ZoneTag: Designing Context-Aware Mobile Media Capture to Increase Participation

Shane Ahern, Marc Davis, Dean Eckles, Simon King,
Mor Naaman, Rahul Nair, Mirjana, Spasojevic, Jeannie Hui-I Yang
Yahoo! Research Berkeley
1950 University Ave, Suite 200
Berkeley, CA 94707, USA
http://research.yahoo.com/location/yahoo_research_berkeley
{sahern, marcd, eckles, simonk, mor, rnair, mirjana, jeanniey}@yahoo-inc.com

ABSTRACT
ZoneTag is a rich mobile client that enables context-aware upload of photographs from cameraphones. In addition to automatically supplying location metadata for each photograph, ZoneTag supports media annotation via context-based tag suggestions. Sources for tag suggestions include past tags from the user, the user's social network, and the public, as well as names of real world entities such as restaurants, events, and venues near the user's location. A seamless interface makes it easy to assign tags to a photo, forming the basis for a richer personal media retrieval and organization system. We believe that lowering the barriers to tagging has great potential for effective retrieval.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces and Presentation (e.g., HCI)**]: Multimedia Information systems; H.5.2 [**Information Interfaces and Presentation (e.g., HCI)**]: User Interfaces

General Terms
Human Factors, Design

Keywords
Photo labeling, tagging, multimedia management, retrieval, tag suggestions

1. INTRODUCTION
Consumer photography has made it exceedingly simple for people to capture images, which they do at an ever-growing rate. The growing rate of photo capture is driven by the proliferation of capture devices (such as digital cameras and cameraphones) as well as decreasing storage costs. At the same time, however, creation of semantic metadata about the photo content remains an elusive goal. This type of metadata is required in order to facilitate retrieval from large collections of photographs and media. Completely automated tools, though improving, have not proven sufficient in providing complete and accurate semantic metadata. Manual annotation interfaces, even on desktop computers, remain time-consuming, often presenting the user with an overwhelming number of photos to annotate. Semi-automatic approaches based on content analysis and often on contextual information have made some headway. Recent emphasis on contextual metadata for photo organization is indicated by a growing number of other projects (e.g. Meaning[1], Context Watcher[2], Lifeblog[3]) that automatically associate contextual metadata with media captured on cameraphones. ZoneTag is unique in its use of this automatic metadata to generate *suggested tags*. We believe that a semi-automatic approach providing suggested tags at the point of capture is a promising step towards lowering the barrier to participation in a media annotation system.

Figure 1. When a photo is taken, a dialog pops up asking the user whether they would like to upload the photo to Flickr.

ZoneTag supports context-based sharing of metadata. Allowing metadata to be shared creates an opportunity to leverage community effort to benefit each of the individuals within the community. Additionally, sharing tags throughout a community presumably leads to some degree of tag convergence, making it easier for individuals to retrieve not only their own media, but others' as well. In shared-metadata systems we can consider three levels of user participation: creators, synthesizers, and consumers. The challenges in designing a metadata-entry system are making it easy for creators to generate metadata, easy for synthesizers to use it, and beneficial for consumers who do not contribute at all. Technically this means easy input (especially challenging on a mobile device) for the creators, good algorithmic support for tag suggestions for the synthesizers, and techniques for leveraging this in a UI that does not require participation.

ZoneTag supports easy photo capture and upload to Flickr [4] in as few as two keystrokes from the mobile device. The immediate benefit for all consumers is that the photo is automatically annotated with location data for browsing by location even without any user effort. For synthesizers, the system offers the user the opportunity to add tags, selected from a searchable categorized list. For creators, tags can be entered on the phone (assisted by T9 predictive text and auto-complete on all previously-selected tags), or entered on the web.

2. SYSTEM DESCRIPTION

The two primary components of ZoneTag are the client application, running on Nokia Series60 cameraphones, and the ZoneTag server, a PHP/MySQL application providing location translation and suggested tags to the client as well as processing uploaded images and metadata from the client and passing the images and tags to Flickr.

The ZoneTag client runs as a background process, monitoring the cell tower to which the phone is currently connected (which provides a rough location) and possibly communicating via BlueTooth to a GPS device, providing much more exact location information (and also providing a mapping between cell tower and physical location.) When the phone switches cell towers (or every 10 minutes if stationary), the client contacts the server for a new set of suggested tags, appropriate to the new context. Suggested tags are pre-fetched because network latency is too great to provide acceptable user experience fetching tags at the moment a photo is captured. The decision to pre-fetch was based on our previous work with the Mobile Media Metadata (MMM) system [5] which pioneered the use of context-sensitive suggested tags but was rather difficult to use due to the considerable latency in its mobile web browser-based tagging interface [6]. When a photo is captured, the ZoneTag application comes to the foreground of the phone's user interface and provides an integrated tagging interface that allows the user to quickly annotate and upload the photo. In addition to the user-selected tags for a photo, the ZoneTag client sends the server current location data (cell info and GPS if available.)

