LS3D: LEGO Search Combining Speech and Stereoscopic 3D

Pedro B. Pascoal, INESC-ID/Técnico Lisboa, Universidade de Lisboa, Lisboa, Portugal
Daniel Mendes, INESC-ID/Técnico Lisboa, Universidade de Lisboa, Lisboa, Portugal
Diogo Henriques, INESC-ID/Técnico Lisboa, Universidade de Lisboa, Lisboa, Portugal
Isabel Trancoso, INESC-ID/Técnico Lisboa, Universidade de Lisboa, Lisboa, Portugal
Alfredo Ferreira, INESC-ID/Técnico Lisboa, Universidade de Lisboa, Lisboa, Portugal

ABSTRACT

The number of available 3D digital objects has been increasing considerably. As such, searching in large collections has been subject of vast research. However, the main focus has been on algorithms and techniques for classification, indexing and retrieval. While some works have been done on query interfaces and results visualization, they do not explore natural interactions. The authors propose a speech interface for 3D object retrieval in immersive virtual environments. As a proof of concept, they developed the LS3D prototype, using the context of LEGO blocks to understand how people naturally describe such objects. Through a preliminary study, it was found that participants mainly resorted to verbal descriptions. Considering these descriptions and using a low cost visualization device, the authors developed their solution. They compared it with a commercial application through a user evaluation. Results suggest that LS3D can outperform its contestant, and ensures better performance and results perception than traditional approaches for 3D object retrieval.

Keywords: 3D Object Retrieval, Immersive Visualization, Multimodal Interaction, Speech Recognition, User Interfaces, Voice Search

INTRODUCTION

The appearance of low-cost technologies that allow scanning of three-dimensional physical objects, such as the Microsoft Kinect\(^1\), Asus Xition\(^2\) or PrimeSense Sensor\(^3\), along with the vulgarization of 3D modeling software, has resulted in a considerable increase of available 3D virtual objects. An implicit consequence of this growth is the increased complexity in searching for a specific 3D model desired by the user, which leads to a slow and tedious retrieval process.

When performing the retrieval of 3D objects and other types of multimedia objects, the intrinsic information contained in them, such as the corresponding files names, has proved insufficient, and new meta-data is often needed (Funkhouser et al., 2003; Smith and Chang, 1997).
With the purpose of overcoming this challenge, several solutions for performing retrieval have been proposed. Some resort to textual annotations (Funkhouser et al., 2003; Smith and Chang, 1997), sketches (Liu et al., 2013; Santos et al. 2008), verbal descriptions (Lee et al. 2010; Wang et al. 2011), gestures (Holz and Wilson, 2011) or using an object as an example (Lavoué, 2011; Paquet and Rioux, 1997). However, all the proposed solutions have drawbacks. For instance, retrieval by example requires the user to have a similar object to use as an example, which may not always be available. The remaining solutions do not properly explore the descriptive power and potential of human interaction. Although some studies (Holz and Wilson, 2011; Kamvar and Beeferman, 2010; Lee and Kawahara, 2012) have followed this direction, albeit not applied in the context of the 3D object retrieval, these do not combine the multimodality of speech and gestures used in natural human interactions. Moreover, the current solutions for searching 3D objects visually present results in a grid of thumbnails (Funkhouser et al., 2003). This may be an inadequate representation, since it loses relevant 3D information.

In order to explore the descriptions of 3D objects, we conducted an experimental session with users in order to understand which verbal and gestural expressions are naturally used. Our aim was to understand if users resort more to speech, gestures or a combination of both. For this purpose, a scenario of building LEGO models was designed, in which one of the subjects had to request the necessary blocks from another subject, describing them as accurately as possible. Based on the results of this experiment we developed a system in which users can search for 3D objects using multimodal interactions. Both exploration and result analysis is performed through a 3D immersive environment, in order to provide a clear representation of the virtual LEGO blocks, as we can see in Figure 1. The use of LEGO blocks has already proven to be a good context to explore new concepts such as interactions (Mendes et al. 2011; Santos et al., 2008) or object tracking (Gupta et al. 2012; Miller et al. 2012). Although we use the LEGO blocks context, our solution can also be extended to different contexts.

In the remaining of the paper we will discuss the state of the art for searching multimedia content, focusing on three-dimension content. Then we present the preliminary study and a discussion regarding its results. After the study it is presented our prototype, followed by an evaluation where we compare our prototype against a commercial application. Finally, we present our conclusions and point out some directions for future work.

RELATED WORK

With the increase of available digital objects of any type, retrieving specific information presents a challenge, and three-dimensional objects, such as any other type of multimedia content, are no exception. One of the traditional ways to perform retrieval consists of using textual queries. However, this method is not trivial, considering that objects do not usually contain sufficient intrinsic information. For instance, files name may not be related to the objects. Generically, search engines often use text associated with objects such as captions, references to the objects or even, when it comes to contents scattered over the Internet, the links of the objects or file names. This concept has already been applied in image retrieval (Smith and Chang, 1997) and 3D object retrieval (Funkhouser et al., 2003). In the work of Funkhouser et al., synonyms of words taken from the texts are also used to of increase the information used to describe the 3D object and resolve vocabulary mismatch.

Despite all these solutions, the information to describe a 3D object is still insufficient, especially regarding its shape. Some of the proposed solutions use query-by-example to ease this process. The goal of search by example is to obtain similar objects in terms of visual aspects such as the
Related Content

Procedural Generation of Road Paths for Driving Simulation
[www.igi-global.com/article/procedural-generation-of-road-paths-for-driving-simulation/147171?camid=4v1a](www.igi-global.com/article/procedural-generation-of-road-paths-for-driving-simulation/147171?camid=4v1a)

Materials of the Data Map
[www.igi-global.com/chapter/materials-data-map/64053?camid=4v1a](www.igi-global.com/chapter/materials-data-map/64053?camid=4v1a)

Behind the Scenes: Service Experience in Traditional Performing Arts Groups in Indonesia
[www.igi-global.com/article/behind-the-scenes/204438?camid=4v1a](www.igi-global.com/article/behind-the-scenes/204438?camid=4v1a)

Picturing Minerals and Rocks
[www.igi-global.com/chapter/picturing-minerals-and-rocks/85390?camid=4v1a](www.igi-global.com/chapter/picturing-minerals-and-rocks/85390?camid=4v1a)