TÉCNICO LISBOA

LSpeak It
Interface Imersiva para Pesquisa de Objectos

Diogo Gil Vieira Henriques

Dissertação para a obtenção de Grau de Mestre em

Engenharia Informática e de Computadores

Orientadores: Prof. Alfredo Manuel dos Santos Ferreira Júnior
Prof. Isabel Maria Martins Trancoso

Júri

Presidente: Prof. Nuno João Neves Mamede
Orientador: Prof. Alfredo Manuel dos Santos Ferreira Júnior
Vogal: Prof. Manuel João Caneira Monteiro da Fonseca

Junho 2014

Para os meus pais e namorada

Agradecimentos

Ao longo desta dissertação, muitos foram os apoios e incentivos a que não posso deixar de manifestar o meu sincero agradecimento. Antes de mais, ao Instituto Superior Técnico que proporcionou inúmeras aprendizagens, não apenas educacionais. Ao professor Doutor Alfredo Ferreira e também à professora Doutora Isabel Trancoso, que orientaram esta dissertação e demonstraram empenho e dedicação ao longo de todo o processo. Quero também agradecer à Fundação para a ciência e Tecnologia (FCT) que através dos projectos CEDAR (PTDC/EIA-EIA/116070/2009), TECTON-3D (PTDC/EEI-SII/3154/2012) e Pest-OE/EEI/LA0021/2013, cedeu suporte financeiro para deslocação e participação nas conferências.

Um agradecimento em particular ao Daniel Mendes, pela disponibilidade, paciência e ajuda nos momentos fulcrais desta dissertação. Ao Carlos Mendes, José David e Hugo Meinedo pela paciência e aprendizagens transmitidas.

Ao Auguste Cunha e Daniela Borges, pelas suas partilhas, discussões e motivação ao longo de todo o percurso desta dissertação. Também ao João Araújo, pelo convívio, ideias e momentos de descontracção, em fases cruciais e de stresse durante o trabalho desenvolvido. A todos aqueles que se disponibilizaram para participar nas diversas sessões de avaliação devo o meu apreço, bem como todos os que colaboraram indirectamente através do interesse demonstrado neste trabalho.

Um agradecimento especial à Mª Helena Domingues, por toda a sua dedicação, motivação e apoio incansável tanto em situações simples, como complexas. Reconheço-lhe também a sua compreensão e interesse que revelou ao longo de toda esta etapa. Ao meu primo Ricardo Morgado, pelo seu apoio, ensinamentos e boleias. A todos os meus amigos e familiares que se esforçaram para a que também houvessem momentos de descontracção, mesmo sendo estes à distância.

Por fim, mas não menos importante, aos meus pais Fernando Henriques e Mª Zélia Henriques sem os quais a minha entrada e frequência no Instituto Superior Técnico não seria possível. Devo-lhes também o apoio incondicional, desde sempre oferecido, sem o qual esta meta não teria sido possível alcançar.

Resumo

O número de objectos tridimensionais em formato digital disponíveis tem aumentado consideravelmente. Com este aumento, a área de recuperação tem sido cada vez mais explorada. No entanto, soluções existentes não exploram interacções naturais. Nesta dissertação, propomos uma interface verbal para recuperação de objectos 3D em ambientes virtuais imersivos. Para o nosso protótipo, LSpeakIt, utilizámos o contexto de blocos de LEGO como *toy-problem*. Para compreender como as pessoas descrevem os objectos físicos, realizámos um estudo preliminar. Verificámos que os utilizadores recorriam sobretudo a descrições verbais. Com base nestes resultados integrámos um sistema de pesquisa por fala, que considerou as descrições utilizadas no estudo realizado.

Tirando partido de um dispositivo de visualização de baixo custo, os Oculus Rift, implementámos quatro modos para visualização de resultados num ambiente imersivo. Estes modos foram avaliados por utilizadores, tendo sido comparado entre eles e com um modo de visualização tradicional. Identificámos que os utilizadores preferiram modos de visualização imersivos à visualização tradicional, a qual os utilizadores estavam mais familiarizados. Para a nossa versão final do protótipo, implementámos um quinto modo de visualização de resultados, que resulta de uma adaptação dos modos preferidos pelos utilizadores.

Comparou-se a solução proposta com uma aplicação comercial, Lego Digital Designer. Os resultados sugerem que o LSpeakIt consegue superar o seu concorrente, oferecendo aos utilizadores uma forma simples e natural de pesquisar objectos virtuais, garantindo uma melhor performance e percepção dos resultados que as abordagens tradicional de recuperação 3D.

Palavras-chave: Interface imersiva, Procura por fala, Objectos 3D, Recuperação

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.

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

Conteúdo

Agradecimentos v
Resumo vii
Abstract ix
Lista de Tabelas xiii
Lista de Figuras xvi
Lista de Abreviaturas 1

1 Introdução **1**
1.1 Motivação 2
1.2 Problema 2
1.3 Contribuições 3
1.4 Publicações 4
1.5 Estrutura do Documento 4

2 Trabalho Relacionado **5**
2.1 Recuperação de informação através de texto 5
2.2 Recuperação de informação através de fala 7
2.3 Recuperação de objectos 3D por exemplo 9
2.4 Recuperação de objectos 3D por esboços ou gestos 11
2.5 Navegação e Visualização 14
2.6 Discussão 17

3 Estudo Preliminar **19**
3.1 Descrição 19
3.2 Resultados 22
3.3 Conclusões 26

4 LSpeakIt **29**
4.1 Arquitectura 30
4.2 Indexação 32
4.3 Especificação de interrogações 35
4.4 Exploração dos resultados 37

4.5 Construção do modelo 43
4.6 Sumário 44

5 Avaliação **45**
5.1 Avaliação de modos para a visualização dos resultados 45
5.2 Validação da solução adoptada 50
5.3 Sumário 57

6 Conclusões e Trabalho Futuro **59**

Bibliografia **66**

A Hardware **67**
A.1 Visualização 67
A.2 Interacção 68
A.3 Discussão 69

B Análise de dispositivos SpacePoint Fusion e Leap Motion **71**
B.1 Metodologia e Participantes 71
B.2 Resultados 72

C Testes sobre os modos de visualização **73**
C.1 Questionário 74
C.2 Resultados 75

D Testes de validação de solução **77**
D.1 Questionário Inicial 77
D.2 Questionário LDD e LSpeakIt 78
D.3 Resultados 78

Lista de Tabelas

3.1 Duração das diferentes etapas de cada sessão da experiência. 20
3.2 Classificação da descrição das peças de forma verbal, gestual e combinada (Mediana, Amplitude interquartil). 24
3.3 Estatística dos subconjuntos de treino e teste. 26

4.1 Os cinco melhores resultados para a interrogação *Inclinada 1 x 2*. 34

5.1 Etapas de avaliação dos modos de visualização. 46
5.2 Classificações para cada modo de visualização (Mediana, Amplitude interquartil). 49
5.3 Etapas das sessões de validação da solução proposta. 51
5.4 Classificação para cada aplicação (Mediana, Amplitude interquartil). 56

B.1 Tempos de conclusão da tarefa em segundos utilizando Leap Motion e SpacePoint Fusion e dispostivo preferido. 72
B.2 Classificação para cada modo de interacção (Mediana, Amplitude interquartil). 72

Lista de Figuras

2.1 Princeton 3D Model Search Engine(Funkhouser et al., 2003; Min et al., 2003). 6
2.2 Sistema de reconhecimento de fala AUDIMUS, usado em (Meinedo et al., 2010). 8
2.3 Sistema de fala comum segundo Moreno et al. (Moreno-Daniel et al., 2007), usando apenas um reconhecedor de fala. 9
2.4 Processo de Indexação e Recuperação dos objectos 3D. 10
2.5 Procura por esboço do projecto LSketchIt (Santos et al., 2008). 12
2.6 Esboços sobre um modelo 3D do trabalho de Xie et al. (Xie et al., 2013). 12
2.7 Modelo com todos os voxels activos. 13
2.8 Modelo com os voxels de peso superior a n. 13
2.9 Ilustração do ambiente do projecto 3D Mars (Nakazato and Huang, 2001). 14
2.10 Ilustração do ambiente de operação espacial g-speak. 15
2.11 Visualização de um objecto 3D, no sistema Navidget (Hachet et al., 2009). 16

3.1 Exemplo de instruções fornecidas ao construtor, semelhante às que acompanham os modelos LEGO originais. 20
3.2 Participante no papel de construtor recebe uma peça do fornecedor. 21
3.3 Percentagem de passos em que os participantes descreveram a peça utilizando gestos. . 22
3.4 Participante solicita uma peça de canto, utilizando uma referência à letra 'L'. 23
3.5 Tempos médios (mm:ss) por sessão para obter as peças (Total), para as descrever (Descrição) e para as procurar (Procura). 24
3.6 Três metáforas utilizadas pelos participantes para descrever a peça representada no canto superior esquerdo. 25
3.7 Conjunto de peças que levaram a descrições peculiares. 26

4.1 Aparato utilizado no protótipo LSpeakIt. 30
4.2 LSpeakIt - Vista das três camadas da arquitectura utilizada. 30
4.3 LSpeakIt - Arquitectura utilizada. 31
4.4 Sistema de indexação do LSpeakIt. 33
4.5 Ligação da biblioteca SAPI com o reconhecedor e sintetizador. 35
4.6 Peça com descrição "redonda 2 x 2". 36
4.7 Peça com descrição "fina 2 x 2". 36

4.8 Modos de apresentar os resultados de interrogações num ambiente imersivo: (a) Rectangular, (b) Quadrático, (c) Cilíndrico, (d) Esférico. 38
4.9 Disposição dos objectos em volta do utilizador. 39
4.10 Distribuição dos objectos de acordo com a sua relevância. 40
4.11 Visualização de um objecto mais próximo utilizando o gesto de apontar com o Leap Motion. 41
4.12 Controlo da mira que permite seleccionar objectos ao longo dos resultados. 42
4.13 Rotação de objecto de acordo com a rotação realizada pelo SpacePoint Fusion. 42
4.14 Vista de modelação do protótipo LSpeakIt onde o utilizador encontra-se mover uma peça. 43

5.1 Peças de lego utilizadas para os utilizadores procurarem nos modos de visualização. . . 47
5.2 Participante durante a sessão, a executar o gesto de apontar. 47
5.3 Tempo despendido nas procuras de cada modo de visualização. 49
5.4 Aplicação comercial LDD à esquerda e LSpeakIt à direita. 51
5.5 Peça utilizada como exemplo. 52
5.6 Peça utilizado como treino. 52
5.7 Conjunto de peças utilizadas para a tarefa. 53
5.8 Participantes na sessão de validação, a visualizar listas de resultados no LDD e no LSpeakIt. 53
5.9 Soma dos tempos em segundos das procuras nas duas aplicações, LDD e LSpeakIt. . . 55
5.10 Soma dos totais de peças erradas das procuras nas duas aplicações. 55

A.1 Z800 3D Vision. 67
A.2 Oculos Rift. 67
A.3 3DVision Wired Glasses. 67
A.4 5DT Data Gloves. 69
A.5 Microsoft Kinect. 69
A.6 Leap Motion. 69
A.7 PS move. 69
A.8 Wii Remote. 69
A.9 SpacePoint Fusion. 69

Lista de Abreviaturas

2D - Two-dimensional

3D - Three-dimensional

CAVE - Computer Assisted Virtual Environment

D2 - D2 Shape Distributions

FOX - Focus sliding surface

GRXML - Speech Recognition Grammar Specification

HMD - Head Mounted Display

HMM - Hidden Markov Models

L2F - Spoken Language Systems Laboratory

LEX - Lexical Model

LDD - Lego Digital Designer

TF-IDF - Term Frequency - Inverse Document Frequency

MLP - Multi-Layer Perceptrons

RO - Reference Object

SAPI - Speech Application Programming Interface

WFST - Weighted Finite-State Transducer

WIMP - Windows Icons Menus and Pointing

Capítulo 1

Introdução

O volume de conteúdo multimédia, nomeadamente imagens, áudios e vídeos, em bases de dados multimédia na Internet ou em colecções *offline* tem vindo a aumentar. Os objectos 3D são um exemplo deste fenómeno. Este aumento deve-se, em parte, ao aparecimento e desenvolvimento de novas aplicações de design, as quais facilitam o processo de criação de modelos 3D. Em simultâneo têm surgido variadas tecnologias de baixo custo, como por exemplo o Microsoft Kinect[1], ASUS Xtion[2] ou PrimeSense Sensor[3] que permitem a captura de objectos reais, transformando-os em objectos 3D digitalizados.

Com o crescimento das bases de dados de objectos 3D, as tarefas de procura, navegação e visualização de resultados tornam-se cada vez mais complexas, devido ao elevado número de elementos que se podem encontrar. Consequentemente, a informação intrínseca nos objectos, como por exemplo o nome do ficheiro, revela-se insuficiente para uma pesquisa simples através de texto, sendo muitas vezes necessário associar nova meta-informação (Smith and Chang, 1997). Com o intuito de solucionar esta questão, têm surgido melhoramentos na pesquisa por texto (Smith and Chang, 1997; Funkhouser et al., 2003), assim como novas abordagens como pesquisa por esboços (Min et al., 2004) ou ainda usando um outro objecto 3D como exemplo de procura. Outra questão pertinente é a visualização dos resultados, uma vez que muitas soluções os apresentam sob a forma de miniaturas em 2D (*thumbnails*) (Funkhouser et al., 2003), podendo conduzir à oclusão de características dos objectos 3D.

Nesta dissertação pretendemos abordar os problemas referidos, fornecendo uma solução que permita uma interface de pesquisa de objectos 3D que seja natural para o utilizador. Como prova de conceito e para validar a abordagem, são utilizadas peças LEGO. Contudo, este trabalho procura realizar uma prova de conceito que se estenda a outros domínios como, por exemplo, a área da indústria automóvel ou construção.

[1] Microsoft Kinect ®, `http://www.xbox.com/en-US/kinect`
[2] ASUS Xtion ®, `http://www.asus.com/Multimedia/Xtion_PRO/`
[3] PrimeSenses Senor ®, `http://www.primesense.com/solutions/sensor/`

1.1 Motivação

A grande maioria dos sistemas de pesquisa actuais permitem apenas pesquisas recorrendo a interrogações por texto. No entanto, com a disseminação do reconhecimento de fala em dispositivos móveis, podemos pensar na hipótese que a fala pode tornar algumas tarefas mais naturais (Kamvar and Beeferman, 2010). Até agora, o reconhecimento automático da fala tem permitido grandes avanços em sistemas de transcrição (Meinedo et al., 2010) ou sistemas de comando e controlo (Moustakas et al., 2006), podendo destacar-se a pesquisa por voz da Google[4], S Voice[5] e Siri[6]. Contudo, no ramo da recuperação de objectos 3D, o uso de fala não tem sido explorado.

A pesquisa de uma forma natural de objectos 3D, a visualização de resultados destes itens tem sido pouco explorada. A grande maioria dos motores de pesquisa de objectos 3D apresenta os resultados numa grelha de *thumbnails*, perdendo assim o factor 3D. Actualmente têm surgido novos dispositivos de baixo custo, tanto de interacção como de visualização que permitem ambientes virtuais sejam cada vez mais imersivos (Basu et al., 2013). Acreditamos que o uso da fala como ferramenta de pesquisa de objectos 3D, num ambiente de realidade virtual imersivo, será uma forma simples e rápida, para o utilizador realizar interrogações. Neste sentido, permite a recuperação dos objectos e simultaneamente a análise dos mesmos, com mais detalhe.

1.2 Problema

Tal como foi referido anteriormente, a maioria dos sistemas existentes para a pesquisa de modelos 3D baseiam-se na procura através de texto, usando esboços ou dando um modelo 3D como exemplo. Todavia, estes sistemas não consideram uma abordagem natural para os utilizadores, como acontece quando a descrição dos objectos é realizada verbalmente.

É importante para os utilizadores não só conseguirem pesquisar os objectos de uma forma simples, mas também conseguirem ter uma boa percepção dos objectos retornados. Isto deve-se ao facto de existirem poucas soluções que exploram uma melhor forma de visualizar os objectos retornados, pois a grande maioria adopta o uso de imagens, deixando parte do objecto oculta. Este facto torna assim mais difícil a distinção de objectos semelhantes (Pascoal et al., 2012).

Posto isto, surge o desafio de criar uma interface para pesquisar objectos 3D, onde se pretende proporcionar aos utilizadores um modo mais natural de pesquisa, bem como, identificar a melhor abordagem para apresentar os resultados.

[4]Google ®, http://www.google.com/mobile/voice-search/
[5]Samsung ®, http://www.samsung.com/global/galaxys3/svoice.html
[6]Apple ®,http://www.apple.com/ios/siri/

1.3 Contribuições

O principal objectivo da solução proposta neste trabalho visa identificar a abordagem mais simples e natural para um utilizador pesquisar objectos 3D num sistema de realidade virtual imersivo. Foi realizado um estudo com utilizadores para compreender as melhores descrições para os objectos. Também foram estudadas as melhores abordagens de apresentar os resultados das pesquisas, sendo estas sujeitas a avaliação com utilizadores. Para que este estudo fosse possível, foi desenvolvida uma interface de pesquisa num ambiente de realidade virtual imersivo, usando os Oculus Rift[7] para a visualização, e o Leap Motion[8] para a captura de gestos, tendo a interacção com o Leap Motion sido substituída por uma interacção utilizando o SpacePoint Fusion[9]. Neste sentido, as principais contribuições deste trabalho são:

- **Estudo para captura das descrições de peças LEGO**

  Como ponto de partida deste trabalho, foram identificadas as descrições utilizadas para as peças de LEGO. Foram estudadas as preferências dos participantes ao descrever as peças de LEGO, através de fala, gestos ou até mesmo a combinação verbal e gestual. Este estudo foi realizado com base numa avaliação feita por utilizadores, assim como um questionário preenchido pelos mesmos. O estudo pode-se mostrar relevante para quem pretenda desenvolver um sistema de pesquisa para objectos no âmbito do LEGO, ou um sistema de pesquisa para objectos 3D de uma forma geral.

- **Definição de abordagens para apresentação de resultados em realidade virtual**

  As soluções existentes para a apresentação dos resultados de uma pesquisa de objectos 3D baseiam-se muito no uso de grelhas *thumbnails*, perdendo assim o 3D dos objectos. Neste trabalho são propostas diversas abordagens para visualizar os resultados num ambiente de realidade virtual imersivo, tanto a nível de disposição dos objectos como para navegar nos resultados através de gestos realizados no ar. Estas abordagens foram testadas com utilizadores, comparando-as com a abordagem tradicional, de forma a percebermos a sua aceitação.

- **Integração de um sistema de pesquisa através da fala**

  Os modos de pesquisa para objectos 3D existentes não consideram a fala como forma de pesquisa, existindo outras soluções. A utilização das descrições verbais recolhidas no estudo inicial do nosso trabalho permitiu-nos usar o reconhecer de fala AUDIMUS (Meinedo et al., 2010), onde foi criada uma nova gramática para este contexto. De forma a tratar e classificar as pesquisas realizadas pelos utilizadores, foi integrado um classificador, para a procura e indexação das pesquisas por fala. Para complementar a pesquisa por fala foi implementado um método que permitisse a

[7]Oculus Rift®, `http://www.oculusvr.com`
[8]Leap Motion ®, `https://www.leapmotion.com/`
[9]PNI Sensor Corporation ®, SpacePoint Fusion,`http://www.pnicorp.com/products/spacepoint-gaming`

procura de objectos escolhendo um destes como exemplo.

1.4 Publicações

No âmbito desta dissertação foram submetidos e aceites dois artigos. O primeiro foi publicado numa conferência nacional (**Interação 2013**), e o segundo numa conferência internacional, "IEEE 9th Symposium on 3D User Interfaces 2014"(**3DUI**). As referências para as publicações são as seguintes:

1. **Diogo Henriques**, Daniel Mendes, Pedro Pascoal, Isabel Trancoso e Alfredo Ferreira, *Evaluation of Immersive Visualization Techniques for 3D Object Retrieval*, 3DUI 2014, IEEE 9th Symposium on 3D User Interfaces 2014 (3DUI 2014), Março 2014

2. Daniel Mendes, **Diogo Henriques**, Isabel Trancoso e Alfredo Ferreira, *Qual Peça? Identificação de Métodos Naturais para a Descrição de Objectos 3D*, 5ª Conferência Nacional em Interacção Humano-Computador (Interação 2013), Novembro 2013

1.5 Estrutura do Documento

Neste documento são apresentadas várias abordagens do estado de arte, para resolver os problemas de recuperação de informação e interfaces em ambientes imersivos. Assim sendo, no Capítulo 2 são apresentados trabalhos de recuperação de informação e de conteúdo multimédia. Ainda neste capítulo aborda-se a interacção em ambientes imersivos e desenvolve-se uma discussão das várias soluções existentes. No capítulo seguinte referimos um estudo preliminar realizado nesta dissertação, de modo a compreender se os participantes recorriam mais a gestos ou à fala na descrição de objectos físicos, e ainda quais as descrições que verbais e gestuais usadas nessa descrição. No Capítulo 4 apresentamos uma solução para o problema, tendo em consideração as abordagens mencionadas no trabalho relacionado e o estudo preliminar. No Capítulo 5 referimos a metodologia de avaliação utilizada e discutimos os resultados obtidos durante as várias iterações. Para finalizar, no Capítulo 6 apresentamos a conclusão e referimos algum trabalho futuro.

Capítulo 2

Trabalho Relacionado

A área de recuperação já é explorada à bastante tempo, no entanto, a maioria da recuperação é relativa a informação textual (Brin and Page, 1998). Actualmente, a quantidade de conteúdo multimédia tem aumentado, tornando cada vez mais necessária a existência de formas de recuperar esse conteúdo. Várias soluções têm surgido para a pesquisa de conteúdo multimédia, onde se incluem os objectos 3D, sendo que as mais recentes têm incidido em esboços e/ou utilizam um objecto 3D como exemplo de pesquisa.

2.1 Recuperação de informação através de texto

Quando se trata de pesquisas por texto, o motor de busca Google, inicialmente criado por Brin e Page (Brin and Page, 1998), é um exemplo, o qual retorna informação de páginas *web*. Quando se realiza uma pesquisa no motor de busca Google, é realizada concretamente uma pesquisa a um conjunto de páginas que foram indexadas graças a um classificador. Este último realiza a procura num grande número páginas através de *hyperlinks* contidos nas páginas que encontra, permitindo aumentar o seu volume de páginas e, concomitantemente, classificar as páginas que encontra. Após a classificação, as páginas são indexadas de acordo com a sua importância, a qual é determinada tendo em conta as citações realizadas a essa página.

No entanto, quando se trata de pesquisa de conteúdo multimédia através de texto, as pesquisas passam a ser menos triviais. No geral, estes motores de pesquisa usam textos associados aos objectos multimédia, como por exemplo legendas, referências aos objectos ou até por vezes os *hyperlinks* dos objectos ou nomes dos ficheiros. Smith e Chang (Smith and Chang, 1997) abordam uma forma de pesquisar imagens e vídeos, recolhendo mais de 500.000 imagens e vídeos de várias páginas web. As páginas web foram verificadas e analisavam o respectivo código fonte para descobrir se estas continham imagens ou vídeos, guardando os respectivos *hyperlinks*. Estes eram processados para obter os

termos necessários para a descrição do respectivo conteúdo multimédia, criando classes e subclasses tendo em conta o número de ocorrência dos termos. A pesquisa realizada nessas classes, fazendo uma intersecção ou união do conteúdo multimédia das classes encontradas.

Com o objectivo de investigar a eficiência da categorização de imagens com o texto associado, Sable e Hatzivassiloglou (Sable and Hatzivassiloglou, 1999) utilizaram uma colecção de mais de 21 000 notícias online, nas quais mais de 1490 continham imagens embebidas com texto associado. Os autores pretendiam classificar as imagens como sendo *indoor* ou *outdoor* e verificaram que produzia melhores resultados ao limitar o texto associado apenas à primeira frase da legenda da imagem. Noutro trabalho, Sable et al. (Sable et al., 2002) usam um processador de língua natural para identificar verbos e os sujeitos na frase da legenda das imagens, classificando as imagens em quatro categorias.

Por outro lado, quando se trata de recuperação de objectos 3D, passamos por problemas similares aos das imagens. Nos trabalhos de Funkhouser et al. (Funkhouser et al., 2003) e Min et al. (Min et al., 2003) os autores implementaram um sistema híbrido, ilustrado na Figura 2.1, que permite fazer pesquisas de objectos 3D por texto, esboços e a união das duas. Relativamente à pesquisa por texto, a informação que é usada baseia-se na informação existente na página de origem do modelo. Este contém o nome do ficheiro, a sua extensão e algumas palavras que possam existir dentro do ficheiro. São estas palavras que auxiliam na elaboração de um documento que descreve o modelo. Os documentos com as descrições dos modelos são processados, removendo-se as palavras comuns que não servem para a descrição e sendo adicionados sinónimos dos nomes dos ficheiros. Na realização da correspondência da informação dos documentos com as palavras chave, especificadas pelos utilizadores, é usado *Term Frequency - Inverse Document Frequency* (TF-IDF). Este é utilizado para classificar e atribuir o peso

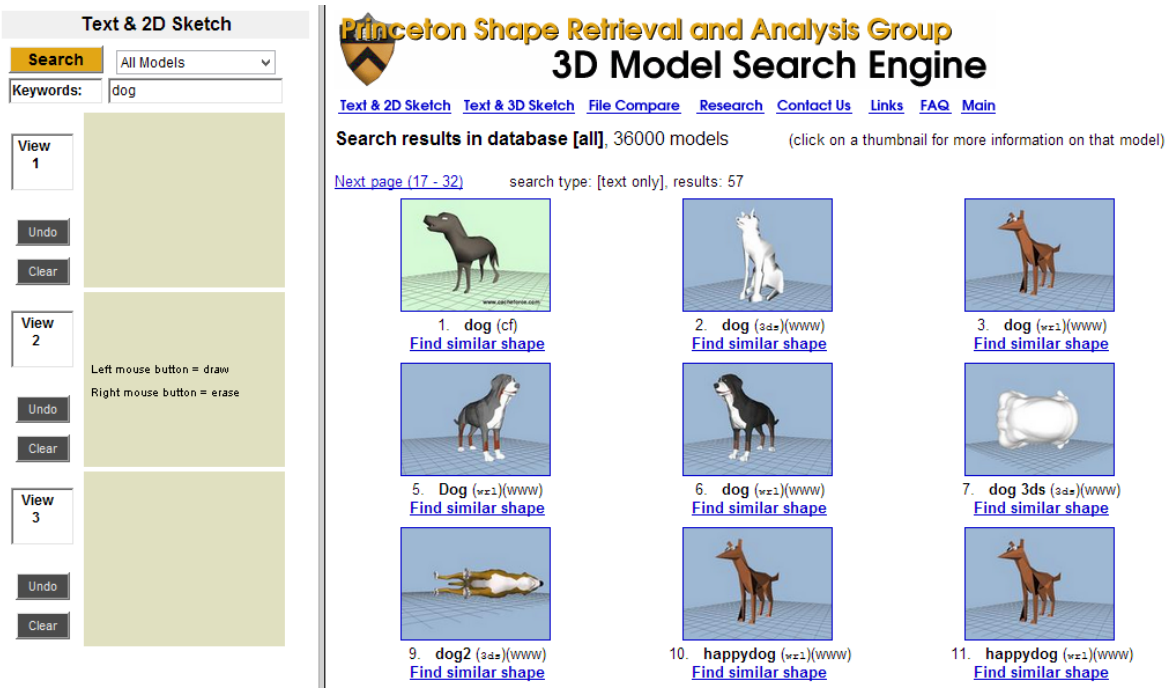

Figura 2.1: Princeton 3D Model Search Engine(Funkhouser et al., 2003; Min et al., 2003).

para documentos de texto, baseando-se em frequências no documento e na frequência inversa em todos os documentos.

À posteriori Min et al. (Min et al., 2004) realizaram um motor de pesquisa de modelos 3D com a capacidade de realizar pesquisas por texto e pesquisas por esboços. A informação utilizada para a descrição dos modelos 3D é semelhante ao trabalho realizado por (Funkhouser et al., 2003) que utiliza informação proveniente do ficheiro, como o nome do ficheiro, nome do ficheiro sem dígitos, conteúdo no ficheiro, bem como um conjunto de sinónimos do nome do ficheiro e a categorização do nome do ficheiro. Para além da informação anterior, utiliza também informação originária da página de onde é retirado o modelo 3D, desde o texto que serve para descrever o link do modelo, o próprio *link*, texto perto do link e ainda o título da página *web*. Toda esta informação é guardada sem as palavras comuns, ficando com as palavras normalizadas. Exemplificando com a palavra "*wheels*", esta é guardada na sua forma normal "*wheel*". Neste sentido, tendo um documento com texto que represente o modelo 3D, os autores usaram uma biblioteca (McCallum, 1996) que disponibilizava vários métodos para efectuar uma análise estatísticas da correspondência das palavras. Realizaram ainda uma comparação de alguns métodos disponíveis pela ferramenta, verificando que o que obtinha melhores resultados era o "TF-IDF log occur". Este é uma variante do método TF-IDF, onde a frequência do documento é o número de documentos em que o termo ocorre pelo menos uma vez.

