LSpeakIt: Immersive interface for 3D object Search

Diogo Gil Vieira Henriques
Instituto Superior Técnico,

ABSTRACT
The number of available three-dimensional digital objects has been increasing considerably. As a result, retrieving such objects from large collections has been subject of research. However, proposed solutions do not explore natural interactions in their interfaces. In this work, we propose a speech interface for 3D object retrieval in immersive virtual environments. For our prototype, LSpeakIt, the context of LEGO blocks was used as a toy-problem. To understand how people naturally describe objects, we conducted a preliminary study. We found that participants mainly resorted to verbal descriptions. Considering these descriptions, we developed our search by speech system. Taking advantage of a low cost visualization device, the Oculus Rift, we implemented four modes for immersive query results’ visualization. These modes were evaluated by users, being compared against each other and a traditional approach. Users favored the immersive modes, despite being more acquainted with the traditional approach. For our final prototype, a fifth mode for result’s visualization was implemented, adapting the users’ preferred modes. We compared the proposed solution with the LEGO Digital Designer commercial application. Results suggest that LSpeakIt can outperform its contestant, while providing users with a simple and natural way for searching virtual objects, and ensuring better performance and results’ perception than traditional approaches for 3D object retrieval.

Author Keywords
Immersive interface, Voice Search, 3D Objects, Retrieval

INTRODUCTION
Virtual environments in computer entertainment are becoming ever more immersive [24]. Indeed, from CAVE like setups [8] to heads-up displays, such as Oculus Rift\(^1\) or the recently announced Project Morpheus by Sony, these immersive environments are getting increasingly common. Several fields can take advantage of this, being able to place the user inside of the virtual world in a more credible way than traditional setups, offering more natural interactions with the virtual content. One possible application for these kind of user immersion is the creation of virtual scenarios [2], which allow the user to place virtual objects mimicking physical world interactions.

However, when facing the challenge of selecting objects in a collection, traditional solutions based on lists and grids of thumbnails may not be feasible, due to changes in the interaction paradigm. Also, it has already been shown that immersive environments can be used to enhance retrieval results explorations over traditional approaches [6]. In this work, we focus on naturally navigate large collections of objects in immersive environments, in order to select a specific one. For this purpose, we use the LEGO building blocks scenario, which has already proven to be a good test bed for research, leading to potentially interesting entertainment applications [22, 15, 16, 5]. We developed a system in which users can search, select, explore and place 3D LEGO blocks through multimodal interactions in a 3D fully immersive environment, as depicted in Figure 1.

In the remaining of the paper we will discuss the state of the art for multimedia content searching, focusing on three-dimensional virtual objects. We then present a preliminary study we conducted to understand how people naturally describe physical LEGO blocks and its results. After, our prototype is described, followed by an evaluation comparing it against a commercial application. Finally, we lay out our conclusions and point out some directions for future work.

RELATED WORK
With the increase of available 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

\(^1\)Oculus Rift: http://www.oculusvr.com/rift/
method is not trivial, considering the objects do not usually contain sufficient intrinsic information. For instance, the files names’ may not be even related to the objects [23, 3]. Generically, search engines often use text associated with objects, such as captions, references to them or even, when it comes to contents scattered over the Internet, links or file names. This concept has already been applied in image retrieval [23] and 3D object retrieval [3]. In the work of Funkhouser et al. [3], synonyms of words taken from the texts are also used to increase the information describing the objects and resolve vocabulary mismatch.

Despite all these solutions, the information to describe a 3D object is still insufficient, specially 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 color [18] or shape [10]. Moreover, this solution requires the user to already have an object identical to the one he is searching, which is usually not the case.

Retrieval by sketching is one way of addressing this problem, offering users the possibility of searching for objects similar to the users’ sketches, for which they do not have a model to serve as an example [21]. In the work of Santos et al. [22], users can make sketches that match the dimensions of a desired LEGO block. Funkhouser et al. [3] also proposed a method of sketching several 2D views of the model. In a different approach, Liu et al. [13] attempted to improve search by sketches taking into account the user profile, i.e. the habits of the users’ when drawing. This lead to improved results as users perform more searches.

