

Qual Peça? Identificação de Métodos Naturais para a Descrição de Objectos 3D

Daniel Mendes Diogo Henriques Isabel Trancoso Alfredo Ferreira
INESC-ID / IST / Universidade de Lisboa

danielmendes@ist.utl.pt, diogo.henriques@ist.utl.pt,
isabel.trancoso@inesc-id.pt, alfredo.ferreira@inesc-id.pt

Resumo

O número de objectos tridimensionais em formato digital tem aumentado consideravelmente nos últimos anos. Com este aumento, intensifica-se o desafio de encontrar um objecto específico em colecções cada vez maiores, de forma natural para o utilizador. Neste trabalho apresentamos um estudo cujo objectivo foi identificar os métodos mais naturais para a descrição de objectos 3D, com especial foco na identificação das expressões verbais e gestuais mais recorrentes. Esta experiência foi realizada utilizando um cenário de construção de modelos LEGO®, no qual um participante constrói o modelo, tendo de requisitar as peças pretendidas a outro participante. Os resultados sugerem que os utilizadores para descrever objectos tridimensionais preferem efectuar descrições verbais, recorrendo raramente a gestos e utilizando-os apenas como complemento. Pretendemos, futuramente, integrar os resultados desta avaliação numa aplicação que ofereça mecanismos naturais para a recuperação de conteúdo tridimensional deste tipo.

Palavras-Chave

Interfaces multi-modais, interacção natural, fala, gestos, descrição de objectos 3D, testes com utilizadores

1. Introdução

O aparecimento de tecnologias de baixo custo que permitem a digitalização de objectos físicos, como o Microsoft Kinect, Asus Xtion ou o PrimeSense Sensor, a par com a vulgarização de software de modelação tridimensional, tem originado um considerável aumento no número de objectos 3D virtuais disponíveis. Uma consequência natural deste aumento é a crescente complexidade da procura de um modelo 3D que se assemelhe a um pretendido por um utilizador, tendo-se tornado esta tarefa num processo cada vez mais demorado e fastidioso.

Para efectuar a recuperação deste e de outros tipos de objectos multimédia, a informação intrínseca nestes, como por exemplo o nome do ficheiro correspondente, tem-se revelado insuficiente, sendo frequente a necessidade de associar nova meta-informação [SC97, FMK⁺03]. Com o objectivo de ultrapassar este desafio, diversas soluções para efectuar pesquisas têm sido propostas. Algumas recorrem à introdução de texto [SC97, FMK⁺03], a esboços [SFDF08, LLJ⁺13], descrições verbais [LRL10, MDPJW07], gestos [HW11] ou usando um objecto como exemplo [PR99, LSF⁺11].

Todas as soluções propostas têm, no entanto, as suas desvantagens. As pesquisas por exemplo requerem que o utilizador já possua um objecto semelhante ao que pretende recuperar, sendo que a ausência desse objecto é precisamente

o que origina a pesquisa. As restantes soluções não exploram convenientemente a riqueza e poder descritivo presente na interacção humana. Embora alguns trabalhos já sigam esta direcção, não são, no entanto, aplicados ao contexto da recuperação de objectos 3D ou não combinam a multi-modalidade de fala e gestos utilizadas em descrições naturais.

Pretendendo explorar descrições de objectos 3D, efectuámos uma sessão experimental com utilizadores para compreender quais as expressões verbais e gestuais naturalmente utilizadas. Era nosso objectivo compreender se os utilizadores recorrem maioritariamente à fala, a gestos ou a uma combinação de ambos. Para isto foi utilizado um cenário de construção de modelos LEGO®, nos quais os participantes tinham de solicitar as peças necessárias a outro participante, descrevendo-as como considerassem adequado.

No resto do documento será analisado o estado da arte relativamente a modalidades para entrada de dados do utilizador com o objectivo de efectuar a recuperação de objectos multimédia, com especial foco em objectos tridimensionais. Apresentamos então os nossos testes com utilizadores, os participantes e discutimos os respectivos resultados. Terminamos com as principais conclusões alcançadas e referimos futuras direcções para o nosso trabalho.

