InfoTouch: An Explorative Multi-Touch Visualization Interface for Tagged Photo Collections

Per Ola Kristensson
Cavendish Laboratory
University of Cambridge
JJ Thomson Avenue, Cambridge, United Kingdom

pok21@cam.ac.uk

Olof Arnell, Annelie Björk, Nils Dahlbäck,
Joackim Pennerup, Erik Prytz, Johan Wikman,
Niclas Åström
Department of Computer and Information Science
Linköping University
Linköping, Sweden

ABSTRACT

We report on a design exploration into how a large multi-touch tabletop display can be used for information visualization. We designed an interface where users explored a tagged photo collection by bi-manual manipulation of the collections' tag cloud. User feedback showed that despite the availability of multi-touch most of the actual interactions were single-touch. However, some particular natural actions, such as grabbing the tag cloud and partitioning it into two parts, were often carried with both hands. Thus our user study indicates that multi-touch can act as a useful complementary interaction method in information visualization interfaces.

Categories and Subject Descriptors

H5.2 [**Information Interfaces and Presentation**]: User Interfaces – *input devices and strategies*.

General Terms

Design, Experimentation, Human Factors

Keywords

Multi-touch, visualization, information visualization, interaction surfaces, tagging, tags, tag clouds, photo browsing, photo collections

1. INTRODUCTION

Two-handed interfaces (e.g. [7]) and multi-touch interaction surfaces have been researched and developed since at least the mid-1960s [3]. We predict that low-cost large multi-touch interfaces will soon be available everywhere. Already, consumers can purchase commoditized capacitance-based touch-screen computers, such as the iPhone. This development paves the way for explorations into how to efficiently exploit everywhere-available large touch-screen displays.

In this spirit, we report of an explorative design for casual

browsing of tagged photo collections. Photo-sharing web services, such as *Flickr*, have popularized the act of uploading pictures to a public web server that is available to everyone. To enable efficient search among millions of photos, users are encouraged to provide descriptive metadata about the uploaded photos. The metadata system in *Flickr* is called "tags" and the process of adding tags to a photo is called "tagging". Tagging a photo involves assigning a set of descriptive words to a photo. A photo can have several tags. For example, a picture of a flower taken in July in Linköping City in Sweden, could be tagged by a particular user as *flower, Linköping, Sweden, summer*, *July*.

Tagging photos has primarily four advantages. First, tagging means it is easy and reliable to search for photos using a textual query. Second, the process of tagging individual photos carries a relatively low-cost for the user. Some tags can also be automatically generated, such as tags corresponding to location data and capture time. Third, photos that share tags form sets that can be manipulated using set algebra to reveal particular subsets of the photo collection. Fourth, tagging allows users to semantically explore the photo space. The (in)famous "tag cloud" is probably the most well-known example (see [5] for a recent overview of different tag cloud designs).

A downside with tagging is that it demands extra effort from users. Ames and Naaman [1] studied users' tagging behavior with *Flickr*. They suggest motivating users by exploiting social incentives and by making tagging easier. They also emphasize the importance of enabling users to tag their photos directly using their mobile camera device at point-of-capture [1].

Assuming users are willing to tag their photos, Yee et al. [9] argues that metadata-based browsing is more efficient for exploration than a standard keyword-based search system, and encourages further investigations of alternative interfaces for photo exploration. Subsequently Porta [8] shows how photo collections can be browsed using a multitude of window-icons-menu-pointing device (WIMP) interfaces. Apted et al. [2] describes a multi-touch system for photo management. Their system is intended to help elderly easily share digital photos in a face-to-face interaction setting.

Here we present InfoTouch – a design exploration where we combined multi-touch interaction with an information visualization interface. In our application users can use all ten fingers to explore a tagged photo collection by manipulating an interconnected tag cloud. We investigate if, and how, multi-touch can complement the WIMP paradigm that is prevalent in

information visualization. To our knowledge, we are the first to explore how to combine multi-touch with an information visualization interface in this way.

In the rest of this paper we first describe the InfoTouch system. Then we present the rationale, setup and results obtained from an explorative user study conducted with our system. Thereafter we discuss lessons learned and present design implications from our findings. Last we conclude and point to future work.

2. MULTI-TOUCH HARDWARE

The back-projected multi-touch display was built using the techniques developed by Han [6]. The display is placed horizontally (tabletop). The screen dimensions are 700×900 mm. The screen resolution is 1024×768 pixels. The display is mounted on top of a table, 920 mm above the ground.

