Kartta: Using Multimedia and Context to Navigate Unfamiliar Environments

Arttu Perttula
Tampere University of Technology
PO Box 300, FI-28101, Pori, Finland
arttu.perttula@tut.fi

Scott Carter, Laurent Denoue
FX Palo Alto Laboratory, Inc.
3400 Hillview Ave., Bldg. 4, Palo Alto, CA 94304
carter,denoue@fxpal.com

ABSTRACT

Most mobile navigation systems focus on answering the question, "I know where I want to go, now can you show me exactly how to get there?" While this approach works well for many tasks, it is not as useful for unconstrained situations in which user goals and spatial landscapes are more fluid, such as festivals or conferences. In this paper we describe the design and iteration of the Kartta system, which we developed to answer a slightly different question: "What are the most interesting areas here and how do I find them?"

Keywords

Mobile, navigation, media

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: Miscellaneous

1. INTRODUCTION

Navigating an unfamiliar, eventful environment such as a festival or conference can be challenging. It can be hard to decide what to see, and when you have decided where to go, it can be difficult to know how to get there. While several research projects have investigated the mobile navigation of physical environments, past work has focused on extracting general landmarks from relatively static sources, such as structured maps. In this paper we describe a system, Kartta, that derives points-of-interest and associated landmarks from user generated content captured onsite. With this approach, Kartta helps users navigate standard environments as well as temporary events, such as festivals and fairs. Using social network and preference information, the system also reflects a more personalized set of points-of-interest. Furthermore, unlike past work, the system supports a diversity of landmark representations, including text tags, images, audio and video clips.

The initial idea for Kartta evolved from a brainstorming session with 15 participants that revealed that people need a mobile tool to help them navigate unfamiliar environments that 1) reflects the interests of their colleagues; 2) can be understood at-a-glance, and 3) provides simple aids to help users find areas of interest. Kartta accomplishes these goals by pooling content generated by groups

Mindtrek 2009, September 30th - October 2nd, 2009, Tampere, FINLAND.

Figure 1: The Kartta mobile interface showing a user (blue dot, upper left) making her way toward a cluster of interest areas. Kartta chose the image shown automatically as the best media to direct users toward this cluster (it is associated with the interest area with a yellow center). Text below the image are unique tags from interest areas in the cluster.

of users into clusters and visualizing those clusters as a translucent layer on top of a traditional map. Kartta also automatically analyzes uploaded media and associated metadata to reveal landmarks.

In the remainder of this paper we review past work and describe the Kartta system in more detail, showing how we used a combination of simulations and field pilots to iterate the tool.

2. RELATED WORK

Past work in mobile location-based media has focused on creating summaries of collected media, descriptions of static environments, or non-spatial visualizations of user-generated content. Jaffe et al. built a system that uses location and other contextual information to select key photos from a collection [6]. However, their work focuses on summarization rather than navigation. Grabler et al. developed a prototype that automatically generates maps based using building textures, road geometry, and external landmark information [5]. In this work, the authors focus on static structures and landmarks rather than temporary environments. Finally, studies by Baus et al. showed that audio landmarks can aid navigation tasks, but their work did not involve user generated content [1].

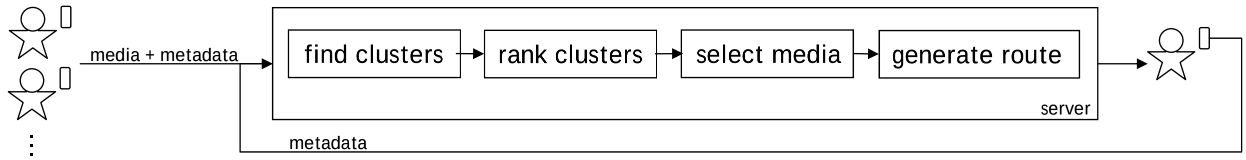

Figure 2: Data flow in the Kartta system. Using the mobile application, participants upload media, positive or negative votes for specific locations, and associated metadata to the server. Synchronously or asynchronously, other participants request a map from the server, optionally sending location and orientation information. The server finds and ranks clusters, selects media within each cluster, generates routes, and sends the updated information back to the mobile client.

Some map-based interfaces are used for retrospective purposes, and in many cases they can also be used for long-term planning. One example is EveryTrail™(`http://www.everytrail.com/`) which allows users to upload GPS tracking information as well as GPS-tagged photos. The system automatically plots that information on Google maps and offers a variety of social interactions around uploaded media (comments, ratings, etc.). However, the tool is not designed to collate and redisplay automatically information from multiple users.