2. Trabalho Relacionado

Com o aumento de qualquer tipo de conteúdo, a recuperação de informação concreta será sempre um desafio, sendo que objectos tridimensionais, ou outro conteúdo multimédia, não são excepção. Uma das formas tradicionais de efectuar esta recuperação consiste em recorrer a interrogações textuais. No entanto, este método não é trivial, uma vez que os objectos usualmente não contêm informação intrínseca suficiente que os descreva, como por exemplo, o nome do ficheiro pode não estar relacionado com os objectos [SC97, FMK⁺03]. De forma genérica, os motores de pesquisa costumam utilizar texto associados aos objectos, como legendas, referências aos objectos ou até, quando se trata de conteúdo disperso pela internet, as hiperligações dos objectos ou nomes dos ficheiros. Este conceito já foi aplicado em recuperação de imagens [SC97] e de objectos 3D [FMK⁺03]. No trabalho de Funkhouser et al. [FMK⁺03], são ainda utilizados sinónimos de termos retirados dos textos mencionados, com o objectivo de aumentar a informação usada para descrever o objecto 3D.

Apesar de tudo, a informação para descrever os objectos 3D continua a ser insuficiente, especialmente no que diz respeito à sua forma. Algumas das soluções propostas utilizam pesquisas por exemplo para facilitar o processo de recuperação por forma. As pesquisas por exemplo consistem em obter objectos semelhantes em termos de aspectos visuais, como a cor [PR99] ou a forma do objecto [LSF⁺11]. Esta solução requer, no entanto, que o utilizador já possua inicialmente um objecto idêntico ao que pretende, o que geralmente não se verifica.

A pesquisa por esboços vem precisamente oferecer aos utilizadores a possibilidade de pesquisarem objectos semelhantes a uma dada forma, para a qual não possuem um modelo para servir como exemplo [PLR05]. No trabalho de Santos et al. [SFDF08], os utilizadores têm a possibilidade de efectuar esboços que correspondem às dimensões dos objectos pretendidos, sendo neste caso pesquisadas peças de LEGO[®] com essas características. Numa abordagem diferente, Liu et al. [LLJ⁺13] procuram melhorar a pesquisa por esboços através da avaliação do perfil do utilizador, verificando os seu hábitos a desenhar os esboços, melhorando os resultados retornados à medida que os utilizadores realizam mais pesquisas.

Holz e Wilson [HW11] seguiram uma abordagem diferente, no sentido de recorrer a um método frequentemente utilizado na descrição de objectos. No seu trabalho, estes autores focaram-se em reconhecer descrições de objectos tridimensionais através de gestos e, com essa descrição, recuperar objectos presentes numa base de dados. Este trabalho consistiu em capturar e interpretar gestos dos utilizadores, explorando a percepção espacial do ser humano. A forma e a movimentação das mãos enquanto os objectos eram descritos foram usadas de forma a criar esboços tridimensionais, como ilustrado na Figura 1. Os autores concluíram que os participantes conseguiam manter as proporções dos objectos físicos relativamente correctas e,

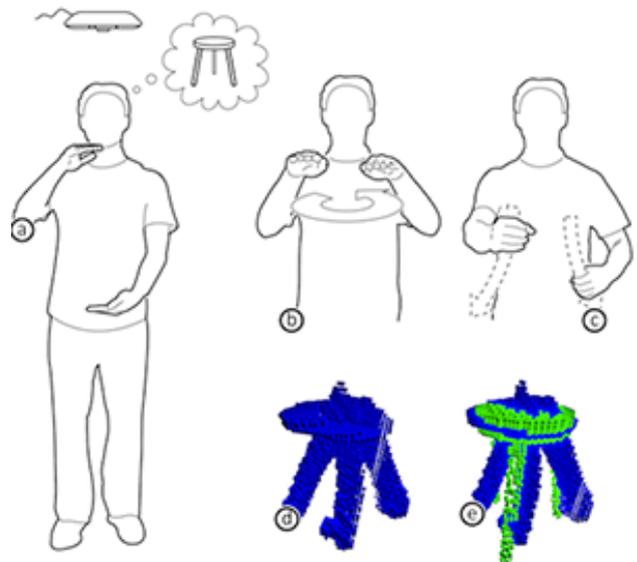


Figura 1. Descrição de objectos 3D através de gestos, no trabalho Data Miming [HW11].