Holz and Wilson [7] followed a different approach in order to apply a method often used to describe physical objects. In their work, the authors focused on recognizing descriptions of three-dimensional objects through gestures. This work consisted of capturing and interpreting gestures and exploring the spatial perception of the users. The shape and movement of the users’ hands when describing the objects are used to create a three-dimensional shape by filling the voxels which users’ hands crossed. The authors concluded that participants were able to keep the correct proportions relatively to the physical objects and, in more detailed areas, users performed gestures more slowly.

However, these works mainly present results relying on the traditional approach of thumbnails’ lists. Nakazako et al. [17] presented 3D MARS, which demonstrates the benefits of using immersive environments for multimedia retrieval. Their work focused mainly on the presentation of query results for a content-base image retrieval system, using a CAVE like setup. Extending 3D MARS’ approach for 3D objects, Pascoal et al. [19] showed that some challenges can be overcome when presenting 3D object retrieval results in this kind of environments. In these works, results are distributed in the virtual space according to the similarity between them. Users can then explore the results by navigating in the immersive environment, which is seen through a head-mounted display. Their system also allows a diversified set of different visual-ization and interaction devices, which were used to test multiple interaction paradigms for 3D object retrieval.

In the area of the information retrieval, we have recently witnessed a wide spread of search-by-voice on mobile devices. This new possibility has led to the preference of this search method over the traditional search-by-text [9]. In most cases, when searching by voice, the speech is converted to text, which is then used as a search parameter [25, 12]. More recently, Lee and Kawahara [11] performed a semantic analysis of the speech queries used to search for books, achieving a greater understanding of what the user desires to retrieve.

The retrieval of multimedia content, in particular for three-dimensional objects, has been subject of previous research. Most solutions, although already started to explore natural methods for describing objects, such as mid-air gestures, do not yet conveniently explore the potential of the interaction between humans and its descriptive power. In other areas, verbal descriptions are already being used for retrieving content. However, they have not yet been applied to 3D objects or complemented with other natural descriptive methods. When the viewing retrieved results, some solutions provide immersive environments. Nevertheless, some of the results may appear overlapped if they are too similar. Traditional approaches do not overlap results, but their visualization based on thumbnails lacks an adequate 3D representation of objects and is not suitable for interacting with in immersive setups.

**PRELIMINARY STUDY**

To understand which methods are most natural for people to describe LEGO blocks, we conducted an experiment. For this experiment, ten pairs of participants were involved. In each pair, one participant had to request specific blocks to the other participant, in order to build a model. Each participant performed the two roles: once as a builder and once as supplier. After a preliminary introduction, it was given to the builder step-by-step instructions to assemble a model, and the supplier was given a box of blocks, containing more than those needed to complete the model. A small barrier between the two participants prevented the builder from seeing the box and the supplier from seeing the instructions, but they could see each other’s faces and hand gestures.

The experimental setup was done with four different models, composed of different blocks but with a similar geometric complexity. The instructions for each model included 20 steps. Figure 2 illustrates some of the steps of one of the models. For each step, the builder had to request the corresponding block from the supplier, describing it as he thought it was more suitable. The supplier had to search for the block and hand it to the builder. After completing their first model, the two subjects changed roles, repeating the process for a different model. After building the two models, they were asked to fill a short questionnaire.

The experiment had 20 participants (7 female), the majority of the participants were college students, whose ages ranged from 18 to 45 years old (55% between 18 and 24). All participants owned at least a bachelors degree. All of them were familiar with LEGO blocks. Except for 4 participants, all
knew their partner prior to the experiment, a fact that origi-
nated very informal communications. All participants were
native portuguese speakers.

As shown in the graphic in Figure 3, 11 out of 20 partici-
pants did not use any gesture to describe blocks. Occasion-
ally, participants used gestures, but always only for comple-
menting the verbal description, without adding any signifi-
cant information. For example, to describe a corner piece,
some participants used the letter ‘L’ as a reference, making
the corresponding gesture while referring the letter, as de-
picted in Figure 4. Another example of how users comple-
mented their speech descriptions from time to time was by
performing sketches in the air, in the shape of the desired
block.

It was often observed that participants made several refine-
ments to the descriptions. So, whereas most blocks took only
a single step to be described, on average 1.4 descriptions were
needed for each step. On average, 4 steps per model had
refinements to the original descriptions, while 2 steps had
refinements made because the supplier provided the wrong
block. Even though half of these refinements were driven by
the wrong block being returned, on several occasions the par-
ticipants in the role of builder added more detail to the initial
description by their own initiative.