Como se pode ver no trabalho anterior, guardar informação de objectos 3D em texto, em termos de informação semântica do objecto, como por exemplo sinónimos, dimensões, entre outras, é algo que gera sempre uma grande quantidade de informação, existindo objectos que partilham a mesma informação. Por esta razão começa a ser frequente recorrer a ontologias, por forma a organizar a informação existente dos objectos 3D. De acordo com Guarino et al. (Guarino and Poli, 1995), uma ontologia é uma especificação formal explícita de uma conceitualização compartilhada de um domínio de conhecimento, e expressa a estruturação e modelagem desse domínio particular. Baseando-se nesta definição de ontologia, Attene et al. (Attene et al., 2009) usaram um sistema de anotação e ontologias para conseguir representar o conhecimento de objectos 3D, ou seja, relações topológicas e aspectos geométricos dos objectos. Este sistema permite aos utilizadores criarem anotações em partes especificas do objecto, melhorando a informação conhecida sobre o objecto 3D.

2.2 Recuperação de informação através de fala

Não faz sentido resumir o estado da arte em reconhecimento de fala neste documento. Todos os sistemas de reconhecimento de vocabulário extenso, no entanto, seguem uma abordagem estatística que tem como principais módulos a extracção de características acústicas, o descodificador e as três principais fontes de conhecimento - o modelo acústico, o modelo lexical (LEX) e o modelo de língua (LM - Language Model). No caso do AUDIMUS Meinedo et al. (2010), o sistema de reconhecimento adoptado neste trabalho e representado na Figura 2.2, o modelo acústico combina as capacidades de

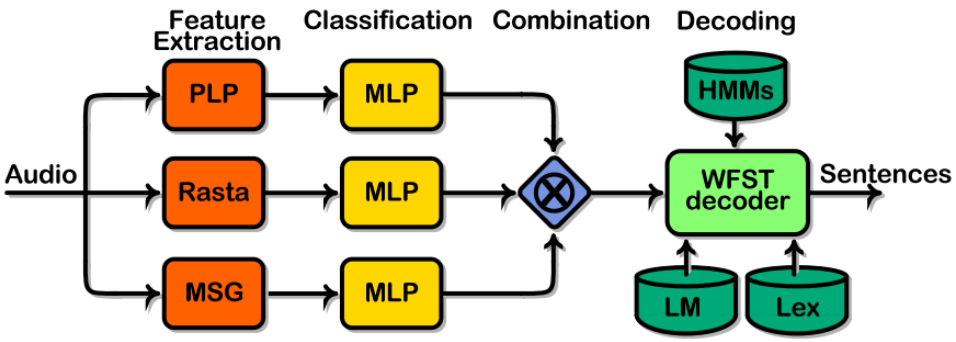

Figura 2.2: Sistema de reconhecimento de fala AUDIMUS, usado em (Meinedo et al., 2010).

modelagem temporais dos modelos ocultos de Markov (HMM)[1] com as capacidades de reconhecimento de padrões dos Perceptrões Multi Camada (MLP)[2] . A extracção de características é realizada por três processos diferentes de classificação fonética, que contribuem para um melhor modelamento da diversidade acústica. Cada classificador MLP contém um contexto local acústico abrangendo várias tramas do sinal.

O processo de descodificação é baseado na abordagem Transdutores Ponderados de Estado Finito (WFST)[3], onde o espaço de procura é um grande WFST, que resulta da composição dos transdutores que representam o modelo acústico, o modelo lexical e o modelo de língua. O sistema de reconhecimento associa a cada palavra uma medida de confiança que combina várias características através de um classificador de máxima entropia, cuja saída representa a probabilidade de cada palavra estar correcta. A fim de gerar uma medida de confiança de palavras, estas características são combinadas através de um classificador de máxima entropia, cuja saída representa a probabilidade de cada palavra estar correcta. O reconhecedor AUDIMUS tem sido sobretudo utilizado na transcrição automática de notícias televisivas em várias línguas entre as quais as variantes Europeia, Brasileira e Africana do Português.

A pesquisa de informação através de fala é algo que tem sido realizado convertendo primeiro a fala em texto. Por sua vez, para alcançar a acepção do texto, este é convertido numa pesquisa por texto, tal como podemos ver ilustrado na Figura 2.3. Todavia, com o objectivo de contornar esta forma de pesquisar, Moreno et al. (Moreno-Daniel et al., 2007) implementaram um motor de pesquisa através de fala para pesquisar filmes, usando como informação sumários dos filmes que estavam disponíveis. Neste trabalho, é realizado um processamento inicial dos documentos existentes, criando um módulo de termos usados em todos os documentos. Quando são realizadas as interrogações usando a fala, o módulo de termos é utilizado como um dicionário de palavras que o reconhecedor de fala tem de retornar, de forma a aproximar as palavras reconhecidas da interrogação, às palavras existentes nos documentos.

[1] Hidden Markov Models
[2] Multi-Layer Perceptrons
[3] Transdutores Ponderados de Estado Finito, em inglês:Weighted Finite-State Transducer

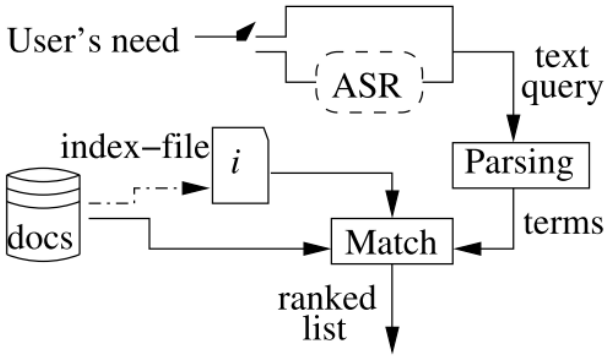

Figura 2.3: Sistema de fala comum segundo Moreno et al. (Moreno-Daniel et al., 2007), usando apenas um reconhecedor de fala.

Mais recentemente Lee et al. (Lee et al., 2010) realizaram um sistema de procura de livros através reconhecimento de fala. A informação que é usada dos livros consiste em: títulos, nomes dos autores e as categorias em que os livros se inserem. A indexação dos livros é feita com base em contagem de termos ocorridos na informação que descreve o livro. Na realização da procura são realizadas três interrogações. A primeira junta a informação dada pelo utilizador relativamente ao titulo, nome de autor e categoria. A segunda interrogação efectua as interrogações nas três categorias em separado. A última realiza uma combinação entre as duas anteriores. Os resultados das interrogações são retornadas, tendo em conta o produto interno das três interrogações.

Moustakas et al. no seu projecto Masterpiece (Moustakas et al., 2006) revelam-nos uma plataforma para interfaces multimodais, que tem como objectivo criar um ambiente virtual imersivo para o utilizador. Os autores juntaram três tipos de reconhecimento: o reconhecimento de gestos para interagir com os objectos 3D; o reconhecimento de esboços 2D e o reconhecimento de fala no uso de comandos. Na descodificação utilizam os HMM e gramática para descrever a sintaxe que o sistema reconhece. Porém, não se pode descurar do contexto, pois cada palavra é obtida pela concatenação da unidade fonética.

A Google tem realizado alguns avanços na pesquisa através de fala. Num estudo realizado por Kamvar et al. (Kamvar and Beeferman, 2010), puderam verificar que a fala é algo que tem lugar nos dispositivos móveis de hoje em dia. No entanto, os autores verificaram que os utilizadores preferem realizar as pesquisas por texto quando se tratam de pesquisas muito longas ou de pesquisas que sejam desconfortáveis de mencionar em voz alta. Acreditamos que estas restrições não acontecerão no nosso trabalho, visto os elementos a pesquisar não necessitarem de pesquisas muito extensas, nem de abordar tópicos desconfortáveis.

2.3 Recuperação de objectos 3D por exemplo

A recuperação de objectos 3D por exemplo tem sido explorada quando se pretende obter um modelo semelhante. Este tipo de abordagem, como as que já analisámos anteriormente, divide-se em duas

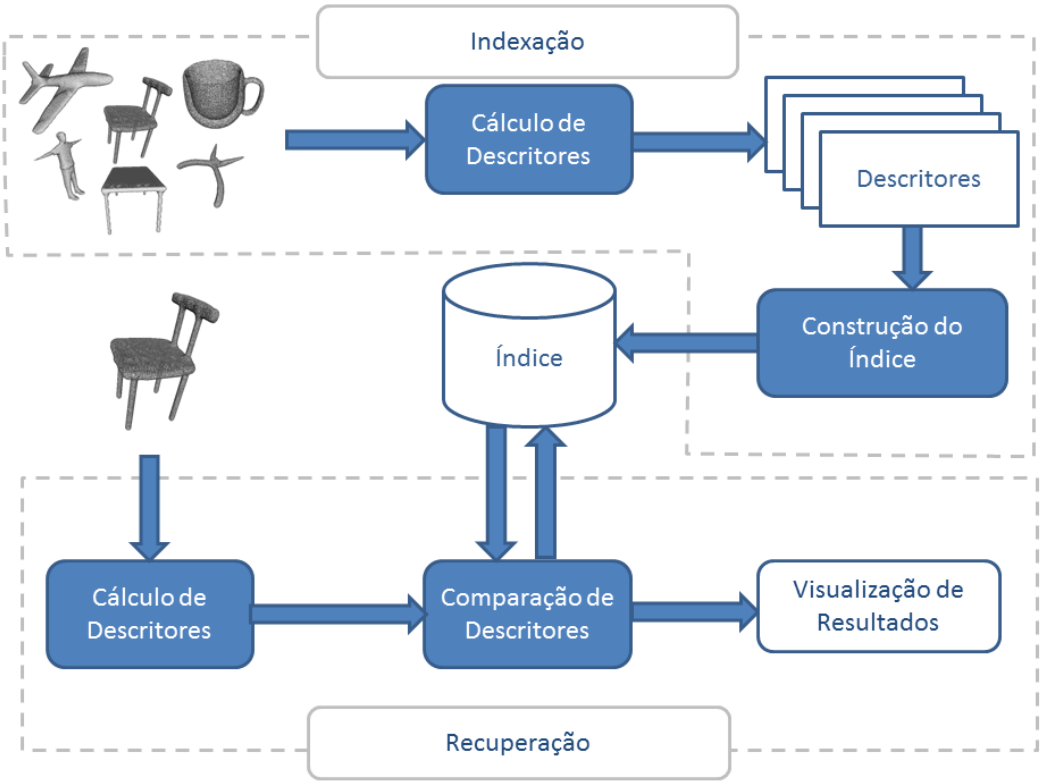

Figura 2.4: Processo de Indexação e Recuperação dos objectos 3D.

fases: a indexação e a interrogação. É possível observar o processo de indexação típico na Figura 2.4. Aqui são extraídos os descritores de cada modelo para construir um índice que seja usado no processo de recuperação. O cálculo de descritores, pode ser efectuado com base na geometria.

O trabalho de Paquet et al (Paquet and Rioux, 1999) é um exemplo disso: não só usava geometria, mas também o aspecto visual dos objectos, isto é a sua cor e textura. Porém, não conseguia encontrar os mesmos objectos com escalas ou rotações diferentes. Ao contrário do anterior, Novotni et al. (Novotni and Klein, 2003), com um descritor baseado em esferas harmónicas, tem capacidade de capturar um objecto na direcção radial, assim como numa direcção ao longo de uma esfera, ou seja os seus descritores são invariantes de rotações de objectos. No seu trabalho Funkhouser et al. (Funkhouser et al., 2003) utilizou o descritor D2 que é invariante a escala e rotação, este representa o objecto através de distâncias Euclidianas. Noutro trabalho, Lou et al.(Lou et al., 2004) construíram os seus descritores após realizado um conjunto de operações de normalização do objecto, voxelização e esqueletização, para conseguir extrair o momento invariante, os rácios da geometria, o momento e os valores próprios. Um outro exemplo mais recente é o trabalho de Assfalg et al (Assfalg et al., 2006), que usa um mapa de curvatura da superfície da forma a subdividirem a superfície, usando as regiões obtidas como histogramas e orientações ponderadas, permitindo ter melhores resultados para descrições de regiões dos objectos.

Todavia, existe um grande volume de dados que têm de ser comparados. A indexação tem como

objectivo optimizar o mecanismo de pesquisa, reduzindo o número de elementos a comparar. Para tal, são usadas estruturas de indexação, como por exemplo a R-Tree e a NB-Tree. Na R-Tree, a divisão é efectuada de acordo com a distribuição dos dados. Isto é, a estrutura de dados divide hierarquicamente o espaço, em "*bounding box*"mínimas, em que cada entrada de um nó não folha terá armazenadas dois tipos de informação, uma maneira de indenficar um nó filho e a "*bounding box*"com todas as entradas do mesmo nó filho. No caso da NB-Tree é realizado primeiro o mapeamento de pontos multidimensionais para uma linha de uma dimensão e, posteriormente, é utilizado uma B^{+}–Tree. Contudo, nem sempre temos um modelo para pesquisar por semelhantes e, muitas vezes, é difícil descrever por palavras o objecto em si.

2.4 Recuperação de objectos 3D por esboços ou gestos

Tendo em conta o âmbito do nosso trabalho, não podemos deixar de mencionar a recuperação de objectos por esboços. Esta veio tentar combater a dificuldade existente por utilizadores em descrever um dado objecto, apesar de conhecerem a sua forma, ou de não poderem usar outros modelos para usar como base para comparação.

Moustakas et al. (Moustakas et al., 2006) usa esboços para recuperar objectos 3D com formas geométricas, para depois serem usados na construção de objectos mais complexos. De forma a conseguir o reconhecimento dos esboços 2D, são processadas primeiro as trajectórias, que não são amostras uniformes devido à velocidade a que o esboço é realizado, levando ao surgimento de pontos que não são desejados. Por isso, é aplicada uma filtragem antes do reconhecimento, a qual irá criar novas amostras dos pontos de contorno uniforme, de forma a manter a distância entre dois pontos sucessivos constante. Para reconhecer os contornos dos esboços 2D após filtragem, é encaixado o esboço nas formas geométricas possíveis com as trajectórias obtidas que usam o método de Levenberg-Marquardt para minimização das linhas dos esboços (Heath, 2002).

Tendo obtido resultados algo semelhantes, realçamos o trabalho de Santos et al.(Santos et al., 2008) em que realizam uma pesquisa por esboço, tal como ilustrado na Figura 2.5. Quando os utilizadores querem procurar por uma peça, podem especificar uma ou mais dimensões enquanto modelam. De seguida, é realizada uma comparação usando os esboços e as dimensões das peças armazenadas. No caso da informação não ser associada às peças, são usadas as dimensões da *bounding box*. No entanto, os autores melhoraram este processo realizando um esboço sobre as partes já devolvidas, para as modificar. Conseguem o reconhecimento de esboços contínuos identificando os pontos que são colineares e quebrando-o em várias linhas, quando a condição muda. Os esboços incrementais são mais usados para especificar largura e profundidade, sendo que a altura é tipicamente alterada após uma peça ter sido devolvida. Neste trabalho os objectos 3D usados são geométricos, daí este método funcionar relativamente bem. Porém, quando os objectos são mais abstractos, uma abordagem deste tipo terá mais dificuldades em encontrar um objecto caso ele exista com as mesmas dimensões

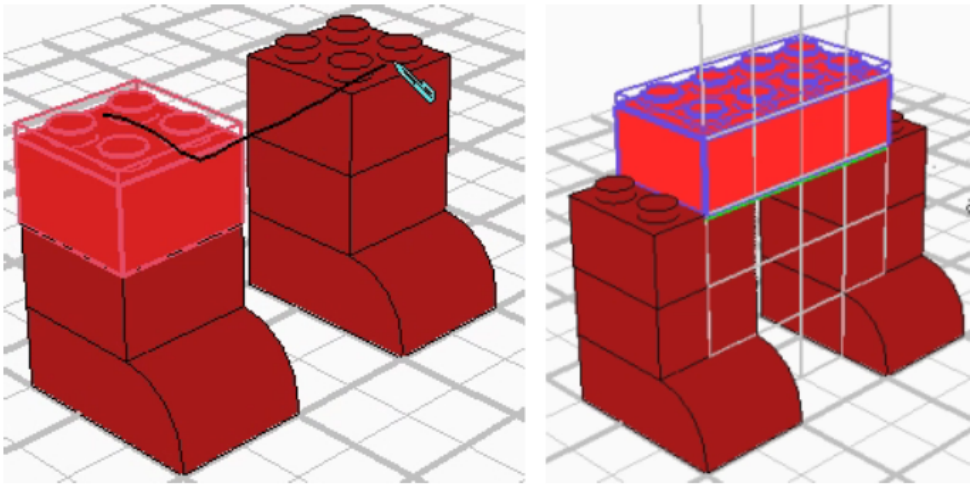

Figura 2.5: Procura por esboço do projecto LSketchIt (Santos et al., 2008).

mas com formas diferentes.

Mais recentemente têm surgido outros trabalhos que permitem a recuperação de modelos 3D através de esboços. Um dos casos é o de Xie et al. (Xie et al., 2013), ilustrado na Figura 2.6, que consiste na construção de elementos 3D usando esboços para ir alterando parte dos objectos, ou seja, o utilizador começa a partir de um modelo 3D, sendo as partes que o utilizador desenha sugeridas com base no contexto que ele inicia. Outros trabalhos como o de Li e Johan (Li and Johan, 2013) e de Eitz et al. (Eitz et al., 2012), realizam primeiro um pré-processamento dos modelos de vários pontos de vista, permitindo assim descrever melhor os objectos. Um trabalho semelhante aos anteriores é o de Saavedra et al. (Saavedra et al., 2012), mas neste caso é determinado qual a melhor projecção para cada modelo 3D, tendo em conta o esboço realizado, e de seguida quais as suas formas chave, mais relevantes, em duas dimensões.

Figura 2.6: Esboços sobre um modelo 3D do trabalho de Xie et al. (Xie et al., 2013).

Com uma abordagem diferente das anteriores, e algo semelhante ao que alguns motores de pesquisa fazem hoje em dia, Liu et al. (Liu et al., 2013), procura melhorar a pesquisa por esboços através da avaliação do perfil do utilizador, isto é, verificando o seu hábito de esboços. Neste sentido, os autores desenvolveram uma modelagem estatística tendo em conta a experiência do utilizador. Esta é usada para criar esboços dos modelos 3D na colecção, que possam ser mais similares aos que o utilizador poderá desenhar.

Holz e Wilson (Holz and Wilson, 2011) preocuparam-se em como descrever os objectos através de gestos e com essa descrição conseguir obter os objectos na base de dados. O trabalho deles, *Data*

Miming, consiste observar e mapear comportamentos dos gestos humanos, explorando a percepção espacial humana, usando a forma e movimentação das mãos enquanto descrevem objectos.

Os autores retiraram, através de estudos, os padrões usados pelas pessoas para descrever os limiares e dimensões dos objectos visualizados. Uma análise *ground truth* demonstrou que as pessoas mantinham relativamente as proporções das partes que compunham os objectos e em que zonas mais significativas dos objectos realizavam gestos mais lentos do que em zonas menos significativas. Em termos de simetria, colocavam as mãos em paralelo, de forma a descreverem os objectos simétricos. Quando descreviam as dimensões dos objectos, usavam as mãos de forma plana, movendo para a frente e para trás, colocando as mãos logo nos limites ou ainda usando os *wireframes* dos objectos. No caso de superfícies grandes, as suas mãos moviam-se como se estivessem a limpar essa superfície. Em superfícies médias, os participantes moviam as mãos tal como na anterior, mas repetiam várias vezes o gesto, quase como se tivessem a acenar. Em superfícies pequenas, usavam as suas mãos para delinear os pequenos componentes dos objectos. Quando se tratava de partes com curvatura, os participantes adaptavam a forma das suas mãos de forma a que correspondesse à curva desejada da parte do objecto.

O protótipo criado continha um *setup* usando o Microsoft Kinect, criando uma representação virtual da descrição do que o utilizador faz, num volume 3D discretizado de $l \times m \times n$ voxels. Cada voxel contém um estado e um peso. Os estados podem ser activos e inactivos, consoante algum gesto tenha passado pelo voxel. O peso é atribuído somente a voxels activos, sendo o valor atribuído consoante o tempo que a pessoa passa no voxel, permitindo assim determinar a significância dele. Nas figuras Figura 2.7 e Figura 2.8 podemos observar a comparação após a filtragem dos voxels menos significativos. Por fim é realizada uma correspondência dos modelos filtrados das descrições, com os modelos dos objectos. A nosso ver, este trabalho exclui objectos de pequena escala, pois os utilizadores não conseguiram descrever esses objectos com detalhe suficiente. Acreditamos que o emprego de gestos se adequará mais com finalidade de interacção com os objectos, tal como iremos descrever na secção seguinte.

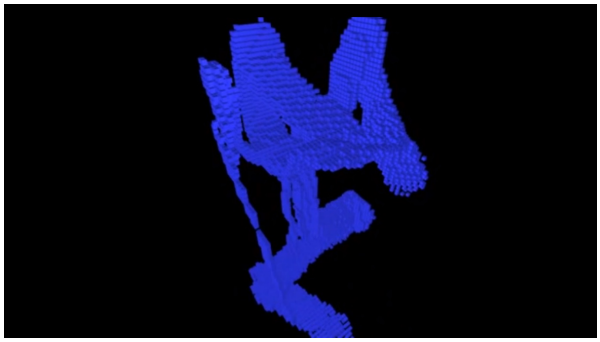

Figura 2.7: Modelo com todos os voxels activos.

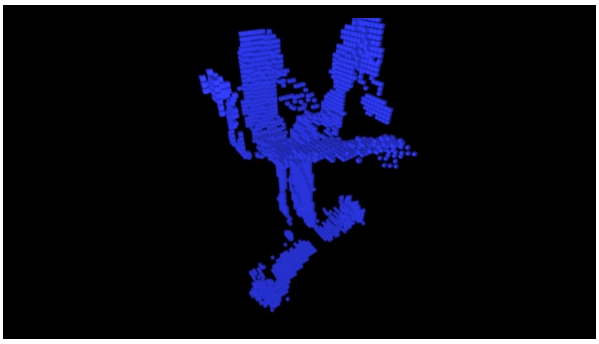

Figura 2.8: Modelo com os voxels de peso superior a n.

2.5 Navegação e Visualização

O uso de interfaces tradicionais que recorrem a rato e teclado, para a visualização e navegação de objectos 3D, faz com que os utilizadores realizem movimentos pouco naturais para conseguir visualizar e interagir com os objectos 3D. A visualização de um mundo virtual através de ecrãs é algo que muitos estão habituados, dado as tecnologias de baixo custo. No entanto, um estudo realizado por Santos et al. (Santos et al., 2009) compara a interacção de utilizadores numa realidade virtual, através da interface tradicional de rato e teclado, e uma usando *Head Mounted Display* (HMD), os quais permitem ao utilizador interagir como se estivesse imerso no realidade virtual. Os autores concluíram que os utilizadores têm melhores resultados com as interfaces tradicionais, mas que à medida que praticam preferem o uso da interface com os óculos HMD.

Quando pretendemos consultar uma colecção de objectos multimédia, a interacção numa realidade virtual pode contribuir para uma melhor análise da colecção. Tendo estes factores em consideração, Nakazato et al. desenvolveu o sistema 3D Mars (Nakazato and Huang, 2001), que consiste num sistema de recuperação de imagens por conteúdo num ambiente imersivo, tal como podemos ver na Figura 2.9. Neste sistema é usado o *Computer Assisted Virtual Environment* (CAVE) e, para explorar o mundo virtual, os utilizadores movem-se dentro do cubo e interagem usando uma "*varinha mágica*". Neste sistema, o utilizador começa por observar inicialmente imagens escolhidas aleatoriamente e, após realização da pesquisa por exemplo, pode visualizar as imagens dispostas no espaço 3D. As imagens são colocadas em localizações especificas tendo em conta as suas características, sendo que as imagens

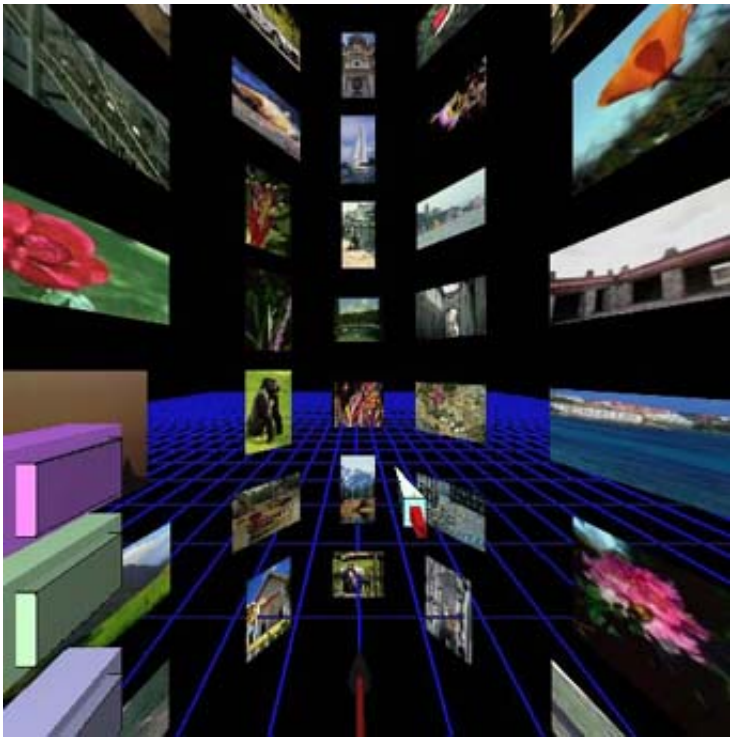

Figura 2.9: Ilustração do ambiente do projecto 3D Mars (Nakazato and Huang, 2001).

mais semelhantes estarão mais próximas do centro. Os autores realizaram ainda um outro método para visualizar as imagens, de forma a permitir uma melhor navegação dos resultados e uma melhor análise. Este consiste numa representação esférica apresentando conjuntos de imagens que estejam próximas de acordo com a semelhança.

Um outro exemplo semelhante é a plataforma g-speak[4] ilustrada na Figura 2.10. O g-speak é um ambiente de operação espacial[5] que usa uma interacção através de gestos para manipulação de conteúdo multimédia em tempo real. Este sistema usa a visualização entre vários monitores e projectores e usa umas luvas com *tracking*, que permitem a manipulação dos objectos visualizados através gestos, como por exemplo de apontar, entre outros movimentos com a mão ou dedos. O sistema g-speak consegue capturar os movimentos com uma precisão de 0.1mm. Usando esta ambiente, Zigeblaum et al. desenvolveram a ferramenta g-salt (Zigelbaum et al., 2010), que permite a interacção com vídeos. Os autores limitaram a expressividade do corpo humano, focando-se apenas nos gestos realizados pelas mãos. Os autores ainda adicionaram alguns gestos para além dos disponibilizados pelo g-speak. Um dos gestos implementados foi o *pinch*, que consiste em tocar a ponta do indicador e do polegar juntos, em que uma das mãos o utilizador agarra o espaço e, movendo a mão, é capaz de navegar no ambiente virtual. Caso ele faça o *pinch* com ambas as mãos consegue não só deslocar-se, mas também rodar no espaço.

Um dos aspectos que pretendemos explorar no nosso trabalho é de oferecer ao utilizador a possibilidade de melhorar a visualização dos objectos 3D e não simplesmente navegar numa lista de objectos.

Figura 2.10: Ilustração do ambiente de operação espacial g-speak.

[4]G-speak © Oblong Industries, `http://www.oblong.com/what-we-do/g-speak`
[5]Ambiente de operação espacial, em inglês: *Spatial operating environment*

De facto, termos uma melhor inspecção de objectos 3D é necessário para comparar os objectos 3D. Apesar de não ter sido com a mesma finalidade de comparar objectos 3D, no seu trabalho, Marton et al. (Marton et al., 2012), pretendem dar uma melhor visualização de objectos 3D, a qual contém uma grande quantidade de detalhe. Neste trabalho os autores utilizam a técnica *Focus sliding surface* (FOX), que consiste em três pontos essenciais: a translação e rotação são obtidas com um movimento de um ponteiro com dois graus de liberdade[6], o que restringe o movimento do ponto de interesse sobre a superfície do modelo; o *zoom* automático depende da velocidade do movimento do utilizador, em que *zoom-in* é utilizado para velocidades mais baixas, e *zoom-out* para velocidades mais altas; por último, a colocação do ponto de interesse de forma automática, ajustando a sua profundidade durante a interacção, é usada de modo a manter a superfície que está a ser manipulada numa boa posição de visualização.

