Gesture Interaction for Content–based Medical Image Retrieval

Antoine Widmer
HES–SO Valais
3960 Sierre, CH
antoine.widmer@hevs.ch

Roger Schaer
HES–SO Valais
3960 Sierre, CH
roger.schaer@hevs.ch

Dimitrios Markonis
HES–SO Valais
3960 Sierre, CH
dimitrios.markonis@hevs.ch

Henning Müller
HES–SO Valais
3960 Sierre, CH
Henning.mueller@hevs.ch

ABSTRACT

Large amounts of medical images are being produced to help physicians in diagnosis and treatment planning. These images are then archived in PACS (Picture Archival and Communication Systems) and usually they are only reused in the context of the same patient during further visits. Medical image retrieval systems allow medical professionals to search for images in institutional archives, the Internet or in the scientific literature. The goal of the search can be in diagnosis but often as well for teaching and research. A large body of research has investigated efficient and effective algorithms to retrieve a set of images to fulfil a specific information need. However, much less research has been done on studying simple and engaging interaction for users of medical image retrieval systems. In this paper we propose an intuitive and engaging web–based interface targeted to be used by a large range of users with gesture control. This interface allows users to retrieve medical images by accessing a system called Parallel Distributed Image Search Engine (ParaDISE), a text– and content–based image retrieval system. Accepting search with keywords and example images, this interface uses simple gestures to get random example images and mark examples as positive and negative relevance feedback with results being updated after each interaction.

Categories and Subject Descriptors

H.5.2 [**Information Systems**]: Information Interfaces and Presentation—*User Interfaces*; H.3.5 [**Information Systems**]: Information Storage and Retrieval—*Online Information Services*

Keywords

web based interface, HCI, gesture interaction

1. INTRODUCTION

Medical images play an essential role in diagnoses and treatment planning for many diseases [5]. The number of images produced by hospitals has been growing rapidly over the past 20 years (for example to an average of 300'000 images per day in the Geneva University hospitals in 2013). To manage this growing number of produced images, content–based image retrieval (CBIR) systems are proposed to search these images using the visual content, so without requiring manual annotation with key words. Studies show that CBIR systems can improve diagnostic quality of radiologists, particularly with little experience [1, 10]. Many papers have been published describing algorithms to retrieve images based on text and/or visual features [10, 13]. Such CBIR systems can be complex to interact with when multiple parameters need to be set in order to refine the search for a particular query session [4]. Only little effort has been spent on simplifying interaction for retrieval systems to make them more intuitive and engaging for users [10]. Using the PARAllel Distributed Image Search Engine (ParaDISE), a retrieval system combining text and content–based medical image retrieval, this paper presents a novel web–based interface allowing users to retrieve medical images with simple gestures and no specific knowledge about image retrieval parameters.

Much has been published on medical image retrieval systems over the past decade and performance as well as speed are improving [4]. Benchmarks exist that measure and compare performance across research teams every year [12], and have done so for the past ten years. However, the majority of available interfaces for image retrieval systems are targeted towards expert users who know how to optimise retrieval quality and set parameters. Therefore, image retrieval systems can sometimes be complex to interact with.

Resulting from human computer interaction research, many new tools can reduce the complexity of image retrieval or other tasks by providing more engaging ways of interaction. For example, researchers have used an eye–tracking system for image retrieval in [3]. Users have to stare at an image for 2 seconds and the system then shows similar images from the database. This system only permits to select a single image to find corresponding images and is thus limited because it does not allow other actions to refine search. As another engaging way for interaction, gesture recognition allows users to interact with a computer without touching a mouse or a keyboard. This is particularly interesting for applications