The specification of each block usually began with its dimen-
sions and color. Regarding dimensions, the unit used was
the number of pins of the block, or the equivalent space for
blocks without pins. Adjectives were used very often in or-
der to avoid counting pins. When presented to a query such
as Now I want a long block, red, with lateral holes, and holes
going from one side to the other, the supplier just searched for
the longest block with these characteristics, without counting
the actual number of of connectors or holes. Likewise, the
height of each block, when differing from normal, it was ei-
er referred to as thin, or high. Blocks with slopes were often
referred to as ramps or roof-shaped, followed by the dimen-
sions at the base and top.

Metaphors were also often used in descriptions. The most fre-
quent ones were related to the already mentioned roof-shaped
or L-shaped blocks, which took part in several models. The
most unusual blocks were the ones that triggered the use of
more creative descriptions, such as letters, teeth and the top
of a trident.

Additionally, at the end of the session, each participant was
asked to fill a questionnaire about the use of the modalities
(speech, gesture, and the combination of both) regarding the
easiness to describe objects. The classification was done us-
ing a Likert scale with four values (1-very difficult and 4-
very easy). The results are presented in Table 1. The anal-
ysis of the answers was made using the Wilcoxon signed-
ranks test. The participants showed a strong preference for
using exclusively speech, compared to using gestures (Z = -
4.018, p = 0.000) or a combination of modalities (Z = - 3.502,
p = 0.000). Furthermore, the combination of gestures and

<table>
<thead>
<tr>
<th>Description Mode</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal</td>
<td>4 (1)</td>
</tr>
<tr>
<td>Gestural</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Combined</td>
<td>2 (1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description Mode</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal</td>
<td>4 (1)</td>
</tr>
<tr>
<td>Gestural</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Combined</td>
<td>2 (1)</td>
</tr>
</tbody>
</table>

Table 1. Classification in terms of ease in describing blocks for three modalities (median, interquartile range).
speech was consistently preferred over using just gestures ($Z = -3456$, $p = 0.001$).

**LSPEAKIT PROTOTYPE**

After verifying that the participants clearly prefer describing the blocks through speech, we developed a prototype, LSpeakIt, that provides a natural way of retrieving 3D LEGO blocks in an immersive environment. To take advantage of speech descriptions, our prototype supports spoken queries, as well as it synthesizes text to speech to give feedback to the user.

**Architecture**

Our prototype was developed accordingly to the architecture presented in Figure 5. We devised three main layers with several modules: Interaction/Visualization, Control and Data. The layer Interaction/Visualization is where we implemented the Voice Interface (Speech Manager module). Here, the spoken queries of the user are recognized and audio feedback is synthesized to inform the user about what was understood by the system. We also implemented interactions with gestures (Gesture Manager module), allowing users to interact using the Leap Motion or the Space Point Fusion devices. The Camera Manager controls the point of view of the user. Is also in this layer that the visualization of the objects as well as the arrangement of the retrieved results is done, in the Visualization module. For this purpose, we use OpenSG, an open source graph system that undertakes the conversion to OpenGL primitives and the rendering process.

The Control layer contains the Query Manager and Objects Manager modules. The Query Manager is where queries are created using users’ input. The Objects Manager creates and displays objects returned as query results and allows the user to interact with them. Finally, the Data layer is where the information of the models is stored and indexed. The Indexation module uses both textual and shape descriptions. For the LEGO blocks’ models we resorted to LDraw\(^2\), an open-source library that contains their geometric information and textual descriptions, and selected a subset of 1810 blocks.

**Indexing**

In the LDraw library, blocks are associated with carefully made textual descriptions, using consistent terms across the library, as exemplified in Figure 6. Since this is a community driven effort, these terms are initially defined by the person who modeled the virtual representation of the block, and then validated by other members. This process assures high quality descriptions. Each block is also defined by its geometrical information, such as lines, triangles, quadrilaterals and colors. To maintain all the information regarding the 3D objects reachable in an easy and fast way, we create indexes and inverted indexes for both textual descriptions and shape matching information. Concerning textual descriptions, each object has its description’s terms, along with the number of times they are repeated, in the descriptions index. Each term has the objects that are described with it associated in the inverted index.