No trabalho de Marton et al. (Marton et al., 2012) conseguimos ter uma boa percepção de objectos 3D com uma grande definição, mas nem sempre temos modelos como os que usaram no seu trabalho. Uma outra técnica que permite dar uma melhor percepção dos objectos, é a Navidget, apresentada por Hachet et al. (Hachet et al., 2009), que permite aos utilizadores indicarem facilmente o que querem ver, através da especificação de uma posição alvo e de uma orientação. Esta técnica, ilustrada na Figura 2.11, é composta por dois passos. No primeiro passo é escolhido um foco através do posicionamento de um raio virtual que intersecte o objecto alvo. No segundo passo, o utilizador poderá definir a posição da câmara, bem como redimensionar o raio da meia esfera, manipulando, desta forma, a distância da câmara ao objecto. A posição da câmara é controlada através da manipulação da posição do cursor que se encontra na superfície da meia esfera.

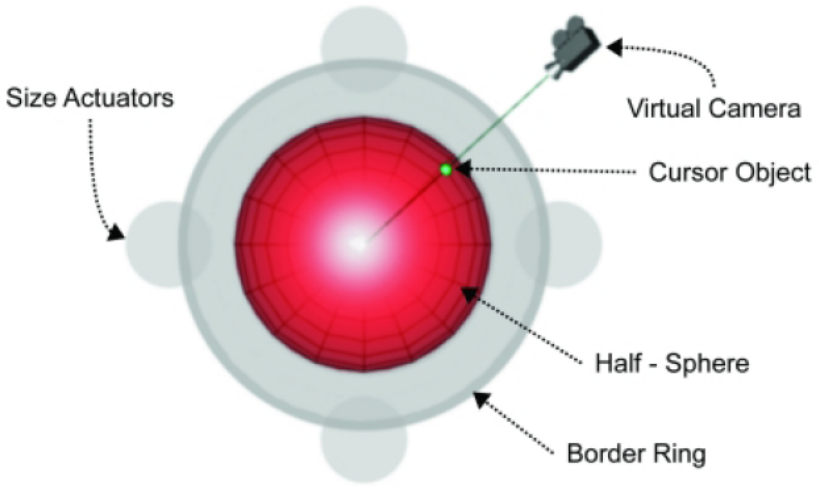

Figura 2.11: Visualização de um objecto 3D, no sistema Navidget (Hachet et al., 2009).

[6]Graus de liberdade em inglês: "*degrees of freedom*"

2.6 Discussão

Trabalhos como os de Smith e Chang (Smith and Chang, 1997) e Funkhouser et al. (Funkhouser et al., 2003), em que a informação é retirada de um conjunto de páginas web, nem sempre conseguem retirar a informação semântica necessária dos objectos que estão a ser pesquisados, podendo por vezes apenas conseguir obter algumas palavras-chave. Contudo, no trabalho de Funkhouser, o facto de serem adicionados sinónimos das palavras melhora esta situação, mas quando se tratam de objectos semelhantes em diferentes posições ou da mesma classe, a distinção pode nem sempre estar retratada no texto.

Em termos de pesquisa por fala, os trabalhos de Moreno et al. (Moreno-Daniel et al., 2007) e de Lee et al. (Lee et al., 2010) focam-se em inicialmente em converter a fala em texto e usar esse texto como objecto de pesquisa. Nestes trabalhos é realizada uma filtragem da informação a recuperar, simplificando bastante as pesquisas. Contudo, eles retiram a informação de resumos das histórias dos livros, sendo que, por vezes, os utilizadores podem não ter conhecimento sobre o conteúdo existente nos resumos.

Ao contrário dos livros e dos filmes, não existem muitas descrições de objectos 3D. Grande parte do desenvolvimento que tem surgido na recuperação de objectos 3D tem sido na construção de descritores de forma, omitindo a informação semântica que poderá facilitar o processo de recuperação. O trabalho Data Miming (Holz and Wilson, 2011) é um caso em que os autores adiantam um pouco mais e descrevem os objectos através de esboços em espaços 3D. Os autores defendem que a fala e os gestos são complementares, no entanto consideram que a descrição por gestos seria suficiente. Acreditamos que a descrição por gestos apenas funcionará para objectos grandes e bastante distintos, dado que é difícil detalhar um objecto 3D usando gestos, como no trabalho anterior. Por outro lado, a fala consegue desviar as limitações existentes nas descrições por esboços, caso os utilizadores tenham toda a informação necessária num objecto. De facto, tal como mencionado no estudo de Kamvar et al. (Kamvar and Beeferman, 2010), as pessoas preferem a pesquisa por fala.

Em termos de visualização, comparando o trabalho de 3D Mars (Nakazato and Huang, 2001) e o ambiente de operação espacial g-speak, estes são bastante semelhantes. No entanto, o g-speak aparenta dar uma grande precisão apenas com a captura de gestos do utilizador. Ainda assim, este ambiente necessita de várias câmeras de infravermelhos de alta precisão para conseguir ter esta precisão, tornando-se assim um sistema bastante dispendioso. Em termos de inspecção dos objectos 3D, com a técnica usada no trabalho de Marton et al. (Marton et al., 2012) é possível inspeccionar a detalhe os objectos. No entanto, para objectos com pouca definição, como é o mais comum de encontrarmos, acreditamos que uma técnica como a de Navidget (Hachet et al., 2009) será suficiente e mais rápida para obter uma visão mais precisa dos objectos 3D para superfícies tácteis. Consideramos que em ambientes imersivos usando dispositivo como os óculos HMD, o utilizador consiga mudar o seu campo de visão ao movimentar a cabeça.

Capítulo 3

Estudo Preliminar

Este capítulo apresenta um estudo preliminar realizado para explorar as descrições de objectos 3D. Para tal, foram realizadas sessões experimentais com utilizadores para compreender quais as expressões verbais e gestuais mais frequentemente utilizadas. Tínhamos como objectivo compreender se os utilizadores recorriam maioritariamente à fala, a gestos (tal como assumido no trabalho de (Holz and Wilson, 2011)) ou a uma combinação de ambos, e ainda quais as descrições mais frequentes. Neste estudo foi utilizado um cenário de construção de modelos LEGO, como o abordado nesta dissertação.

3.1 Descrição

Nesta experiência recorreu-se ao cenário da construção de modelos LEGO, sendo que pretendemos, futuramente, utilizar os métodos identificados para a recuperação de peças virtuais neste contexto. A experiência tinha como objectivo a construção de um modelo físico e cada sessão envolvia dois participantes. A um dos participantes era dado o papel de construtor e a outro o papel de fornecedor. O construtor tinha de montar o modelo LEGO solicitando, para isso, as peças ao fornecedor. Idealmente deveríamos ter feito um estudo da interacção homem-máquina do tipo *Wizard of Oz* (Salber and Coutaz, 1993). No entanto, era importante ter exemplos de interacção entre humanos para comparar. Cada participante desempenhava apenas uma vez cada papel, para evitar o factor de aprendizagem.

As sessões de teste dividiram-se em três etapas: introdução à experiência, execução da tarefa e preenchimento de questionário. Foram realizados testes pilotos por forma a ajustar e verificar a adequação e duração da tarefa. Na Tabela 3.1 encontram-se os objectivos para a duração de cada sessão de teste, assim como as suas componentes, sendo que a duração deveria ter um limite aproximado de 30 minutos.

Etapa	Descrição	Duração
1	Introdução à experiência	5 minutos
2	Execução da tarefa	20 minutos
3	Preenchimento do questionário	5 minutos

Tabela 3.1: Duração das diferentes etapas de cada sessão da experiência.

3.1.1 Introdução à experiência

No início da sessão da experiência, procedeu-se a uma breve apresentação sobre os objectivos do teste, seguida da explicação e distribuição dos papeis a desempenhar. Posteriormente, era dada a possibilidade aos participantes de se familiarizarem com um exemplo das instruções passo-a-passo para a construção de um modelo, ilustradas na Figura 3.1 e diferentes daquelas que iriam seguir na sessão.

3.1.2 Execução da tarefa

Terminada a introdução e tendo explicado como se iria processar a sessão, eram dadas as instruções para a construção do modelo ao participante com o papel de construtor, assim como as caixas contendo as diversas peças ao fornecedor. As caixas tinham um vasto conjunto de peças, não apenas as necessárias para a construção. O fornecedor não conseguia ver as instruções da construção, existindo uma separação entre os participantes. Analogamente, o construtor não conseguia ver as caixas com as peças.

Existiam quatro modelos compostos por peças distintas, mas com complexidade geométrica semelhante entre elas. As instruções para cada construção continham cerca de vinte passos, sendo cada

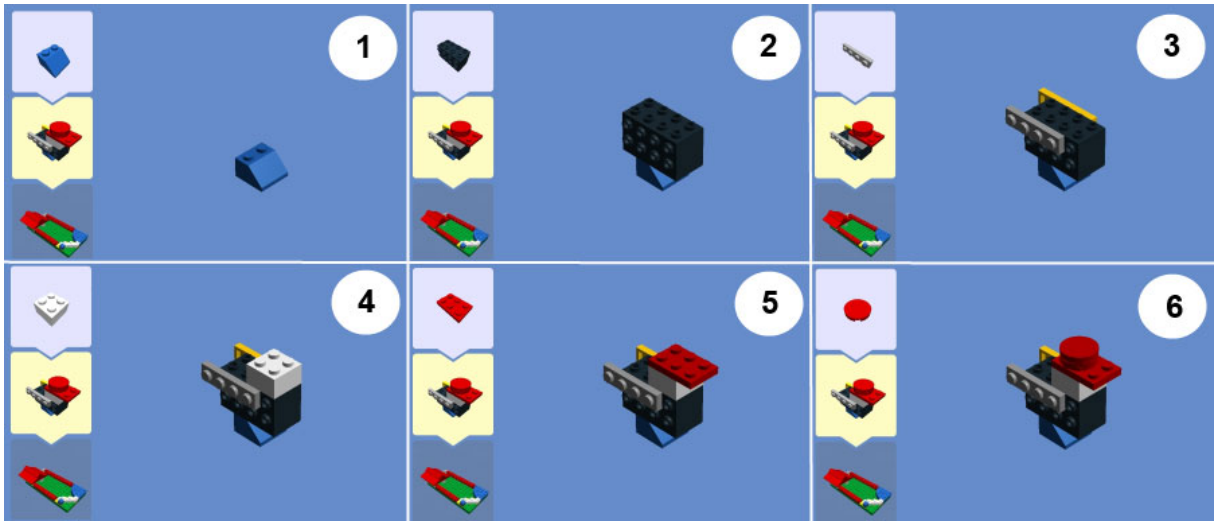

Figura 3.1: Exemplo de instruções fornecidas ao construtor, semelhante às que acompanham os modelos LEGO originais.

passo iniciado pelo pedido de uma peça e terminado pela entrega da peça correcta. O construtor realizava o pedido de peça necessária ao fornecedor, descrevendo-a como considerasse adequado. Feita a descrição, o fornecedor procurava a peça, entregando-a ao construtor para que este pudesse continuar, como visível na Figura 3.2. Após terminada a construção, os participantes trocavam de papéis, repetindo o processo para um modelo diferente.

3.1.3 Preenchimento dos questionários

No final da sessão, era solicitado aos participantes que preenchessem um questionário. Pretendíamos com este questionário fazer um levantamento do perfil dos participantes, assim como identificar as suas preferências relativamente às modalidades (fala, gestos e combinação de fala e gestos) para as descrições das peças, assim como aquelas que tiveram mais dificuldade em descrever.

3.1.4 Participantes

Realizaram-se um total de dez sessões de teste, contando com um total de vinte participantes, sete dos quais do sexo feminino. As idades dos participantes situaram-se entre os dezoito e os quarenta anos, sendo que a maioria tinha entre os dezoito e os vinte e quatro anos (55%). Todos tinham frequência universitária, em diferentes áreas, possuindo pelo menos uma licenciatura. Todos os participantes afirmaram também estarem familiarizados com os brinquedos da marca LEGO, tendo-os utilizado alguma vez na vida. À excepção de dois pares de participantes, todos os outros conheciam o seu par antes da experiência, algo que originou uma interacção informal o que se reflectiu nas suas descrições. Todos

Figura 3.2: Participante no papel de construtor recebe uma peça do fornecedor.

os participantes eram falantes nativos de português.

3.2 Resultados

Durante a realização dos testes, foram recolhidos diversos dados. A análise destes dados foi efectuada segundo duas componentes separadas. Por um lado utilizámos medidas objectivas recolhidas durante a construção do modelo, tendo sido registadas as descrições realizadas pelos participantes, através da gravação de vídeo com áudio. E ainda foi gravado o áudio com um microfone posicionado ligeiramente a baixo do queixo dos participantes, sendo os áudios gravados a 16KHz, 16-bits por amostra. Foram também cronometrados, durante cada sessão, os tempos que os participantes levaram a descrever e a procurar as peças. Por outro lado recorremos a medidas subjectivas, provenientes das respostas dos participantes aos questionários. Adicionalmente, apresentamos um conjunto de observações pertinentes que recolhemos durante as sessões de testes, relativas ao modo como os utilizadores descreveram as peças pretendidas.

3.2.1 Medidas objectivas

Durante a execução da tarefa, registámos o modo como os participantes no papel de construtor solicitavam as peças aos fornecedores. Observámos que, nos diversos passos da construção, os participantes descreveram a peça necessária recorrendo essencialmente à fala. Tal como se pode ver no gráfico na Figura 3.3, onze participantes não utilizaram qualquer gesto em nenhum dos passos da sessão. Certos participantes utilizaram gestos, mas sempre com o objectivo de complementar a descrição verbal, não

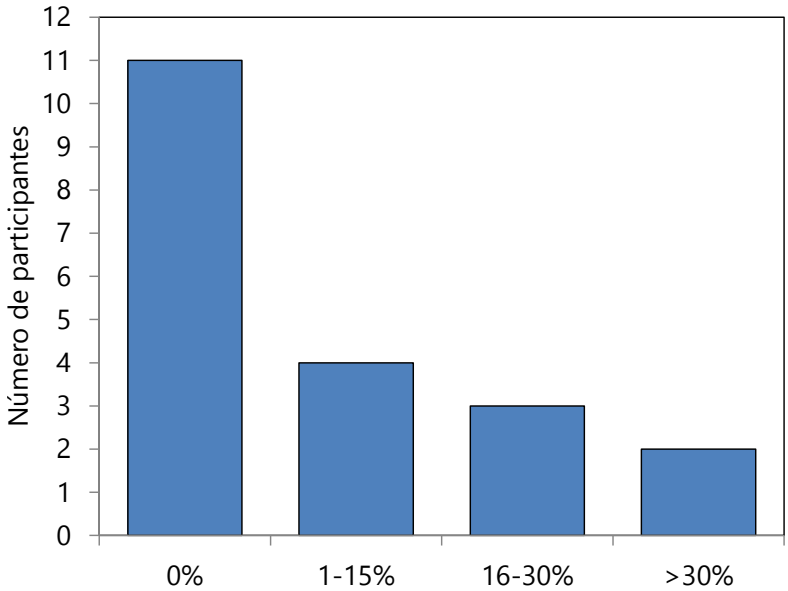

Figura 3.3: Percentagem de passos em que os participantes descreveram a peça utilizando gestos.

acrescentado informação significativa. Por exemplo, ao descrever uma peça de canto, alguns participantes usaram como referência a letra 'L', fazendo o gesto correspondente enquanto a referiam (Figura 3.4). Outro exemplo de como os utilizadores complementavam, pontualmente, a descrição por fala foi a realização de esboços no ar, com a forma da peça pretendida, usando uma palavra a acompanhar os gestos *"assim"*, como por exemplo *"Uma de três por um, assim."*.

Foi frequente observar que os participantes efectuaram diversos refinamentos às descrições. A maioria dos passos necessitava de apenas uma descrição, em média eram necessárias 1.4 descrições para cada passo. O número de passos médios que necessitaram refinamento foi quatro, e dois para o número de refinamentos realizados por ter sido apresentada a peça errada. Embora a metade dos refinamentos tenham tido origem com a peça errada devolvida, por diversas ocasiões os participantes no papel de construtor acrescentaram mais detalhe à descrição inicial por iniciativa própria.

A última das medidas objectivas que foram registadas foi o tempo utilizado na recuperação das peças. Para cada passo da construção anotou-se o tempo despendido na descrição de cada passo, assim como o tempo necessário para a conclusão de cada passo (desde o início do passo, até à entrega da peça correcta). As médias dos tempos obtidos estão visíveis no gráfico da Figura 3.5. O facto de o tempo total ser inferior à soma dos tempos da procura e da descrição deve-se aos fornecedores iniciarem a procura da peça antes do construtor terminar a descrição. Além disso, os construtores, por diversas ocasiões, refinaram a sua descrição antes do fornecedor apresentar uma peça. Pode-se ainda constatar que é despendido sensivelmente o dobro do tempo à procura da peça do que a descrevê-la, tornando-se claro que a existência de um sistema de recuperação é totalmente justificada.

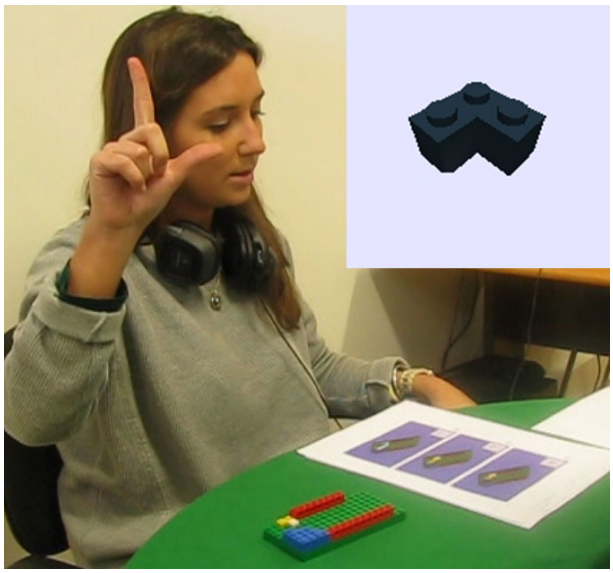

Figura 3.4: Participante solicita uma peça de canto, utilizando uma referência à letra 'L'.

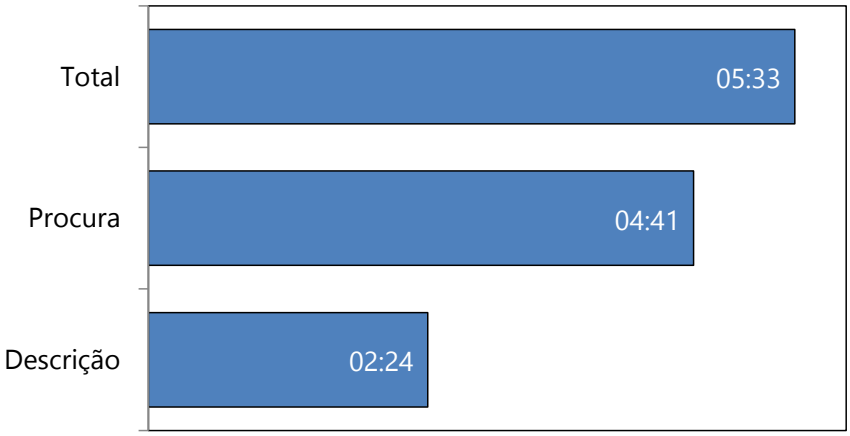

Figura 3.5: Tempos médios (mm:ss) por sessão para obter as peças (Total), para as descrever (Descrição) e para as procurar (Procura).

3.2.2 Medidas subjectivas

No final das construções era solicitado aos participantes que preenchessem um questionário para classificar cada uma das modalidades relativamente à sua facilidade para descrever os objectos. Para a classificação recorreu-se a uma escala de Likert com quatro valores, sendo que 1 corresponde a muito difícil e 4 a muito fácil. Os resultados obtidos estão presentes na Tabela 3.2.

A análise das respostas aos questionários foi feita com recurso ao teste de Wilcoxon. Com este teste foi possível concluir que os participantes concordaram fortemente que a descrição exclusivamente por fala é a modalidade preferida em relação a gestos (Z=-4.018, p=0.000) e à combinação das modalidades (Z=-3.502, p=0.000). Além disso, a utilização combinada de gestos e fala foi consistentemente preferida sobre a utilização de apenas gestos (Z=-3.456, p=0.001).

3.2.3 Observações

Ao longo dos testes registámos os diversos aspectos sobre as descrições das peças efectuadas pelos participantes, bem como após os testes através das gravações de áudio e vídeo. Contudo as gravações de áudio ficaram com bastante ruído devido aos fornecedores mexeram nas peças que estavam dentro da caixa e de ainda se mexerem muito fazendo com que o microfone se desviasse da sua posição

Descrição	Classificação
Verbal	4 (1)
Gestual	1 (1)
Combinada	2 (1)

Tabela 3.2: Classificação da descrição das peças de forma verbal, gestual e combinada (Mediana, Amplitude interquartil).

ideal. Constatámos que geralmente começavam pelas dimensões e cor da peça. Relativamente às dimensões, a unidade utilizada era o número de pinos de encaixe que as peças possuíam, ou o espaço equivalente em peças sem pinos. Para peças com declive, os construtores salientavam este facto, por vezes denominando-as de "rampa"ou "em forma de telhado", explicitando de seguida as dimensões na base e no topo. A altura era normalmente resumida a fina, normal ou alta. Alguns participantes complementavam a descrição da altura com gestos.

Observámos que a palavra *"agora"* foi frequentemente utilizada como começo de uma nova descrição, seguindo por *"quero"*, *"preciso"* ou *"dá-me"*. No entanto, estas últimas duas eram também usadas como início de uma descrição. Verificámos ainda, que as descrições apenas com a descrição da peça, sem qualquer forma verbal. Como por exemplo *"Peça branca, fininha, dois por oito"*. Este último exemplo ilustra as descrições directas, as quais foram mais frequentes com participantes que não se conheciam, sendo também verificado o uso de diminutivos (neste caso o sufixo *-inha"*). Os diminutivos foram usados não para transmitir o tamanho pequeno das peças ou insignificância, mostrando ser muito comuns na linguagem informal utilizada, ocorrendo um em cada sete pedidos. Os participantes utilizaram muito frequentemente confirmações. Algumas das quais foram, *"Okay"*, *"sim"*, *"exacto"*, *"correcto"*, *"isso"* ou equivalentes como *"É esta"* foram também usadas para o mesmo fim.

Foi frequente a utilização de metáforas por parte dos participantes para a descrição das peças. Para além da já mencionada forma de telhado para peças inclinadas, houve descrições peculiares para peças menos vulgares. A peça que contou com mais criatividade por parte dos participantes está ilustrada na Figura 3.6. Alguns participantes descreveram esta peça como sendo a proa de um barco, tendo a forma da letra 'A' ou até mesmo alegando semelhanças com um *Space Invader*[1].

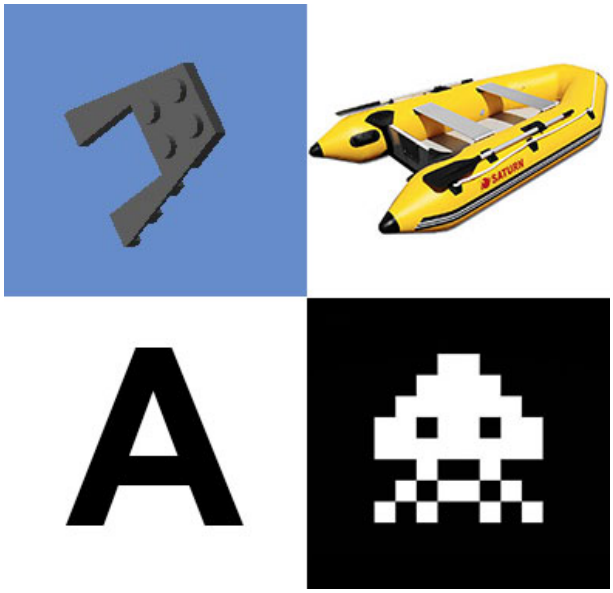

Figura 3.6: Três metáforas utilizadas pelos participantes para descrever a peça representada no canto superior esquerdo.

[1] *Space Invaders* é um dos jogos de vídeos mais icónicos de sempre. Foi lançado em 1978, tendo sido desenvolvido pela Taito Corporation e distribuído pela Midway Games.

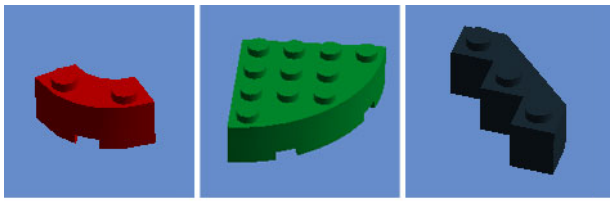

Figura 3.7: Conjunto de peças que levaram a descrições peculiares.

Existiram ainda outros exemplos de peças que originaram descrições interessantes. Para a peça da Figura 3.7-a, alguns participantes descreveram-na como um quarto da letra 'O' ou do muro de um poço. A peça da Figura 3.7-b foi comparada com um triângulo de um queijo, ou, mais frequentemente, com um quarto de um circulo ou um canto com rebordo redondo. Para a peça da Figura 3.7-c surgiram metáforas originais como dentes de tubarão ou o topo de um tridente.

Para efeitos de desenvolvimento do nosso protótipo, este corpus foi dividido em subconjuntos de treino e teste. A Tabela 3.3 descreve a estatística principal dos dois subconjuntos. Os comentários dos participantes que não estavam relacionados com a tarefa foram excluídos das transcrições. Estas transcrições tiveram que ser realizadas manualmente, dada a quantidade de ruído existente nos ficheiros de áudio e a desadequação do modelo de língua ao domínio em causa.

	Treino	Teste	Total
Número participantes	16	4	20
Número pedidos	306	66	372
Número pedidos com gestos	26	11	37
Número de palavras	6088	1015	7103
Número de palavras diferentes	503	167	542
Número de pausas preenchidas	119	21	140
Número de diminutivos	130	15	145
Número de confirmações	156	32	188
Número de metáforas	46	11	57

Tabela 3.3: Estatística dos subconjuntos de treino e teste.

3.3 Conclusões

Com a maior facilidade de criação de conteúdo 3D virtual, a recuperação de objectos específicos em grandes bibliotecas torna-se um desafio cada vez mais relevante. Uma das etapas da recuperação consiste, evidentemente, na descrição do objecto que o utilizador pretende obter, sendo que as soluções existentes na literatura não exploram convenientemente métodos naturais para o fazer, não tirando partido da riqueza descritiva presente na interacção humana.

Com o objectivo de compreender os métodos mais naturais para descrever objectos tridimensionais, conduzimos uma experiência com vinte participantes. Na experiência, um dos participantes tinha de construir um modelo LEGO, tendo de solicitar as peças, descrevendo-as, a outro participante. Os resultados dos testes revelaram que os participantes preferem efectuar descrições exclusivamente verbais. Estas descrições podem ser ocasionalmente complementadas com gestos, não adicionando nova informação com grande relevância. Outro aspecto interessante que observámos prende-se com as expressões utilizadas. O uso destas expressões metafóricas por parte dos participantes pode levantar problemas relativos à sua interpretação por parte de um sistema de reconhecimento, sendo que essa será, sem dúvida, uma área a explorar.

Capítulo 4

LSpeakIt

Neste capítulo iremos apresentar primeiro a arquitectura do protótipo LSpeakIt. Apresentaremos também o processo de indexação realizado para criar os vários índices para procurar os objectos, seguido da especificação das interrogações verbais suportadas no protótipo e como são reconhecidas. Para finalizar, apresentamos o modo como a exploração dos resultados foi desenvolvida no ambiente imersivo e interacção com os resultados. Após verificada a preferência pela descrição verbal pelos participantes no estudo preliminar, decidimos implementar o nosso protótipo, LSpeakIt, representado na Figura 4.1, suportando descrições verbais. Neste protótipo, o utilizador faz pedidos por intermédio de descrições verbais, e recebe o resultado do reconhecimento do sistema, isto é, o que o sistema interpreta ou compreende da interrogação do utilizador. Para esta interacção verbal utilizámos auscultadores com microfone.

Para que fosse transmitida uma boa percepção dos objectos a pesquisar, o nosso protótipo ilustra os objectos num ambiente de realidade virtual imersivo. Após a análise de alguns dos dispositivos existentes, apresentada em Apêndice A, optámos por utilizar no nosso protótipo os Oculus Riftt[1]. Para que os utilizadores pudessem interagir com os objectos apresentados utilizámos no nosso protótipo implementámos duas alternativas usando dois dispositivos dos analisados (em Apêndice A): Leap Motion[2] e SpacePoint Fusion[3].