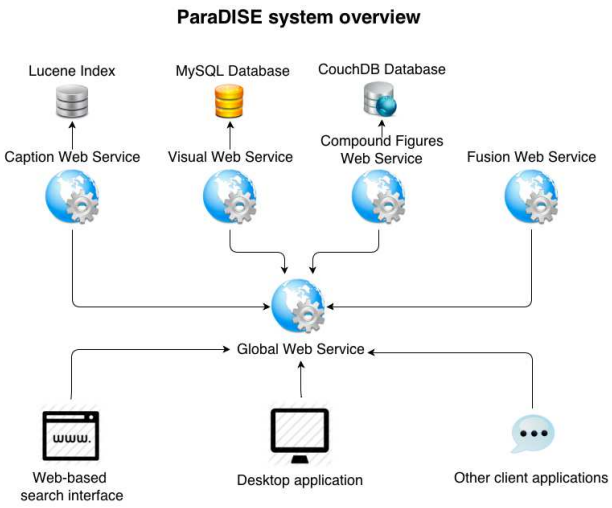

Figure 1: The ParaDISE service layer architecture.

that need to be used in public settings or in hospitals where sterility is essential. However, this kind of interaction could also be useful for more general navigation in big data such as in oil and gas exploration. Most gesture interaction research has been carried out using the Kinect from Microsoft [6, 9]. The most important drawback of using Kinect in medical image retrieval systems appears to be the relatively distant range (between 40 cm and 300 cm) in which the Kinect can recognise gestures [6]. This range does not allow the Kinect to be used for close computer interaction frequently needed in retrieval systems. Our approach uses a different device (called Leap motion (tm)) tracking fine motion of up to 10 fingers in a very close range (between 1 cm and 50 cm) [14]. This device allows the user to finely interact with a computer with a precision of below 1 mm. Currently, very little research involving the device has been published, due to its recent availability. With this device, we present a novel and engaging way to browse through and view medical images returned by a medical image retrieval system in an intuitive way.

2. SYSTEM DESCRIPTION

In this paper, we focus on the description of a novel web–based interface displaying images returned by the ParaDISE retrieval system and allowing gesture interaction. We interface with ParaDISE via a simple access composed of multiple underlying web services accessible through the Hyper Text Transfer Protocol (HTTP). The system structure with its components is illustrated in Figure 1. ParaDISE includes the following services:

- The *caption* web service is responsible for search by keywords and uses Lucene.
- The *visual* service extracts visual information of the images in order to retrieve similar matches using CBIR.
- The *compound figures* web service is used to retrieve links between compound figures (multiple images in a single figure) and their parts (obtained by separating the compound figure into its subparts), as well as getting the coordinates of each child within the figure.

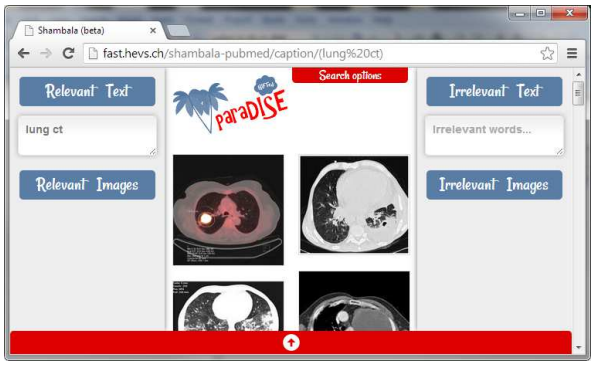

Figure 2: The web–based interface of Shambala.

- The *fusion* web service combines results from various sources (text and image search, for example) into a single list based on a given fusion rule.
- The *global* web service acts as a proxy for client applications, allowing hiding the complexity of calling several services by providing a general image search method taking text, images or both as input.

ParaDISE was used to ImageCLEF 2013 data [7] containing more than 300'000 images of the medical literature (obtaining best results in modality classification and second best for visual image retrieval). The CBIR tecniques develped by medGIFT [11] was used, as it is consistently among the top runs in ImageCLEF [11, 8]. A filtering by image modality was included using the modality classification proposed in [8]. Since ParaDISE is accessible as a web service, it is possible to use it via a variety of interfaces such as Khresmoi for professionals[1]. In this context, we created Shambala, a simplified web–based interface for search using the ParaDISE services. Our primary goal in developing Shambala was to make it as simple and engaging as possible to interact for a large range of users with simple drag and drop.