em zonas com detalhes mais significativas, realizavam os gestos de forma mais lenta.

Recentemente tem-se assistido a uma disseminação da utilização de pesquisa através de fala em dispositivos móveis. Esta nova possibilidade tem originado que este modo de pesquisa comece a ser preferido à tradicional pesquisa por texto [KB10]. Geralmente, em pesquisas por fala, a fala é convertida em texto, sendo esse texto utilizado como parâmetro de pesquisa [MDPJW07, LRL10]. Nestes trabalhos, a informação a recuperar é filtrada, simplificando consideravelmente as pesquisas. No entanto, a informação é retirada de resumos, sendo que, ocasionalmente, pode não existir informação suficiente para a correcta descrição dos objectos a pesquisar. Mais recentemente, Lee e Kawahara [LK12] realizaram uma análise semântica às interrogações por fala efectuadas pelos utilizadores, conseguindo uma compreensão acrescida daquilo que o utilizador pretende recuperar.

A recuperação de conteúdo multimédia, nomeadamente para objectos tridimensionais, tem sido alvo de investigação. A maioria das soluções, embora já começando a explorar métodos mais naturais para a descrição dos objectos, como por exemplo gestos, não exploram ainda convenientemente a riqueza da interacção entre seres humanos e o seu poder descritivo. Noutras áreas, descrições verbais são já utilizadas para a recuperação de conteúdo, não tendo ainda, no entanto, sido aplicadas a objectos 3D nem complementadas com outros métodos descritivos naturais para os utilizadores.

Com o objectivo de compreender quais os métodos mais naturais para a descrição de objectos tridimensionais, conduzimos uma experiência com utilizadores.