O LSpeakIt utiliza uma colecção de peças de LEGO, onde para obter a informação destas utiliza o standard definido pelo LDraw[4]. LDraw é um standard aberto para programas de criação de modelos de LEGO virtual e que disponibiliza livremente uma vasta colecção de peças para a comunidade. Decidimos abordar este *toy-problem* como prova de conceito no nosso protótipo. Contudo, acreditamos que pode ser transposto para uma outra área, como por exemplo na indústria automóvel, em que as peças de um veículo automóvel são descritas de forma semelhante às peças de LEGO.

[1] Oculus Rift®, `http://www.oculusvr.com`
[2] Leap Motion ®, `https://www.leapmotion.com/`
[3] PNI Sensor Corporation ®, SpacePoint Fusion,`http://www.pnicorp.com/products/spacepoint-gaming`
[4] LDraw: `http://www.ldraw.org/`

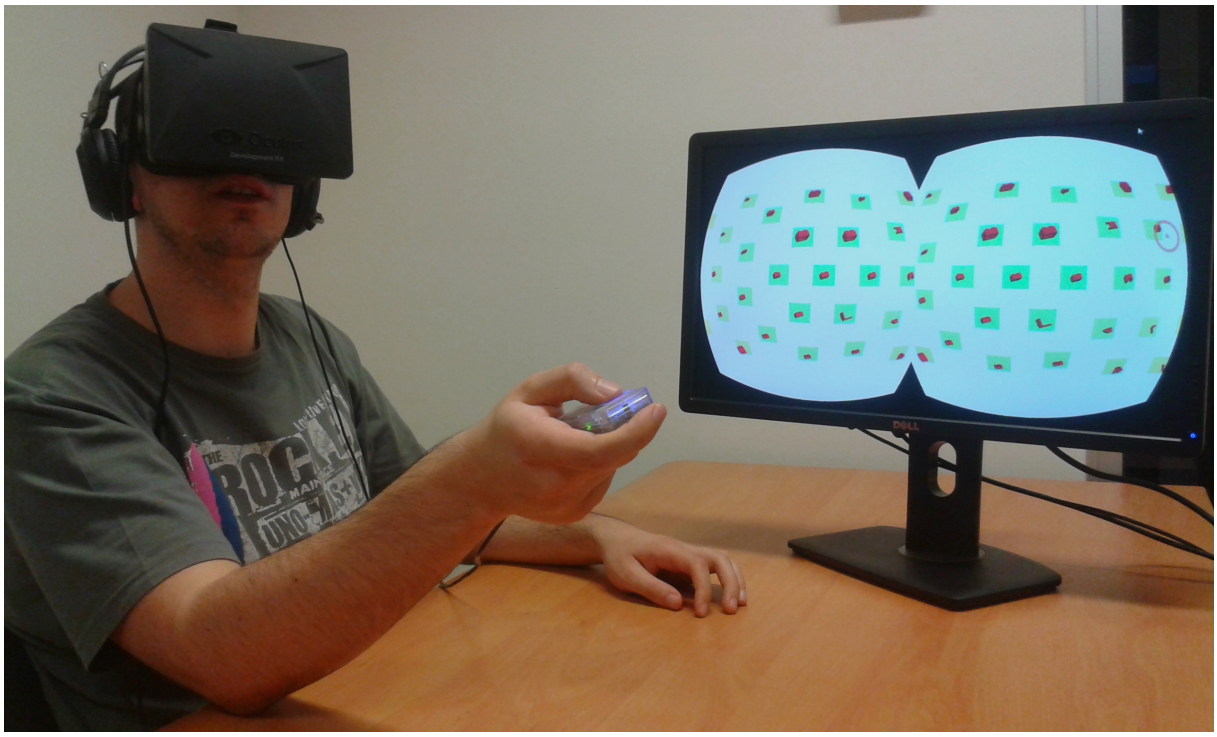

Figura 4.1: Aparato utilizado no protótipo LSpeakIt.

4.1 Arquitectura

De forma a assegurar modularidade, o protótipo LSpeakIt foi desenvolvido em três camadas (Figura 4.2): **Interacção/Visualização**, **Controlo** e **Dados**. A camada de topo é responsável por todas as interacções e visualizações para o utilizador. A camada **Controlo** é responsável pela criação dos modelos no ambiente virtual assim como gerir os pedidos vindos da camada **Interacção/Visualização**. Por fim a camada mais baixa, **Dados**, é responsável pela parte da gestão dos modelos na colecção e criação dos índices.

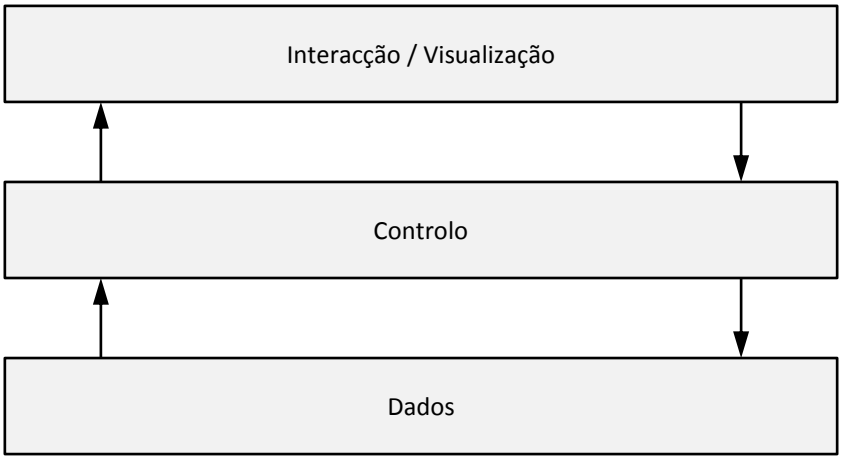

Figura 4.2: LSpeakIt - Vista das três camadas da arquitectura utilizada.

A camada de **Interacção/Visualização** é composta por quatro módulos distintos: Gestor de Fala, Gestor de Gestos, Gestor de Câmara e Visualização (Figura 4.3). O primeiro módulo tem como objectivo reconhecer e compreender as descrições realizadas pelos utilizadores e de transmitir o que foi compreendido. Neste módulo o reconhecimento da fala é realizado utilizando o sistema de reconhecimento de fala AUDIMUS (Meinedo et al., 2010) e a sintetização da fala é realizada através do sistema de sintetização fala DIXI (Paulo et al., 2008). O **Gestor de Gestos** recebe os movimentos realizados com as mãos do utilizador, no caso do Leap Motion, ou orientações, no caso do SpacePoint Fusion, sendo aqui analisados os movimentos realizados e determinado que gestos foram executados pelo utilizador. O **Gestor de Câmara** recebe a orientação da cabeça do utilizador através dos Oculus Rift e determina para onde ele deve estar olhar no ambiente virtual. O módulo **Visualização** tem como objectivo tratar da ilustração dos ambiente virtual.

A camada central é composta por dois módulos: Gestão de Interrogações e Gestão de Objectos. A **Gestão de Interrogações** recebe as descrições do módulo Gestão de Fala e as interacções do módulo Gestão de Gestos, e trata de organizar essa informação para realizar as respectivas interrogações, podendo estas representar pedidos de objectos semelhantes ou relacionados com as descrições recebidas. O módulo **Gestão de Objectos** é onde se constrói todas as peças de LEGO disponíveis, assim como é o local onde contém os detalhes de cada peça.

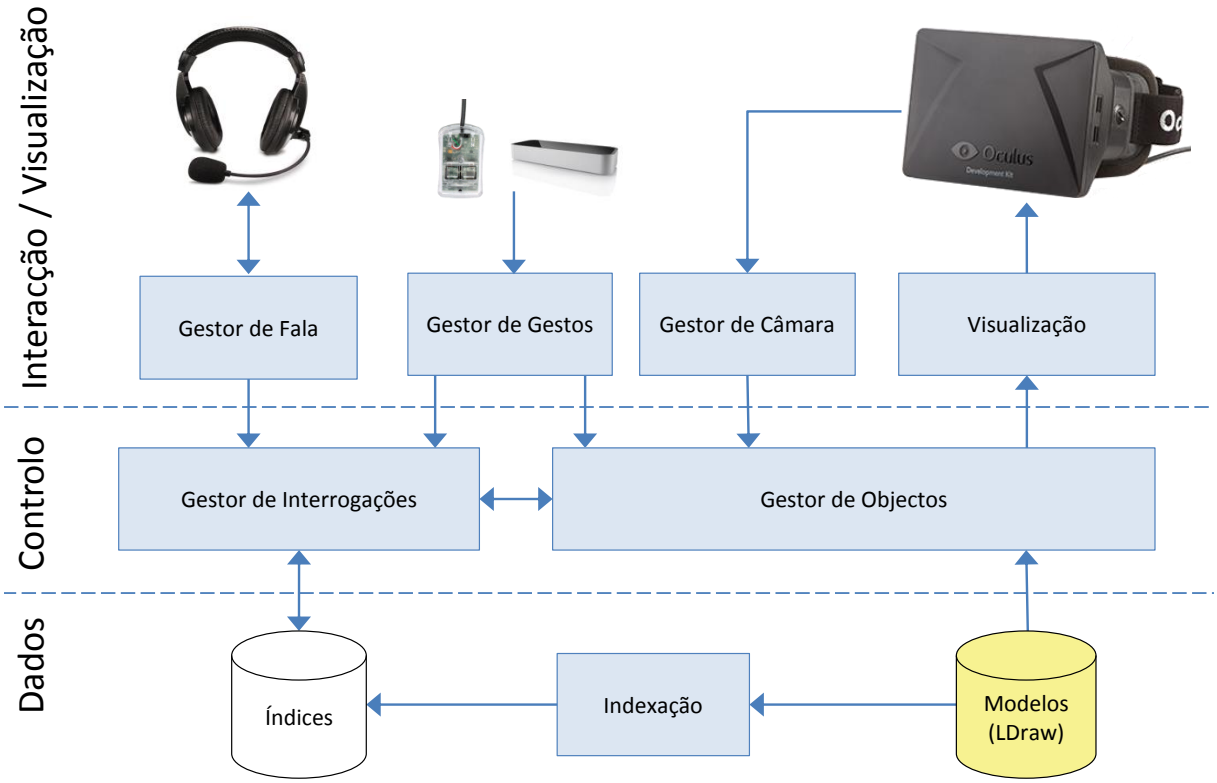

Figura 4.3: LSpeakIt - Arquitectura utilizada.

Por fim, a camada inferior, **Dados**, contém o módulo de **Indexação**, os **Índices** e a colecção de modelos de peças LDraw. É no módulo de **Indexação** que são analisadas as descrições, tanto geométricas como textuais dos modelos, de forma a construir os **Índices** que são usados pelo Gestor de Interrogações. Os modelos das peças contêm a informação geométrica das peças a serem utilizadas pelos módulos Gestor de Objectos e Indexação, bem como as respectivas descrições de cada peça a ser utilizada pelo módulo Indexação. A arquitectura com interacção entre os módulos apresentados encontra-se ilustrada na Figura 4.3.

4.2 Indexação

De modo a tirar partido das soluções existentes na correspondência por texto, implementámos a abordagem proposta por (Gennaro et al., 2010), mas no contexto de objectos 3D. Para isso, utilizou-se a colecção de peças LEGO da LDraw como objectos 3D para o nosso protótipo. Neste sentido foi utilizado um subconjunto de 1810 peças desta colecção. Cada objecto desta colecção contém uma descrição textual, com termos consistentes em toda a colecção. Cada bloco é composto por informação geométrica, como linhas, triângulos e quadriláteros.

Para manter toda a informação dos objectos acessível de uma forma fácil e rápida, construímos índices e índices invertidos para as correspondentes informações da descrição textuais e de forma, como ilustrado na Figura 4.4. Relativamente às descrições textual, cada objecto tem termos da sua descrição, juntamente com o número de vezes que eles se repetem, no índice de descrições. Cada termo tem os objectos que são descritos com ele associado no índice invertido.

4.2.1 Representação textual para objectos 3D

De forma a sermos capazes de utilizar motores de busca de texto, para recuperação de objectos 3D, é necessário termos representações por texto dos objectos. Para tal usámos um método semelhante ao proposto por (Gennaro et al., 2010), adaptado para o conteúdo virtual em 3D.

Para extrair as características da forma optámos por usar o descritor D2, pois mostrou resultados promissores em avaliações anteriores (Funkhouser et al., 2003). Este descritor necessita de muito menos esforço computacional e de armazenamento, quando comparado aos métodos mais discriminativos, como o descritor *lightfield* (Shen et al., 2003). O descritor D2 representa o objecto 3D pela distribuição de distâncias Euclidianas entre pares de pontos aleatórios da sua superfície. Após aplicado o D2 a cada objecto, obtemos os vectores de características de 128 dimensões para cada objecto.

Com base na ideia que objectos semelhantes terão uma visão semelhante do espaço envolvente, seleccionamos um conjunto de objectos de referência (RO)[5] da colecção. É calculada a distância euclidiana

[5]Objecto de referência, "Reference Object"

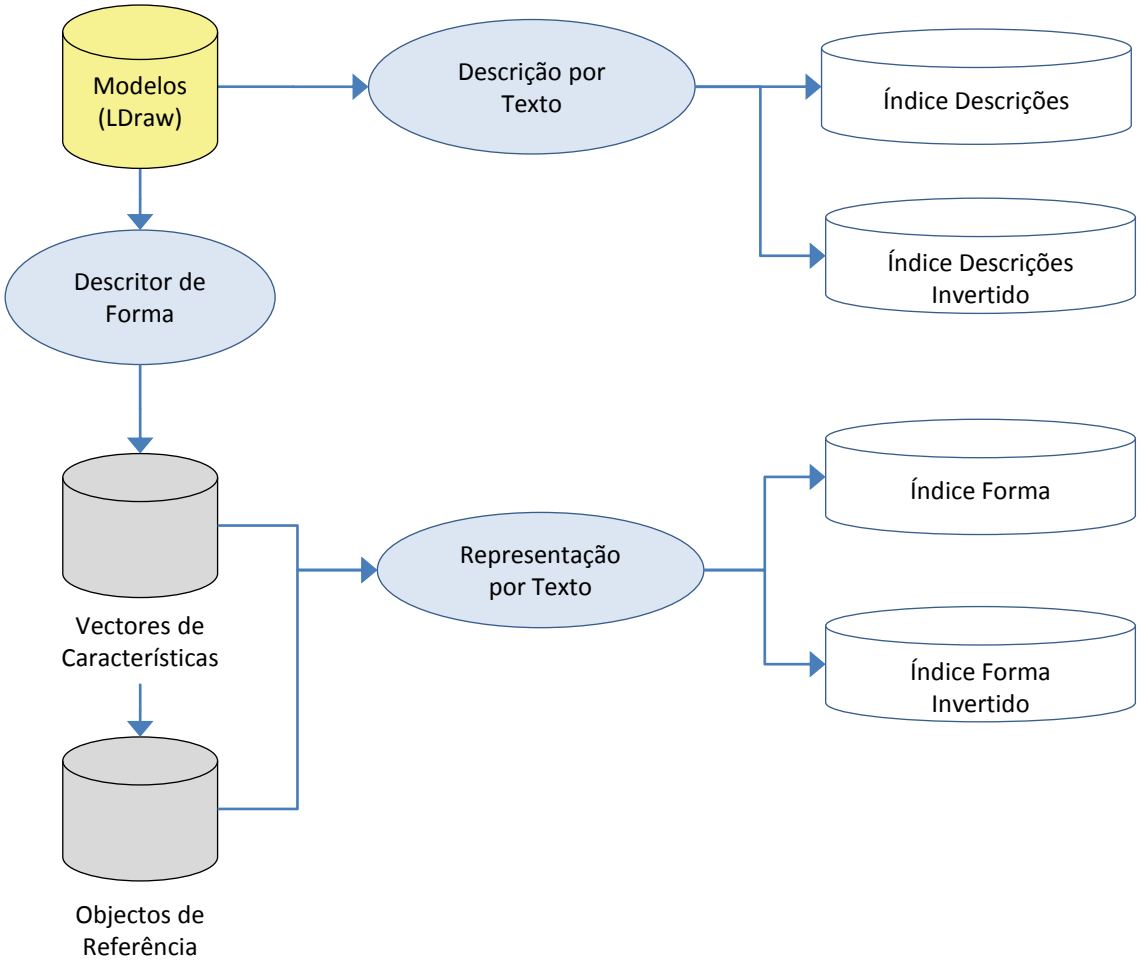

Figura 4.4: Sistema de indexação do LSpeakIt.

entre a assinatura de cada objecto na colecção e assinatura de cada um destes RO. É então utilizada esta distância para ordenar todos os RO para cada objecto. Atribuindo um termo único para cada RO, criamos uma representação textual de um objecto, repetindo estes termos mais vezes quando o RO está mais próximo do objecto. Por exemplo, vamos considerar um objecto A e três RO, os termos associados são RO1, RO2, RO3. A:(RO3, RO2, RO1) é a ordem dos RO, ou seja são ordenados de acordo com sua proximidade a A, sendo o mais próximo RO3. Neste caso A terá a seguinte representação textual: RO3 RO3 RO3 RO2 RO2 RO1.

No nosso protótipo, usámos 85 RO seleccionados aleatoriamente da colecção para representar os 1810 objectos. Este número veio da equação (Equação 4.1) sugerida por (Amato and Savino, 2008), onde X são todos os objectos na colecção:

$$\#RO \geq 2 \cdot \sqrt{\#X} \tag{4.1}$$

Para alcançar um compromisso aceitável entre a indexação, os tempos de interrogação e resultados precisos, utilizámos 30 RO para gerar a representação textual. Desta forma, embora cada objecto seja

comparado com 85 RO, consideramos apenas os 30 mais próximos, reduzindo significativamente o tamanho das representações textuais.

As representações geradas dos objectos, já em forma de texto podem, em seguida, ser processadas de uma forma semelhante às descrições textuais. Para cada objecto, os RO mais próximos de si estão associados no índice de forma, dando maior relevância aos objectos mais próximos. Para cada RO, os objectos da colecção que lhe estão mais próximos são associados no índice de forma invertida.

4.2.2 Interrogações ao sistema

Para procurar objectos na colecção, oferecemos dois tipos de interrogações: interrogação por texto e interrogação por exemplo. Ambos contam com a abordagem TF-IDF para calcular os pesos dos resultados. Interrogação por texto é uma aplicação directa de um motor de busca por texto padrão para vários documentos. Para cada termo da interrogação, o sistema irá reunir documentos que contenham tais termos das descrições no índice invertido. A relevância de cada documento para a interrogação é calculado ponderando a relevância dentro da colecção dos termos contidos na interrogação e ainda a relevância dos mesmos termos, tanto na interrogação como no documento. A relevância de um termo dentro da colecção será maior, quanto menos documentos o termo ocorrer, enquanto que a relevância dos termos na interrogação e no documento será maior quanto mais vezes ele é repetido. A Tabela 4.1 um exemplo de uma interrogação por texto no nosso sistema.

Objectos	Relevância	Descrição
	0.589498	Inclinado Bloco Curvo 2 x 1
	0.574664	Inclinado Bloco 45 2 x 1
	0.534796	Inclinado Curvo Topo 2 x 2 x 1
	0.521746	Inclinado Liso 45 2 x 1
	0.519484	Inclinado Bloco 45 2 x 2

Tabela 4.1: Os cinco melhores resultados para a interrogação *Inclinada 1 x 2*.

Na procura de objectos semelhantes, isto é, na interrogação por exemplo, são comparadas tanto as descrições textuais como a informação da forma. Assim, a interrogação por exemplo combina a interrogação de descrição por texto com uma interrogação de informação da forma, ambas processadas de forma idêntica. No primeiro caso consideramos as descrições do objecto seleccionado e usando-as para realizar uma interrogação por texto. No segundo caso utilizamos a representação textual da informação da forma do objecto e utilizamos novamente uma interrogação por texto, mas neste caso para os

índices de forma. Para calcular a relevância de um objecto usando esta combinação o sistema recorre às outras abordagens descritas. Uma vez que cada abordagem gera um valor de relevância contido no intervalo [0, 1], pode ser calculada a média destes, mantendo-se o valor de semelhança definitivo neste intervalo. A ideia principal é que os objectos semelhantes ao objecto seleccionado usando tanto as descrições por texto e informação da forma será mais relevante, e os objectos que são semelhantes apenas em relação a um método serão penalizados.

4.3 Especificação de interrogações

Apesar do sistema realizar interrogações por texto, o que o utilizador irá fazer serão pedidos por fala. Neste caso existe um reconhecimento e compreensão do que este está a descrever no seu pedido. Para esse reconhecimento foi necessário utilizar a biblioteca SAPI da Microsoft[6], para podermos alcançar não só o sistema de reconhecimento AUDIMUS (Meinedo et al., 2010) para Português Europeu, mas também o sistema de sintetização DIXI (Paulo et al., 2008). Podemos visualizar esta informação na Figura 4.5.

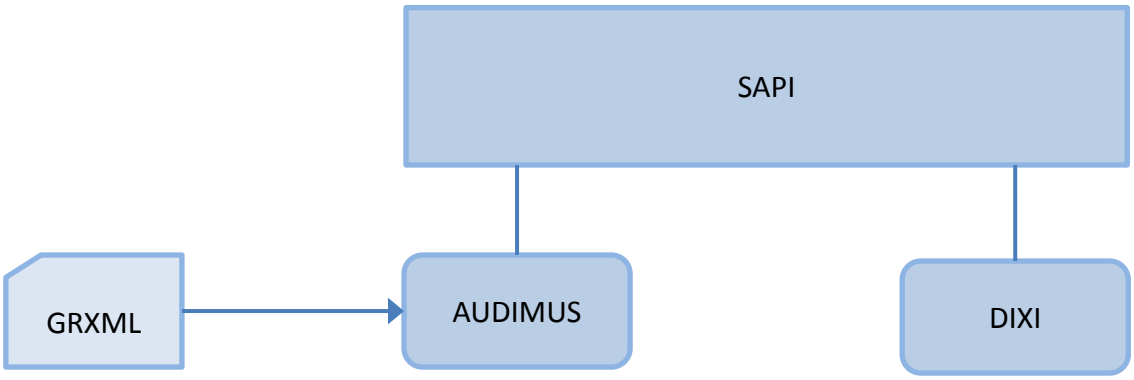

Figura 4.5: Ligação da biblioteca SAPI com o reconhecedor e sintetizador.

Dada a desadequação do contexto das descrições das peças com o modelo de língua utilizado para o contexto de noticiários em (Meinedo et al., 2010), foi necessário abordar uma forma alternativa. Tendo isso em conta, utilizámos as gravações recolhidas no estudo apresentado no Capítulo 3, as quais foram divididas em subconjunto de treino e subconjunto de teste, de modo a podermos construir uma gramática neste contexto. A gramática foi construida utilizando especificação GRXML[7], sendo atribuída ao sistema de reconhecimento AUDIMUS, limitando o contexto com as descrições do *toy-problem*. O vocabulário desta gramática incluía mais de 500 palavras diferentes, dada a variação de palavras entre género e singular ou plural (como por exemplo *inclinado/inclinada/inclinados*). Este vocabulário permitiu cobrir as descrições das 1810 peças.

[6] *Microsoft Speach API*,http://msdn.microsoft.com/en-us/library/ee125663(v=vs.85).aspx
[7] "Speech Recognition Grammar Specification", http://www.w3.org/TR/speech-grammar/

A gramática permitia descrever uma peça utilizando várias descrições. Um exemplo disso é o caso da peça apresentada na Figura 4.6, pertencente à classe de peças LEGO redondas. O utilizador poderia descrever esta peça utilizando 14 adjectivos diferentes. Um outro exemplo é o caso da peça apresentada na Figura 4.7, pertencente à classe peças finas. Neste caso o utilizador tinha 21 adjectivos que corresponderiam a esta classe. Estas correspondências estão explicitadas no GRXML, onde está definido a que classes um adjectivo ou descrição pertencem.

Para iniciar uma interrogação foi necessário ter em conta algumas das possíveis abordagens para a interacção com o sistema: estar sempre a reconhecer as descrições; pressionar para falar; ou acordar o sistema com uma frase específica. No primeiro caso, o que sucede é o sistema estar sempre à escuta de possíveis interrogações da parte do utilizador. Este caso não permite que o utilizador possa fazer algum comentário durante a sua experiência, podendo resultar em alguma interrogação indesejada. No caso da alternativa de pressionar para falar existem algumas variações. Um dos casos seria o utilizador pressionar o botão enquanto faria o seu pedido. Um outro caso seria utilizar o botão como ponto de início e/ou fim do pedido realizado pelo utilizador. No entanto, esta abordagem exigia que o utilizador necessitasse de um botão especifico para controlar os pedidos que seriam realizados. A última alternativa permite que haja uma interacção controlada da parte do utilizador sem ser necessário um botão físico. Neste caso, é dado o início do reconhecimento, terminando com o reconhecimento da interrogação verbal realizada pelo utilizador. Sendo esta última a opção que desenvolvemos no nosso protótipo, o utilizador precisaria de introduzir a frase "Acorda Lego", por forma a activar o sistema. O sistema responderia "Sim", estando, a partir desse momento, pronto a receber a interrogação

Posto isto, um exemplo de uma interrogação simples seria "Acorda LEGO. Quero uma peça fina, 2 por 2.", sendo de seguida sintetizado o que foi compreendido pelo sistema "Ilustrando peça fina, 2 por 2.". Esta situação ilustra uma clara diferença em relação aos pedidos que os participantes do estudo preliminar realizavam na utilização de cor como na descrição. Isto deve-se ao facto de, ao contrário do que foi verificado no estudo neste ambiente virtual a cor não era um factor de descrição, pois todas as peças poderiam ser ilustradas com todas as cores disponíveis. Deste modo, se os utilizadores

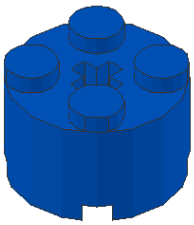

Figura 4.6: Peça com descrição "redonda 2 x 2".

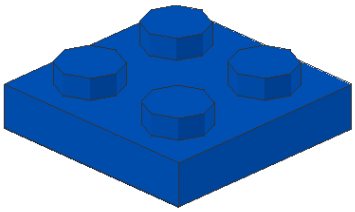

Figura 4.7: Peça com descrição "fina 2 x 2".

pretenderem alterar a cor da peça após a sua selecção, têm essa hipótese (por exemplo, "Acorda LEGO. Pinta de amarelo.").

Dado o grande número de peças disponíveis, os utilizadores têm a possibilidade de refinar os resultados e, para tal, podem filtrar e excluir peças. A filtragem dos resultados consiste em filtrar as peças que contenham uma descrição em particular. Um exemplo de interrogação será "Acorda LEGO. Filtra peças por inclinadas.", sendo respondida por "Filtrados resultados por peças inclinadas". As exclusões são o inverso do filtrar, ou seja são excluídas peças com uma dada descrição. Um exemplo desta interrogação é "Acorda LEGO. Exclui peças com inclinação, respondendo o sistema "Excluídas peças com inclinação". Para o utilizador realizar uma interrogação por exemplo, é necessário seleccionar o objecto pretendido e realizar uma interrogação como, "Acorda LEGO. Dá-me parecidas esta.".

4.4 Exploração dos resultados

Para que os utilizadores possam visualizar e explorar os resultados retornados, o nosso protótipo tira proveito de um ambiente imersivo. Este é transmitido para os utilizadores através de um par de óculos HMD, tendo sido utilizado o modelo Oculus Rift. Deste modo, os utilizadores conseguem ter uma visualização na primeira pessoa no ambiente à semelhança do trabalho apresentado por Pascoal et al. (Pascoal et al., 2012), dando um verdadeira percepção do 3D dos objectos ao contrário dos sistemas tradicionais (Funkhouser et al., 2003).

4.4.1 Apresentação de resultados

Ao contrário do trabalho de Pascoal et al. (2012) mencionado anteriormente, a solução proposta apresenta a disposições dos objectos envolvendo o utilizador, e utilizando a relevância dos resultados para a colocação das peças. Deste modo, analisámos primeiro um conjunto de modos de visualização imersivos e de seguida estes foram adaptados de acordo com preferências dos utilizadores dando origem a um novo modo.

Modos de visualização

Para proporcionar uma melhor visualização dos resultados ao utilizador desenvolvemos um conjunto de modos de visualização num ambiente imersivo: Rectangular, Quadrático, Cilíndrico e Esférico. A Figura 4.8 ilustra estes quatro modos de visualização.

O modo Rectangular, consiste em apresentar uma grelha de objectos na forma de rectângulo na frente do utilizador, tal como apresentado na Figura 4.8-a. Neste modo os objectos com maior relevância estão apresentados no canto superior esquerdo e no canto inferior direito o menos relevante.