The complete description of how the multi-touch display works is out of scope for this paper (see [6] and accompanying references for a complete description). What follows here is a very brief summary. Our system projects infrared light through a 9 mm thick acrylic glass surface covered by a Rosco Grey PVC screen. When the user touches the acrylic glass via the PVC screen, light diffraction patterns appear within the acrylic glass. These patterns are observed using an infrared camera mounted below the acrylic glass and detected as finger-touch coordinates by a pattern recognition algorithm.[1] The screen surface is projected on top of the PVC screen using an off-the-shelf projector.

3. USER INTERFACE

The screen is divided into two areas, the *Tag Cloud Area* and the *Thumbnail Area*. The entire interface is shown in Figure 1 and demonstrated in actual use in Figure 2.
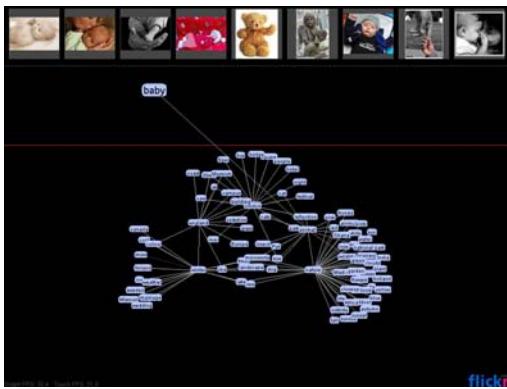

Figure 1. The thumbnail area at the very top displays photos that match the user's current query. The bottom area is the tag cloud area. Tags that are part of the user's query are shown in the search field between the tag cloud and the thumbnail area. In the photo one tag "baby" is present in the search field.

3.1 Tag Cloud Area

The bottom part hosts the interactive graph visualization of the photo tags. This area occupies around 80% of the available

_____

[1] http://www.nuigroup.com/touchlib/

display area. The user interface consists of a four layered stack: the graph, the physics model, the rendering and basic interaction model, and the multi-touch interaction model. We describe each layer in turn. Thereafter we describe the behavior of the search field.

Figure 2. A user using both hands to simultaneously drag two tags from the graph into the *Search Field*.

3.1.1 Graph

Each tag in the photo collection is represented as a vertex in an undirected weighted graph. Each edge represents a relationship between two vertices (tags). Two vertices are related by an edge if there exists at least one photo that includes both tags. An edge's weight $w(t_1, t_2)$ between two tags t_1 and t_2 is determined as:

$$w(t_1, t_2) = \frac{C(t_1, t_2)}{C(t_1) + C(t_2) - C(t_1, t_2)}, \qquad (1)$$

where $C(t_1, \ldots, t_n)$ returns the number of photos that contains all tags $t_i, i \in [1, n]$.

3.1.2 Physics Model

To make the graph interactive we attached a physics model to the mathematical graph formulation above. Each vertex is modeled as a body with a certain inertia. Each edge is modeled as a spring that can repel and attract the two vertices connected to it.

The length of a spring is determined by the weight of its edge. Each vertex is configured with a small degree of positive gravitational force that causes the springs connected to the vertex to contract slightly. To avoid stiffness a degree of elasticity is used to make it easier to drag tags (vertices) around. Elasticity acts as a stabilizer and causes the physical model of the graph to quickly converge to a steady-state when the user has been moving around the tags (vertices).

3.1.3 Rendering and Basic Interaction Model

The tags are rendered as filled light blue rectangles with rounded corners. The name of the tag is rendered as black text centered within an enclosing rectangle (Figure 2). The edges are rendered as thick gray lines. The color intensity and length of a line are proportional to the edge's weight, which in turn is proportional to the corresponding spring's length. The visualization of the

physical graph model is rendered using the *Prefuse* visualization toolkit[2] in real-time with no perceptible delay.

When the user drags a tag, the force exhibited by the extended spring is propagated over time throughout the physical model. By experimentation we configured the elasticity in the system to feel natural and fluid.

3.1.4 Multi-Touch Interaction Model
Prefuse only supports single-selection (e.g. WIMP). We extended *Prefuse* by implementing an additional multi-touch interaction layer. This layer is essentially a generalization of managing a single contact point to managing multiple contact points in *Prefuse*.

3.1.5 Search Field
The search field resides between the graph and the thumbnail area. The search field is initially empty in the application. When the user drags nodes to the search field, the system performs a search to find the intersection of photos in the collection that are tagged by the tags present in the search field. For example, if the user has dragged the tags *flower* and *Sweden* to the search field, the thumbnail area shows photos that contain both tags.