Capitalizing on existing solutions for text matching, we applied the approach proposed by Gennaro et al. [4] in the context of 3D object retrieval. To extract the shape features we used the D2 descriptor that had show promising results in previous works [3]. We extracted the shape features for each object, creating its own signature in a 128 dimensional feature vector. Based on the idea that similar objects will have similar view of the descriptors’ surrounding space [4], we selected a set of reference objects (RO) from the collection. The euclidean distance of the signature of every object in the collection to each one of these RO is calculated. This distance is then used to sort all RO for each object. Assigning a unique term for each RO, we create a textual representation for an

\(^2\)LDraw: http://www.ldraw.org/
object repeating these terms more times the closer the respective RO is to the object for an object repeating these terms more times the closer the respective RO is to the object.

We use 85 randomly selected RO from the collection to represent the 1810 objects, accordingly to the equation suggested by Amato and Savino [1]: \( \#RO \geq 2 \cdot \sqrt{N} \). To achieve an acceptable compromise between indexing and querying time and accurate results, we represent each object with the 30 closest RO. This way, although comparing each object with 85 RO, we only use 30 in its representation, thus significantly reducing the size of the textual representations. The generated representations of the objects in textual form can be processed in a similar fashion as textual descriptions. For each object, its closest RO are associated in the shape index, giving a higher relevance to closer objects. For each RO, the objects of the collection that are closer to it are associated in the shape inverted index.

To retrieve objects in the collection, our system relies on the TF-IDF approach to calculate results’ relevance. We developed two types of queries: query-by-text and query-by-example. The query-by-text follows a direct implementation of a standard search engine for textual documents. For each term in the query, the system will gather objects that contain the term from the corresponding inverted index, combining these results. Then, the similarity of each object to the query is calculated. The query-by-example uses both textual description and shape information. With this information, we perform queries-by-text to the respective indexes. Since both approaches generate a similarity value contained in the interval \([0,1]\), they can be averaged, while keeping the final similarity value also on this interval. The main idea is that objects that appear similar to the query using both textual description and shape information will be more relevant.

**Query specification**

To query the system users can use spoken queries. For this purpose, the spoken interface was done integrating speech recognition [14] and synthesis [20] modules. We built a grammar using using GRXML\(^3\), based on the most common expressions from the recordings of the preliminary study. The selected vocabulary included more than 500 different words, due to the inclusion of inflected forms of the words from the training data.

Our grammar allows users to describe a single block with more than one description. Users can describe blocks through their dimensions as using adjectives. For example, considering the block in Figure 6 (left), which belongs to the class of rounds LEGO blocks, could be described with 14 different adjectives (e.g. redondo / round, circular / circular, cilíndrica / cylindrical, circunferência / circumference). The block in Figure 6 (right), which belongs to the plates class, can be described by 21 different adjectives (e.g. fina / thin, baixinha / short, espalmada / flattened, plana / flat).

In our system, every query starts with the keyword, Acorda LEGO / Wake up LEGO, to which the system verbally replies with a prompt Sim / Yes. An example of a query might be Acorda LEGO. Quero um bloco fina, dois por dois / Wake up LEGO. I want a thin block, two by two. In our system, as every block can have a multitude of colors, users can concentrate in describing the blocks’ shape and dimension in the first query. Once the desired block is selected, they can change its color in a new query, as in Acorda LEGO. Pinta de encarnado / Wake up LEGO. Paint it red. Given the very large amount of LEGO blocks, three additional strategies have been implemented to restrict the number of pieces shown. Users can verbally filter by a given characteristic (e.g. Filtra por curva / Filter by curved), or exclude a type of blocks (e.g. Exclui DUPLO / Exclude DUPLO) or get blocks similar to the selected one (e.g Dá-me semelhante a esta / Give me similar to this one).

After the user specified a query, the system will promptly repeat what it understood. For example, for the query Acorda LEGO. Quero um bloco dois por dois / Wake up LEGO. I want a two by two block the system will synthesise Ilustrando peças dois por dois / Showing blocks with two by two. After this, the system will display the blocks that correspond to the user’s query.

**Exploration of Results**

To explore queries’ results, we followed the work presented by Henriques et al. [6]. We devised a visualization approach combining the Cylindrical and Spherical modes, which were the users’ preferred, accordingly to the authors. Our approach, Barrel, shows results distributed in the front half of a barrel (Figure 7), situating the user inside the barrel and placing the higher ranked blocks in front of the user, expand-

---

\(^3\)Speech Recognition Grammar Specification, http://www.w3.org/TR/speech-grammar/
Figure 8. Arrangement of objects by rank.