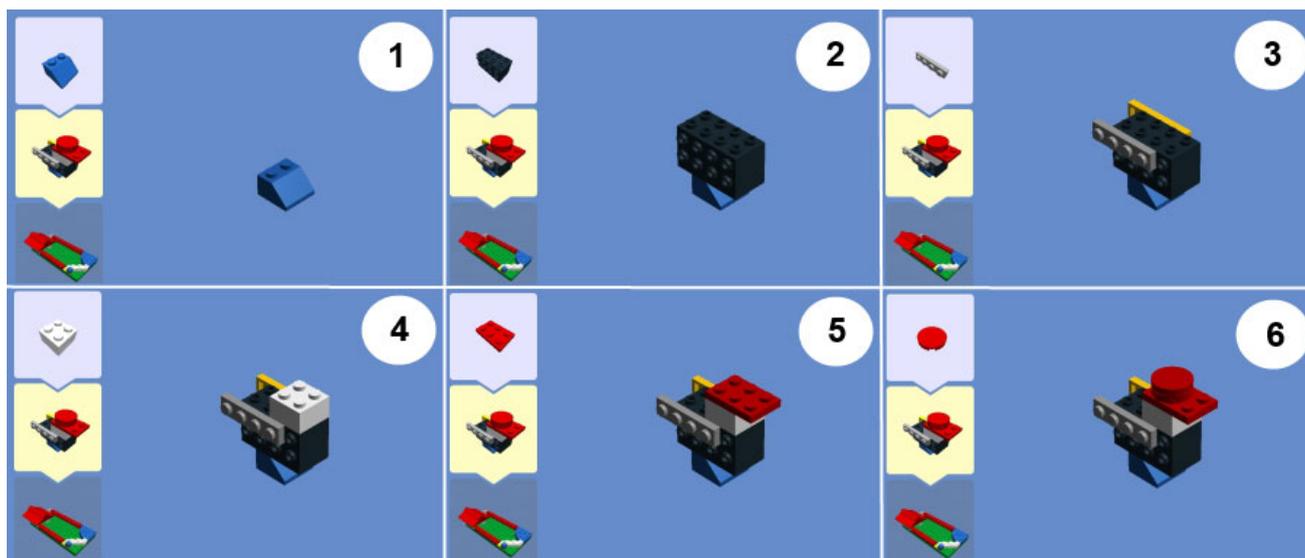


Figura 2. Exemplo de instruções fornecidas ao construtor, semelhante às que acompanham os modelos LEGO® originais.

3. Metodologia da Experiência

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 meta 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.

Este estudo seguiu uma abordagem semelhante ao Wizard of Oz [SC93], sendo que, no entanto, existiam participantes a desempenhar ambos os papéis. Cada participante desempenhava apenas uma vez cada papel, para evitar o factor de aprendizagem. Cada sessão foi dividida em três etapas: introdução à experiência, execução da tarefa e preenchimento de questionário. Com a realização de um teste piloto verificou-se a adequação da tarefa, sendo a duração de cada etapa definida de acordo com a Tabela 1. Cada sessão com dois utilizadores tinha a duração aproximada de 30 minutos.

3.1 Introdução à Experiência

Inicialmente, a experiência era introduzida aos participantes. Foram-lhes apresentados os objectivos da experiência,

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

Tabela 1. Duração das diferentes etapas de cada sessão da experiência.

os papéis que cada um assumiria e qual a sua função. Nesta fase, era também possibilitado aos participantes familiarizarem-se com um exemplo das instruções passo-a-passo para a construção de um modelo, ilustradas na Figura 2 e diferentes daquelas que iriam seguir na sessão.

3.2 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 em relação aos demais. As instruções para cada construção continham 23 passos, sendo cada passo o acto 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 possa continuar, como visível na Figura 3. Após terminada a construção, os participantes trocavam de papéis, repetindo o processo para um modelo diferente.

3.3 Preenchimento dos Questionários

No final da sessão, era solicitado aos participantes que preenchessem um questionário. Este questionário tinha como objectivos, por um lado, definir o perfil dos participantes e, por outro, identificar as suas preferências relativamente



Figura 3. Participante no papel de construtor recebe uma peça do fornecedor.

às modalidades para as descrições das peças, assim como aquelas que tiveram mais dificuldade em descrever.

4. Participantes

A experiência foi feita em 10 sessões, contando com um total de 20 participantes, 7 dos quais do sexo feminino. As idades dos participantes situaram-se entre os 18 e os 45 anos, sendo que a maioria tinha entre os 18 e os 24 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.

5. 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. 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.

5.1 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

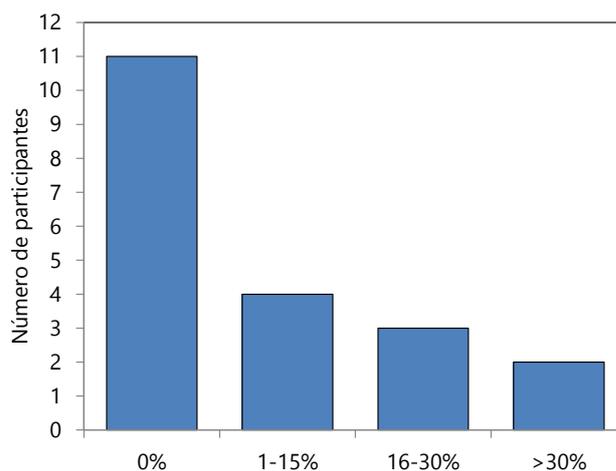


Figura 4. Percentagem de passos em que os participantes descreveram a peça utilizando gestos.

construção, os participantes descreveram a peça necessária recorrendo essencialmente à fala. Tal como se pode ver no gráfico na Figura 4, 11 participantes não utilizaram qualquer gesto em nenhum dos passos da sessão. Ocasionalmente os participantes utilizaram gestos, mas sempre com o objectivo de complementar a descrição verbal, não acrescentando nenhuma 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 5). 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.

