Back to MARS: The unexplored possibilities in query result visualization

Alfredo Ferreira
INESC-ID/IST/TU Lisbon
Lisbon, Portugal
alfredo.ferreira@ist.utl.pt

Pedro B. Pascoal
INESC-ID/IST/TU Lisbon
Lisbon, Portugal
pmbp@ist.utl.pt

Manuel J. Fonseca
INESC-ID/IST/TU Lisbon
Lisboa, Portugal
mjf@inesc-id.pt

ABSTRACT

A decade ago, Nakazato proposed 3D MARS, an immersive virtual reality environment for content-based image retrieval. Even so, the idea of taking advantage of post-WIMP interfaces for multimedia retrieval was no further explored for content-based retrieval. Considering the latest low-cost, off-the-shelf hardware for visualization and interaction, we believe that is time to explore immersive virtual environments for multimedia retrieval. In this paper we highlight the advantages of such approach, identifying possibilities and challenges. Focusing on a specific field, we introduce a preliminary immersive virtual reality prototype for 3D object retrieval. However, the concepts behind this prototype can be easily extended to the other media.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Interaction Styles, Input Devices and Strategies*

Keywords

Multimedia Information Retrieval, 3D Object Retrieval, Immersive Virtual Environment

1. INTRODUCTION

Despite advances on multimedia information retrieval (MIR), this field still on its infancy. Especially when compared to its textual counterpart. Actual textual search engines are maturely developed and its widespread use makes them familiar to most users. The current scenario in MIR is quite different. Indeed, existing content-based MIR solutions are far from being largely used by the common user.

A few exceptional systems were able to strive with relative success, such as **Retrievr**[1], a search tool for **Flickr**[2] based on visual queries. However, most existing solutions still face major drawbacks and challenges to be tackled. Among others, extensively identified in Datta's survey [5], we highlight two. First, queries rely mostly on meta-information, often keyword-based. This means that, in a closer analysis, searches can be reduced to text information retrieval of multimedia objects. Second, the result visualization follows the traditional paradigm, where the results are presented as a list of items on a screen. These items are usually thumbnails, but can be just filenames or metadata. Such methodology greatly hinders the interpretation of query results on collections of videos or 3D objects.

Notably, a decade ago, a new visualization system for content-based image retrieval(CBIR) was proposed by Nakazato and Huang from the University of Illinois. The 3DMARS [11] was an immersive virtual reality (VR) environment to perform image retrieval. It worked on the NCSA CAVE [4] which provided fully immersive experience and later on desktop VR systems. However, despite this ground-breaking work and recent developments in the interaction domain, little advantages have been taken by the multimedia information retrieval community from immersive virtual environments.

In this paper we bring up the work of Nakazato and Huang as a starting point to the exploration of new possibilities for result visualization in multimedia information retrieval. With the spreading of stereoscopic viewing and last generation interaction devices outside lab environment and into our everyday lives, we believe that in a short time users will expect richer results from multimedia search engines than just a list of thumbnails. Following this rationale, and despite it could be applied to any type of media, we will focus our approach on 3D object retrieval (3DOR).

2. TRADITIONAL 3DOR APPROACHES

The first and most noticeable 3D search engine, at least within researchers working on this area, is the Princeton 3D Model Search Engine[8]. This remarkable work provide content-based retrieval of 3D models from a collection of more than 36000 objects. Four query specification options are available: text based; by example; by 2D sketch; and by 3D sketch. The results of this queries are presented as an array of model thumbnails.

Additionally to queries by example and sketch-based queries, the FOX-MIIRE search engine[1] introduced the query by

[1]http://labs.systemone.at/retrievr/
[2]http://www.flickr.com/

photo. This was the first tool capable of retrieve a 3D model from a photograph of a similar object. However, and similarly to Princeton engine, the results are displayed as a thumbnail list.

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

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

These traditional approaches to query specification and result visualization do not take advantage of latest advances of neither computer graphics or interaction paradigms. Current hardware and software are capable of handling millions of triangles per frame and generating complex effects in real-time. Additionally, the growingly common use of new human-computer interaction (HCI) paradigms and devices brought new possibilities for multi-modal systems.