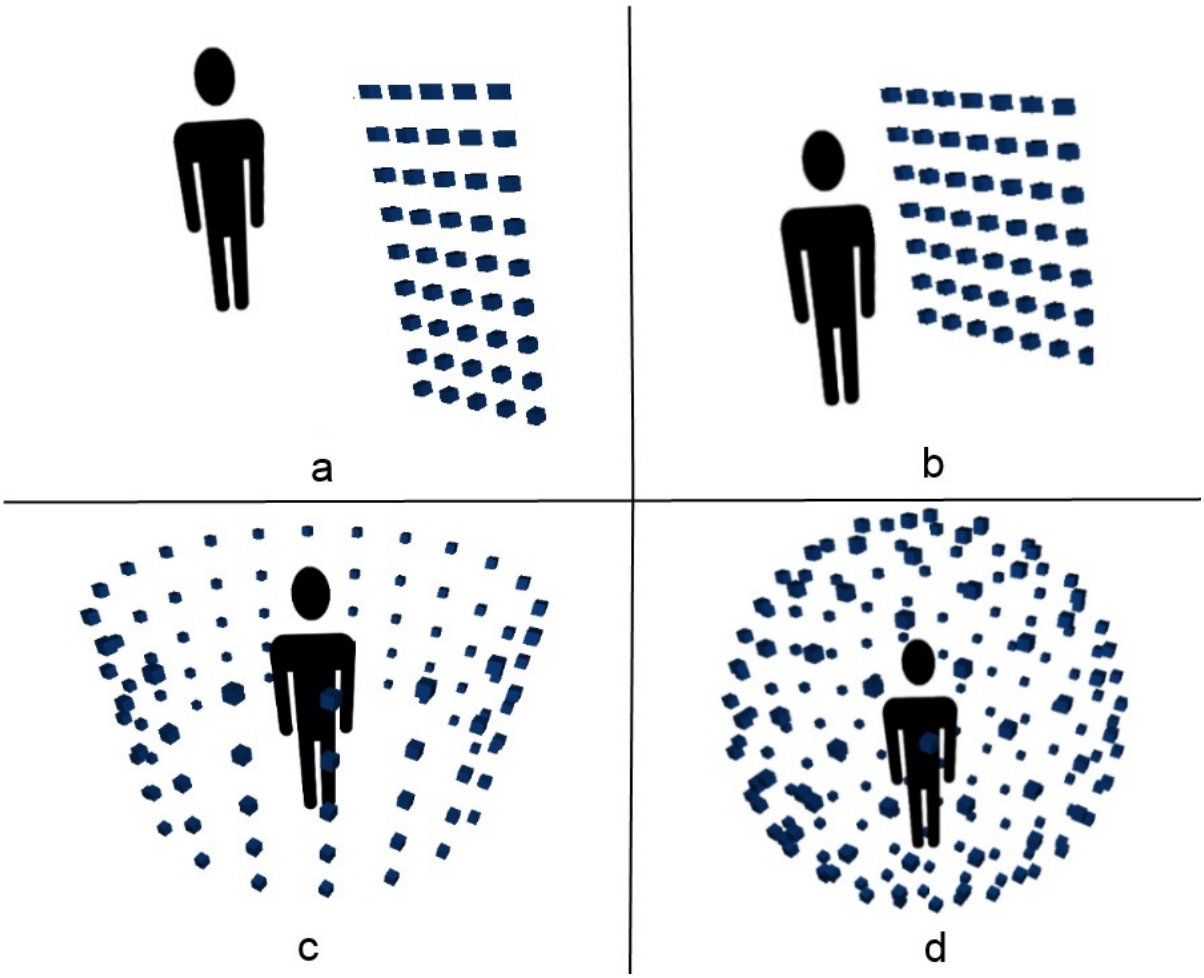

Figura 4.8: Modos de apresentar os resultados de interrogações num ambiente imersivo:
(a) Rectangular, (b) Quadrático, (c) Cilíndrico, (d) Esférico.

Estando os objectos mais relevantes apresentados à mesma altura que o olhar do utilizador.

O modo Quadrático, tal como o Rectangular consiste numa grelha de objectos, mas neste caso na forma de um quadrado, como podemos ver na Figura 4.8-b. Ao contrário da anterior os objectos mais relevantes encontram-se no centro da grelha e os menos relevantes no limiar desta, sendo a grelha preenchida de forma espiral. Outra diferença para o anterior, é que grelha começa um pouco mais em cima de forma que o olhar do utilizador comece de encontro o centro da grelha, onde estarão os objectos mais relevantes.

Por fim o modo Cilíndrico e Esférico, ao contrário dos anteriores, pretende envolver o utilizador com os resultados como se estes estivessem numa superfície cilíndrica ou esférica, respectivamente. Aqui tal como na anterior o resultados mais relevantes começam na direcção do olhar do utilizador, isto é no meio do cilindro e esfera, expandido ao longo da superfície.

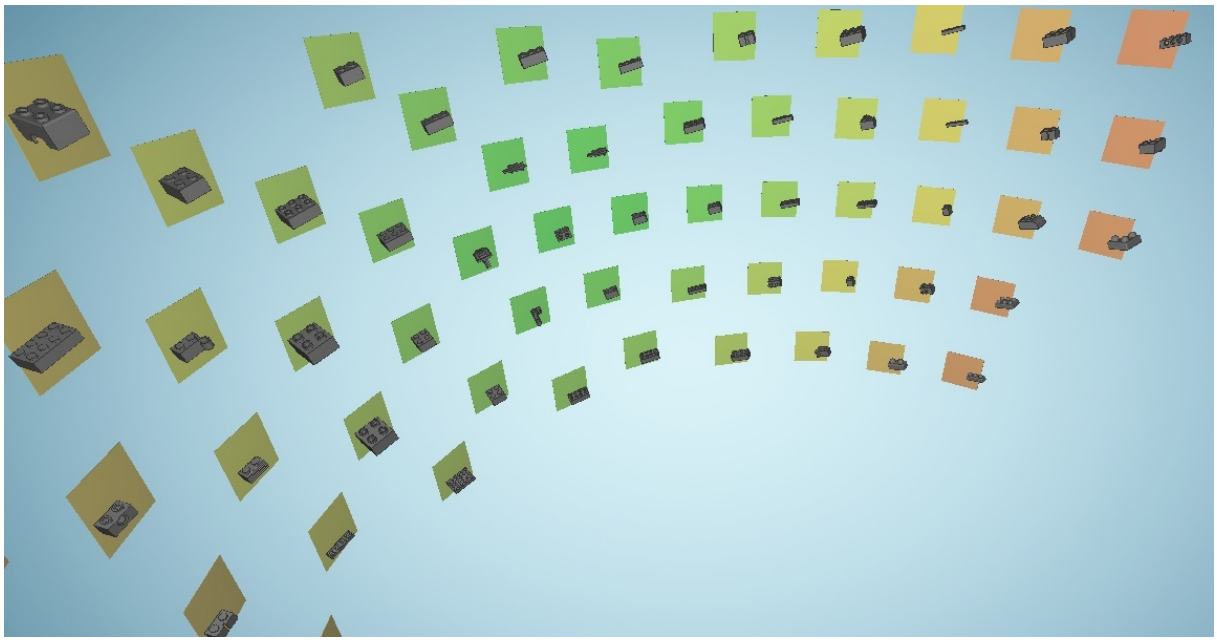

Figura 4.9: Disposição dos objectos em volta do utilizador.

Modo de visualização Barril

A disposição dos objectos foi avaliada com utilizadores numa fase inicial do protótipo, estando esta avaliação apresentada no Capítulo 5.1. Esta avaliação levou a que a disposição dos resultados final originasse da combinação de dois modos de visualização dos resultados preferidos pelos utilizadores, cilíndrica e esférica. Podemos ver a disposição das peças na Figura 4.9. Aqui nesta situação os objectos são colocados numa superfície semelhante a um cilindro arredondado na vertical, ou seja quase uma superfície esférica com dois cortes em cima e em baixo, tendo uma forma semelhante à de um Barril.

Tirando partido do ambiente imersivo conseguimos apresentar 80 objectos, sem que o utilizador se sinta sobrecarregado com a quantidade apresentada. Os objectos retornados são colocados de modo a que o utilizador não necessite de rodar a sua cabeça mais do que noventa graus para cada lado. Estes objectos são apresentados ao utilizador a rodar sobre os seus eixos verticais, com uma ligeira inclinação, 45º no eixo vertical e 22.5º no eixo horizontal. Esta inclinação nas peças permite ao utilizador uma melhor percepção das peças e das suas dimensões.

A colocação dos objectos é realizada de acordo com a sua relevância ao pedido efectuado. Isto é, a peça que se encontra imediatamente em frente ao campo de visão do utilizador é a peça mais relevante para o pedido efectuado. As seguintes irão sendo distribuídas na superfície de acordo com a sua relevância, sendo posicionadas segundo uma forma espiral, tal como podemos ver na Figura 4.10. Com a variação do olhar dos utilizadores, era frequente perderem a noção de onde se encontrava a peça mais relevante. De modo a guiar o utilizador, foi colocado um quadrado por trás do objecto, em que a sua cor simboliza a relevância do mesmo, sendo os objectos mais relevantes representados com

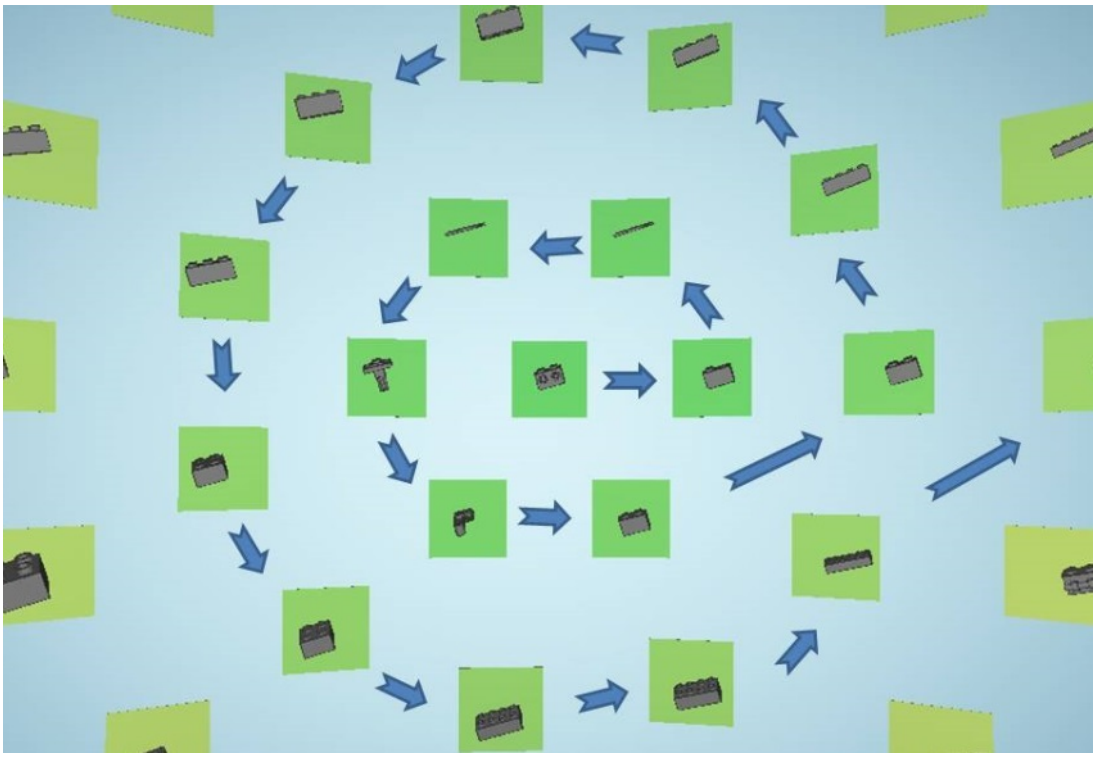

Figura 4.10: Distribuição dos objectos de acordo com a sua relevância.

cor verde e a vermelho os menos relevantes.

4.4.2 Interacção com os objectos

Para visualizarem os objectos com mais detalhe as peças, com o intuito de se destinguir peças semelhantes, foi realizada uma interacção com objectos, podendo seleccioná-los e interagir com estes. A interacção foi realizada numa primeira versão com o dispositivo Leap Motion, para os quatro modos de visualização.

Interacção com Leap Motion

Na primeira interacção implementada, utilizámos o dispositivo Leap Motion, um sensor óptico que permite a captura de gestos realizados por mãos e dedos. Para todos os modos de visualização existem dois gestos, que permitem os utilizadores navegar nos resultas, apontar e controlo dos resultados. O apontar é realizado utilizando o dedo indicador esticado enquanto os restantes dedos da mão estão encolhidos, tal como podemos ver na Figura 4.11. Para o utilizador saber onde está apontar na área dos resultados, colámos uma mira que irá seguir a orientação do dedo indicador. Uma peça é seleccionada quando colidir com a mira, movendo a peça mais perto do campo de visão do utilizador, começando a rodar sobre o seu próprio eixo vertical para o utilizador a ver de diferentes pontos de vista. O gesto

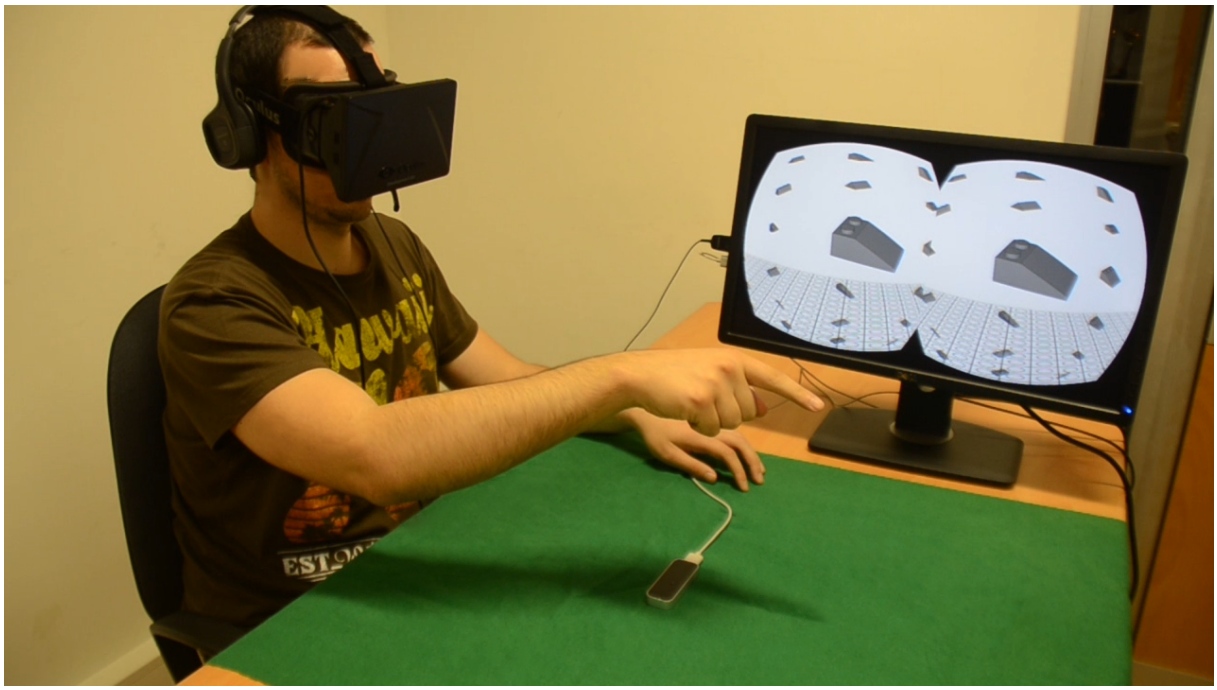

Figura 4.11: Visualização de um objecto mais próximo utilizando o gesto de apontar com o Leap Motion.

de controlo é realizado utilizando os cinco dedos de uma mão estendidos. Este gesto permite mover os resultados em redor do utilizador, variando para cada modo de visualização, a forma como são movidos. No caso dos modos Rectangular e Quadrático os resultados eram movidos na horizontal ou vertical havendo um mapeamento directo com o movimento da mão do utilizador. No modo Cilíndrico, o movimento seria semelhante no caso vertical. Quanto ao movimento horizontal das mãos, este origina uma rotação dos resultados, ou seja mover a mão para a esquerda causa uma rotação do conjunto de resultados para esquerda, em redor do utilizador, e para a direita, caso mova a mão para esta direcção. No modo Esférico, o movimento horizontal das mãos é o mesmo que o modo Cilíndrico, no movimento vertical aplica também uma rotação ao conjunto dos resultados, mas neste caso estes rodam para cima e para baixo, respectivamente, com o movimento da mão.

Interacção com SpacePoint Fusion

Após a avaliação dos modos de visualização, apresentada na Secção 5.1, reparámos que os utilizadores demonstraram ter algumas dificuldades em utilizar gestos com o Leap Motion. Houve inclusivamente situações em que os utilizadores perdiam a noção de onde se encontrava o dispositivo, procurando usar o mínimo possível de gestos para controlar os resultados, dado poderem visualizá-los com uma rotação do ponto de vista, rodando a cabeça. Devido a estes factos decidimos utilizar outro dispositivo, o qual contém um mapeamento directo e preciso, o SpacePoint Fusion. Este dispositivo está equipado com um giroscópio e dois botões posicionados lado a lado, no topo do dispositivo.

Tirou-se partido do giroscópio do SpacePoint Fusion, para facilitar a acção de apontar para os objec-

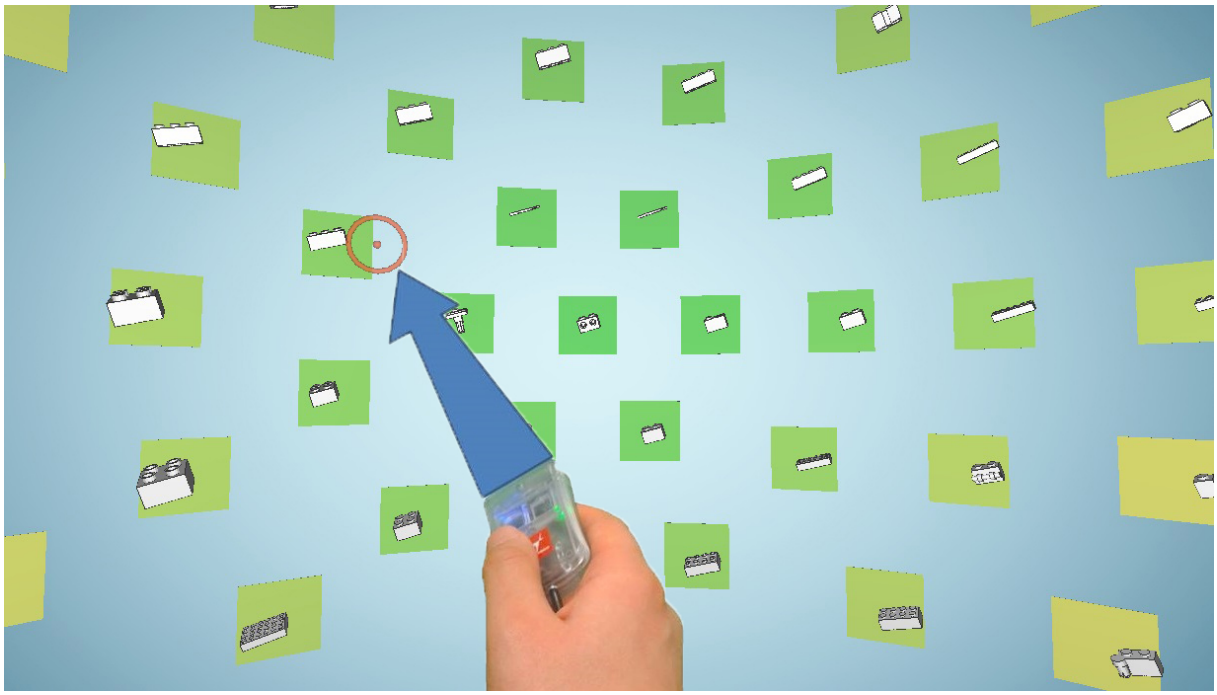

Figura 4.12: Controlo da mira que permite seleccionar objectos ao longo dos resultados.

tos apresentados, como se encontra representado na Figura 4.12. Deste modo, o utilizador consegue controlar a mira de uma forma simples, matendo sempre a noção de onde se encontra o dispositivo, ao contrário do verificado no Leap Motion. A precisão deste dispositivo faz com que a peça pretendida seja seleccionada de forma exacta, mesmo quando os objectos seleccionados se encontram distantes do ponto de vista inicial.

Ao pressionar o botão esquerdo, o objecto é seleccionado para uma melhor análise, o botão direito pode ser usado para cancelar essa mesma tarefa. A análise do objecto consistia numa aproximação deste para frente do ponto de vista do utilizador e em seguida rodá-lo livremente consoante a rotação realizada com o dispositivo. A Figura 4.13 ilustra o mapeamento directo realizado da rotação SpacePoint Fusion com o objecto a ser analisado. Esta análise permitia aos utilizadores terem uma vista

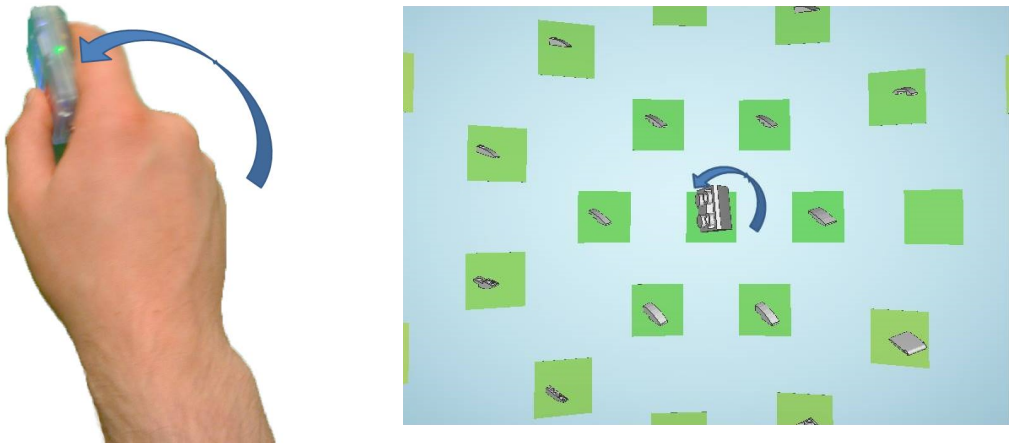

Figura 4.13: Rotação de objecto de acordo com a rotação realizada pelo SpacePoint Fusion.

mais pormenorizada dos objectos, revelando detalhes não visíveis das peças quando apresentadas em conjunto com as outras. Esta análise também conferia uma melhor noção do tamanho da peça.

De modo a escolhermos qual os dispositivos de interacção utilizar para o protótipo final, realizámos uma pequena análise qualitativa com utilizadores para tirar sua preferência. Realizámos esta análise com 8 utilizadores, onde dos quais 50% nunca tinha utilizado ambos os dispositivos, 25% ambos os dispositivos e os restantes 25% já tinha experiência com apenas um deles. Foi pedido aos participantes que seleccionassem um conjunto de objectos que apareciam com cor diferente o mais rápido possível e no final mencionar a sua preferência relativamente aos dispositivos experimentados. Verificámos que não só estes preferiram a interacção com o SpacePoint Fusion, como todos os utilizadores foram mais rápidos com este dispositivo. Esta análise encontra-se mais detalhada em Apêndice B.

4.5 Construção do modelo

No nosso protótipo implementámos um modo onde o utilizador pode criar modelos com as peças pesquisadas. Este modo é o primeiro que o utilizador encontra no protótipo, sendo enviado para a vista de resultados quando realizar algum pedido. Foi adicionada uma grelha nesta vista para o utilizador poder colocar os objectos que procurou. Os objectos podem ainda ser colocados sobre outros objectos, sendo usado um raio que intersecta a grelha ou outros objectos, com a direcção dada pelo SpacePoint Fusion, para definir a posição dos objectos a colocar.

Na Figura 4.14 está ilustrada a vista de modelação, na qual o utilizador se encontra a mudar a posição de uma peça. Para seleccionar um objecto o utilizador terá que controlar a mira disponível e quando esta estiver a sobre a peça o utilizador poderá seleccioná-lo pressionando o botão esquerdo

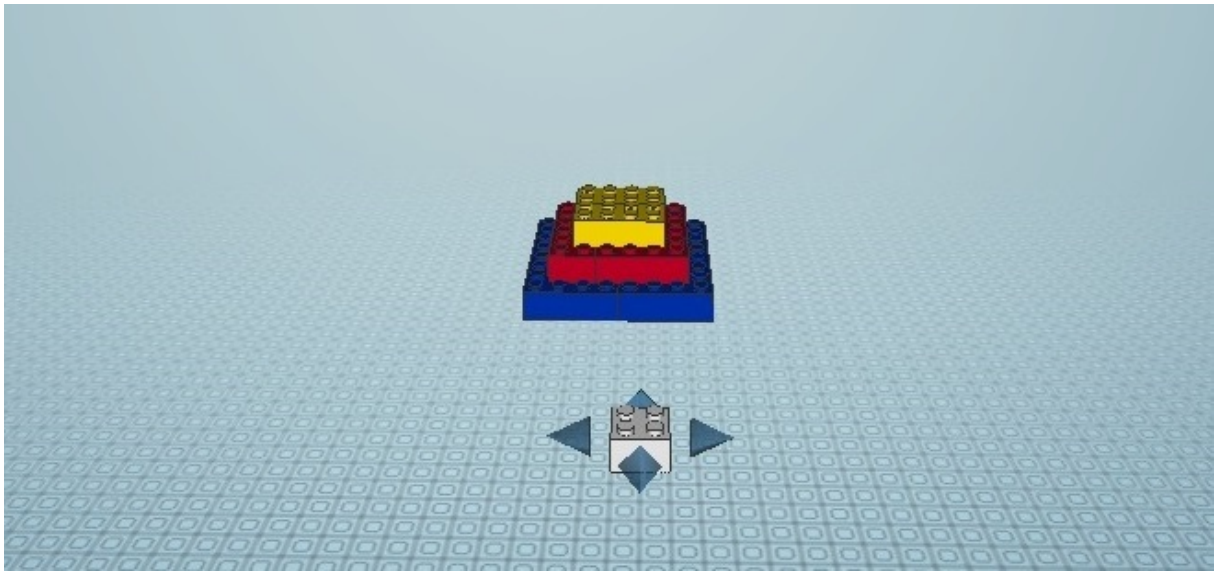

Figura 4.14: Vista de modelação do protótipo LSpeakIt onde o utilizador encontra-se mover uma peça.

do SpacePoint Fusion. Posteriormente, o utilizador poderá colocá-lo novamente pressionando o botão esquerdo, enquanto que o botão direito neste ambiente é usado para o remover.

4.6 Sumário

Este capítulo começou por apresentar uma visão global do protótipo LSpeakIt e da sua arquitectura. O protótipo foi desenvolvido tendo em conta o *toy-problem* de LEGO, onde o utilizador pode pesquisar peças de LEGO, descrevendo-as verbalmente. Posteriormente apresentamos o processo de indexação dos objectos utilizados no LSpeakIt, bem como, a forma de se realizarem interrogações aos índices. Seguidamente, apresentamos como são reconhecidas as descrições verbais, assim como as potencialidades das interrogações verbais no nosso protótipo. Ainda neste capítulo, é abordada a forma de refinar os resultados apresentados, adicionando filtragens ou exclusões à pesquisa. Neste sentido, o utilizador pode ver e interagir com os resultados apresentados, bem como analisá-los com mais detalhe, averiguando possíveis partes ocultas dos objectos.

Capítulo 5

Avaliação

Ao longo desta dissertação, envolveram-se utilizadores em diferentes etapas, captando as suas preferências e opiniões. Deste modo, foi possível ir ao encontro das suas expectativas, permitindo assim identificar a interacção mais natural possível para o objectivo que nos propusemos. Contou-se com a participação de utilizadores em três etapas: Estudo das descrições, Avaliação dos modos de visualização dos resultados e Validação da solução proposta.

Primeiramente, foi realizado um estudo para compreender como é que os participantes descreviam as peças apresentadas. Com este estudo, já apresentado no Capítulo 3, foi possível verificar que os utilizadores preferiram descrições verbais a gestuais. Este estudo permitiu-nos também ver quais as descrições utilizadas pelos participantes. De seguida, numa primeira versão do protótipo, foi testado como é que se deveriam apresentar e interagir com os resultados. Aqui foram testados quatro modos de visualizar os resultados, comparando-os com uma apresentação tradicional.

Após estas avaliações foram feitas alterações no protótipo de forma a ter um único modo de apresentar os resultados e ainda foram incluídas as interrogações verbais. Esta versão do protótipo foi sujeita a um processo de validação junto dos utilizadores, na qual foi avaliada relativamente à sua facilidade de interacção, tendo sido comparada com uma aplicação comercial de pesquisa de peças e construção de modelos LEGO.

