracking system with ten NaturalPoint TrackIR cameras, for the tracking of the head and hands positions.

These evaluation was conducted in laboratory João Lourenco Fernandes² on the Taguspark campus, and counted with the participation of ten users. The users were asked to conduct simple searches and then navigate and manipulate the query results. In the end, we asked the participants to fill answered a questionnaire that aimed to observe the ease of using each of this interaction paradigms.

5.3.2 Result and Discussion

This evaluation allowed us to gauge the ease of learning and using different interaction paradigms with our prototype. From the questionnaires we observed that, for the easiest interaction paradigm to use, the participant were divided between Screen + Mouse and the LEMe Wall + Spacepoint. In the case of the Screen + Mouse, we concluded that is due being an interaction in the daily use of the computer, and most participants already being used to such interface.

The selection of the LEMe Wall + Spacepoint, was mainly thanks to the motion-tracking system of the SpacePoint device which auto-calibrates and maintains a precise accuracy of the location pointed by the user. The fact that the object were displayed in a large screen, also provided the user with a better view of the query result, without need of navigating to a closer position of the object.

For the easiest interaction paradigm to learn, most participants chose the HMD + 5DT Data Glove and a smaller number divided themselves between the use Screen + Mouse and the LEMe Wall + Spacepoint.

The fact that the HMD + 5DT Data Glove takes advantage of the user’s body movements, such as walking and rotating the head to set the virtual camera position, as well as the use of simple gestures to manipulate the objects in the 3D world, made it very easy to learn. However, it was pointed out by the participants, that the great number of wires used by these devices hindered their movement and interaction, and forcing the constant awareness in order to avoiding the wires.

The usability rating of TV Screen + Wiimote, faded in comparison with the LEMe Wall + Spacepoint. There were two main reason for this. First, the LEMe Wall provided a more wide view of the virtual world than the TV Screen. Second, the Wiimote use of the accelerometer and gyroscope, is limited in comparison to the SpacePoint who additionally has a magnetometer, that provides a more direct mapping.

6. CONCLUSIONS AND FUTURE WORK

The use of lists of thumbnails for the query result visualization in current 3D search engines, greatly hinders the interpretation of query results on collections of 3D objects. In this thesis we have proposed a novel approach for the query result visualization of 3D object retrieval. In this context

---

we studied 3D object retrieval, virtual environments, and post-WIMP interaction. Taking advantage of each fields benefits, we developed a system which merges 3D object retrieval techniques with virtual reality environments, using new off-the-shelf devices.

The Im-O-Ret, presents the query results as 3D objects, instead of the list of thumbnails presented in current traditional systems. The query results are scattered in a 3D space according to their similarity. Each coordinate value will be determined by an assigned shape matching algorithm to that axis. These matching algorithms can be replaced, in order to produce more precise results for specific domains.

The Im-O-Ret uses a post-WIMP interface, which can use a set of different interaction devices and displays. For instance, Im-O-Ret supports combining hand gesturing and speech commands to provide a more natural interaction. From a practical point-of-view, thanks to our modular architecture, it is possible to add new input devices with minimal effort.

To validate our solution we conducted a set of user tests, where a panel of users, evaluated a traditional 3D search engine against our 3D retrieval approach. The results gathered allow us to conclude that Im-O-Ret provides better visualization and interpretation of the query results. However it requires more time to perform search tasks. This is mainly due to overhead of navigation through the results. Nevertheless, thanks to the better visualization and interpretation of the query results provided, with the Im-O-Ret the search tasks are done in less steps and with fewer errors. In a final evaluation we tested the usability of our system with multiple interaction scenarios, using different sets of visualization and interaction devices.

With the validation of our approach, we raises new challenges for 3DOR using immersive virtual environments. For instance, the usage of different shape matching algorithms can generate different results, specific to certain domains. Also, the spatial organization of results in our approach, was implemented as a simple solution by assigning a different shape descriptor to each axis. However, we concluded that such solution not always provide a distribution of query results that is meaningful to the user.

In the navigation, we observed that the use of a simple wayfinding technique, was limited and allowed little control. The Navidget technique [18, 21], which extends the techniques by points of interest, and offers the possibility to control the distance to the target as well as direction for its visualization, illustrates an example that would improve the exploration in our system.

Another major challenge that should be tackled is the query specification. For our prototype only a simple query-by-example was implemented, but similar to the Princeton 3D model search engine [16], with the 2D and 3D sketches, a more query interface is required. For instance, the integration of gesture-based query specification, such as the one proposed by Holz and Wilson [19].

These wide range of different features could provide a better visualization, navigation and interaction to our prototype, which alone, has given a huge step regarding 3DOR in immersive environments. With the knowledge and achievements accumulated in our work, we hope that it may be a reference for future researches on these subjects.

7. ACKNOWLEDGMENTS

The work described in this paper was partially supported by the Portuguese Foundation for Science and Technology (FCT) through the project 3DORuS, reference PTDC/EIA-EIA/102930/2008 and by the INESC-ID multianual funding, through the PIDDAC Program funds.

8. REFERENCES