3. NEW PARADIGMS IN HCI

The recent dissemination among common users of new HCI paradigms and devices (e.g. Nintendo Wiimote[4] or Microsoft Kinect[5]) brought new possibilities for multi-modal systems. For decades, the "windows, icons, menus, pointing device" (WIMP) interaction style prevailed outside the research field, while post-WIMP interfaces were being devised and explored [16], but without major impact in everyday use of computer systems.

Particularly, the use of gestures to interact with system has been part of the interface scene since the very early days. A pioneering multimodal application was "Put-that-there" [2], by Bolt. In "Put-that-there", the user commands simple shapes on a large-screen graphics display surface. This approach combined gestures and voice commands to interact with the system. However, just recently such interaction paradigm have been introduced in off-the-shelf commodity products.

Recent technological advances allowed development of low-cost, lightweight, easy to use systems. With limited resources, novel and more natural HCI can be developed and explored. For instance, Lee [10] used a Wiimote and took advantage of its high resolution infra-red camera to implement multipoint interactive whiteboard, finger tracking and head tracking for desktop virtual reality displays. Post-WIMP finally arrived to the masses.

[3]http://sketchup.google.com/3dwarehouse/
[4]http://www.nintendo.com/wii/console/controllers
[5]http://www.xbox.com/en-US/kinect

Figure 1: The interface of 3D MARS.

Generally, post-WIMP approaches abandoned the traditional mouse and keyboard combination, favouring devices with six degrees of freedom (DoF). Unlike traditional WIMP interaction style, where it is necessary to map the inputs from a 2D interaction space to a 3D visualization space, six DoF devices allow straightforward direct mapping between device movements and rotations and corresponding effects on the three-dimensional space. This represents an huge leap to the concept of direct manipulation, which, according to Shneiderman [14], rapidly increments operations and allows the immediate visualization of effects on an manipulated object. This helps making the interaction more comprehensible, predictable and controllable.

Combining six DoF devices with stereoscopy, it is possible to make a multi-modal immersive interaction with direct and natural manipulation of objects shapes within virtual environments. This may be experienced using immersive displays (e.g., HMDs, CAVEs) [7] or desktop [15].

Despite the growing interest around the application of this new paradigms in HCI, no relevant efforts were made to explore the latest technological advances for multimedia information retrieval. Indeed, to the extent of our knowledge, there has not been presented any research or new solution that take advantage of immersive virtual environments for information retrieval since Nakazato's 3D MARS [11] .

4. 3D MARS

The 3D MARS system demonstrates that the use of 3D visualization in multimedia retrieval has two benefit. First, more content can be displayed at the same time without occluding one another. Second, by assigning different meanings to each axis, the user can determine which features are important as well as examine the query result with respect to three different criteria at the same time.

Nakazato focused his work on query result visualization. Thus 3D MARS supports only query-by-example mechanism to specify the search. The user select one image from a list and the system retrieves and displays the most similar images from the image database in a 3D virtual space. The image location on this space is determined by its distance to the query image, where more similar images are closer to the origin of the space. The distance in each coordinate axis depend on a pre-defined set of features. The X-axis, Y-axis and Z-axis represent color, texture and structure of images respectively.

The interaction with the query results is done through a wand that the user holds while freely walking around the CAVE, as depicted in Figure 1. By wearing shutter glasses, the user can see a stereoscopic view of the world, which provides a full immersive experience. In such solution, visualizing query results goes far beyond scrolling on a list of thumbnails. The user navigates among the results in a three-dimensional space.

The 3D MARS was a catalyst for the incitement proposed in this paper: explore immersive visualization systems for multimedia information retrieval. Following that idea, we devised an immersive 3D virtual reality system for the display of query results of queries for 3D object Retrieval.

5. IMMERSIVE 3DOR

Taking advantage of the new paradigms in HCI, we propose an immersive VR system for 3D object retrieval (**Im-O-Ret**). The version of the system presented in this paper relies on a large-screen display, the LEMe Wall [6], and the a six DoF interaction device, the SpacePoint Fusion, an off-the-shelf device developed by *PNI Sensor Corporation*. However, minimal effort is required in order to have the system working in a context with HMD glasses or stereoscopic glasses, as well as using other input devices, such as Wiimote or Kinect.

