GAZE-BASED IMAGE RETRIEVAL SYSTEM USING DUAL EYE-TRACKERS

James Coddington, Junxia Xu, Srinivas Sridharan, Manjeet Rege, and Reynold Bailey

Department of Computer Science, Rochester Institute of Technology

{jxc6857, jsx4189, sxs9716, mr, rjb}@cs.rit.edu

ABSTRACT

In this paper we present a novel gaze-based image retrieval application. The application is designed to be run on a dual monitor setup with a separate eye tracking device dedicated to each monitor. A source image is displayed on one monitor and the retrieved images are displayed on the second monitor. The system is completely gaze controlled. The user selects one or more objects or regions in the source image by fixating on them. The system then retrieves images containing similar objects from an image database. These are displayed in a grid on the second monitor. The user can then fixate on one of these images to select it as the new source image and the process can be repeated until a satisfactory image is found.

Index Terms— Content-based image retrieval, eye-tracking

1. INTRODUCTION

Within the field of content-based image retrieval (CBIR), a wide variety of solutions have been proposed to perform efficient image retrieval [1]. While many of these solutions focussed on performing retrieval based on low-level feature similarity of images [2, 3], it was soon realized that the performance of these systems was limited due to the semantic gap [4] as they were unable to infer the interest of the user. To overcome this, most content-based image retrieval systems typically utilize mouse-clicks and other traditional forms of input to identify the regions or objects of interest. In this paper, we instead utilize eye-tracking as a control mechanism for content-based image retrieval. Figure 1 shows a user interacting with our gaze-based image retrieval system.

2. BACKGROUND

Eye tracking systems first emerged in the early 1900s [5, 6] (see Jacob and Karn [7] for a review of the history of eye-tracking). Until the 1980s, eye trackers were primarily used to collect eye movement data during psychophysical experiments. This data was typically analyzed after the completion of the experiments. During the 1980s, the benefits of real-time analysis of eye movement data were realized as eye-trackers evolved as a channel for human-computer interaction [8, 9]. Real-time eye tracking has also been used in interactive graphics applications [10, 11, 12, 13] and large scale display systems [14] to improve computational efficiency and perceived quality.

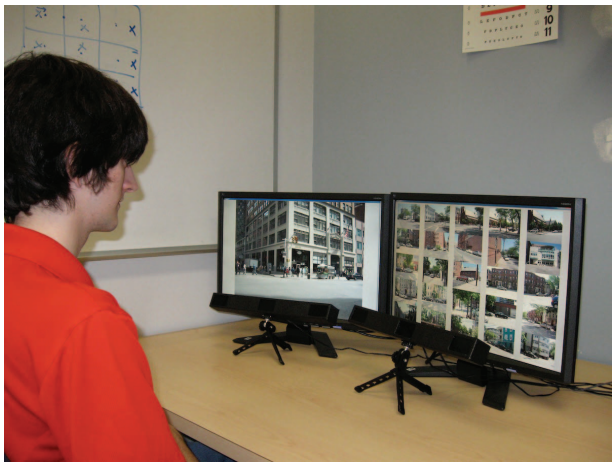

Fig. 1. User interacting with our gaze-based image retrieval system. The monitor on the left displays the source image and the monitor on the right displays the retrieved images. Each monitor has a dedicated eye-tracking device.

There have been several previous efforts which use eye tracking within the context of image retrieval. Oyekoya and Stentiford [15] used eye-tracking during an image search task to find a target image in a database of 1000 images with pre-computed similarity measures. Klami et al. [16] and Zhang et al. [17] explored whether eye movement measures such as fixation duration, fixation count, and number of revisits could be used to infer the relevance of images during image retrieval. Zhen [18] combined eye-tracking with visual features of segmented regions to perform content-based image retrieval. Our approach also uses segmented images, however in our case the images come from a large database of images which were manually segmented and labeled by humans. Different databases or search algorithms can be easily be incorporated into our system making it a useful platform for gaze-based image retrieval research.