3.2 Thumbnail Area
This area displays thumbnails of photos that match users' queries in the search field. If the user touches a thumbnail the system shows the full photo centered over the entire display.

4. USER STUDY
Since InfoTouch is designed for casual exploration there are no easily identifiable dependent variables (such as task time or error rate) that are meaningful indicators of the appropriateness of the design. Therefore, instead of conducting traditional controlled experiments, we decided to do an open ended "probe" on how typical users would interact and reflect over InfoTouch. We chose to develop a couple of scenarios with different goals, in the spirit of the scenarios described by Dahlbäck et al. [4]. We want to emphasize that our initial probe is not intended to replace or exclude a traditional experimental design or an open-ended study of an actual deployment of the system. Rather the data collected and reported here is best seen as indicative of how well the design *could* work, and how multi-touch interaction may interplay with an information visualization interface.

4.1 Data Source
We used the open *Flickr* application programming interface (API) to retrieve the 50,000 most popular photos and their associated tags. Photos in *Flickr* are sometimes tagged with tags that do not describe the photo, such as camera brand, or who shot a particular photo. Such noise was manually removed.

4.2 Setup

4.2.1 Participants
We recruited 15 participants from the university campus. Their ages ranged from 19 to 35 (mean $\approx$ 23, sd $\approx$ 5). Four were women and 11 men. Eight participants had used *Flickr* before. All

---

[2] http://www.prefuse.org

participants except one had used a touch-screen before. Three participants had used a multi-touch display before.

4.2.2 Tasks
As this is an open-ended initial user study we decided to give the participants a rather broad set of tasks that covered a spectrum of activities possible with InfoTouch.

The first task (*Task 1*) was a goal-oriented retrieval task. We showed participants a specific photo from the dataset and asked them to retrieve it.

The second task (*Task 2*) was a general exploration task. We asked participants to use InfoTouch to explore suitable desktop wallpapers for their own computers at home.

The third task (*Task 3*) was also a general exploration task. We posed participants the following scenario: *"Imagine you are going to travel for a couple of weeks. Use InfoTouch to explore potential travel destinations."* [paraphrased from Swedish]

4.3 Procedure
We first demonstrated to the participants how to use InfoTouch. Thereafter we asked them to perform each of the three tasks. The order of the tasks was balanced across participants.

After all the tasks were completed we let the participants fill out a questionnaire and thereafter we conducted a 10 minute interview with them.

5. RESULTS AND DISCUSSION
5.1 Multi-Touch Usage
On average participants used multi-touch 16.3 times (sd $\approx$ 18.9) across all three tasks. An average multi-touch action lasted 862 ms (sd $\approx$ 460 ms). Two participants did not use multi-touch at all, and one participant only used multi-touch a single time. In relation to all interaction events, multi-touch was used approximately 10% of the time. Observations of actual use revealed that multi-touch was mainly used to organize the graph and separate many nodes at once. For example, one common action was to "grab" the graph with both hands and pull it apart – essentially separating the tag cloud laterally into two halves.

5.2 User Ratings
In the questionnaire we asked participants to rate a series of statements on a 1-5 Likert scale. Figure 3 summarizes all the responses. Among the positive ratings, 14 out of 15 participants agreed or partially agreed that using multi-touch was fun and the navigation method was inspiring to use. 12 out of 15 agreed or partially agreed that InfoTouch was good way of exploring *Flickr* photos. However, only two completely agreed, and 9 partially agreed, that multi-touch was well-suited for photo exploration. Among the remaining four, two were neutral, and two partially disagreed with this statement.

5.3 Interview Data
In the interview part, many participants thought the task of finding a specific photo the most difficult. Participants stated they found it difficult to imagine what other users tagged a specific photo as. Many also expressed they thought the interface more suitable for casual exploration of the "tag space", rather than to use it as a search interface.