Regardless of the hardware details, the **Im-ORet** allows the user to browse the results of a query to collection of 3D objects in an immersive virtual environment. The objects are distributed in the virtual 3D space according to their similarity. This is measured by the distance of each result to the query, which stands in the origin of the coordinates. To each of the three axis is assigned a different shape matching algorithm. The similarity to the query returned by the corresponding algorithm determines the coordinate. Current version of **Im-O-Ret** uses the Lightfield Descriptors [3] on the X-axis, the Coord and Angle Histogram [13] for the Y-axis, the Spherical Harmonics Descriptor [9] for the Z-axis. Figure 2 illustrates a user browsing the results of a query.

5.1 Possibilities

Similar to the 3D MARS, this work opens a myriad of new possibilities. By assigning different shape matching algorithms to each axis, one can adapt the query mechanism to specific domains, producing more precise results. Applying transparency to results, it is possible to overlay results of distinct queries. Adding effects to results, such as glow or special colors, it order to convey additional information.

Since query results are not images or thumbnails, but three-dimensional models, it is possible to navigate around them in the virtual environment and even manipulate them. Moreover, instead of a static view of the result, displaying it as a 3D object that can be rotating over one axis, offers a better perception of the model. Adding stereoscopy will improve

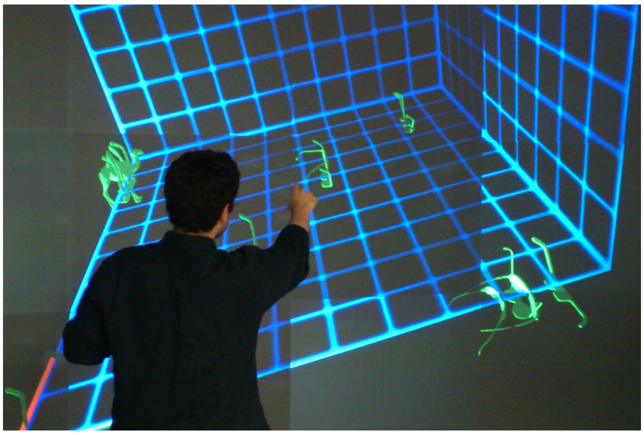

Figure 2: User exploring query results in Im-O-Ret

even more the visualization since the user gains depth perception over the environment.

The combined use of VE and devices with six DoF, provides a more complete visualization and makes interaction more natural, comprehensible and predictable. Their use, will also add some challenges to the implementation of such system.

5.2 Challenges

While in traditional 3DOR systems the query results are represented and ordered as a list of thumbnails ordered by a given similarity measure, when we move to a virtual environment, the distribution of results in a 3D space becomes a challenge. How query results should arranged in 3D space to be meaningful to the user remains an open question. In our approach we select three shape descriptors and assigned each one to a coordinate axis, but this is a preliminary approach. We believe that a final solution is more complex that this. Further investigation on this topic is clearly required.

On the other hand, the way users navigate and interact with objects in an immersive environment and interact with it still an open issue. Norman[12] stated that gesturing is a natural, automatic behaviour, but the unintended interpretations of gestures can create undesirable states. Having this in mind, it is important to aim for an interface that is both predictable and easy to learn.

Above all, an important challenge remains open. No easy query specification mechanism has been presented, neither in traditional search engines, nor with new HCI paradigms. Although sketch-based queries apparently provide good results, they greatly depend on the ability of the user to draw a 3D model, which hinders the goal of a widely used, content-based, 3D search engine.

6. CONCLUSIONS

We believe that recent advances in low-cost, post-WIMP enabler technology, can be seen as an opportunity to overcome some drawbacks of current multimedia information retrieval solutions. Combined with the dissemination of stereoscopic visualization as a commodity, these interaction paradigms will acquaint common users with immersive virtual reality environments.

In this paper we highlight that such scenario is a fertile ground to be explored by search engines for multimedia information retrieval. In that context, we identified two major research topics: query result visualization and query specification. While the latest requires further study, we already started tackling the first one.