3. SYSTEM DESCRIPTION

Our content-based image retrieval system uses eye tracking to identify scene content that a user considers to be important. A large database of images is then searched to find similar content. Our application is designed to be run on a dual monitor setup with the source image on one monitor and the retrieved images displayed on the second monitor.

3.1. Working With Two Eye-Trackers

A search of literature on gaze-controlled applications revealed that little work has been done with dual monitors. Räihä and Špakov [19] demonstrated a two-monitor setup using a single eye-tracker. In their case, the monitors were positioned side-by-side with no angle between them as shown in Figure 2 (a). Unfortunately, such a setup suffers from the well established problem of degraded accuracy in the extreme peripheral regions of the field of view [20]. To overcome this limitation we propose a configuration where each monitor has a dedicated eye-tracker as shown in Figure 2 (b). In this configuration, the user also has the freedom to angle the monitors for more comfortable viewing.

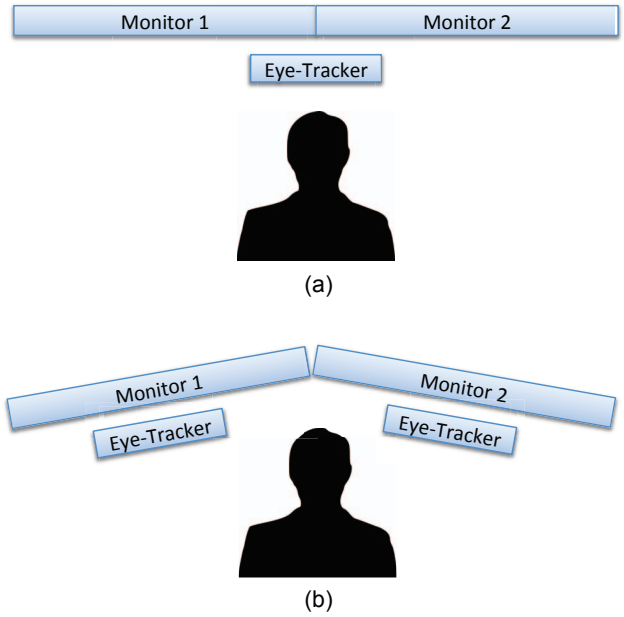

Fig. 2. Two possible configurations for eye-tracking with two monitors. (a) Two monitors and a single eye-tracker. (b) Two monitors with dedicated eye-trackers.

We utilize two Mirametrix S1 eye-trackers each operating at 60 Hz with gaze position accuracy less than 1 degree. The eye-trackers use infrared illumination and an infrared camera to record video of the observer's eyes. The video is analyzed to locate the corneal reflection and pupil center and this information is used to determine the location on the screen where the observer is looking. The eye-trackers were connected to a desktop computer with a 2.8 GHz Intel Core i7 4-core processor and 12 GB of RAM. Our image retrieval application was written using Matlab.

For the Mirametrix S1 (and most eye-trackers, in general) only one instance of the eye-tracking software is allowed to run at a time. Part of the reason for this is that there has never been a real demand for multiple eye-trackers connected to a single system. We overcome this problem by loading one instance of the eye-tracking software within a user account on Microsoft Windows 7. We then switch users, leaving the ports active, and load another instance in the second user account with different port numbers. This allows us to access both eye-trackers from the second account.

Fig. 3. This image contains: car, window, door, tree, building, manhole, car occluded, sidewalk, sky, street sign, awning, street light, sign, tractor, fire hydrant, telephone pole, street lamp, security screen, guard rail, balcony, man, cooler, potted plant, cone, van, door

3.2. LabelMe Database

The LabelMe image database [21] used for this project contains several thousand annotated images that span many categories. Visitors to the LabelMe website are asked to view images in the database and draw and label polygons around objects that they see in the image. This results in several annotations associated with each image. This annotation can be searched to see if a label exists in the image. For example, Figure 3 shows the polygons overlaid onto the image and the corresponding items labeled. The LabelMe database was chosen as the testbed for our image retrieval platform since it provides ground-truth labels. Different databases or search algorithms can be easily be incorporated into our system.