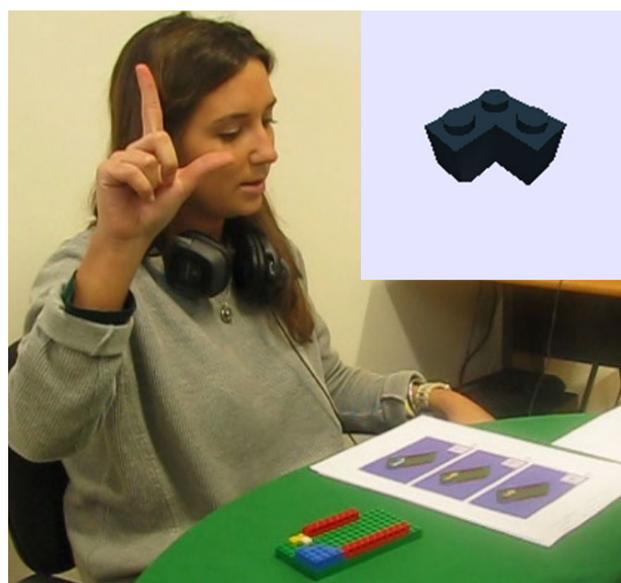


Figura 5. Participante solicita uma peça de canto, utilizando uma referência à letra 'L'.

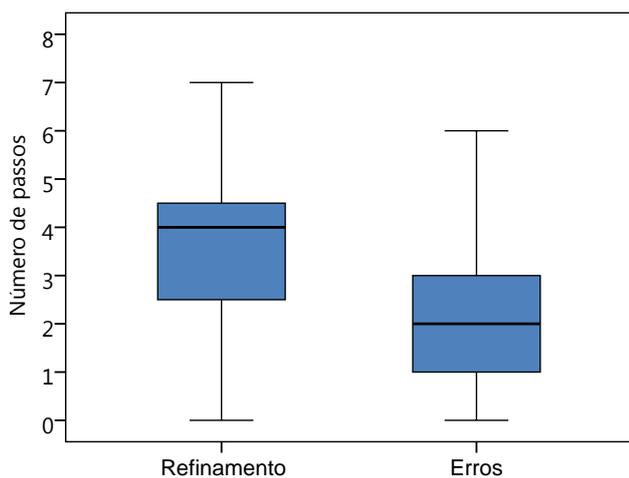


Figura 6. Número de passos por construção em que houve refinamento da pesquisa (esquerda) e erro na peça devolvida (direita).

Foi frequente observar que os participantes efectuaram diversos refinamentos às descrições. No gráfico da Figura 6 estão representados os números de passos da construção em que os participantes efectuaram pelo menos um refinamento da descrição, assim como o número de passos em que o fornecedor devolveu uma peça diferente da solicitada. Embora a grande maioria 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 total desde que o construtor começou a descrever a peça até ao momento em que o fornecedor lhe entregou a peça correcta, assim como o tempo despendido a descrever a peça pelo construtor e o tempo necessário pelo fornecedor para a encontrar. As médias dos tempos obtidos estão visíveis no gráfico da Fi-

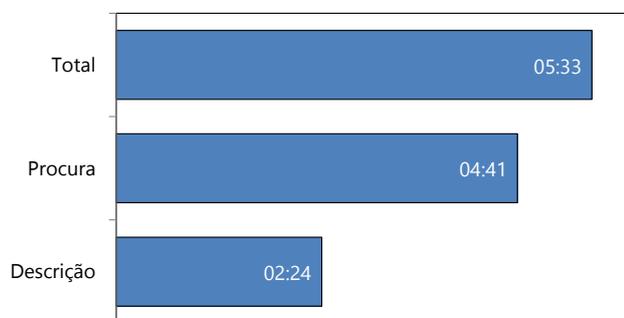


Figura 7. Tempos médios (mm:ss) por sessão para obter as peças (cima), para as descrever (baixo) e para as procurar (meio).