5.1 Avaliação de modos para a visualização dos resultados

A primeira versão do protótipo tinha como objectivo avaliar as quatro visualizações de resultados implementados, onde os resultados são apresentados em ambiente imersivo. Os modos de visualização, **Rectangular**, **Quadrático**, **Cilíndrico** e **Esférico**, encontram-se apresentados na Secção 4.4.1, assim como a respectiva interacção (apresentada na Secção 4.4.2) para cada um dos modos referidos com o

dispositivo Leap Motion.

Foi inserido neste protótipo um modo de visualização **Tradicional**, onde os objectos eram representados por *thumbnails*. A interacção com este modo consistia na utilização de uma barra para mover os resultados para cima e para baixo através do rato, ou apenas utilizando o *scroll* do próprio rato.

5.1.1 Metodologia

As sessões de testes destinadas a avaliar os modos de apresentação dos resultados num ambiente imersivo tiveram uma duração de vinte minutos. Cada sessão estava dividida em três etapas, apresentadas na Tabela 5.1.

Etapa	Descrição	Duração
1	Introdução à experiência	5 minutos
2	Execução da tarefa	5 x 3 minutos
3	Preenchimento do questionário	5 minutos

Tabela 5.1: Etapas de avaliação dos modos de visualização.

1. **Introdução à experiência e modos de visualização dos resultados**
   Ao iniciar a sessão de testes com um novo utilizador, foi explicada a metodologia da sessão, seguida por uma apresentação e explicação geral do protótipo. Assim como os diferentes modos de visualização de resultados. Esta primeira etapa terminava com um período de adaptação ao protótipo, onde eram colocados objectos aleatoriamente para que o utilizador tivesse oportunidade de interagir com os mesmos. Esta adaptação servia para introduzir o ambiente imersivo, bem como os gestos que os utilizadores poderiam realizar.

2. **Tarefas a executar**
   Após explicação dos modos de visualização e do período de adaptação, foi fornecida uma peça de LEGO física ao utilizador. Em seguida foi solicitado que encontrasse essa peça no nosso protótipo. Este procedimento foi repetido para cada modo de visualização. É possível ver quais as peças utilizadas em cada sessão na Figura 5.1. Antes dos utilizadores realizarem a primeira pesquisa, foi-lhes solicitado que priorizassem a forma da peça em detrimento da cor, pois nos resultados apresentados todas as peças teriam a mesma cor. A peça física foi escolhida aleatoriamente de modo que os utilizadores procurassem peças diferentes para cada modo de visualização. A ordem em que os modos de visualização foram avaliados por cada utilizador foi também aleatória para evitar resultados tendenciosos devido ao efeito de aprendizagem.

3. **Questionário**
   Terminadas as tarefas pedimos aos utilizadores que preenchessem um questionário, onde classificavam as várias técnicas segundo uma escala de Likert de cinco valores, bem como um conjunto de questões no intuito de identificar o perfil do utilizar. Este questionário encontra-se disponível

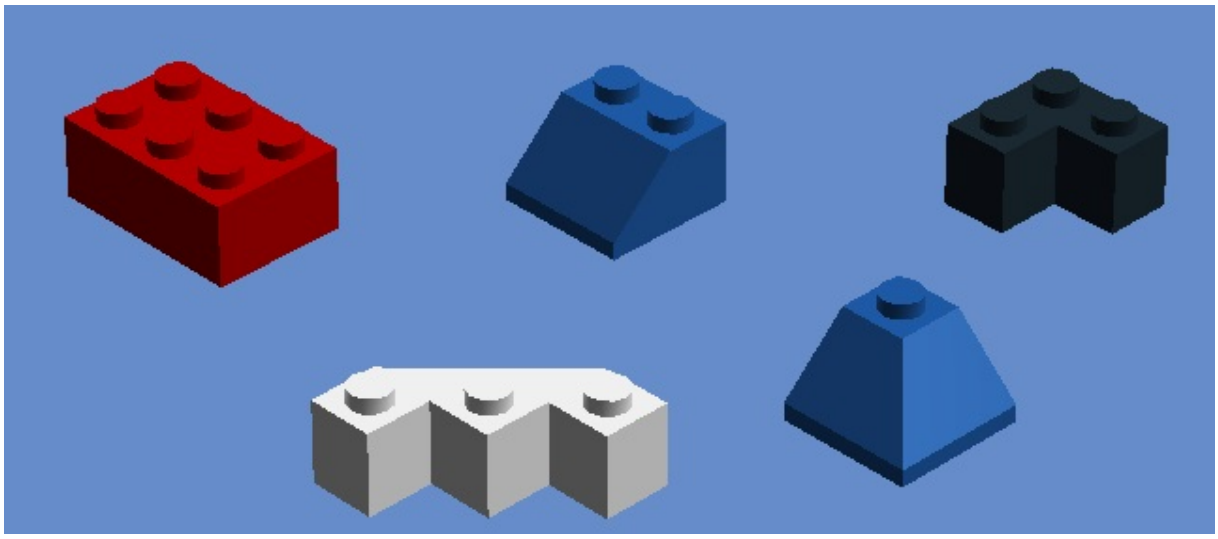

Figura 5.1: Peças de lego utilizadas para os utilizadores procurarem nos modos de visualização.

para consulta no Apêndice C.1.

5.1.2 Aparato e participantes

Para avaliar os modos de visualização implementados, realizámos a experiência num ambiente controlado, sem influências externas, seguindo a metodologia apresentada anteriormente. Esta experiência foi executada numa primeira versão do nosso protótipo com os diversos modos, utilizando os Oculus Rift e Leap Motion tal como é possível observar na Figura 5.2. Foi ainda utilizado um rato e ecrã, exclusivamente para o modo **Tradicional**.

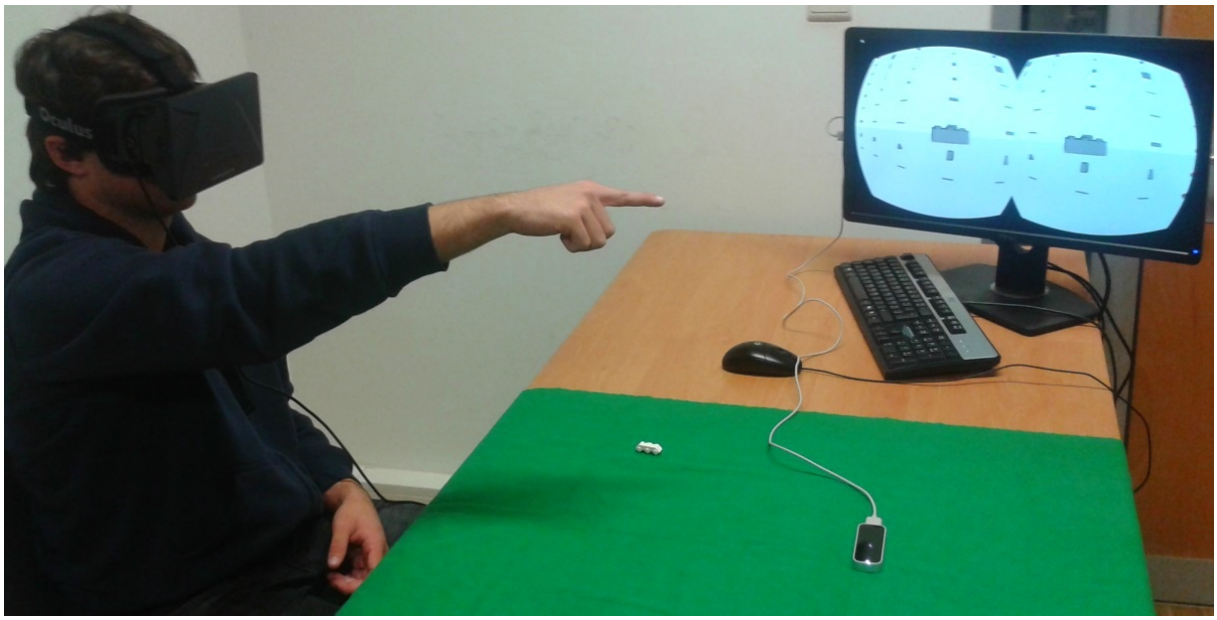

Figura 5.2: Participante durante a sessão, a executar o gesto de apontar.

Esta avaliação contou com 29 participantes, 22 do sexo masculino e 7 do sexo feminino. Estes tinham idades compreendidas entre os 17 e os 40 anos, sendo que a grande maioria situava entre os 18 e os 29 anos. Cerca de 90% eram estudantes de Engenharia Informática. Todos os participantes consideravam-se experientes na utilização de motores de busca de imagens e em explorar *thumbnails* em ecrãs. 20% dos participantes afirmaram ter experimentado sistemas de pesquisa de objectos 3D, como Google 3D Wharehouse. Cerca de 50% dos participantes tinham experiência prévia com ambientes virtuais estereoscópicos. Apenas 27% dos participantes já tinham experimentado óculos HMD em pequenas demonstrações e cerca de 10% já tinham experimentado o Leap Motion. No entanto todos os que afirmaram ter experimentado os óculos HMD e o Leap Motion, apenas o fizeram com demonstrações de jogos.

5.1.3 Análise dos resultados

Com o objectivo de perceber qual era o modo de visualização mais fácil de utilizar e simultaneamente qual o preferido pelos utilizadores. Nesta última, o tempo foi cronometrado a fim de conhecer o tempo levado para concluir a tarefa em cada um dos modos. Após a conclusão das tarefas, os utilizadores também preencheram um breve questionário com vista a avaliarem os diferentes modos de visualização. Desta forma será apresentada em primeiro lugar uma análise quantitativa sobre os tempos ao realizar as pesquisas e em seguida, uma análise qualitativa com base nos questionários e, por fim, um conjunto de observações retiradas durante as sessões. Os valores medidos estão apresentados no Apêndice C.2.

Análise quantitativa

Para cada modo de visualização foi cronometrado o tempo despendido pelos utilizadores para realizarem a procura. Analisando os tempos obtidos pelo teste *One-Way ANOVA*, não foram encontradas diferenças estatisticamente relevantes entre os cinco modos de visualização.

O tempo despendido nas procuras para cada modo de visualização está apresentado no gráfico da Figura 5.3. Dado os utilizadores considerarem-se experientes na utilização do modo **Tradicional**, isto é utilizando rato e ecrãs tradicionais na procura de conteúdo multimédia, por exemplo imagens, foi constatado como o modo mais rápido. No caso dos modos **Rectangular** e **Quadrático**, estes estiveram muito próximos da abordagem tradicional, isto pode-se dever à semelhança da estrutura apresentada, ou seja, os objectos estão organizados numa grelha. No modo **Cilíndrico**, apesar de ter o tempo mais rápido, houve alguma dispersão entre os utilizadores, estando a sua performance em média um pouco mais baixa em relação ao modo **Esférico**.

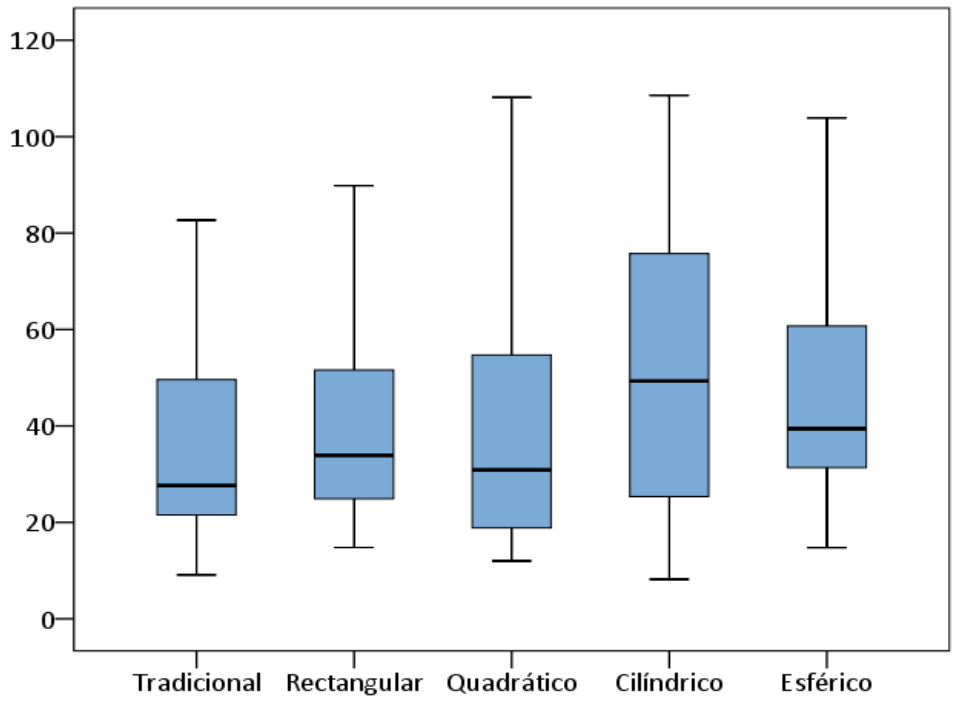

Figura 5.3: Tempo despendido nas procuras de cada modo de visualização.

Análise qualitativa

No questionário foi solicitado aos utilizadores que classificassem o modo de visualização utilizando uma escala de Likert de 5 valores (1 - muito mau e 5 - muito bom), no que diz respeito ao quão divertido e fácil foi utilizar. As respectivas classificações estão apresentadas na Tabela 5.2. O teste de Wilcoxon signed-ranks foi utilizado para verificar as diferenças estatisticamente relevantes.

Relativamente à facilidade de utilização os utilizadores não chegaram a um consenso de qual o modo de seria mais fácil de usar. No geral os utilizadores consideraram que todos os modos de visualização eram fáceis de utilizar, mesmo havendo alguma variação de opinião. Maioria dos utilizadores considerou que o modo de visualização **Tradicional** era mais fácil de utilizar.

No entanto no que diz respeito à diversão, os utilizadores concordaram que o **Tradicional** era, entre todos, o menos divertido (**Rectangular**: Z=-3.744,p=0.000; **Quadrático**: Z=-3.430,p=0.001; **Cilíndrico**:

Modo de visualização	Facilidade	Diversão
Tradicional	5 (2)	3 (2)
Rectangular	4 (1)	4 (1)
Quadrático	4 (1)	4 (1)
Cilíndrico	4 (1)	4.5 (1)
Esférico	4 (2)	5 (1)

Tabela 5.2: Classificações para cada modo de visualização (Mediana, Amplitude interquartil).

Z=-3.542,p=0.000; **Esférico**: Z=-2.863,p=0.004). Adicionalmente os utilizadores concordaram que o modo **Cilíndrico** era mais divertido que o modo **Quadrático** (Z=-2.236,p=0.025). O modo **Cilíndrico** e **Esférico**, foram considerados os modos mais divertidos pelos utilizadores. Os comentários dos utilizadores revelaram preferência nestes dois modos de visualização, nos quais são envolvidos por objectos em seu redor.

Observações

Ao longo das sessões de testes foi observado que o modo de visualização **Tradicional** trazia muita confusão aos utilizadores. Isto devia-se ao facto dos objectos estarem representados por uma imagem que fazia com que os utilizadores muitas vezes ficassem na dúvida se seria a peça correcta. Esta representação de imagens fez com que os utilizadores percorressem a lista mais que uma vez à procura do objecto causando alguma frustração aos utilizadores. Isto não foi verificado nos outros modos, mesmo quando estes levavam mais tempo.

Alguns dos comentários feitos pelos utilizadores mostraram que os modos de visualização imersivos facilitavam a apreensão de um elevado número de objectos, principalmente no modo **Cilíndrico**, pois simplesmente teriam que rodar a cabeça para verem os resultados, não necessitando de usar muitas vezes os gestos. Verificámos que apesar de os utilizadores saberem em que zona se encontravam os objectos mais relevantes houve alguma dificuldade de os situarem novamente, quando moviam os objectos. Isto observou-se sobretudo nos modos **Esférico** e **Quadrático**. Alguns utilizadores sugeriram que deveria de haver algo que os permitisse facilmente perceber onde estariam os objectos mais relevantes.

Relativamente aos gestos, houve alguma dificuldade da parte dos utilizadores, principalmente no gesto de apontar. Isto deveu-se às limitações tecnológicas do Leap Motion e da sua combinação com os Oculus Rift, pois fazia com que o utilizador perdesse a noção da volume que teria para realizar os gestos. Contudo, os utilizadores mostraram que este gesto era importante, uma vez que com ele conseguiam uma melhor percepção dos objectos. No entanto, os participantes preferiam serem eles próprios a controlar o objecto em vez deste rodar automaticamente. Estes comentários foram tidos em consideração no desenvolvimento da iteração seguinte do protótipo.

5.2 Validação da solução adoptada

Com o objectivo de validar a solução adoptada para pesquisa de objectos tridimensionais num ambiente imersivo, foi levada a cabo uma nova sessão de avaliação com utilizadores. Nesta sessão foi comparada a versão final do protótipo LSpeakIt, apresentada no Capítulo 4, com uma aplicação comercial de pesquisa de peças e construção de modelos de LEGO, o Lego Digital Designer (LDD), de modo a

verificar adequalibilidade e eficiência da nossa solução.

5.2.1 Aplicações testadas

A aplicação utilizada como base de comparação com o LSpeakIt, foi o LDD. Esta aplicação recorre ao paradigma *Windows Icons Menus and Pointing devices* (WIMP) e utiliza um modo de visualização de resultados tradicional, ou seja uma grelha de resultados onde os objectos são representados por *thumbnails*. No entanto, o LDD contém a lista de objectos dividida em categorias tal como podemos ver na Figura 5.4.

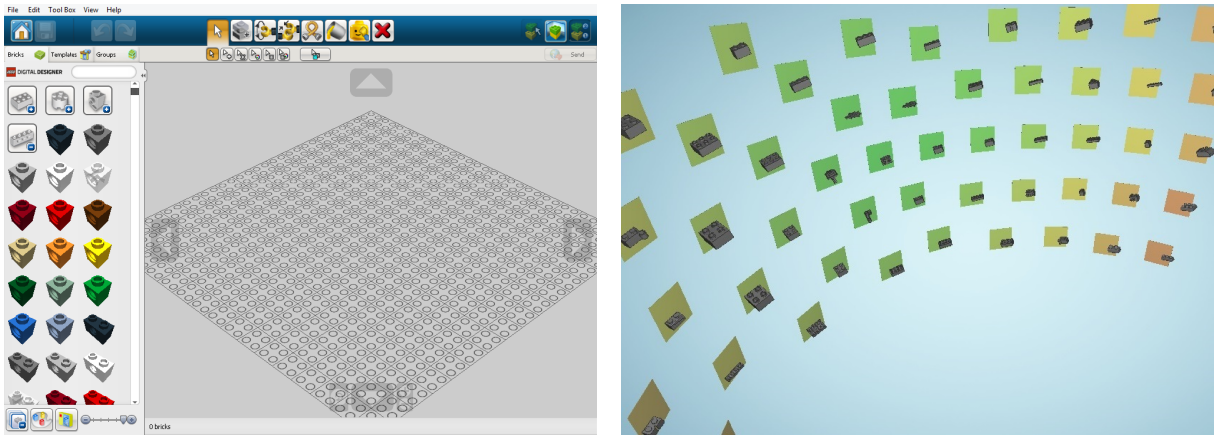

Figura 5.4: Aplicação comercial LDD à esquerda e LSpeakIt à direita.

Nesta fase foi utilizada a versão final do LSpeakIt, apresentada no Capítulo 4, onde os participantes podiam fazer interrogações por fala (Secção 4.3), e a lista de resultados era ilustrada num ambiente imersivo com os resultados envolvendo o utilizador (Secção 4.4).

5.2.2 Metodologia

As sessões de testes destinadas a validar a solução proposta tiveram uma duração de cinquenta minutos, estando cada sessão dividida em três etapas (Tabela 5.3).

Etapa	Descrição	Duração
1	Preenchimento do questionário inicial	5 minutos
2	Explicação sobre a sessão de teste	5 minutos
3	Execução de tarefas	2 x 20 minutos

Tabela 5.3: Etapas das sessões de validação da solução proposta.

1. **Questionário inicial**

   Antes de iniciar a sessão, foi solicitado aos participantes o preenchimento de um questionário, o qual está disponível em Apêndice D.1. Este tinha como objectivo identificar o perfil relativamente à

sua experiência com sistemas de pesquisa de conteúdo multimédia, assim como com pesquisas verbais e com dispositivos de interacção e visualização.

2. **Explicação sobre a sessão de teste**
   Após o preenchimento do questionário, ocorreu uma breve explicação sobre o que iria consistir o resto da sessão de teste. Eram também apresentados a aplicação LDD e o nosso protótipo, com os quais teriam de interagir.

3. **Avaliação de protótipo**
   Tal como mencionado anteriormente, o nosso teste pretendia avaliar o no nosso protótipo comparando-o com uma aplicação comercial. Nesta avaliação, foi seleccionado aleatoriamente o sistema pelo qual os utilizadores iriam começar, sendo para ambos os casos realizado o seguinte processo:

   (a) **Habituação ao sistema**
   Após a escolha da primeira aplicação foi demonstrado um exemplo de como se interagia com ela. De seguida foi apresentada uma peça física ao participante e solicitou-se ao mesmo que experimentasse o sistema de modo a conseguir encontrá-la, sendo para todos os participantes apresentadas as mesmas peças, tanto na situação de exemplo (Figura 5.5) como na situação de treino (Figura 5.6).

Figura 5.5: Peça utilizada como exemplo.

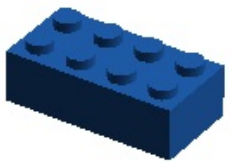

Figura 5.6: Peça utilizado como treino.

   (b) **Tarefa**
   Terminado o treino do utilizador, foi solicitado que este procurasse um conjunto de oito peças, ilustradas na Figura 5.7, na aplicação. Ao ser fornecida uma peça de cada vez ao participante, pedia-se, em primeiro lugar, que este a analisasse e em seguida a procurasse. As oito peças seleccionadas foram sempre as mesmas para todos os participantes e a sua ordem aleatória. No entanto assegurou-se que dois participantes diferentes não realizavam a mesma sequência, para evitar influência da ordem das peças a serem apresentadas.

   (c) **Preenchimento de questionário**
   No final de todas as pesquisas para a aplicação a testar, foi solicitado ao utilizador que preenchesse um questionário, no qual classificava a aplicação segundo diversos parâmetros, com uma escala de Likert de quatro valores. Os questionários encontram-se no Apêndice D.2. O utilizador era informado que poderia repensar e alterar sempre que quisesse as suas cota-

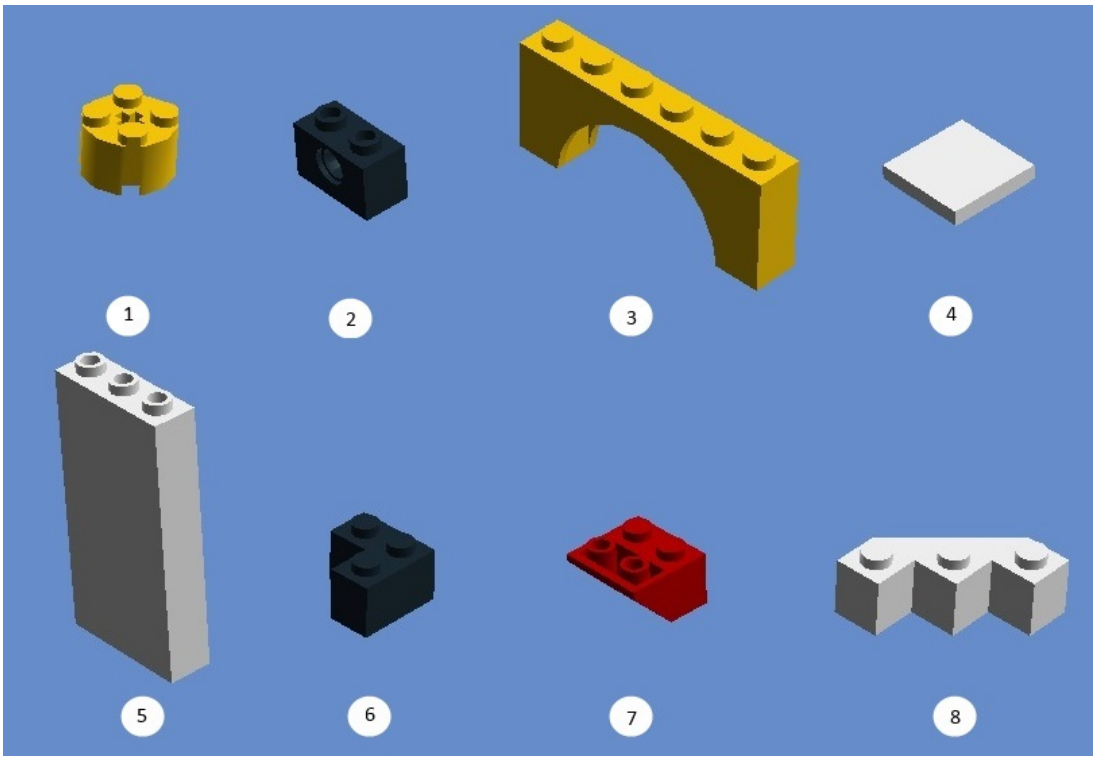

Figura 5.7: Conjunto de peças utilizadas para a tarefa.

ções da aplicação anterior, se achasse pertinente.

5.2.3 Aparato e participantes

As sessões de testes foram realizadas num ambiente controlado, sem influências externas, seguindo a metodologia apresentada. Esta experiências foram executadas utilizando rato e ecrã tradicional, no caso do LDD, e, no caso do nosso protótipo foram utilizados o SpacePoint Fusion para interagir com os resultados, os Oculus Rift para a visualização dos mesmos e um *HeadSet* para a interacção verbal, como visível na Figura 5.8.

Figura 5.8: Participantes na sessão de validação, a visualizar listas de resultados no LDD e no LSpeakIt.

Esta avaliação foi realizada por 20 utilizadores, 16 do sexo masculino e 4 do sexo feminino. Estes tinham idades compreendidas entre os 18 e 50 anos, com a maioria entre os 24 e os 29 anos (70%). Grande parte dos participantes viveram mais tempo no distrito de Lisboa, tendo um sotaque semelhante entre si. Relativamente à sua experiência, nenhum deles tinha utilizado o LDD. No entanto todos eram utilizadores experientes em sistemas de pesquisa de imagem, utilizando-os, pelo menos, semanalmente. Cerca de 30% dos participantes já tinham utilizado motores de busca de objectos 3D, nos quais os objectos são representados por *thumbnails*. 60% dos utilizadores já tinham utilizado sistemas de pesquisa por fala, no entanto comentários dos utilizadores relevaram que não utilizaram mais que uma vez. Relativamente à sua experiência com estereoscopia, 60% afirmou já ter experimentado, cerca de 50% já experimentou óculos HMD em demonstrações de jogos.

5.2.4 Análise dos resultados

Para avaliar o nosso protótipo cronometrámos o tempo despendido para cada uma das oito procuras, e contámos o número de peças erradas seleccionadas pelos participantes até acertarem na peça correcta. Após as procuras estarem realizadas solicitámos aos utilizadores que preenchessem um breve questionário para avaliar o sistema testado, com base nos seguintes parâmetros: facilidade de visualização dos resultados, facilidade de utilização e diversão. Os resultados obtidos da avaliação encontram-se no Apêndice D.3. Em seguida iremos apresentar uma análise quantitativa, uma análise qualitativa dos questionários e, por fim, observações que realizámos ao longo das sessões.

Análise quantitativa

De modo a analisarmos os testes em termos de tempo foi utilizado o teste de Wilcoxon signed-ranks. O qual podemos verificar o diferenças estatisticamente relevantes (Z=-3.509, p=0.000) na qual o LSpeakIt mostrou ser mais rápido. A Figura 5.9 apresenta os tempos das procuras em ambos os sistemas, onde podemos ver a diferença existente no tempo despendido nas procuras de ambos os sistemas

Para o caso do número peças erradas, recorremos novamente ao teste de Wilcoxon signed-ranks. Aqui foi claro que o LSpeakIt obteve melhor resultado (Z=-6.697, p=0.000). Através da Figura 5.10, podemos observar que o número de erros existentes no nosso protótipo foi bastante inferior em relação ao LDD. Comentários dos utilizadores revelaram que muitas vezes não conseguiam distinguir as peças no LDD, ao contrário do que foi constatado no LSpeakIt, afectando assim o número de peças erradas.