One participant thought multi-touch was unnecessary and that the application would work just as well with a standard mouse. Another participant stated that the application would work with a mouse, but multi-touch was more satisfying and interesting. Overall, it appears the multi-touch capability was used more as a complementary convenience feature in InfoTouch rather than as an absolute necessity. This is also similar to how, for example, the iPhone device uses multi-touch for auxiliary actions, such as zooming in and out of a photo.
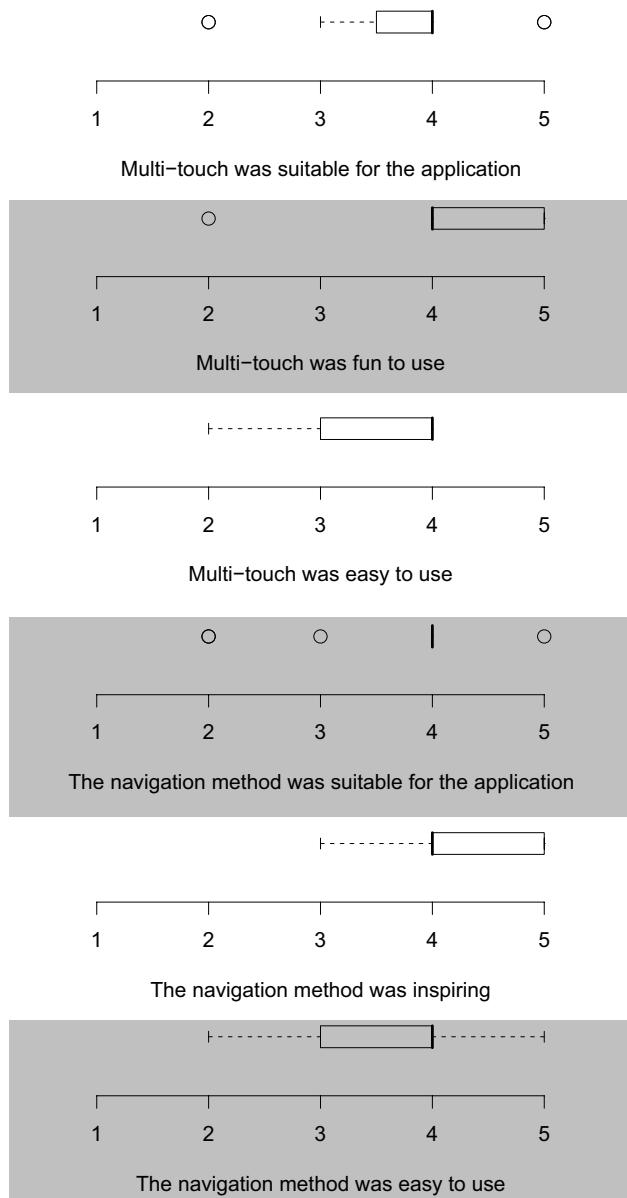
Figure 3. Box-and-whisker plots of user ratings (1 = Strongly Disagree, 5 = Strongly Agree).

6. CONCLUSIONS

In this paper we have presented InfoTouch – a multi-touch visualization interface for browsing tagged photo collections. As an explorative design project, we set out to investigate how recently available low-cost multi-touch technology could be used as a walk-up photo collection browsing interface. From our initial design study we conclude that participants generally liked the concept of using multi-touch to casually browse through photo collections. However, participants treated the multi-touch capability as a complementary feature that was used approximately 10% of the time. We conclude that with the advent of large low-cost multi-touch displays, multi-touch is a promising complement to traditional single-point interaction in information visualization applications.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Ames, M. and Naaman, M. Why we tag: motivations for annotation in mobile and online media. *Proc. CHI 2007*, ACM Press (2007), 971-980.

[2] Apted, T., Kay, J. and Quigley, A. Tabletop sharing of digital photographs for the elderly. *Proc. CHI 2006*, ACM Press (2006), 781-790.

[3] Buxton, B. Multi-touch systems that I have known and loved. http://www.billbuxton.com/multitouchOverview.html. Manuscript. [Accessed August 19, 2008]

[4] Dahlbäck, N., Jönsson, A. and Ahrenberg, L. Wizard of oz-studies: why and how. *Knowledge-Based Systems 6*, 4 (1993), 258-266.

[5] Halvey, M.J. and Keane, M.T. An assessment of tag presentation techniques. *Proc. WWW 2007*, ACM Press (2007), 1313-1314.

[6] Han, J.Y. Low-cost multi-touch sensing through frustrated total internal reflection. *Proc. UIST 2005*, ACM Press (2005), 115-118.

[7] Hinckley, K., Pausch, R., Proffitt, D. and Kassell, N.F. Two-handed virtual manipulation. *ACM Transactions on Computer-Human Interaction 5*, 3 (1998), 260-302.

[8] Porta, M. Browsing large collections of images through unconventional visualization techniques. *Proc. AVI 2006*, ACM Press (2006), 440-444.

[9] Yee, K, Swearingen, K, Li, K. Hearst, M. Faceted metadata for image search and browsing. *Proc. CHI 2003*, ACM Press (2003), 401-408.