g-Search

Using Gestural Input for 3D Information Retrieval

Sheng Kai Tang
Tangible Media Group
MIT Media Lab
tonytang@media.mit.edu

ABSTRACT
In this paper, we propose the idea of using bare-hand gestures as a way to search for online 3D shape models. This idea is based on the survey of current development of search and 3D technologies. We believe that with more and more 3D data stored on the cloud and the coming Google Glass, a new and better way to search for required 3D data is needed. In this project, we proposed three scenarios and implemented them as demos of concepts and found some research issues which can be further discussed and developed in the future.

Author Keywords
Gestural search, 3D Model, Kinect

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors; Design; Measurement.

INTRODUCTION

"Search" has become a popular way for people to obtain information online nowadays. By using text keywords to describe concepts and ideas, "Search Engine" automatically matches these keywords with existing online contents and shows the sorted results to users. A common difficulty encountered while searching is that users always forget exact words or word combinations to describe ideas they've seen or heard before. Although the semantic search technique is developed and applied in search engines, it could only do simple correlation between properties of words and is far from natural language recognition. Hence, the ability of a user to convert complex thoughts into brief keywords is still the key to a successful search (Figure 1).

Figure 1. Google Text Search.

"A picture is worth a thousand words" refers to the notion that visual representation is easier to convey intricate idea than verbal one. That's why, in 2009, Google launched the "similar images" function enabling users to search images containing similar information. This function keeps users from translating rich visual information to limited textual description by simply uploading a target image for search. Although current mechanism merely looks for images with similar color patterns, the directness of using an image to relate to other images prevents users from complicated cognitive translation to achieve intuitive interaction (Figure 2).

Figure 2. Google Image Search

From text to image, the search technologies were driven by the increasing richness of digital content. Due to the invention of depth camera, which is capable of turning any physical objects into three-dimensional (3D) digital format, as well as the release of Google Earth platform, which crowdsources to build the 3D digital world, searching 3D information online is going to be a potential need of the future. Furthermore, with the popularity of computationally and visually augmented wearable devices, such as Google Glass, a new way of device-less input for describing 3D information is also required. Hence, we propose the idea of g-Search, a system enabling users to retrieve 3D information with intuitive gestural input (Figure 1).

Figure 3. Our prediction: Google Object Search with hand gestures

GESTURAL INPUT FOR 3D INFORMATION RETRIEVAL

In this project, we use iconic gestures, which are suitable for describing shape, size and orientation of an object, as a way to setup the criteria of search. We create three scenarios focusing on retrieving handheld 3D objects including cellphones, books and packages as demos of concepts.

1D Search—Cellphone

Although a cellphone is a 3D object, users mainly feel its width when holding it. In other words, width is the key parameter for shaping an iconic hand gesture. So, when a user shapes a cellphone-holding gesture, the system will recognize the gesture, measure the width, retrieve cellphone models with the same width, and then augment the cellphone image visually on the gesture (Figure 4).

Figure 4. Cellphone Search Example

2D Search—Book

People usually use two hands to hold a book before reading. The distance between two hands defines the width parameter of a book. When holding a book, people can also feel the thickness of it. So thickness can be the second parameter used for search. When a two-hand holding gesture is formed and recognized, the system will measure the distance between two hands and match it with the width of books. The system will also measure the distance between the thumb and middle finger and match it with the thickness of books (Figure 5).

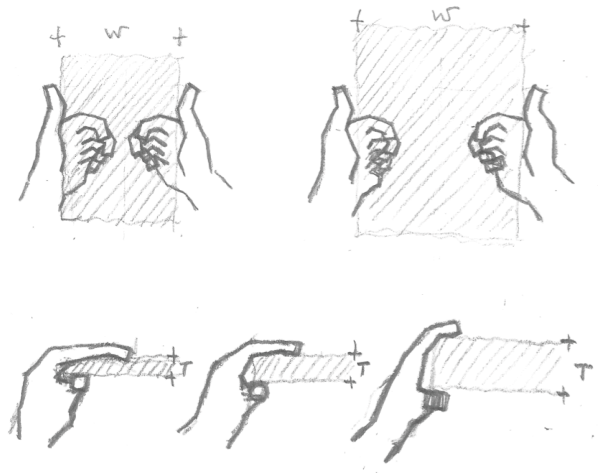

Figure 5. Book Search Example

3D Search—Package

When holding a package, people always use two hands to feel the size and shape of it. The distance between two hands determines the width, while the distance between thumb and the rest of fingertips shows the height and depth. So, in this scenario, a user can freely shape two-hand package-holding gesture to form a virtual box. The system will retrieve required parameters including width, height and depth by recognizing the position of fingertips. By changing the distances between two hands and fingertips, a user dynamically change the parameter combination for search (Figure 6).

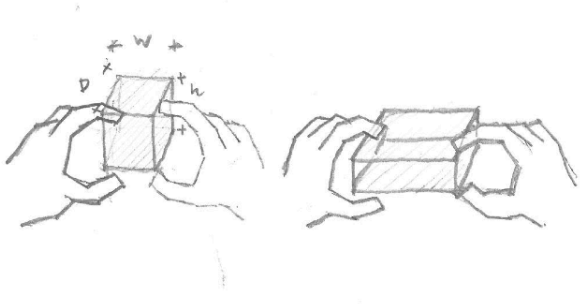

Figure 6. Package Search Example.