We developed a novel visualization approach for 3D object retrieval. The **Im-O-Ret** offers the users an immersive virtual environment for browsing results of a query to a collection of 3D objects. The query results are displayed as 3D models in a 3D space, instead of the traditional list of thumbnails. The user can explore the results, navigating in that space and directly manipulating the objects.

Looking back to 3D MARS, the initial work proposed by Nakazaro, we realize it was a valid idea that fell almost into obliviousness. We expect that our preliminary work, which lies over concepts introduced by 3D MARS, could prove the goodness of our incitement to explore the possibilities offered by immersive virtual environments to the multimedia information retrieval.

7. ACKNOWLEDGMENTS

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

8. REFERENCES

[1] T. F. Ansary, J.-P. Vandeborre, and M. Daoudi. 3d-model search engine from photos. In *Proceedings of the 6th ACM international conference on Image and video retrieval*, CIVR '07, pages 89–92, New York, NY, USA, 2007. ACM.

[2] R. A. Bolt. Put-that-there: Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '80, pages 262–270, New York, NY, USA, 1980. ACM.

[3] D.-Y. Chen, X.-P. Tian, Y. te Shen, and M. Ouhyoung. On visual similarity based 3d model retrieval. volume 22 of *EUROGRAPHICS 2003 Proceedings*, pages 223–232, 2003.

[4] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the cave. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '93, pages 135–142, New York, NY, USA, 1993. ACM.

[5] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40:5:1–5:60, May 2008.

[6] B. R. de AraÃžjo, T. Guerreiro, R. J. Costa, J. A. P. Jorge, and J. M. Pereira. Leme wall: Desenvolvendo um sistema de multi-projecção. 13Âž Encontro PortuguÃłs de ComputaÃğÃčo GrÃąfica, Vila Real, Portugal, 2005.

[7] T. DeFanti, D. Acevedo, R. Ainsworth, M. Brown, S. Cutchin, G. Dawe, K.-U. Doerr, A. Johnson, C. Knox, R. Kooima, F. Kuester, J. Leigh, L. Long, P. Otto, V. Petrovic, K. Ponto, A. Prudhomme, R. Rao, L. Renambot, D. Sandin, J. Schulze, L. Smarr, M. Srinivasan, P. Weber, and G. Wickham. The future of the cave. *Central European Journal of Engineering*, 1:16–37, 2011. 10.2478/s13531-010-0002-5.

[8] T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, and D. Jacobs. A search engine for 3d models. *ACM Trans. Graph.*, 22:83–105, January 2003.

[9] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Rotation invariant spherical harmonic representation of 3d shape descriptors. In *Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, SGP '03, pages 156–164, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association.

[10] J. Lee. Hacking the nintendo wii remote. *Pervasive Computing, IEEE*, 7(3):39 –45, july-sept. 2008.

[11] T. S. H. Munehiro Nakazato. 3d mars: Immersive virtual reality for content-based image retrieval. In *Proceedings of 2001 IEEE International Conference on Multimedia and Expo (ICME2001)*, 2001.

[12] D. A. Norman. Natural user interfaces are not natural. *interactions*, 17:6–10, May 2010.

[13] E. Paquet and M. Rioux. Nefertiti: a query by content software for three-dimensional models databases management. In *NRC 97: Proceedings of the International Conference on Recent Advances in 3-D Digital Imaging and Modeling*, page 345, Washington, DC, USA, 1997. IEEE Computer Society.

[14] B. Shneiderman. Direct manipulation for comprehensible, predictable and controllable user interfaces. In *Proceedings of the 2nd international conference on Intelligent user interfaces*, IUI '97, pages 33–39, New York, NY, USA, 1997. ACM.

[15] B. Sousa Santos, P. Dias, A. Pimentel, J.-W. Baggerman, C. Ferreira, S. Silva, and J. Madeira. Head-mounted display versus desktop for 3d navigation in virtual reality: a user study. *Multimedia Tools Appl.*, 41:161–181, January 2009.

[16] A. van Dam. Post-wimp user interfaces. *Commun. ACM*, 40:63–67, February 1997.