From there. Additionally, we used a colored square behind each block to illustrate its rank, being greener for the higher ranked and redder to the least ranked, as shown in Figure 8. Also, to provide a better perception of the objects, all blocks are slowly rotating. This allows users to perceive some additional details. When the user selects one object, the remainder will stop, in order to unnecessarily draw user’s attention.

Similarly to the modes developed by Henriques et al. [6], we started by using the Leap Motion device, but after some brief pilot experiences we noticed that people often lost track of the volume where they could interact. Because of that, we explored different interactions using the Space Point Fusion device, which didn’t present this restriction. After the integrating this device with the system, we compared interactions with the two devices through a brief user evaluation. This evaluation were undertaken by 8 participants, 50% of them didn’t have any experience with neither these devices, 25% had experienced with both devices and the remaining 25% just one of the devices. We verified that not only did users prefer to use SpacePoint Fusion, they were also faster performing the testes with it than with Leap Motion.

Resorting to the gyroscope of the Space Point Fusion, users can point to a specific block (Figure 9). After pointing to a block, the left button of the device can be pressed to bring the object closer to the point of view of the user. Then, the user can freely rotate the device in his hand to rotate the object with a 1:1 mapping, allowing the selected block to be viewed in different angles, as illustrated in Figure 10.

Building a LEGO Model

After browsing the collection and finding the desired block, we allow the user to build a custom virtual LEGO model. After selecting the block, the user can press once again the left button of the Space Point Fusion, hiding the query’s result and changing to build mode, while keeping the selected block. Here, a LEGO grid is displayed, as depicted in Figure 11. The user can then point at any desired position to place the block, pressing once again the left button of the device.

In build mode, it is also possible to manipulate blocks already placed in the model, simply by pointing and selecting them. While keeping a block selected, the user also can search by similar blocks or color it, using the aforementioned voice instructions, or even discard it, by pressing the right button of the Space Point Fusion.

USER EVALUATION

To validate our solution, we compared LSpeakIt against a commercial application, Lego Digital Designer (LDD), through an user evaluation. LDD uses a Windows Icons...
Menus and Pointing devices (WIMP) paradigm and presents the blocks collection in a 2D traditional grid of thumbnails divided in categories.

Participants
The evaluation was carried out by 20 participants (4 female), whose ages ranged from 18 to 50 years old (70% between 24 and 29). All participants were native Portuguese speakers, being most of them from Lisbon. Concerning previous experience, none of them had used LDD before. However, they were experienced with image search systems, using them on a daily basis. Only 30% of participants were familiar with 3D objects search engines, where all systems represented objects through thumbnails. 60% of participants had already used systems with spoken querys. Also 60% of participants had already experienced some kind of stereoscopic visualization, however only 50% had previously tried head-mounted displays.

Apparatus
The evaluation sessions were conducted in a controlled environment, without external influences. To run LDD, a computer with a standard screen with mouse was used, as shown in Figure 12 (left). LSpeakIt was experienced using the SpacePoint Fusion for interactions, Oculus Rift for the visualization and a headset for voice interaction, as can also be seen in Figure 12 (right).

Methodology
User tests were structured in four stages: pre-test questionnaire to evaluate user profile and previous experience; briefing of the purpose of the tests; execution of the two tasks in the two systems; and questionnaires regarding the used systems. To ensure even test distribution, 50% of the users started the test with LDD and the other 50% with our prototype.

The experience with each system started with a brief adaptation time, where users could experience its interface. After this, users were asked to search for eight blocks, which were shown to them using physical LEGO blocks. The order of the blocks was selected randomly, but different users would never do the same sequence, in order to ensure an even test distribution. After finishing searching for the eight blocks, users were asked to answer a small questionnaire about their experience regarding the used systems.

Results
We conducted three different perspectives to analyse the results from our user evaluation. In the first place, we present a quantitative analysis regarding objective measures, namely time and errors. Then, we also present qualitative analysis based on questionnaire answers. Finally, we discuss several observations captured over the test sessions.