Many other projects have explored synchronous mobile media sharing. Fono's and Counts' Sandboxes application displays collaboratively captured multimedia on mobile phones [4]. However, this work does not address organization (i.e., it does not utilize context to structure captures). Furthermore, as the interface is *temporal* rather than spatial it does not provide any navigation or map-based support. Similarly, in Clawson et al.'s Mobiphos system users capture pictures and share content with other proximate users, but the application does not visualize data spatially [2].

In contrast to all of these systems, Kartta helps users navigate unfamiliar environments using information culled from multiple participants in real time.

3. KARTTA

Kartta provides users a map showing them nearby areas they are likely to find interesting. To do this it gathers and analyzes both explicit and implicit recommendations of specific locations.

Kartta is composed of a server and a mobile application built using Mobile Python. The mobile application can be used to capture content and context data which are sent to the server automatically. The server uses this information to create a map of the immediate region around a user, highlighting points-of-interest as well as landmarks to help the user navigate. The mobile application polls the server regularly for a new map and set of landmarks.

The mobile application visualizes not only the augmented map returned by the server but also the users' current location (Figure 1). Users can configure the interface either to show media captured by an explicit list of friends, or to show only their own captures. Because, as Fleck et al. point out [3], mobile users often feel that capturing media is too distracting, the application includes a simple, one-button interface for recommending that other users visit an area (a positive vote) or avoid it (a negative vote). A positive vote is represented on the map as a green dot and a negative vote a red dot. Over time, votes implicitly create an interest-map of a place.

Mobile users can also launch media capture applications with another button. The mobile application sends media to the server along with context information (if available) including phone tilt (up, level, or down), compass orientation, time-of-day, and location. Context data can also be recorded continuously. The application currently senses location information via either embedded or attached GPS devices. While GPS currently affords only a gross estimate of location, we believe that it is sufficient since our application depends on aggregated data (votes). Furthermore, users can add tags to disambiguate areas of interest. Finally, media captured at a location automatically corresponds to a positive vote for that place (similar to the approach in [6]).

The server ingests media and votes sent from mobile devices, finds clusters of media, selects representative captures for each cluster, and generates routes (see Figure 2). The server saves mobile media in a database, and a set of ingest and access services produce multiple downsampled representations of uploaded media: thumbnails of photos, keyframes and lower resolution copies of videos, and copies with lower sample rates for audio clips. The mobile application can use these different representations to provide tiered access to media.

To organize captured media, the server first finds and ranks clusters of interest areas, ranks media within each cluster, and then generates route information.

3.1 Finding and ranking clusters of interest areas

In this step, media captures and votes, or *interest areas*, are grouped spatially using a simple connected components approach (see Figure 3a). After finding clusters of interest areas, the server calculates the importance of clusters. This calculation ranks higher clusters that the user will likely be interested in and that are also close to the user. Specifically, clusters are ranked higher if they include more net positive captures (after subtracting negative votes), include more distinct users who contributed to the cluster (similar to the approach in [6]), are closer to the current location of the user, and are not areas the user has already visited. The mobile tool uses this rank to vary the visual intensity of each cluster (see Figure 3b). Lastly, tags from all interest areas are pooled into one list.

3.2 Selecting media within clusters

In this step, the system selects a representative media capture in each cluster. This is an important step toward avoiding excessive download times. Because our focus is on mobile navigation, not media summarization, this algorithm selects media that helps the user orient to and find the cluster. Specifically, media is more likely to be selected if at the time of capture the camera was directed upwards (detected by the accelerometer), the night/day information matches the current user's context (as determined by time-of-day and sunrise/sunset information), and the compass direction matches user's current context. Also, if the user has contributed media to a cluster, all other media will be removed from this ranking. This further reduces bandwidth while helping the user to recognize areas she has already visited.

3.3 Generating routes and sending information to the client

In the final step the server maps trails to connect the most important clusters. Since a completely automated approach would be extremely difficult and error-prone, the server generates smoothed routes from the continuous location capture information from each user. The server exposes the route, cluster (and associated tag lists), and media metadata as an XML file via a web service. Mobile clients can poll the server with updated location and compass information to gather new clusters and routes. Clients download selected media from the server using URLs embedded in the XML.

3.4 Graceful degradation

While the server can utilize context information to rank clusters, select media from clusters, and select routes, no metadata is *required* other than the location of media captures. If a metadata is missing (such as the compass or orientation), it is simply assigned an average value. Similarly, if GPS logging is not available, route information is removed, but clusters are still calculated. This graceful degradation allows the server to provide useful information to a variety of mobile *and* non-mobile users. For example, a user could download route and capture information to his desktop and forward that data to an external service, such as EveryTrail™, to either plan an event or review captures from a past event.