3.3. User Interaction

In order for a user to interact with our system, we first perform two standard 9-point calibrations - one for each eye-tracker. The Mirametrix S1 trackers tolerate some degree of head movement but we still ask users of our system to try to limit their head movement to only rotations between the two screens in order to ensure accurate eye-tracking.

A source image from the database is displayed on one monitor and the retrieved images are displayed on the second monitor. When the application is launched, twenty five randomly selected images from the LableMe database are displayed on the second monitor. The user can choose to fixate on one of these images for 2 seconds to load it as the source image on the first monitor.

We initially experimented with two possible modes of gaze-based interaction to select regions or objects of interest in the source image:

- **Fixation duration:** In this mode, the user is given 8 seconds to look at the image. The object receiving the largest percentage of fixation time is chosen as the target for the image retrieval process.
- **Dwell-time:** Fixations during scene exploration typically last for 200-300 ms. For the dwell-time mode of interaction, the user can select objects or regions in the image by fixating for a slightly longer duration of 1 second. This can be repeated to select multiple regions or objects in the scene. The user can fixate anywhere on the screen for two seconds to end input and begin the image retrieval process.

We eventually abandoned the fixation duration mode since the dwell-time approach provided better control. Alternative methods of gaze-based interaction can easily be incorporated into our system.

Once the user has selected objects or regions of interest in the image, the system performs a search based on the labels associated with those objects. As matching images are found, they are sequentially added to a 5 X 5 grid which is displayed on the second monitor. The search stops after twenty five images have been found or the database is exhausted. The user can then fixate on one of these images to select it as the new source image and the process can be repeated until a satisfactory image is found. Figure 4 shows a source image and Figure 5 shows the twenty five retrieved images. In this case the user selected a region containing an object labeled as 'building'.

Fig. 4. Screenshot of one monitor showing an example source image from the LabelMe database. The user selects objects or regions of interest using the dwell-time technique.

Fig. 5. Screenshot of second monitor showing the layout of the twenty five retrieved images from the LabelMe database.

4. CONCLUSIONS AND FUTURE WORK

The novel dual-eye-tracker image retrieval system presented in this provides an excellent platform for conducting research on gaze-based image retrieval and as well as content-based image retrieval in general. We have identified several areas for improvement in the near future:

- Although the LabelMe database provides ground-truth labels for images. We have found that the labels are not always consistent. For example, one person might label a region 'man' while another labels a similar region as 'guy'. The current implementation cannot tell that the two are similar. We plan to modify the search process to also utilize synonyms of the labels corresponding to the selected regions. While this approach will not eliminate the problem entirely, it should improve the quality and number of successful matches.
- There are also opportunities to improve the speed of the image retrieval process. Currently we perform a naive sequential search through the database using a single CPU thread which can become a performance bottleneck as the size of the database and the number of target

labels increase. We partially overcome this issue by terminating the search after the first twenty five matches have been retrieved. However, this means that we may be omitting better or more pleasing images from among the set of retrieved images. We plan to utilize GPU hardware to increase the speed of the search process.

- Related to the last point, is the need for a more effective interface for displaying the retrieved images so they are not limited to some fixed number. One idea would be to reserve the top left and bottom right grid position for 'backward' and 'forward' button that display the previous and next screens of retrieved images respectively. We will also explore other gaze-based interfaces such as the one proposed by Kozma et al. [22] which uses concentric rings of images and allows the user to zoom in to reveal more images.

- Finally, since the LabelMe database provides multiple ground-truth exemplars of various objects we plan to utilize this to explore the automatic retrieval and annotation of images from other databases including web-based databases such as Google Images [23] and Flickr [24].

5. REFERENCES

[1] Ritendra Datta, Dhiraj Joshi, Jia Li, James, and Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, pp. 5:1–5:60, 2008.

[2] Myron Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele, and Peter Yanker, "Query by image and video content: The qbic system," *Computer*, vol. 28, pp. 23–32, September 1995.

[3] Wei-Ying Ma and B. S. Manjunath, "Netra: a toolbox for navigating large image databases," *Multimedia Syst.*, vol. 7, pp. 184–198, May 1999.