To achieve our goal, we decided to keep the number of user–adjustable options and parameters as small as possible and have them hidden for more advanced users. However, the user still has several ways of searching for example images that are relevant such as "keyword search", "image search" and "caption information search". "Keyword search" allows the user to provide both "relevant" and "irrelevant" terms in text boxes to find images related to a given subject such as pathology, body part, etc. "image search" allows the user to provide relevant and irrelevant images directly from the search results, from other databases accessible via the Internet or the user can upload his/her own images via a simple drag and drop into the interface. With "Caption information search", the user can choose to use the caption information of positive and negative example images as additional input for the relevance feedback in the form of text.

Figure 2 depicts the user interface of Shambala (in this case on a small screen). The user interface is composed of 3 main areas in vertical direction. The central area presents the set of images returned by ParaDISE in an efficient way. The left and right areas consist of the drop zones for keywords and images that are relevant and irrelevant to the cur-

[1] `http://professional.khresmoi.eu/`

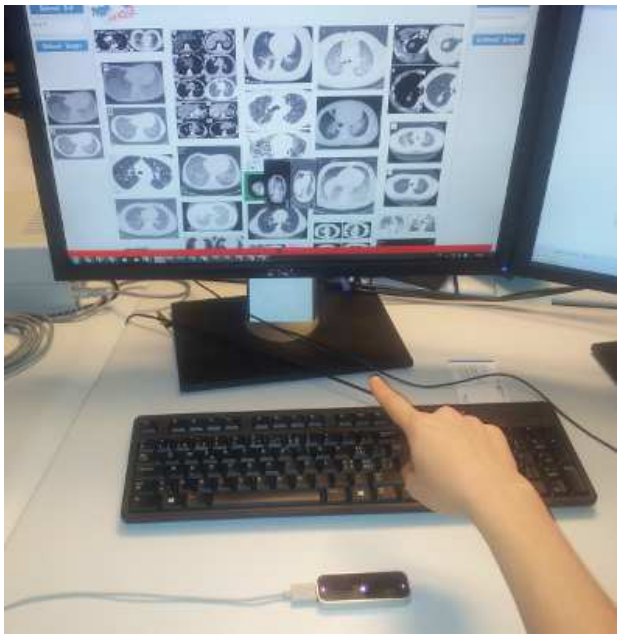

Figure 3: Gesture interaction with Shambala.

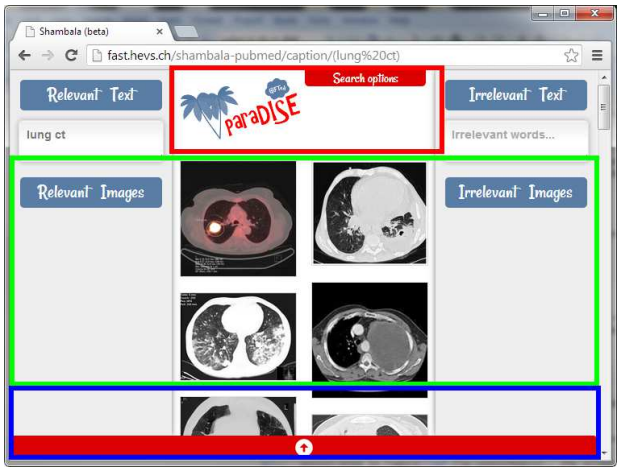

Figure 4: Logical regions to navigate and interact with returned images from ParaDISE. When the cursor is placed in the top red region or the bottom blue region, the list of images is scrolled up or scrolled down, respectively. When the cursor is within the green region, the user can interact with the images (zoom + selection).

rent image search. To create an intuitive and engaging interaction, a hand and finger motion sensor (Leap Motion(tm)) tracks the user gestures. Nevertheless, a keyboard is still used to enter keywords for relevant and irrelevant terms. The caption feedback makes the use of text possible without a keyboard.