4. SIMULATIONS AND FIELD PILOTS

We built a desktop-based simulation of the server algorithms to test their viability and a simulated mobile application to pilot test before conducting field experiments. Our desktop-based simulator takes as input a map and a set of parameters specifying high-level capture information, including the number of captures, the number of users, and the percent positive votes. The simulator extrapolates pseudo-random captures from this initial information. The user is given a location and orientation on the map that is adjustable via a mouse. The simulator can also accept additional points on the map that act as "attractors", allowing it to simulate conditions in which there is a main thoroughfare or gathering point (such as a food court at a festival). The simulator can then selectively apply the ranking and selection algorithms.

This approach allowed us to tune our algorithms rapidly and uncovered some important design decisions for the mobile client. Based on experimentation with the simulation we decided to utilize non-linear drop-off for ranking clusters to reduce clutter; adjusted weights so that popular areas retain their high ranking even if the user is far away; smoothed user-generated routes using splines, and allowed mobile users to select clusters in rank order.

We also conducted an informal critique of an early version of the mobile application running in a simulator with two experts in mobile application design. Both participants found usability issues, and one expert pointed out that we should have a backup location sensor in case one became too slow or stalled altogether. For this reason, we deployed the application on phones with embedded GPS sensors (Nokia N95s) while also providing an external sensor to each participant in the studies described below.

4.1 Field pilots

Massimi et al. used a scavenger hunt for evaluating mobile collaborative systems [7]. We felt this approach would be particularly useful because it blends the control of a laboratory study with the realism of a field experiment via a lightweight and well-understood game. We adapted this approach to test our system, focusing on understanding how well the interface helps users find specific waypoints in recommended areas as well as avoid poorly rated areas.

4.1.1 Initial pilot

In our first pilot, four users recruited from our lab were divided into two teams of two, and we gave every user a device running our system. We used a simplified version of the mobile application that did not include route information, showed one tag per capture rather than several per cluster, and inferred orientation from a GPS trace rather than relying on an external sensor. The goal of the exercise was to find colored balls (30 in total) that had been scattered around an office complex. We wanted to incentivize teams to split up to find items while still encouraging them to use the map to see what their teammates had found, and furthermore to communicate information about found items via messaging, taking a picture, or tagging a place. To that end, teams were given 10 points if either member found and took a picture of a ball, and another 6 if both members took a picture of the same ball. Two researchers followed participants during the study, taking notes and helping if users were completely stuck. Also, we held a focus group session with all participants after the study focusing on potential features.

Results.

Participants collectively captured 66 total targets and submitted 45 tags over one hour. Ultimately, we found that the interface was not yet at a stage to judge its usability for groups, but participants provided a host of recommendations in the focus group. Participants overall felt that the interface required too much attention, and requested vibratory and auditory alerts. They also suggested that the map may not be necessary at all, and that the visual interface might consist only of hotspots as well as paths to those spots. They also suggested audio tagging, list views of recorded tags, and orientation controls and views.

Methodologically, we found that our attempts to encourage both individual search and group communication had mixed results. While participants clearly used the tagging feature to signal information about a target to the other group member, in other cases participants simply met up physically to discuss the location of targets, and in still other cases ignored the map altogether.

4.1.2 Follow-up pilot

Given the difficulties we encountered in the first study, we designed another, more focused experiment that eliminated collaboration issues. Furthermore, inspired by Baus' et al.'s finding that audio can be useful for navigation tasks [1], we wanted to include a variety of media types. In this experiment, six single users recruited from our lab were asked to use Kartta to find and record objects of interest in a semi-familiar, semi-urban environment. We pre-recorded objects using a mixture of media – four photos (two tagged), two audio with no tags, two video with no tags, two tagged negative votes, and two tagged positive votes. Participants were given a device with the Kartta application and all but three of the votes and media pre-loaded and were asked to take a photo of the object they thought was being recorded or tagged at each location (for example, one hotspot was linked to an audio clip of the chimes of a clock, of which participants were to take a photo). Media captured by participants was saved to the device locally and uploaded to the server in a separate thread. A few minutes into the study, the application downloaded and displayed the remaining votes (we added a vibrated alert to signal updates). Participants were also asked to avoid areas marked with negative votes.