IMPLEMENTATION

Hardware

In this system, we only adopt an Xbox Kinect to obtain both RGB and depth image data. The Kinect is hanged above the user's hands with 45-degree view angle looking toward the two hands (Figure 7).

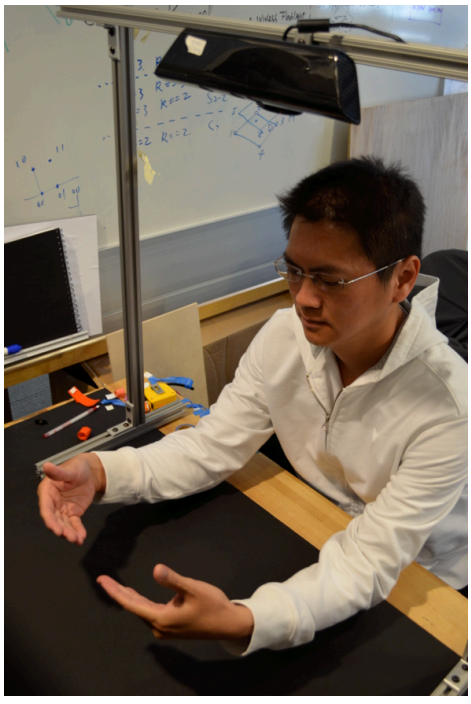

Figure 7. A Kinect is hanged over the user's two hands with 45 degree view angle.

Software

After observing users holding objects of the proposed scenarios, we figure out that there are three key points, thumb tip, middle finger tip and thenor which are required to form the object-holding gestures (Figure 8).

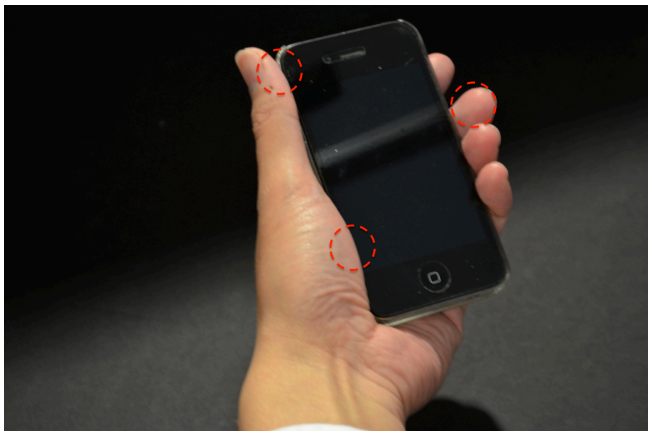

Figure 8. Key points for forming a holding gesture.

Instead of doing hand recognition based on point cloud of Kinect, we decide to do a simple trick to retrieve the 3D position of these three points. We, first of all, put color markers on these three points. Second, use RGB image to color track the markers. Then, get the 3D position of color-tracked position from depth image. Lastly, according to these three 3D positions, we not only generate a model matrix for 3D augmentation, but calculate all required distances defined in the scenario sessions (Figure 9).

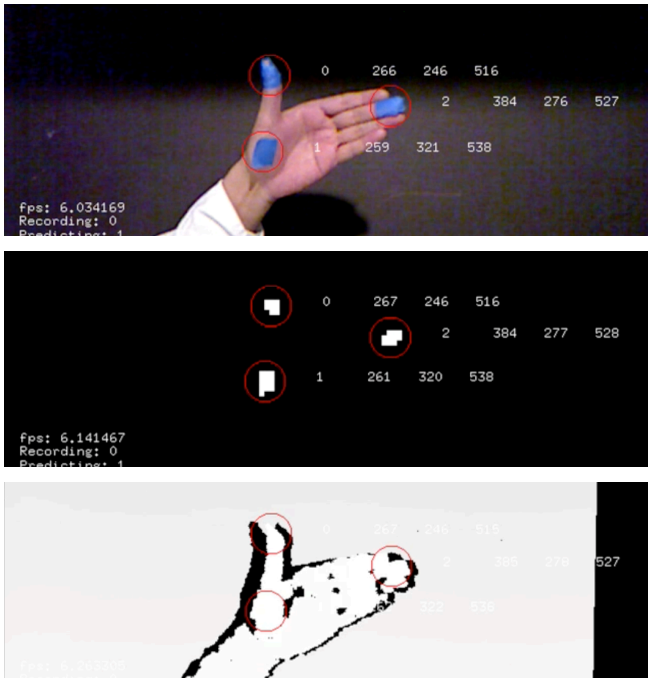

Figure 9. Image processing for points recognition and tracking.

RESULTS

We successfully implement three propose scenarios. As shown in figure 10, when the user gradually increases the distances between the middle fingertip and thumb tip, different cellphone models are dynamically matched and augmented on top of the hand gestures.

Figure 10. Cellphone Search by single-hand-holding gesture.