2.1 Gesture interaction

To engage the user to interact with Shambala, we developed a set of easy–to–learn gestures to allow browsing through images, to show the image in full size and give the focus to the text areas to enter keywords as illustrated in Figure 3. Ensuring the simplicity of interaction, this set of gestures was derived from another set of gestures widely used in touchscreen interaction [2]. To increase the number of actions that a user can perform, we decomposed gestures into the two following states:

Navigation/browsing images state: In this state, the user can browse the list of images returned by ParaDISE by moving a cursor on the webpage using a finger. As illustrated in Figure 4, the screen is divided into 3 horizontal parts. Placing the cursor in the top and bottom parts allows the user to scroll up and down in the list of results returned by ParaDISE. When the cursor is located in the central part, the user can zoom in on the image he/she is pointing at by pulling his/her finger towards him. This state allows the user to quickly navigate through the set of images, check selected images with instant zooming. In addition, the user is allowed to give the focus to text boxes by moving the cursor over them to enter "relevant" or "irrelevant" keywords. As soon as the text is typed, a new search with the new keywords is performed automatically.

Image selection state: A feature of ParaDISE is the search of images using visual features of positive and negative examples. For this, the user needs to select images that are relevant and irrelevant to a current search. The selection of images is achieved by pointing the finger at the chosen image in the middle area depicted in Figure 4 and pushing the finger towards the screen until the frame of the image turns green. At this stage the image can be dragged into the left area (relevant images) or into the right area (irrelevant images) to refine the search as shown in Figure 5. As soon as the image is dropped on one side of the screen, a new search is performed and the interface goes back to the navigation/browsing image state. Finally, the selection can be canceled by simply pulling the finger close to the body and the interface goes back to the Navigation/browsing image state.

3. ANALYSIS AND USER TEST

Shambala, the presented web–based interface, offers a simple way to search for medical images with intuitive interaction. In principle, Shambala can be used with a combination of the keyboard to enter keywords and a mouse to browse the set of images and to drag and drop selected images. In this combination, Shambala already achieves simplicity for user interaction. Clear separations of the dropping and viewing zones as illustrated in Figure 2 allows the user to quickly understand how the interface is working. The addition of gesture interaction enables to investigate a new and more engaging interaction for the users that has also advantages in sterile environments where touching anything should be avoided, such as in an operation room. In the field of image retrieval, Leap Motion allows interactions that are more difficult to achieve with a mouse and a keyboard, e.g., fast visual comparison of two images using both hands to manipulate both images simultaneously.

A preliminary analysis explores the advantages and drawbacks associated with the use of gesture interaction in an interface for content–based image retrieval systems. As first indication, we asked 2 colleagues working in a different research area to search images using the simple gestures described earlier. Their feedback was very positive and they

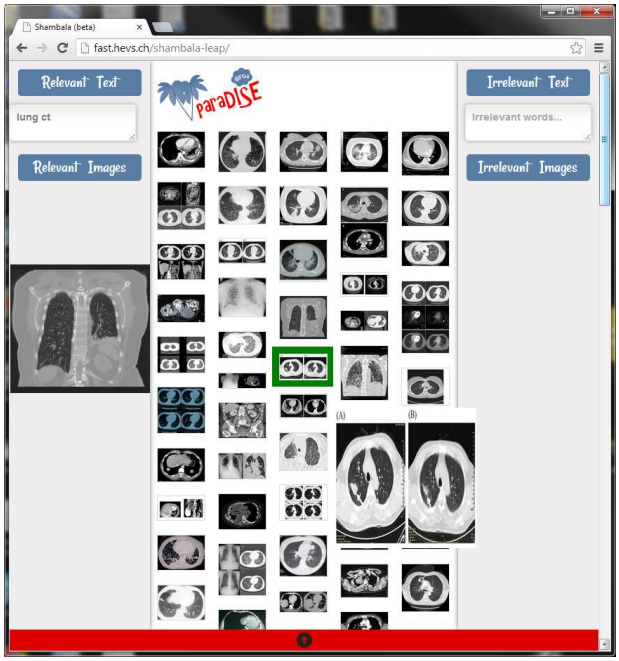

Figure 5: Placing an image in the irrelevant images drop zone on the right.

quicked learned the interaction principles. The main drawbacks consisted of the fatigue induced by holding one arm up in the air above the motion sensor when using it over a longer period and intermittent slow responses from ParaDISE. Nonetheless, a more extensive study involving a larger number of participants has to be carried out to validate the current interface and to investigate possible improvements in terms of user satisfaction or speed of the interaction.