[4] Changhu Wang, Lei Zhang, and Hong-Jiang Zhang, "Learning to reduce the semantic gap in web image retrieval and annotation," in *proc. of ACM SIGIR*.

[5] Raymond Dodge and Thomas Sparks Cline, "The angle velocity of eye movements," *Psychological Review*, vol. 8, no. 2, pp. 145 – 152, 152a, 153–157, 1901.

[6] E. B. Huey, *The Psychology and Pedagogy of Reading*, Cambridge, MA: MIT Press., 1968, (Originally published 1908).

[7] R. J. K. Jacob and K. S. Karn, *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, chapter Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary), pp. 573–605, Elsevier Science, Amsterdam, 2003.

[8] J. L. Levine, "An eye-controlled computer," Research Report RC-8857, IBM Thomas J. Watson Research Center, Yorktown Heights, N.Y., 1981.

[9] T. E. Hutchinson, K. P. White, W. N. Martin, K. C. Reichert, and L. A. Frey, "Human-computer interaction using eye-gaze input," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 19, no. 6, pp. 1527–1534, 1989.

[10] David Luebke, Benjamin Watson, Jonathan D. Cohen, Martin Reddy, and Amitabh Varshney, *Level of Detail for 3D Graphics*, Elsevier Science Inc., New York, NY, USA, 2002.

[11] Marc Levoy and Ross Whitaker, "Gaze-directed volume rendering," *SIGGRAPH Comput. Graph.*, vol. 24, no. 2, pp. 217–223, 1990.

[12] A. T. Duchowski, "A breadth-first survey of eye-tracking applications.," *Behav Res Methods Instrum Comput*, vol. 34, no. 4, pp. 455–470, November 2002.

[13] Carol O'Sullivan, John Dingliana, and Sarah Howlett, "Eye-movements and interactive graphics," *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movement Research*, pp. 555–571, 2003, J. Hyona, R. Radach, and H. Deubel (Eds.).

[14] P. Baudisch, D. DeCarlo, A. Duchowski, and W. Geisler, "Focusing on the essential: considering attention in display design," *Commun. ACM*, vol. 46, no. 3, pp. 60–66, 2003.

[15] Oyewole Oyekoya and Fred Stentiford, "An eye tracking interface for image search," in *In: Proceedings of the 2006 Symposium on Eye Tracking Research & Applications*, 2006.

[16] Arto Klami, Craig Saunders, Teófilo E. de Campos, and Samuel Kaski, "Can relevance of images be inferred from eye movements?," in *proc. of ACM international conference on Multimedia information retrieval*.

[17] Yun Zhang, Hong Fu, Zhen Liang, Zheru Chi, and David Dagan Feng, "Eye movement as an interaction mechanism for relevance feedback in a content-based image retrieval system," in *ETRA*, 2010, pp. 37–40.

[18] Zhen Liang, Hong Fu, Yun Zhang, Zheru Chi, and David Dagan Feng, "Content-based image retrieval using a combination of visual features and eye tracking data," in *ETRA*, 2010, pp. 41–44.

[19] Kari-Jouko Räihä and Oleg Špakov, "Disambiguating ninja cursors with eye gaze," in *Proceedings of the 27th international conference on Human factors in computing systems*, New York, NY, USA, 2009, CHI '09, pp. 1411–1414, ACM.

[20] Andrew T. Duchowski, *Eye Tracking Methodology: Theory and Practice*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.

[21] Bryan C. Russell, Antonio Torralba, Kevin P. Murphy, and William T. Freeman, "LabelMe: a database and web-based tool for image annotation," *Int. J. Comput. Vision*, vol. 77, pp. 157–173, May 2008.

[22] László Kozma, Arto Klami, and Samuel Kaski, "GaZIR: gaze-based zooming interface for image retrieval," in *Proceedings of the 2009 international conference on Multimodal interfaces*, New York, NY, USA, 2009, ICMI-MLMI '09, pp. 305–312, ACM.

[23] Google Inc., "Google™ Image Search," http://images.google.com/.

[24] Yahoo! Inc., "flickr™," http://www.flickr.com/.