Análise qualitativa

No questionário foi solicitado aos utilizadores que classificassem os sistemas utilizando uma escala de Likert de quatro pontos (1 - muito mau e 4 - muito bom), no que diz respeito ao aspecto da diver-

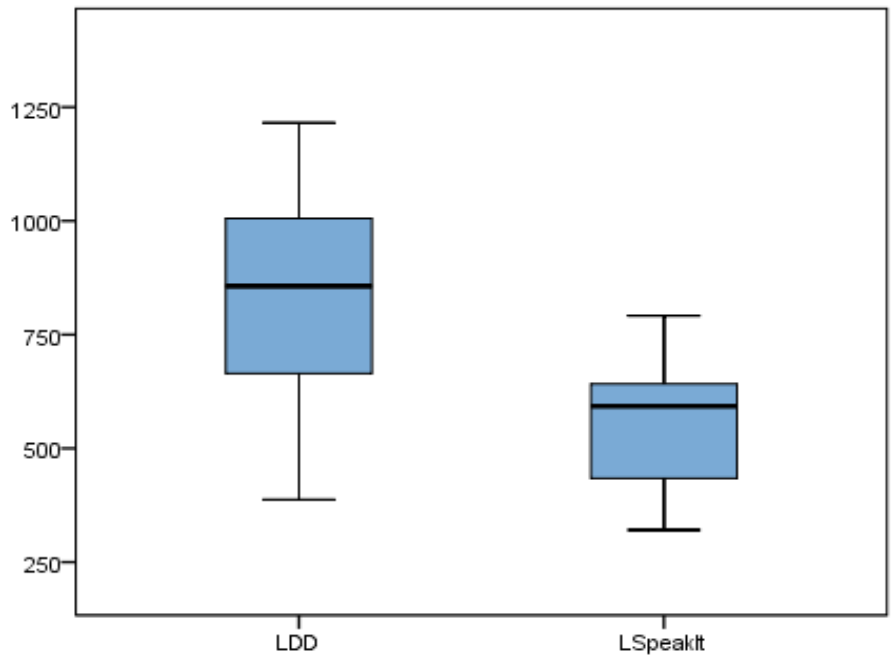

Figura 5.9: Soma dos tempos em segundos das procuras nas duas aplicações, LDD e LSpeakIt.

são, facilidade de visualizar as peças e facilidade de utilização. As respectivas classificações estão apresentadas na Tabela 5.4. O teste de Wilcoxon signed-ranks foi utilizado para verificar as diferenças estatisticamente relevantes.

Houve consenso entre os utilizadores que o **LSpeakIt** foi mais fácil de utilizar (Z=-2.441, p=0.015), revelando que mesmo com utilizadores inexperientes em termos de pesquisas por fala, estes conseguiram com facilidade utilizar o sistema. Relativamente à facilidade de visualização dos objectos, os utilizadores concordaram que o **LSpeakIt** era o preferido (Z=-2.780, p =0.005). Quanto à diversão o **LSpeakIt** foi novamente o preferido (Z=-3.626, p=0.000).

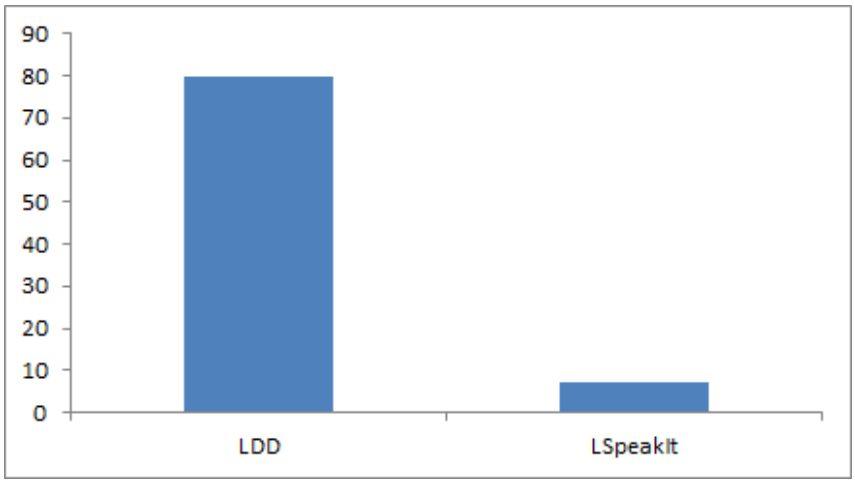

Figura 5.10: Soma dos totais de peças erradas das procuras nas duas aplicações.

Sistema	LDD	LSpeakIt
Facilidade de utilização	2 (1)	3 (1)
Facilidade de visualização	2 (1)	3 (0)
Diversão	2 (1)	4 (0)

Tabela 5.4: Classificação para cada aplicação (Mediana, Amplitude interquartil).

Observações

Ao longo da avaliação os utilizadores demonstraram que a divisão por categorias do LDD foi bastante útil na procura. No entanto, esta mesma divisão por categorias falhou em algumas peças como no caso da peça 8 da Figura 5.7, onde a imagem que representava a categoria não era de todo semelhante à peça a procurar. Este facto levou a que houvesse uma dificuldade acrescida na procura, fazendo com que os utilizadores abrissem muitas vezes todas as categorias para facilitar as próximas procuras. Enquanto na peça número 1 da Figura 5.7 um a imagem da categoria era a própria peça a procurar, o que facilitava muito a sua procura. Verificou-se que os utilizadores que procuravam esta peça primeiro do que a peça 8 da Figura 5.7, faziam mais comentários demonstrando uma maior frustração na sua procura no LDD. Este tipo de comentários não se verificaram em nenhum dos utilizadores ao testarem LSpeakIt, mesmo nas procuras mais longas.

A grande maioria dos utilizadores mencionavam que não tinham a certeza quando seleccionavam uma peça no LDD, ao contrário do que aconteceu com o LSpeakIt. Isto constatou-se principalmente em peças que continham partes não visíveis ou quando não era possível distinguir o seu tamanho, ou seja, quando as peças não tinham pinos como no caso da peça quatro da Figura 5.7. Esta mesma peça foi confundida bastantes vezes no LDD por uma peça semelhante, mas com dimensões 1 x 1.

Comentários dos participantes mostraram que o LSpeakIt era muito fácil de utilizar na procura dos objectos pelas suas dimensões. Contudo, alguns dos participantes divergiram na forma como descreviam as dimensões, havendo erros como mencionar uma peça com 2 pinos de largura e 4 de comprimento, como uma peça 2 x 8. Geralmente este erro era detectado pelos participantes quando recebiam o *feeback* sonoro do sistema. Verificámos que as descrições de uma peça por adjectivos era algo que ocorria quando os utilizadores já estavam à algum tempo a utilizar o sistema. Porém, no início existia a dúvida nos utilizadores sobre quais os adjectivos que poderiam utilizar.

Relativamente aos pedidos por fala no LSpeakIt, os utilizadores realizaram um total de 782 pedidos, dos quais cerca de 32% foram refinamentos. Apesar de haver uma clara preferência para realizarem novos pedidos, os utilizadores consideraram que os refinamentos foram úteis e que facilitavam a pesquisa. Dentro dos refinamentos houve uma preferência dos utilizadores pela filtragem dos resultados, tendo sido cerca de 55% dos seus refinamentos. Por outro lado, a exclusão foi a menos utilizada tendo sido apenas 9% dos refinamentos. Comentários dos utilizadores mostraram que muitas vezes não ne-

cessitavam de usar a exclusão porque a filtragem cobria a necessidade de excluir resultados. Por fim, a procura por exemplo correspondeu a cerca de 36% dos refinamentos, era utilizada não só quando os utilizadores encontravam peças semelhantes, mas também como auxilio quando estes não sabiam como descrever as peças.

5.3 Sumário

Neste capítulo foram apresentados os dois momentos de avaliação sobre o trabalho desenvolvido. O primeiro focou-se nos modos de visualização, isto é, como é que os utilizadores poderiam visualizar os resultados num ambiente imersivo e a sua comparação com um sistema tradicional. Comparando as quatro abordagens imersivas com o modo Tradicional, verificámos que os utilizadores foram quase tão rápidos utilizando os modos imersivos, apesar da diferença na experiência, havendo inclusivamente casos em que os modos imersivos foram mais rápidos que a tradicional. Relativamente à opinião dos participantes, houve uma clara preferência pelos modos imersivos. Dos modos de visualização dos resultados destacaram-se o Cilíndrico e o Esférico, onde comentários dos utilizadores mostraram que serem envolvidos pelos objectos foi a situação preferida, uma vez que transmitia uma boa noção dos objectos existentes, assim como permitia facilmente observá-los com uma simples rotação da cabeça.

Após estes primeiros testes foi desenvolvido, no protótipo final uma combinação dos modos de visualização Cilíndrico e Esférico, juntamente com interrogações verbais. Para comprovar adequabilidade deste conceito aplicado no protótipo final, este foi validado por utilizadores, comparando-o com a aplicação comercial Lego Digital Designer. Com base nos resultados da validação podemos concluir que o LSpeakIt foi capaz de competir com uma aplicação comercial que usa um modo de visualização tradicional, no qual os utilizadores são experientes na sua utilização. De salientar que houve uma clara preferência dos utilizadores pelo LSpeakIt em termos de diversão, facilidade de visualização dos objectos e na utilização do sistema de procura por fala.

Capítulo 6

Conclusões e Trabalho Futuro

O avanço tecnológico dos últimos anos permitiu que a criação de conteúdo tridimensional virtual fosse cada vez mais acessível. Como consequência, existem, cada vez mais, grandes colecções deste conteúdo, tornando a tarefa de encontrar um objecto específico um desafio. As soluções existentes partem de uma interface semelhante a outros sistemas de pesquisa convencionais, tais como motores de pesquisa de imagem. As interfaces destes sistemas de procura descartam os avanços em termos de interacção e visualização que têm surgido, apresentando os resultados numa grelha de *thumbnails* e realizando as procuras por texto. Contudo, existem alguns trabalhos que tentam abordar diferentes formas de visualizar os elementos da colecção ou resultados de pesquisas recorrendo a ambientes virtuais imersivos. No entanto, estes últimos normalmente realizam procuras tendo como base um objecto 3D como exemplo para a sua procura, não havendo um solução para os utilizadores descreverem estes objectos num ambiente imersivo sem terem um objecto semelhante à partida.

Procurou-se, no trabalho desenvolvido nesta dissertação, realizar uma interface para pesquisa de objectos 3D que seja natural para o utilizador. Pretendia-se encontrar uma solução que tirasse partido dos avanços do reconhecimento de fala, bem como dos novos paradigmas de interacção que têm surgido com o aparecimento de novo hardware de baixo custo para interacção e visualização imersiva. Para além disso, pretendia-se oferecer uma interface de pesquisa capaz de ser executada por utilizadores sem experiência em ambientes virtuais imersivos.

Para o efeito, começámos por realizar um estudo preliminar com utilizadores, com o intuito de compreender como é que estes descreviam os objectos no mundo físico quando não recorrem a qualquer instrumento. Isto é, se utilizam descrições verbais, gestuais, ou uma combinação de ambas. Neste estudo utilizámos o *toy-problem* de LEGO, onde foi solicitado aos participantes que descrevessem as peças que necessitavam para a construção de um modelo a outro participante. Verificámos que os participantes preferiam as descrições verbais, as quais foram sempre usadas quer acompanhadas ou não por gestos. Adicionalmente, verificámos uma preferência nas descrições através de dimensões e metáforas.

Com base nas descrições recolhidas no estudo preliminar desenvolvemos o nosso protótipo (LSpeakIt) abordando o *toy-problem* de procura peças de LEGO como prova de conceito. Desenvolvemos assim um sistema de indexação utilizando as descrições recolhidas bem como descrições de forma que foram realizadas por um descritor, mas mantidas com uma representação textual. Tomando partido de um sistema imersivo, desenvolvemos quatro modos de visualização no nosso protótipo. Estes foram comparados, numa avaliação com utilizadores, entre si e com um modo de visualização tradicional usando rato e ecrã. Com esta avaliação concluiu-se que mesmo os utilizadores com pouca experiência em ambientes imersivos conseguiam visualizar e controlar os resultados facilmente. Comprovou-se ainda que houve uma clara preferência pelos modos de visualização em que os utilizadores eram envolvidos com os resultados em seu redor, os modos Cilíndrico e Esférico.

Tendo em conta estes resultados, a versão mais recente do protótipo inclui um novo modo de visualização que integra ambas as vantagens dos modos preferidos pelos utilizadores. Este novo modo permite que os utilizadores vejam todos os resultados sem terem que rodar a cabeça mais de noventa graus para cada lado. Foi incluído ainda neste protótipo um reconhecedor de fala capaz de reconhecer as interrogações verbais dos utilizadores no contexto de LEGO. Esta versão foi sujeita a uma outra avaliação envolvendo utilizadores, no propósito de o comparar com uma aplicação de LEGO comercial, a qual recorre a uma solução por categorias com *thumbnails*. Quando comparado com uma aplicação cujo o paradigma de interacção os utilizadores estavam mais familiarizados, o LSpeakIt mostrou-se capaz de concorrer com esta aplicação em termos de utilização, destacando positivamente o modo de pesquisa, bem como a visualização dos resultados.

Como trabalho futuro consideramos que seria interessante investir não só no desenvolvimento de modelos de língua mais sofisticados, específicos para um sistema de pesquisa de conteúdo multimédia, bem como em gestores de diálogo com estratégias de recuperação de erros que permitam ao sistema dialogar com os utilizadores sem exigir que estes recomecem a pesquisa do início em situações de erro. O sistema desenvolvido pode ser de facto um excelente ponto de partida para a investigação neste domínio, assim como de estratégias de *"entrainment"* que procurem aumentar a empatia entre o sistema e o utilizador.

Adicionalmente, seria pertinente um estudo e desenvolvimento de agrupamentos dos resultados apresentados, ou até mesmo de toda a colecção, tal como verificado no Lego Digital Designer, recorrendo, no entanto, a mais que um objecto para os representar. Seria igualmente interessante verificar quais os resultados quando utilizando outros descritores de forma mais recentes.

Em termos de interacção seria relevante estudar alternativas num cenário em que o próprio utilizador se pudesse deslocar de pé. Os novos Oculus Rift SDK2, que estão a ser desenvolvidos neste momento, poderão ser uma solução para esta questão. Estes capturam a posição da cabeça do utilizador, contêm uma resolução de 1080p e uma resistência superior em termos do *"motion blur"*. Acreditamos que esta nova versão irá enriquecer a visualização dos resultados, podendo evitar possíveis tonturas que acontecem ocasionalmente com a versão corrente em cenários de movimento.

Por fim, apesar destas soluções terem sido aplicadas para o contexto do *toy-problem* de LEGO, acreditamos que estas são portáveis para outras áreas. Exemplos destas áreas poderão ser a indústria automóvel e a construção, entre outras. Com este trabalho mostrou-se que a visualização de objectos tridimensionais enriquece a percepção dos mesmos, assim como as interrogações verbais se adequam em ambientes imersivos.

Bibliografia

Giuseppe Amato and Pasquale Savino. Approximate similarity search in metric spaces using inverted files. In *Proceedings of the International Conference on Scalable Information Systems*, 2008.

Jürgen Assfalg, Alberto Del Bimbo, and Pietro Pala. Content-based retrieval of 3d models through curvature maps: a cbr approach exploiting media conversion. *Multimedia Tools and Applications*, 31 (1):29–50, 2006.

Marco Attene, Francesco Robbiano, Michela Spagnuolo, and Bianca Falcidieno. Characterization of 3d shape parts for semantic annotation. *Computer-Aided Design*, 41(10):756–763, 2009.

Ravin Balakrishnan, Thomas Baudel, Gordon Kurtenbach, and George Fitzmaurice. The Rockin'Mouse. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '97*, pages 311–318, New York, NY, USA, 1997.

Aryabrata Basu, Kyle Johnsen, and Kenneth Bogert. Poster: Evolution and usability of ubiquitous immersive 3d interfaces. In *3D User Interfaces (3DUI), 2013 IEEE Symposium on*, pages 131–132. IEEE, 2013.

Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1):107–117, 1998.

Carolina Cruz-Neira, Daniel J Sandin, Thomas A DeFanti, Robert V Kenyon, and John C Hart. The cave: audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6): 64–72, 1992.

Thomas A DeFanti, Daniel Acevedo, Richard A Ainsworth, Maxine D Brown, Steven Cutchin, Gregory Dawe, Kai-Uwe Doerr, Andrew Johnson, Chris Knox, Robert Kooima, et al. The future of the cave. *Central European Journal of Engineering*, 1(1):16–37, 2011.

Mathias Eitz, Ronald Richter, Tamy Boubekeur, Kristian Hildebrand, and Marc Alexa. Sketch-based shape retrieval. *ACM Transactions on Graphics*, 31(4):31, 2012.

Thomas Funkhouser, Patrick Min, Michael Kazhdan, Joyce Chen, Alex Halderman, David Dobkin, and David Jacobs. A search engine for 3d models. *ACM Transactions on Graphics (TOG)*, 22(1):83–105, 2003.

Maia Garau, Mel Slater, Vinoba Vinayagamoorthy, Andrea Brogni, Anthony Steed, and M. Angela Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 529–536, New York, NY, USA, 2003.

Claudio Gennaro, Giuseppe Amato, Paolo Bolettieri, and Pasquale Savino. An approach to content-based image retrieval based on the lucene search engine library. In *Research and Advanced Technology for Digital Libraries*, pages 55–66. Springer, 2010.

N Guarino and R Poli. Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human and Computer Studies*, 43(5):625–640, 1995.

Martin Hachet, Fabrice Decle, Sebastian Knödel, and Pascal Guitton. Navidget for 3D interaction: Camera positioning and further uses. volume 67, pages 225–236, 2009.

Michael T Heath. Scientific computing: an introductory survey, 2002.

Christian Holz and Andrew Wilson. Data miming: inferring spatial object descriptions from human gesture. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 811–820. ACM, 2011.

Maryam Kamvar and Doug Beeferman. Say what? why users choose to speak their web queries. In *Interspeech*, 2010.

Cheongjae Lee et al. Let's buy books: finding ebooks using voice search. In *2010 IEEE Spoken Language Technology Workshop*, pages 85–90, 2010.

Bo Li and Henry Johan. Sketch-based 3d model retrieval by incorporating 2d-3d alignment. *Multimedia tools and applications*, 65(3):363–385, 2013.

Yong-Jin Liu, Xi Luo, A. Joneja, Cui-Xia Ma, Xiao-Lan Fu, and Dawei Song. User-adaptive sketch-based 3-d cad model retrieval. *Automation Science and Engineering, IEEE Transactions on*, 10(3):783–795, 2013.

K Lou, K Ramani, and S Prabhakar. Content-based three-dimensional engineering shape search. pages 754–765, 2004.

Fabio Marton, Marco Agus, Enrico Gobbetti, Giovanni Pintore, and Marcos Balsa Rodriguez. Natural exploration of 3d massive models on large-scale light field displays using the fox proximal navigation technique. Elsevier, 2012.

Andrew Kachites McCallum. Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering. 1996.

Hugo Meinedo, Alberto Abad, T Pellegrini, I Trancoso, and J Neto. The L2F Broadcast News Speech Recognition System. pages 93–96, 2010.

Patrick Min, John A Halderman, Michael Kazhdan, and Thomas A Funkhouser. Early experiences with a 3d model search engine. In *Proceedings of the eighth international conference on 3D Web technology*. ACM, 2003.

Patrick Min, Michael Kazhdan, and Thomas Funkhouser. A comparison of text and shape matching for retrieval of online 3D models. pages 209–220, 2004.

A. Moreno-Daniel, S. Parthasarathy, B. H. Juang, and J. G. Wilpon. Spoken Query Processing for Information Retrieval. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, pages IV–121–IV–124. IEEE, 2007.

K. Moustakas, M.G. Strintzis, D. Tzovaras, S. Carbini, O. Bernier, J.E. Viallet, S. Raidt, M. Mancas, M. Dimiccoli, E. Yagci, S. Balci, and E.I. Leon. Masterpiece: physical interaction and 3d content-based search in vr applications. volume 13, pages 92–100, 2006.

M. Nakazato and T.S. Huang. 3d mars: immersive virtual reality for content-based image retrieval. In *Multimedia and Expo, 2001. ICME 2001. IEEE International Conference on*, pages 44–47. IEEE, 2001.

Marcin Novotni and Reinhard Klein. 3d zernike descriptors for content based shape retrieval. In *Proceedings of the eighth ACM symposium on Solid modeling and applications*, pages 216–225. ACM, 2003.

Eric Paquet and Marc Rioux. Nefertiti: a query by content system for three-dimensional model and image databases management. volume 17, pages 157–166, 1999.

Pedro Pascoal, Alfredo Ferreira, and Joaquim Jorge. Towards an immersive interface for 3d object retrieval. In *Proceedings of the 5th Eurographics conference on 3D Object Retrieval*, EG 3DOR'12, pages 51–54, Aire-la-Ville, Switzerland, Switzerland, 2012. Eurographics Association.

Sérgio Paulo, Luís C Oliveira, Carlos Mendes, Luís Figueira, Renato Cassaca, Céu Viana, and Helena Moniz. Dixi–a generic text-to-speech system for european portuguese. In *Computational Processing of the Portuguese Language*, pages 91–100. Springer, 2008.

Chris Rooney, Alex Endert, Jean-Daniel Fekete, Kasper Hornbæk, and Chris North. Powerwall: int. workshop on interactive, ultra-high-resolution displays. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 3227–3230, New York, NY, USA, 2013.

Jose M Saavedra, Benjamin Bustos, Tobias Schreck, S Yoon, and Maximiliam Scherer. Sketch-based 3D Model Retrieval using Keyshapes for Global and Local Representation. pages 47–50, 2012.

Carl Sable, Kathleen McKeown, and Kenneth W. Church. NLP found helpful (at least for one text categorization task). In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - EMNLP '02*, volume 10, pages 172–179, Morristown, NJ, USA, 2002. Association for Computational Linguistics.

Carl L. Sable and Vasileios Hatzivassiloglou. Text-based approaches for the categorization of images. pages 19–38. Springer, 1999.

Daniel Salber and Joëlle Coutaz. Applying the wizard of oz technique to the study of multimodal systems. In *In Human Computer Interaction, 3rd International Conference EWHCI'93*, pages 219–230. Springer Verlag, 1993.

Beatriz Sousa Santos, Paulo Dias, Angela Pimentel, Jan-Willem Baggerman, Carlos Ferreira, Samuel Silva, and Joaquim Madeira. Head-mounted display versus desktop for 3d navigation in virtual reality: a user study. volume 41, pages 161–181. Springer, 2009.

Tiago Santos, Alfredo Ferreira, Filipe Dias, and Manuel J. Fonseca. Using sketches and retrieval to create lego models. In *Proceedings of the Fifth Eurographics Conference on Sketch-Based Interfaces and Modeling*, SBM'08, pages 89–96, Aire-la-Ville, Switzerland, Switzerland, 2008. Eurographics Association.

Yu-Te Shen, Ding-Yun Chen, Xiao-Pei Tian, and Ming Ouhyoung. 3d model search engine based on lightfield descriptors. In *Proc. eurographics*, 2003.

J.R. Smith and Shih-Fu Chang. Visually searching the web for content. *MultiMedia, IEEE*, 4(3):12–20, 1997.

Robert Wang, Sylvain Paris, and Jovan Popović. 6d hands: markerless hand-tracking for computer aided design. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 549–558, New York, NY, USA, 2011.

Xiaohua Xie, Kai Xu, Niloy J Mitra, Daniel Cohen-Or, Wenyong Gong, Qi Su, and Baoquan Chen. Sketch-to-design: Context-based part assembly. In *Computer Graphics Forum*, volume 32, pages 233–245. Wiley Online Library, 2013.

Ying Yin and Randall Davis. Toward natural interaction in the real world: real-time gesture recognition. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, ICMI-MLMI '10, pages 15:1–15:8, New York, NY, USA, 2010.

Jamie Zigelbaum, Alan Browning, Daniel Leithinger, Olivier Bau, and Hiroshi Ishii. G-stalt: a chirocentric, spatiotemporal, and telekinetic gestural interface. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, pages 261–264. ACM, 2010.

Apêndice A

Hardware

Nos últimos anos têm surgido avanços que permitem explorar novas possibilidades de visualização e interacção, algumas mencionadas na Secção 2.5. Na secção seguinte iremos apresentar alguns dispositivos de visualização e, posteriormente, vamos referir alguns dispositivos de interacção. Por fim, iremos discutir os dispositivos apresentados.

A.1 Visualização

Ambientes de realidade virtual podem ser classificados em duas categorias, imersiva e não imersiva, ou seja, com a percepção de se estar dentro do ambiente ou não. A realidade virtual imersiva consiste tipicamente no uso de capacetes ou de salas de projecção. Por outro lado, a não imersiva consiste apenas no uso de monitores (Garau et al., 2003).

Head-mounted Displays (HMD) consistem em óculos com um ecrã LCD para para cada olho. Os HMD's podem desenhar o mundo consoante a orientação da cabeça do utilizador através de um acelerómetro. Estes óculos permitem apresentar a mesma imagem em ambos os olhos ou criarem uma imagem estereoscópica. Um exemplo disto são os z800 3D Vision[1], ilustrados na Figura A.1. Estes óculos permitem não só uma visão estetoscópica com uma resolução 800 por 600 e um campo de visão

Figura A.1: Z800 3D Vision.

Figura A.2: Oculos Rift.

Figura A.3: 3DVision Wired Glasses.

[1] z800 3D Vision ® eMagin Corporation, http://www.3dvisor.com/

de 40 graus. Um outro exemplo mais recente são os óculos da Figura A.2, os Oculus Rift[2]. Tal como os z800, os Oculus Rift permitem uma visão estereoscópica, mas ao contrário do anterior, tem uma resolução de 1280x800 e o seu campo de visão é mais de 90 graus.

A CAVE (Cruz-Neira et al., 1992; DeFanti et al., 2011), consiste num ambiente rodeado de vistas e som onde, através de projecções estereoscópicas, se tenta dar a ilusão ao utilizador de estar completamente imerso num ambiente de realidade virtual. A *Power Wall* (Rooney et al., 2013), é um exemplo que utiliza vários ecrãs para conseguir aumentar a resolução de visualização.

Liquid Crystal shutter glasses contêm um líquido cristalino que bloqueia ou deixa passar luz de forma sincronizada com a frequência de frames do ecrã. Os óculos deixam passar a imagens de diferentes perspectivas para cada olho, usando o conceito de sequência de frames alterados. Um exemplo são os óculos na Figura A.3, os 3D Vision Wired Glasses da Nvidia[3].

A.2 Interacção

Quanto aos dispositivos de interacção iremos considerar em dispositivos de mapeamento directo ou parcialmente directo, classificação utilizada em (Balakrishnan et al., 1997). Isto é, iremos considerar dispositivos com três ou mais graus de liberdade e se dependem de um conjunto de esquemas para mapeamento das entradas de 2D para 3D.

5 DT Data Glove[4], tal como podemos ver na Figura A.4, consiste numa luva com sensores de fibra óptica flexível que captura gestos e movimentos da mão. Esta possui seis sensores que detectam o dobrar de dedos, assim como a rotação e fechar da mão. Um outro tipo de luva, que permite também uma boa precisão é o caso *colored glove* (Yin and Davis, 2010), que consiste capturar os movimentos dos dedos e mão seguindo a cor das luvas.

O Wii Remote[5], ilustrado na Figura A.8, é equipado com uma câmera com sistema de seguimento de infravermelhos e um acelerómetro de três eixos, que oferece 6 graus de liberdade.

O PS Move[6], utiliza a câmera Playstation Eye e o comando Playstation Move, ilustrados na Figura A.7, para a captura dos movimentos. O comando, tal como na Wii Remote, possui um acelerómetro de três eixos. A diferença está em descobrir a posição do comando graças a uma esfera luminosa presente no topo do comando. A luz da esfera é escolhida pelo sistema baseado pelas cores ambientes detectadas pela câmera, de forma a que o comando seja facilmente distinguível do resto do ambiente. É possível ainda calcular a distância à câmera de acordo com o tamanho da esfera.

[2]Oculus Rift®, `http://www.oculusvr.com`

[3]3D Vision Wired Glasses da Nvidia ®, `http://www.nvidia.com/object/product-geforce-3d-vision-wired-glasses-us.html`

[4]5DT Data Glove, mindflux ®, http://www.mindflux.com.au/products/5dt/glove.html

[5]Wii Remote ®, `http://www.nintendo.com/wii/console/controllers`

[6]Playstation Move ®, `http://us.playstation.com/ps3/playstation-move/product-information/`

Figura A.4: 5DT Data Gloves.

Figura A.5: Microsoft Kinect.

Figura A.6: Leap Motion.

Figura A.7: PS move.

Figura A.8: Wii Remote.

Figura A.9: SpacePoint Fusion.

O SpacePoint Fusion[7], usa magnetómetro, giroscópio e acelerómetro para os três eixos e mantém a sua precisão, permitindo uma melhor experiência em termos de imersão.

O Microsoft Kinect, ilustrado na Figura A.5, é um dispositivo com capacidade de captura de movimentos 3D, reconhecimento de voz e corporal, está equipado com um sensor de profundidade, com uma câmera RGB e um microfone. O sensor de profundidade consiste na combinação de um projector de infravermelhos com um sensor CMOS monocromático, que permitem uma visão tridimensional do local. A câmera permite a detecção de cor facilita o reconhecimento facial. O sistema de captura de som consiste numa série de quatro microfones que podem isolar vozes dos utilizadores do barulho ambiente. Recentemente surgiu um trabalho, de Wang et al. Wang et al. (2011), que utiliza o Microfone Kinect para seguimento dos dedos do utilizador, contudo necessita de um *setup* rigoroso e usando dois Microsoft Kinect.