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

Tabela 2. Classificação da descrição das peças de forma verbal, gestual e combinada (Mediana, Amplitude interquartil).

gura 7. 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.

5.2 Medidas Subjectivas

No final das construções era solicitado aos participantes que preenchessem um questionário para classificar cada uma das modalidades - fala, gestos e combinação de fala e gestos - 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 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 ($Z=-3.456, p=0.001$).

5.3 Observações

Ao longo dos testes registámos os diversos aspectos sobre as descrições das peças efectuadas pelos participantes. 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 com a 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, como demonstrado na Figura 8.

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 participan-

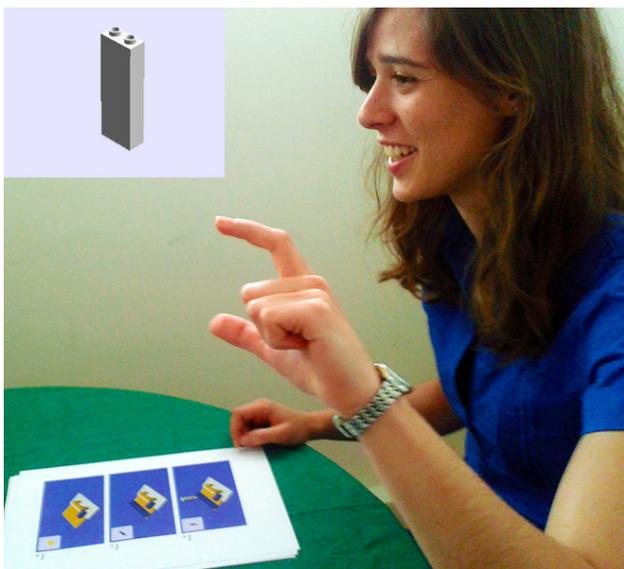


Figura 8. Participante complementa a descrição de uma peça alta com um gesto.

tes está ilustrada na Figura 9. 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. Existiram ainda outros exemplos de peças que originaram descrições dignas de nota. Para

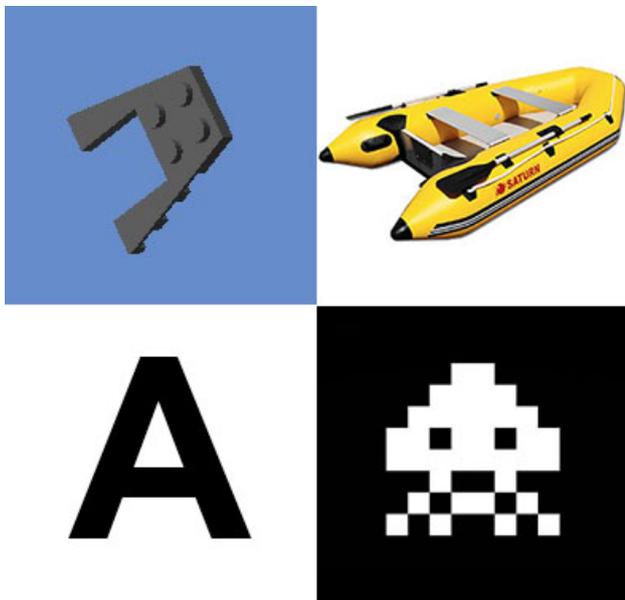


Figura 9. Três metáforas utilizadas pelos participantes para descrever a peça representada no canto superior esquerdo.

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.