One researcher shadowed the participant both to record behavior and to answer questions about the interface in case the participant got stuck. Since the task took place in an uncontrolled environment we did not incentivize the participants to complete the study quickly. Instead, we asked them to signal to the shadower

(a) (b) (c) (d)

Figure 3: Simulated clustering example. (a) Original interest areas. (b) Interest areas clustered, ranked, and captures selected. (c) Routes added. (d) User cursor moved from center to upper-right – note the change in cluster rank.

when they believed they had completed the task. At the end, we asked participants several follow-up questions designed to determine which media was most helpful for navigation, the ease of navigating the map, and their ability to avoid negative areas.

Results.

Despite there being no incentive, all participants completed the study in the time allotted. However, only one participant successfully captured photos of all objects, and one captured only five. On average, participants captured eight out of ten objects. Media in the set that only appeared after the start of the experiment were those most likely to be left out. All participants chose photos as either the first or second most useful media for navigation. Four chose tags as the first or second most useful. One participant commented that it was easier to find those media that he could at least “roughly make out from a distance.” Overall, participants found it useful that the interface reflected their orientation as well as location (four on a five-point scale on average), and they found it easy to navigate the map (also four out of five).

Participants had few questions for the shadower during the task. Though it did not affect participant’s ability to complete the task, network reliability was also an issue, and we relied on media retrieved from the device for analysis. In general, participants had fun with the task, one noting that it was “like virtual geocaching.”

5. DISCUSSION

Our overall experience with the simulations and field pilots indicate that user generated content can help people navigate unfamiliar events. In particular, participants seemed to find it natural to move toward “hot spots” on the map given rich representations of perceptible landmarks. However, more work is necessary to support unconstrained environments. Our experiments revealed a tension between aggregating content on one hand and reducing clutter and transparency on the other. For example, it may not always be possible to track a user’s location because of interference. Other sensors may also be available only sporadically. In these cases, the server should automatically adjust the weights used to calculate cluster importance. We were also surprised that users in the field study preferred photos and tags considerably more than audio and video. However, past work used audio in an ambient way to indicate that the user was along the right route [1], rather than using audio as a representation of the end point.

Also, we found simulating behavior to be a crucial step in rapidly iterating not only our algorithms but also the resulting visualizations. While historically considered a first-class methodology [8], the approach is not often used to evaluate mobile applications. Given the difficulty of conducting realistic evaluations, though, we are convinced that simulations are a key tool.

6. CONCLUSION AND FUTURE WORK

While our experiments are limited, we have shown that it is possible to gather interest areas and landmarks from user generated data. Also, because of our focus on fluid, informal environments we made design decisions that allow users to derive some value from the system with only minimal input, including one-click location voting, automatically capturing metadata, and generating visualizations of captured media.

Additionally, we found simulations to be a critically valuable tool in prototyping this social mobile application. After running more complex simulations, the next step for this work is a full-fledged field experiment with a group or friends or colleagues in an unconstrained environment.

7. ACKNOWLEDGEMENTS

We thank Jari Multisilta as well as our study participants for their time and feedback.

8. REFERENCES

[1] J. Baus, R.Wasinger, I. Aslan, A. Krüger, A. Maier, T. Schwartz. Auditory perceptible landmarks in mobile navigation. In *IUI ’07*, pp. 302–304.

[2] J. Clawson, A. Voida, N. Patel, and K. Lyons. Mobiphos: a collocated-synchronous mobile photo sharing application. In *MobileHCI ’08*, pp. 187–195.

[3] M. Fleck, M. Frid, T. Kindberg, E. O’Brien-Strain, R. Rajani, M. Spasojevic. From informing to remembering: Ubiquitous systems in interactive museums. *IEEE Pervasive Computing*, 1(2):13–21, 2002.

[4] D. Fono and S. Counts. Sandboxes: supporting social play through collaborative multimedia composition on mobile phones. In *CSCW ’06*, pp. 163–166.

[5] F. Grabler, M. Agrawala, R. W. Sumner, and M. Pauly. Automatic generation of tourist maps. In *Siggraph ’08*, pp. 1–11.

[6] A. Jaffe, M. Naaman, T. Tassa, and M. Davis. Generating summaries and visualization for large collections of geo-referenced photographs. In *MIR ’06*, pp. 89–98.

[7] M. Massimi, C. H. Ganoe, and J. M. Carroll. Scavenger hunt: An empirical method for mobile collaborative problem-solving. *IEEE Pervasive Computing*, 6(1):81–87, 2007.

[8] J. E. McGrath. Methodology Matters: Doing Research in the Behavioral and Social Sciences, In *Human-computer interaction: toward the year 2000*, pp. 152–169, Morgan Kaufmann, 1995.