Quantitative Analysis
During our user evaluation, we registered the time taken for users to search for the eight blocks, as well as the number of wrong blocks selected. These values are depicted in the graph of Figure 13. To analyze the results regarding time spent searching for the eight blocks, we used the Wilcoxon signed-ranks test, with which we concluded that statistically significant differences exist: users did faster searches with LSpeakIt than with LDD (Z=-3.509, p=0.000). Regarding the number
of wrong blocks selected, we also used the Wilcoxon signed-ranks test. We concluded that statistically significant differences exist. The LSpeakIt solution performed better than LDD (Z=-6.697, p=0.000), preventing more errors.

**Qualitative Analysis**

After completing the task in each system, users were asked to classify the search interface for both systems, using a 4 values Likert scale, concerning: how fun it was to use; how easy it was to view the blocks; and how easy it was to use. The users’ ratings are presented in Table 2. The Wilcoxon signed-ranks test was again used to find statistically significant differences. Users strongly agreed that LSpeakIt was the easiest to use (Z=-2.441,p=0.015). In our system, even participants that were not used to spoken queries managed to use it successfully. Concerning objects’ visualization, users strongly agreed that LSpeakIt was the easiest (Z=-2.780, p=0.005). Finally, they also agreed that LDD was less fun than LSpeakIt (Z=-3.626, p=0.000).

**Observations**

During the evaluations we noticed some relevant aspects from users behavior as well as we registered some comments. Throughout the evaluation users mentioned that the division by categories of LDD was helpful. However, this division did not help in two of the selected blocks, where blocks that represented the categories weren’t similar to the desired object. These caused users to open all categories and browsing through them, which eased searching for the next blocks. Users that didn’t start with one of these two blocks made comments reflecting some frustration while searching in the LDD, something that did not occurred in LSpeakIt, even on longer searches.

In LDD, users shown to have problems understanding the blocks depicted, which did not happen in LSpeakIt. Comments from users suggested that our prototype gave them a better perception of objects’ size and details, mainly due to the blocks’ rotation when listed. This was particularly noticeable in blocks without pins, where users often selected a block with a different size in LDD.

Furthermore, users commented that it was easy to use LSpeakIt to search for objects by their dimension, despite some users mentioned their dimensions wrong during the search. These mistakes where all noticed and corrected when the search results were shown and the system told what it had understood. We also noticed that users resorted more to metaphors and adjectives at their latter searches. When asked why this happened, they said that they didn’t know what words could be used at the beginning and at the end they already known the potential of our prototype.

Users noted that the refinements available in LSpeakIt helped them to find the objects. Moreover, users also mentioned that the exclusion refinement was the less needed, because the other two possibilities for refinement, example and filter, were enough to fulfill the tasks. This was also noticed by the distribution of the refinements, where 55% of the these were to filter, 36% examples and only 9% exclusions.

**CONCLUSION AND FUTURE WORK**

With immersive virtual environments becoming more common, allowing users to be placed in a more credible way inside a virtual world, new challenges arise. One example of this is the selection of objects from a collection within an immersive environment, when considering applications for the assembly of virtual models, such as LEGO models. With this motivation, we developed a prototype that allows the user to build virtual LEGO models, exploring natural ways for retrieving objects in such environments.

In order to understand the most natural and simple methods to describe three-dimensional objects, we conducted a preliminary study, where we verified that participants preferred to use exclusively verbal descriptions. Based on the knowledge attained, we developed our LSpeakIt prototype using spoken queries for retrieving the virtual LEGO blocks. To browse the blocks collection, we devised a new barrel like visualization mode, in which the user is surrounded by virtual objects.

We conducted an evaluation with 20 users, comparing our prototype with a commercial application of the LEGO company. Results suggest that our prototype is able to overcome a traditional application that uses a 2D grid with thumbnails, which is a more familiar interface to users. Through the results, we could conclude that our solution gave a better perception of the objects, being less prone to errors than the commercial application. Moreover, through users’ comments and questionnaire analysis, we can conclude that most participants found it easy to use our search solution, even without extensive training.

As future work, we consider that would be interesting to invest in more sophisticated language models. As for showing queries’ results, improvements could be done using object clustering, as available in LDD, but with more than one object illustrating each cluster. While our prototypes gives an experience where the user is always in the same position, the new version of Oculus Rift, coming up with head tracking, may offer an increased user immersion without restricting user’s position.

Although we explored LEGO blocks in our prototype, we believe that our solution can be applied to other scenarios, namely for virtual models assembly in automotive and construction industries.

**REFERENCES**