In figure 11, according to the distance between two hands, images of books which width match the distance are shown and augmented on the two hands. When the user increases the two-hand distance, he can navigate among the books stored in the database. When a user holds an augmented book with a specific width, changing the distance between the thumb tip and middle fingertip can navigate books with the same width but different thickness.

Figure 11. Book Search by specifying the width and thickness.

In figure 12, the user shapes a holding-box gesture to trigger a 3D box augmentation. The left hand controls the depth of the box, the right hand defines the height of the box, and the two-hand distance decides the width of the box. When moving two hands to shape different shape, size and orientation, the user can manipulate the augmented box. As soon as the dimensions fit a box in the database, the image will be augmented on top of the virtual box.

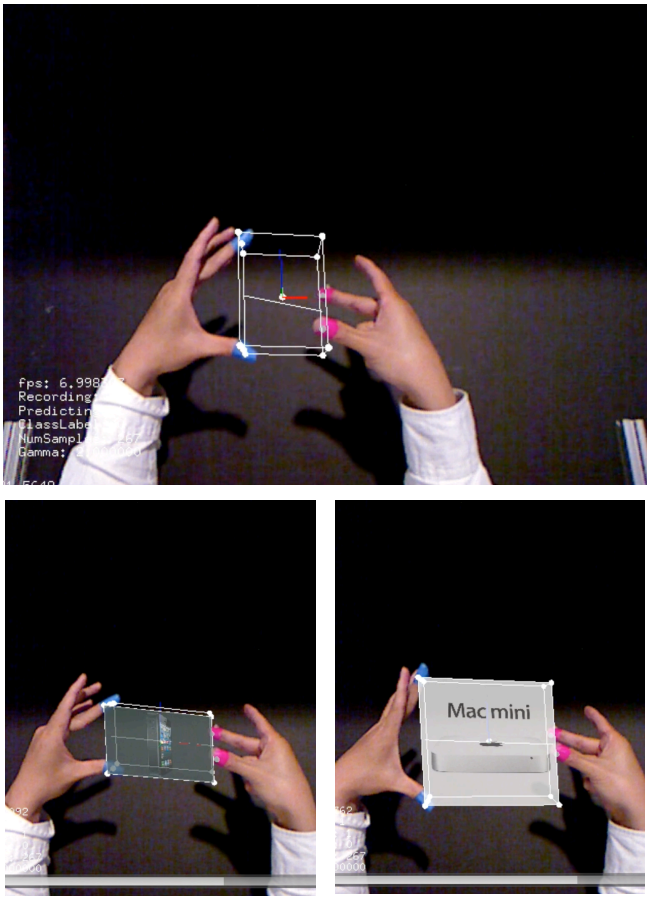

Figure 12. Box Search by two hand specifications of width, height and depth.

CONCLUSION

After manipulating and playing with the three scenarios, we have three reflections which need to be further clarified and discovered.

- Currently, there are only a few items stored in our database, so when the system does the parameter match, it can always find out a perfect match result. However, when we need to include plenty of objects in the database for search, it will create ambiguity of search.
- We propose the idea of using hand gestures to describe the shape. However, different users have different impression at the same object. The diversity of describing objects in gesture will cause some problems for the recognition and matching.

RELATED WORKS

3D Search

Due to the popularity of depth camera and Google Earth platform, large collections of 3D models are going to become as common as 2D image collections in the coming future. The need of efficiently searching for 3D model in a huge repository is a research issue which many researchers have already devoted into. The Graphics & Geometry Group of CS Department of Princeton University started the 3D search research since 2001 and has lots of research results [1]. The most popular one is using 2D sketch to search 3D digital model. The University of Edinburgh developed a system which enables users to draw a rough shape or shape combination in the system to search for more detail 3D models in the repository [2].

Gestural Input

According to Rime and Schiaratura's definition, there are four kinds of gesture popularly recognized and used in our daily life, which are symbolic, deictic, iconic and pantomimic [3]. While symbolic gestures have a single meaning for each of them, deictic gestures mostly are used to point and direct. While iconic gestures are describing the appearances of objects, pantomimic gestures are illustrating the interactionship between users and objects. Currently, the popular projects focusing on the development of Human Computer Interaction (HCI) adopted symbolic and deictic gestures to interact with digital information. For example, g-Stalt [4], T(ether) [5] and SpaceTop [6] adopted symbolic and deictic gestures for manipulating the digital contents. This phenomenon might be because existing GUI systems require only commanding and pointing. There are actually some research project begin to use iconic and pantomimic gestures as ways to interact with digital information.

REFERENCES

1. The Graphics & Geometry Group of CS Department of Princeton University. http://gfx.cs.princeton.edu/proj/shape/#Publications
2. PartBrowser: Immersive 3D-based geometry search. http://develop3d.com/blog/partbrowser-immersive-3d-based-geometry-search
3. Gesture interface. http://www.billbuxton.com/input14.Gesture.pdf
4. gStalt. http://tangible.media.mit.edu/project/gstalt/
5. T(ether) http://tangible.media.mit.edu/project/tether/
6. SpaceTop http://leejinha.com/SpaceTop