4. CONCLUSIONS

The novel web–based interface for medical image retrieval presented in this paper creates a way of interacting with gesture control when searching for medical images. However, we have already identified several areas for improvements: (1) a keyboard is currently still needed for starting text–based search. A better way to engage users could consist of providing a short list of keywords that can be dragged to the relevant or irrelevant side and that can be linked to the context of a query session, making the interface usable without a keyboard. (2) As highlighted by the pilot study, there is room to improve the reactivity of the system. Users feel that response times are slow when result display takes over 5 seconds. (3) A wider user study needs to be carried out to investigate in detail the advantages and drawbacks of using gesture–based interaction in a medical image retrieval system. Particularly in the lexical field such gesture control can be extremely useful in sterile environments, where this can have a real added value.

5. REFERENCES

[1] A. M. Aisen, L. S. Broderick, H. Winer-Muram, C. E. Brodley, A. C. Kak, C. Pavlopoulou, J. Dy, C.-R. Shyu, and A. Marchiori. Automated storage and retrieval of thin–section CT images to assist diagnosis: System description and preliminary assessment. *Radiology*, 228(1):265–270, July 2003.

[2] A. Bragdon, E. Nelson, Y. Li, and K. Hinckley. Experimental analysis of touch-screen gesture designs in mobile environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 403–412, New York, NY, USA, 2011. ACM.

[3] J. Coddington, J. Xu, S. Sridharan, M. Rege, and R. Bailey. Gaze-based image retrieval system using dual eye–trackers. In *IEEE International Conference on Emerging Signal Processing Applications (ESPA)*, pages 37–40, 2012.

[4] A. Depeursinge, B. Fischer, H. Müller, and T. M. Deserno. Prototypes for content–based image retrieval in clinical practice. *The Open Medical Informatics Journal*, 5:58–72, 2011.

[5] K. Doi. Current status and future potential of computer–aided diagnosis in medical imaging. *British Journal of Radiology*, 78:3–19, 2005.

[6] L. Gallo, A. P. Placitelli, and M. Ciampi. Controller–free exploration of medical image data: Experiencing the kinect. In *24th International Symposium on Computer–Based Medical Systems (CBMS)*, pages 1–6, 2011.

[7] A. García Seco de Herrera, J. Kalpathy-Cramer, D. Demner Fushman, S. Antani, and H. Müller. Overview of the ImageCLEF 2013 medical tasks. In *Working Notes of CLEF 2013 (Cross Language Evaluation Forum)*, September 2013.

[8] A. García Seco de Herrera, D. Markonis, I. Eggel, and H. Müller. The medGIFT group in ImageCLEFmed 2012. In *Working Notes of CLEF 2012*, 2012.

[9] C. Kirmizibayrak, N. Radeva, M. Wakid, J. Philbeck, J. Sibert, and J. Hahn. Evaluation of gesture based interfaces for medical volume visualization tasks. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*, pages 69–74, Hong Kong, China, 2011. ACM.

[10] L. R. Long, S. Antani, T. M. Deserno, and G. R. Thoma. Content-based image retrieval in medicine: Retrospective assessment, state of the art, and future directions. *International Journal of Healthcare Information Systems and Informatics*, 4(1):1–16, January 2009.

[11] D. Markonis, A. García Seco de Herrera, I. Eggel, and H. Müller. The medGIFT group in ImageCLEFmed 2011. In *Working Notes of CLEF 2011*, 2011.

[12] H. Müller, A. García Seco de Herrera, J. Kalpathy-Cramer, D. Demner Fushman, S. Antani, and I. Eggel. Overview of the ImageCLEF 2012 medical image retrieval and classification tasks. In *Working Notes of CLEF 2012 (Cross Language Evaluation Forum)*, September 2012.

[13] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content–based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.

[14] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler. Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, 13(5):6380–6393, 2013.