O Leap Motion[8], ilustrado na Figura A.6, foi lançado recentemente e tem uma tecnologia similar ao do Microsoft Kinect, para a captura de movimentos. O Leap Motion está equipado com duas câmeras VGA e três projectores de infravermelhos que permitem capturar gestos numa área hemisférica aproximadamente de um metro. Este dispositivo foi desenhado para capturar os movimentos de dedos, ou de objectos semelhantes que atravessem a área observada, com uma precisão espacial de 0,01 mm.

A.3 Discussão

Analisando os dispositivos de visualização, podemos retirar que um ambiente que use um sistema semelhante a uma CAVE e *powerwall* será sempre mais dispendioso e pode não trazer a imersividade que pretendemos. Por outro lado, os 3D Vision Wired Glasses da Nvidia necessitarão sempre de um

[7]PNI Sensor Corporation ®, SpacePoint Fusion, `http://www.pnicorp.com/products/spacepoint-gaming`
[8]Leap Motion ®, `https://www.leapmotion.com/`

ecrã, quebrando um pouco a imersividade que queríamos atingir. Por fim, existem os óculos HMD, tendo referido dois em cima. Segundo as características mencionadas, os Oculus Rift superam os z800 3D Vision em praticamente todos os aspectos de visualização, dando uma melhor resolução, um campo de visão maior e uma menor latência a determinar a orientação da cabeça.

Quanto aos dispositivos de interacção, ao comparar o Wii Remote, com o PS Move, podemos concluir que são bastante semelhantes, mas o PS Move consegue ter uma melhor precisão, graças à sua esfera luminosa. Contudo, acreditamos que estes dispositivos não são simples de manusear para utilizadores inexperientes. Analisando o SpacePoint Fusion, está num nivel bastante semelhante ao PS Move no entanto consegue ter uma melhor precisão que este. O uso do Microsoft Kinect permite ao utilizador realizar os seus próprios gestos, dando assim, uma maior imersividade aos utilizadores. No entanto, a sua baixa resolução torna difícil o manuseamento mais preciso. Por sua vez, o Leap Motion cobre em parte falta de precisão, mas com a desvantagem do seu campo de visão estar restrito, o utilizador não se afastar muito do dispositivo.

Apêndice B

Análise de dispositivos SpacePoint Fusion e Leap Motion

Após terem sido realizados os testes dos modos de visualização, apresentados no Secção 5.1, apercebemo-nos que os utilizadores frequentemente perdiam a noção do local onde o dispositivo Leap Motion se encontrava. Houve inclusivamente vários utilizadores que tentavam apontar para peças que se encontravam na sua lateral, colocando a sua mão completamente fora do volume de captura do dispositivo. Deste modo realizámos um novo modo de interacção utilizando um dispositivo SpacePoint Fusion, apresentado na Secção 4.4.2, de modo a verificar se este conseguia um desempenho superior.

B.1 Metodologia e Participantes

Realizámos os testes com utilizadores, os quais levavam cerca de 10 minutos a completar, estando divididos em três etapas:

- **Introdução ao teste** - Esta etapa consistia numa breve explicação no que estaria ser testado e do que era pedido a estes para realizarem.
- **Tarefa** - Era solicitado aos utilizadores que seleccionassem, apontando para elas, cinco peças de LEGO que seriam apresentadas no modo de visualização cilíndrico. As cinco peças a seleccionar encontravam-se em posições sempre diferentes umas das outras. Este teste era realizado para ambas os dispositivos de interacção, com o Leap Motion e com SpacePoint Fusion, sendo a ordem de começo escolhida aleatoriamente para cada utilizador.
- **Questionário** - Após a realização da tarefa com ambos os modos de interacção, os utilizadores responderam a um curto questionário sobre a sua experiência anterior e relativamente aos

dispositivos testados.

Esta avaliação contou 8 participantes, 25% dos quais eram do sexo feminino. Estes tinham idades compreendidas entre os 18 e os 40 anos, dos quais 63% encontravam-se entre 24 e os 30 anos. Cerca de 50% dos utilizadores nunca tinham utilizado nenhum dos dispositivos, 25% já tinham experimentado ambos e os restantes 25% interagido com apenas um dos dispositivos. O aparato da avaliação contou, para além dos dispositivos já mencionados, contou também com os Oculus Rift para a visualização do ambiente virtual.

B.2 Resultados

Analisando os tempos despendidos para conclusão da tarefa, apresentados na Tabela B.1 é clara a diferença do tempo despendido entre os dispositivos. Foi aplicado o teste de Shapiro-Wik a estes tempos, verificando-se que possuíam uma distribuição normal. Todos os utilizadores demoram menos tempo a completar a tarefa com o SpacePoint Fusion.

Utilizador	Leap Motion	SpacePoint Fusion	Dispositivo Preferido
1	94,17	18,94	SpacePoint
2	89,89	25,57	SpacePoint
3	58,63	26,77	SpacePoint
4	78,97	12,48	SpacePoint
5	88,85	15,49	SpacePoint
6	103,61	22,72	SpacePoint
7	73,17	25,26	SpacePoint
8	70,59	28,87	SpacePoint

Tabela B.1: Tempos de conclusão da tarefa em segundos utilizando Leap Motion e SpacePoint Fusion e dispostivo preferido.

Todos os utilizadores preferiram utilizar o dispositivo SpacePoint Fusion, e afirmaram que era bastante simples a sua utilização. Os utilizadores foram questionados sobre quão fácil foi seleccionar uma peça com ambos os dispositivos, utilizando uma escala de Likert de 6 valores (onde 6 era muito fácil). Os resultados das classificações em termos de facilidades encontram-se na Tabela B.2. O teste de Wilcoxon signed-rank foi utilizado para verificar as diferenças estatisticamente relevantes. Com este teste verificámos que os utilizadores concordaram que o modo de interacção que utiliza o SpacePoint Fusion era mais fácil (Z=-3.482, p=0.000).

Dispositivo	Facilidade de selecção
Leap Motion	3 (1)
SpacePoint Fusion	6 (0)

Tabela B.2: Classificação para cada modo de interacção (Mediana, Amplitude interquartil).

Apêndice C

Testes sobre os modos de visualização

C.1 Questionário

1 - Género
- a) Masculino
- b) Fiminino

2 - Idade
- a) Menos de 18
- b) 18 - 23
- c) 24 - 29
- d) 30 - 40
- e) 41 - 50
- f) Mais de 50

3 - Com que frequencia utiliza dipositivos "touch"?
- a) menos de 1 vez por semana
- b) 1 vez por semana
- c) 3 vezes por semana
- d) diariamente

4 - Que dispositivo utiliza?
- a) Smartphones
- b) Tablet
- c) Mesa multi toque
- d) Monitores com toque

5 - Já experimentou ver um filme 3D?
- a) Sim
- b) Não

6 - Para além dos filmes 3D já teve alguma experiência com esteroescopia?
- a) Sim
- b) Não

7 - Caso tenha respondido sim à pergunta anterior, em que situação?

8 - Já tinha alguma experiência com Head Mounted Display (HMD) ?
- a) Sim
- b) Não

9 - Caso tenha respondido sim na pergunta anterior, em que situação?

10 - Já alguma vez utilizou o Microsoft Kinect?
- a) Sim
- b) Não

11 - Caso tenha respondido sim na pergunta anterior, em que situação?

12 - Já alguma vez utilizou o Leap Motion?
- a) Sim
- b) Não

13 - Caso tenha respondido sim na pergunta anterior, em que situação?

14 - Já usou algum destes programas?
- a) Mikes Lego CAD (MLCAD)
- b) Lego Digital Designer(LDD)
- c) LeoCad
- d) SR 3D Builder

15 - Já tinha utilizado algum sistem de pesquisa de imagens?
- a) Sim
- b) Não

16 - Se respondeu sim na questão anterior, qual(is) o(s) que usou?
- a) Flickr
- b) Google
- c) WeSee
- d) Compfight
- e) CC Search
- f) TinyEye

17 - Já tinha utilizado algum sistema de pesquisa de objectos 3D?
- a) Sim
- b) Não

18 - Caso tenha respondido sim na pergunta anterior, em que situação?

19 - Preencha a seguinte tabela de acordo com facilidade de utilização dos modos.
(1 - Difícil e 5 - Fácil)

	1	2	3	4	5
Tradicional					
Rectangular					
Quadrático					
Cilíndrico					
Esférico					

20 - Preencha a seguinte tabela de acordo com diversão em cada modo.
(1 - Nada divertido e 5 - Divertido)

	1	2	3	4	5
Tradicional					
Rectangular					
Quadrático					
Cilíndrico					
Esférico					

21 - Classifique a sua experiência relativamente a imersão que sentiu ao utilizar o sistema.
(1 - Nada imerso e 4 - Imerso)

1	2	3	4

22 - Sentiu alguma indisposição depois de utilizar os Oculus Rift?
a) Sim
b) Não

23 - Relativamente ao controlo da lista de resultados com gestos, sentiu ser:
(1 - Difícil e 4 - Fácil)

1	2	3	4

24 - Visualizar a peça mais perto e a rodar, com o gesto de apontar para analisar a peça foi:
(1 - Irrelevante e 4 - Relevante)

1	2	3	4

25 - Relativamente a acertar na peça pretendida usando o gesto de apontar, sentiu ser:
(1 - Difícil e 4 - Fácil)

1	2	3	4

C.2 Resultados

Utilizador	1	2	3	4	5	6	7
1	a	c	d	a, b	a	b	
2	b	c	d	a, b	a	a	Nintendo 3DS
3	a	b	d	a	a	b	
4	a	e	d	a, b	a	b	
5	a	b	d	a, b	a	b	
6	a	c	d	a, b	a	a	Jogos 3D (Uncharted 3)
7	a	c	d	a	a	b	
8	a	b	d	a, b	a	b	
9	a	b	d	a, b, c	a	a	mesa multitoque
10	a	b	d	a, c	a	b	
11	a	c	d	a	a	a	Estereogramas
12	a	c	d	a	a	a	Trabalhos de modelação 3D
13	a	c	d	a, b	a	b	
14	a	e	d	a, b	a	b	
15	b	b	d	a	a	a	Oculus Rift
16	a	c	d	a, b	a	a	Capacete de Realidade Virtual
17	a	c	d	a, b	a	b	
18	b	b	d	a	a	b	
19	a	a	d	a	a	a	oculos vermelhos e azuis- nvidia
20	a	c	d	a	a	a	A testar teses de outros colegas
21	a	b	d	a, b	a	a	Minecraft com óculos estéreoscopicos
22	b	c	d	a	a	b	
23	a	c	d	a	a	a	Jogos de realidade aumentada, nomeadamente FPS
24	b	b	d	a	a	a	Óculos vermelhos e azuis, Wiggle stereoscopy
25	a	c	d	a, b	a	a	Mesa multtoque com esteroescopia
26	b	c	d	a, b	a	b	
27	b	c	d	a	a	b	
28	a	b	d	a, b	a	b	
29	a	e	d	b	a	a	Capacete de Realidade Virtual

Utilizador	8	9	10	11	12	13
1	b		b		b	
2	b		a	jogos Xbox	b	
3	b		b		b	
4	b		a	jogos Xbox	b	
5	b		b		b	
6	b		b		b	
7	b		b		b	
8	b		a	jogos Xbox	b	
9	a	demos do oculus rift	a	jogos Xbox	b	
10	b		a	jogos Xbox	b	
11	b		a	jogos Xbox	b	
12	a	demos do oculus rift	b		b	
13	b		b		b	
14	b		a	jogos Xbox	b	
15	a	demos do oculus rift	a	jogos Xbox	b	
16	b		b		b	
17	b		b		a	demos Leap
18	b		b		b	
19	b		a	jogos Xbox	b	
20	a	demos do oculus rift	a	jogos Xbox	a	demos Leap
21	a	demos do oculus rift	a	jogos Xbox	a	demos Leap
22	b		a	jogos Xbox	b	
23	a	demos do oculus rift	a	jogos Xbox	b	
24	a	demos do oculus rift	a	jogos Xbox	b	
25	b		b		b	
26	b		b		b	
27	b		a	jogos Xbox	b	
28	b		a	jogos Xbox	b	
29	a	z800	b		b	

Utilizador	14	15	16	17	18
1	b	a	a, b	b	
2		a	a, b	b	
3		a	a, b	a	
4		a	a, b	a	Unity Asset Stor
5		a	a, b	b	
6		a	a, b	b	
7		a	a, b	b	
8	b	a	a, b	b	
9		a	b	b	
10		a	a, b	b	
11		a	a, b	a	Google 3D Warehouse
12		a	b	b	
13		a	a, b	b	
14		a	a, b	b	
15		a	a, b	b	
16		a	b	b	
17		a	a, b	b	
18		a	a, b	b	
19		a	a, b	b	
20		a	b	b	
21		a	a, b	b	
22		a	a, b	a	Google 3D Warehouse
23		a	a, b	b	
24		a	a, b	b	
25		a	a, b	b	
26		a	a, b	b	
27		a	a, b	b	
28		a	a, b	b	
29	b	a	a, b	a	SketchUp

Utilizador	19					20				
	T	R	Q	C	E	T	R	Q	C	E
1	5	4	3	5	3	4	5	4	4	3
2	4	5	5	4	2	3	5	5	4	3
3	4	5	3	4	3	3	5	3	4	2
4	5	5	5	4	5	4	4	4	5	5
5	5	5	5	4	5	3	5	5	5	5
6	5	4	4	4	4	4	5	5	5	5
7	4	3	4	5	5	2	3	4	5	5
8	3	4	5	4	4	2	5	4	4	5
9	5	5	4	4	3	3	5	4	5	4
10	3	4	4	5	4	1	4	4	5	5
11	3	3	3	4	5	2	4	4	4	5
12	5	5	5	5	5	3	5	5	5	5
13	5	4	4	3	5	4	5	5	4	5
14	4	5	5	5	5	3	5	5	5	5
15	5	4	4	4	3	3	4	4	5	3
16	4	3	3	4	5	3	3	3	4	5
17	5	4	4	3	2	4	3	3	4	2
18	5	4	4	4	5	2	3	3	4	5
19	3	3	3	5	5	3	5	5	5	5
20	5	5	5	4	4	4	4	3	5	5
21	5	4	4	3	2	4	4	4	3	2
22	5	4	4	3	3	4	4	4	3	3
23	1	3	3	5	4	1	4	4	5	5
24	5	3	4	4	4	4	4	4	4	4
25	5	5	5	5	4	4	5	5	5	5
26	3	4	4	4	4	3	4	4	4	4
27	3	4	4	5	4	2	4	4	5	4
28	3	5	5	5	5	1	4	4	4	5
29	5	4	4	4	4	4	4	5	5	5

T - Tradicional
R - Rectangular
Q - Quadrático
C - Cilíndrico
E - Esférico

Utilizador	21	22	23	24	25
1	4	a	2	3	3
2	4	a	4	4	3
3	4	b	3	4	4
4	4	b	4	4	3
5	3	b	3	4	3
6	3	b	3	4	4
7	3	b	3	4	3
8	4	b	3	2	2
9	4	b	4	4	4
10	4	b	4	4	4
11	4	b	3	4	3
12	4	b	4	4	4
13	4	b	3	4	3
14	3	b	4	4	4
15	3	b	3	4	4
16	3	b	3	4	2
17	3	b	3	4	3
18	4	b	3	4	4
19	3	a	4	3	4
20	3	b	3	4	3
21	3	b	3	4	2
22	4	b	3	4	1
23	4	b	4	4	3
24	4	b	3	4	3
25	4	b	4	4	4
26	3	b	4	4	3
27	4	b	4	4	4
28	3	b	4	4	3
29	3	b	4	3	4

Tempos

Utilizador	T	R	Q	C	E
1	60.831	41.258	95.882	63.228	95.366
2	45.619	17.836	15.625	101.119	56.864
3	82.666	87.438	74.416	123.068	65.916
4	23.207	89.838	66.708	39.808	37.807
5	30.197	110.643	15.436	101.779	64.548
6	11.139	28.663	108.155	100.14	46.405
7	61.822	45.002	12.002	30.073	41.284
8	26.682	55.083	53.752	10.275	20.296
9	20.99	19.343	19.985	70.02	23.935
10	141.792	216.802	39.064	124.711	33.941
11	46.306	33.881	18.611	38.388	84.141
12	37.601	27.707	51.815	37.992	22.289
13	35.18	24.901	187.314	12.99	14.759
14	62.854	106.647	53.266	47.073	19.441
15	12.236	51.616	19.648	51.982	31.001
16	62.637	31.528	30.864	108.54	55.719
17	18.89	81.666	52.066	100.183	71.886
18	73.379	29.732	71.08	75.739	103.848
19	33.637	39.673	55.671	55.852	40.873
20	26.341	36.09	29.784	9.6	42.719
21	28.629	15.306	47.974	26.886	187.329
22	13.335	17.956	19.091	8.18	34.949
23	21.31	28.439	15.387	51.564	24.838
24	21.759	14.782	30.413	25.349	39.428
25	22.148	180.276	192.837	51.625	35.495
26	23.4	51.601	21.645	76.85	100.901
27	26.009	39.729	16.571	22.352	31.736
28	52.93	62.849	18.315	11.22	116.12
29	9.102	18.521	82.108	167.33	37.909

Apêndice D

Testes de validação de solução

D.1 Questionário Inicial

1 - Género
- a) Masculino
- b) Fiminino

2 - Idade
- a) Menos de 18
- b) 18 - 23
- c) 24 - 29
- d) 30 - 40
- e) 41 - 50
- f) Mais de 50

3 - Distrito onde viveu mais tempo?

4 - Com que frequencia utiliza dipositivos "touch"?

	a	b	c	d
Smartphones				
Tablet				
Mesa multi toque				
Monitores com toque				

Legenda:
- a) menos de 1 vez por semana
- b) 1 vez por semana
- c) 3 vezes por semana
- d) diariamente

5 - Já experimentou ver um filme 3D?
- a) Sim
- b) Não

6 - Para além dos filmes 3D já teve alguma experiência com esteroescopia?
- a) Sim
- b) Não

7 - Caso tenha respondido sim à pergunta anterior, em que situação?

8 - Já tinha alguma experiência com Head Mounted Display (HMD) ?
- a) Sim
- b) Não

9 - Caso tenha respondido sim na pergunta anterior, em que situação?

10 - Já usou algum destes programas?
- a) Mikes Lego CAD (MLCAD)
- b) Lego Digital Designer(LDD)
- c) LeoCad
- d) SR 3D Builder

11 - Com que frequencia pesquisa imagens?
- a) menos de 1 vez por semana
- b) 1 vez por semana
- c) 3 vezes por semana
- d) diariamente

12 - Já utilizou algum sistema de pesquisa de objectos 3D?
- a) Sketchup
- b) Unity asset store
- c) Princeton Search Engine
- d) outro:

13 - Já utilizou algum destes sistemas de pesquisa por fala?
- a) Google Now
- b) Apple -Siri
- c) Samsung - S Voice
- d) outro:

D.2 Questionário LDD e LSpeakIt

LDD

1 - Relativamente à dificuldade de encontrar a peça pretendida na lista
(em que 1 - difícil de encontrar e 4 - fácil de encontrar)
1 2 3 4

2 - Achou que era fácil de perceber as peças na lista?
(Em que 1 - difícil de perceber e 4 - fácil de perceber)
1 2 3 4

3 - Foi fácil de utilizar a procura do Lego Digital Designer ?
(Em que 1 - Difícil e 4 - Fácil)
1 2 3 4

4 - Foi relevantes a divisão por categorias?
(Em que 1 - Nada relevante e 4 - Relevante)
1 2 3 4

5 - Achou que o método de procura do Lego Digital Designer foi divertido?
(Em que 1 - Nada divertido e 4 - Divertido)
1 2 3 4

LSpeakIt

1 - Classifique a sua experiência relativamente à imersão que sentiu ao utilizar o sistema LSpeak It
(Em que 1 - Nada imerso e 4 - Completamente imerso)
1 2 3 4

2 - Sentiu alguma indisposição depois de utilizar os Oculus Rift?
a) Sim
b) Não

3 - Considera que foi relevante as peças estarem a rodar?
(Em que 1 - Nada relevante e 4 - Relevante)
1 2 3 4

4 - Considera que foi relevante poder :
(Em que 1 - Nada relevante e 4 - Relevante)

	1	2	3	4
Filtrar resultados				
Excluir resultados				
Procurar semelhantes				

5 - Achou que era fácil de perceber as peças na lista?
(Em que 1 - difícil de perceber e 4 - fácil de perceber)
1 2 3 4

6 - Foi fácil descrever as peças?
(Em que 1 - Difícil e 4 - Fácil)
1 2 3 4

7 - Foi fácil de utilizar a procura do LSpeakIt?
(Em que 1 - Difícil e 4 - Fácil)
1 2 3 4

8 - Achou que o método de procura do LSpeakIt foi divertido?
(Em que 1 - Nada divertido e 4 - Divertido)
1 2 3 4

D.3 Resultados

Questionário Inicial

Utilizador	1	2	3	5	6	7	8	9	10	11	12	13
1	a	c	Lisboa	a	b		a	Demos Rift		c		a,b
2	a	c	Lisboa	a	a	Demos	b			b		a
3	b	c	Lisboa	a	b		b			c		
4	a	b	Lisboa	a	a	Demos	b			d	a	a
5	b	c	Lisboa	a	a	Demos	b			c		b
6	a	c	Lisboa	a	b		b			d		
7	a	b	Lisboa	a	a	Futuroscope	a	Demos Rift		d		a,c
8	a	c	Lisboa	a	a	Nitendo 3Ds	b			c		
9	a	c	Lisboa	a	a	Demos	a	Demos Rift		b		
10	a	c	Santarém	a	a	Demos	a	Demos Rift		d	a,b	a,b
11	a	c	Lisboa	a	a	Demos	a	Demos z800		c	a,b	a,b
12	a	c	Lisboa	a	a	Demos	a	Demos Rift		d	a	b
13	a	b	Setúbal	a	a	Nitendo 3Ds	a	Demos Rift		d	b	
14	a	c	Lisboa	a	b		a	Demos Rift		d		a
15	a	b	Lisboa	a	a	Estereograms	b			b		a
16	a	c	Lisboa	a	a	Estereograms	a	Demos Rift		d	a,b	
17	b	b	Lisboa	a	b		a	Demos Rift		d		c
18	a	c	Lisboa	a	a	Estereograms	b			d		
19	a	e	Lisboa	a	a	Estereograms	a	Demos z800		c		a
20	b	c	Lisboa	a	b		b			b		

Utilizador	4			
	S	T	MM	MoM
1	d	d	b	d
2	d	b	a	a
3	d	d	a	a
4	d	a	a	a
5	d	d	a	a
6	d	b	a	b
7	d	a	a	a
8	d	d	a	a
9	d	a	c	a
10	d	d	a	a
11	d	d	a	d
12	d	d	a	a
13	d	a	a	a
14	d	a	a	a
15	d	b	a	b
16	d	a	a	c
17	d	a	a	a
18	d	d	a	a
19	d	d	a	a
20	d	a	a	a

S - SmartPhones
T-Tablet
MM - Mesas multi toque
MoM- Monitores multi toque

LDD

Utilizador	1	2	3	4	5
1	3	3	2	2	2
2	3	3	4	4	4
3	4	2	1	4	1
4	3	4	4	3	4
5	2	3	2	4	2
6	3	3	2	3	2
7	2	2	2	2	2
8	4	3	4	4	4
9	2	3	2	2	2
10	2	2	4	2	4
11	2	3	3	2	3
12	2	4	2	3	2
13	2	4	3	3	3
14	2	3	1	3	1
15	2	3	1	3	1
16	2	2	3	3	3
17	2	2	2	2	2
18	2	2	2	4	2
19	2	2	2	3	2
20	2	3	2	2	2

LSpeakIt

Utilizador	1	2	3	5	6	7	8
1	3	b	3	3	3	4	3
2	3	b	4	4	3	4	4
3	4	b	4	4	3	4	4
4	3	b	4	3	3	3	4
5	4	b	3	3	3	3	4
6	4	b	3	3	4	4	4
7	3	b	4	3	3	3	4
8	3	b	4	2	4	2	4
9	4	b	4	3	3	3	3
10	4	b	4	4	4	4	4
11	4	b	4	3	2	3	3
12	3	b	4	3	3	4	4
13	4	b	3	4	3	4	4
14	3	b	4	3	4	3	4
15	4	b	4	3	3	3	4
16	3	b	4	3	2	3	4
17	3	b	4	2	3	4	3
18	4	b	4	3	3	3	4
19	3	b	3	3	2	4	4
20	3	b	3	3	3	3	4

Utilizador	4		
	F	E	S
1	4	3	4
2	4	2	2
3	4	4	4
4	4	4	4
5	2	2	3
6	4	3	3
7	3	2	4
8	4	1	3
9	3	2	3
10	1	1	2
11	4	2	4
12	4	4	4
13	4	4	4
14	3	1	1
15	4	3	4
16	4	4	4
17	4	1	1
18	3	1	3
19	3	3	3
20	4	4	4

F - Filtrar Resultados
E - Excluir Resultados
S - Procurar Semelhantes

Tempos LDD

Utilizador	Peças							
	1	2	3	4	5	6	7	8
1	18.22	37.74	59.02	41.63	82.41	47.69	51.78	537.86
2	15.09	413.05	139.54	54.69	34.06	18.72	32.80	490.33
3	24.29	49.08	270.20	120.04	97.31	23.17	31.17	476.64
4	22.31	27.54	99.20	25.03	23.47	21.12	49.66	786.86
5	53.52	47.79	62.46	28.59	97.26	159.05	41.72	639.64
6	23.60	22.50	50.21	30.20	26.49	14.19	51.78	168.48
7	43.83	39.63	156.47	63.03	116.75	19.77	23.93	149.85
8	43.83	66.80	73.10	50.28	355.22	52.83	34.25	174.24
9	24.79	41.26	42.01	27.17	226.86	166.85	20.26	209.98
10	13.14	35.79	47.06	41.59	30.33	51.01	22.76	342.08
11	37.19	26.78	173.82	47.31	85.79	82.00	98.98	255.47
12	25.13	66.50	361.17	151.45	80.89	124.48	132.49	273.32
13	28.01	46.98	205.25	54.37	46.53	53.13	39.47	242.16
14	38.84	117.59	83.97	38.84	137.38	18.90	50.84	438.51
15	19.22	23.20	124.46	24.84	47.06	64.01	49.99	510.15
16	28.73	70.82	129.86	53.27	99.19	61.13	48.79	399.70
17	26.93	148.73	129.99	43.90	34.21	57.57	41.92	472.59
18	18.72	15.78	40.53	42.43	172.78	39.37	44.05	442.96
19	10.21	16.51	199.64	20.26	24.19	37.74	29.89	180.39
20	13.61	14.37	68.63	74.29	61.39	22.68	49.92	266.20

Tempos LSpeakIt

Utilizador	Peças							
	1	2	3	4	5	6	7	8
1	69.86	114.86	38.25	108.24	60.78	119.03	118.01	274.48
2	28.69	38.43	136.10	77.04	184.71	28.42	133.73	18.04
3	28.65	41.65	67.28	118.49	140.01	12.23	89.30	103.78
4	102.04	172.03	128.26	168.04	96.03	125.09	62.68	190.90
5	37.28	29.16	58.07	26.77	97.56	225.04	70.82	73.84
6	31.26	18.05	55.25	43.58	85.07	12.01	21.43	70.33
7	17.08	35.12	101.85	113.24	119.62	56.24	75.79	20.96
8	40.80	61.53	48.22	69.20	147.34	50.77	130.09	42.48
9	17.08	63.34	55.45	156.68	62.72	22.72	50.24	166.18
10	40.00	49.24	81.81	23.33	85.51	11.61	14.09	14.92
11	22.33	27.49	70.30	114.88	126.35	167.61	27.21	235.58
12	18.50	27.09	114.15	366.22	43.05	29.01	27.99	13.65
13	18.31	39.31	13.80	71.07	22.93	15.78	72.41	102.94
14	35.27	49.51	37.82	54.51	46.82	49.33	63.50	97.54
15	24.12	117.38	71.93	159.38	17.92	54.24	132.97	63.83
16	32.99	72.08	61.14	18.50	86.19	50.98	30.38	76.82
17	80.21	135.11	47.49	67.92	81.44	23.83	61.44	220.84
18	32.69	33.96	111.37	148.64	141.95	61.75	22.71	25.78
19	37.76	84.90	76.05	85.88	80.71	34.76	40.98	75.47
20	45.44	67.64	21.99	111.48	68.02	36.04	139.04	157.54