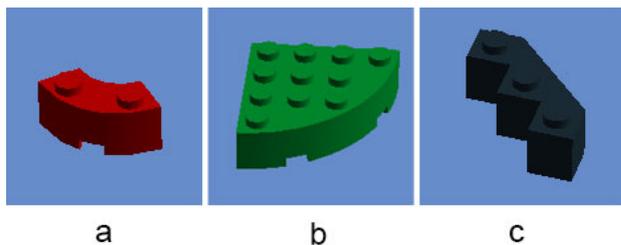


Figura 10. Conjunto de peças que levaram a descrições peculiares.

a peça da Figura 10-a, alguns participantes descreveram-na como um quarto da letra 'O' ou do muro de um poço. A peça da Figura 10-b foi comparada com um triângulo de um queijo, ou, mais frequentemente, com um quarto de um círculo ou um canto com rebordo redondo. Para a peça da Figura 10-c surgiram metáforas originais como dentes de tubarão ou o topo de um tridente.

6. 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. Outro aspecto interessante que observámos prende-se com as expressões utilizadas. Muito frequentemente os participantes recorreram a metáforas imaginativas, comparando as características mais relevantes das peças com outros objectos.

Como trabalho futuro, pretendemos aplicar os conhecimentos obtidos no desenvolvimento de um sistema de construção de modelos LEGO® virtuais, de forma a oferecer uma interface natural para a recuperação das peças desejadas. No entanto, as metáforas utilizadas pelos participantes, não obstante de se terem revelado uma grande fonte descritiva, poderão originar desafios acrescidos. Apesar de termos recorrido ao cenário concreto de peças LEGO®, acreditamos que as lições aprendidas com este trabalho poderão ser aplicadas em contextos diferentes que utilizem descrições naturais de objectos.

7. Agradecimentos

O trabalho apresentado neste artigo foi parcialmente financiado pela Fundação para a Ciência e a Tecnologia (FCT) através dos projetos CEDAR (referência PTDC/EIA-EIA/116070/2009) e Pest-OE/EEI/LA0021/2013.

Referências

- [FMK⁺03] 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*, 22(1):83–105, January 2003.
- [HW11] Christian Holz and Andrew Wilson. Data miming: inferring spatial object descriptions from human gesture. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 811–820, 2011.
- [KB10] Maryam Kamvar and Doug Beeferman. Say what? why users choose to speak their web queries. In *Interspeech*, 2010.
- [LK12] Cheongjae Lee and Tatsuya Kawahara. Hybrid vector space model for flexible voice search. In *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, pages 1–4. IEEE, 2012.
- [LLJ⁺13] Yong-Jin Liu, Xi Luo, Ajay Joneja, Cui-Xia Ma, Xiao-Lan Fu, and Dawei Song. User-Adaptive Sketch-Based 3-D CAD Model Retrieval. *IEEE Transactions on Automation Science and Engineering*, pages 1–13, 2013.
- [LRL10] C Lee, Alexander Rudnicky, and GG Lee. Let's buy books: finding ebooks using voice search. *Spoken Language Technology Workshop (SLT), 2010 IEEE*, pages 85–90, 2010.
- [LSF⁺11] H Laga, T Schreck, A Ferreira, A Godil, I Pratikakis, and R Veltkamp. Bag of words and local spectral descriptor for 3d partial shape retrieval. 2011.
- [MDPJW07] 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.
- [PLR05] Jiantao Pu, Kuiyang Lou, and Karthik Raman. A 2 d sketch-based user interface for 3 d cad model retrieval. *Computer-Aided Design and Applications*, 2(6):717–725, 2005.
- [PR99] Eric Paquet and Marc Rioux. Nefer-titi: a query by content system for three-dimensional model and image databases management. *Image and Vision Computing*, 17:157–166, 1999.
- [SC93] 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.
- [SC97] JR Smith and SF Chang. Visually searching the web for content. *MultiMedia, IEEE*, 1997.
- [SFDF08] Tiago Santos, Alfredo Ferreira, Filipe Dias, and Manuel J. Fonseca. Using sketches and retrieval to create lego models. *Proceedings of the Fifth Eurographics conference on Sketch-Based Interfaces and Modeling*, 2008.