When the server receives a new photo it translates the location information from the client into human-readable labels (i.e., city, state, country, ZIP code) which propagate to Flickr as tags. If there is no known location mapping for a given cellID, users can 'teach' the server new locations via a web interface. These new locations are then propagated throughout the user's social network (as determined by Flickr family, friends, and contacts) with preference given to locations obtained from users with less social distance. Other tags are shared amongst users in a similar fashion, with tags from socially nearby users ranked higher in the tag suggestion list (and with the user's own tags given highest priority.) In addition to providing suggested tags to the ZoneTag client, the server also provides a web interface to tag suggestions, linked from the Flickr website. Finally, the ZoneTag server periodically crawls ZoneTag images on the Flickr website for tag changes made via Flickr's web interface, which can impact the suggested tags propagated back to the handset client. This allows users to enter new tags on the web and simply select them on the mobile interface without the hassle of text entry via T9 or triple-tap.

3. USER EXPERIENCE

Using the ZoneTag application, a user can upload a newly captured image from her cameraphone to Flickr in two quick clicks. After an image is captured, ZoneTag displays a "Post to Flickr? Yes/No" dialog over the newly captured image (Figure 1). At this stage, the user can review the image and decide whether she wants to upload the image (click "Yes"). If the user declines, the cameraphone returns to its regular state. If the user accepts, more options are shown, (the user is given the chance to upload the image immediately, keeping the settings and tags entered for the previous photo.) The option screen allows the users to control various features, and add or change the metadata tags associated with the image. The controls include the option to set the image privacy settings: following the Flickr model a photo can be completely private; viewable only to Flickr users marked as 'friends' or 'family' by the photo owner; or public. The user has the option to specify a title for the photo, which appears on the Flickr photo page. Finally, the user can choose whether to expose the photo's location data. If she chooses to include location data, the zip code and city names are added as tags to the photo's page on Flickr. Most importantly, the user has an option to select or type in tags (textual labels) to appear on the Flickr photo page. Tags on Flickr are used for personal organization and public sharing; both generate incentives for some people to tag.

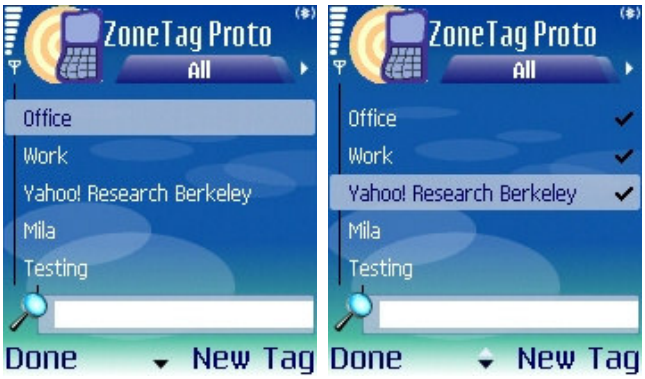

Figure 2. Users can choose to add tags to the photo before it is uploaded. A list of suggested tags is shown (left), and the user can select them (right) or type in new tags.

On the tagging screen (Figure 2) the user is presented with a list of tags pre-fetched from the ZoneTag server and presented in order of likelihood of being selected for the current context. The tags are grouped into categories, graphically organized as tabs. Drawing on other lessons learned from MMM, ZoneTag metadata tags are drawn from a flat tag space, more conceptually accessible to users than the more complex facetted tags used in MMM[6]; ZoneTag tag categories reflect tag sources rather than semantic category or facet.

Currently the categories are:

- **All**: tags from all other categories, as well as any tag the user has ever entered on the phone or applied to a ZoneTag photo on the Flickr website
- **Local**: tags created by the user or his social network likely to be used in the current location
- **Recent**: tags the user has used within 24 hours – useful when the photographer's subject is changing context with her (e.g., traveling with friends)
- **Venues**: nearby locations hosting events from upcoming.org
- **Events**: event names of nearby events from upcoming.org
- **Eats**: nearby restaurants
- **Terms**: generated by running term extraction on nearby events
- User customizable via RSS

Within each category a user can search through the available tags by entering the first few letters of a tag in the search box. When a photo is uploaded, the tags associated with that photo remain selected for the next photo allowing subsequent images with similar content to be uploaded with little interaction.

4. DEPLOYMENT

ZoneTag is currently deployed to 250 users (100 active) in 18 countries. Early results indicate that the tag suggestion interface is increasing point-of-capture tagging, as well as limiting the number of keystrokes required in the tagging process. For instance, Figure 3 shows the number of tags per photo, averaged over all photos from all users, and split by the ZoneTag version. The *SuggestedTags* version is the one described above; the *Basic* version is an earlier version of ZoneTag that did not have suggested tags: the user had to type in tags in full. The data represents roughly 4000 photographs for each version. Notice that the data does not represent a bucketized inter-subject test: most users have started using *Basic* and later upgraded to *SuggestedTags*, others only started using the system with *SuggestedTags*, and some never upgraded.

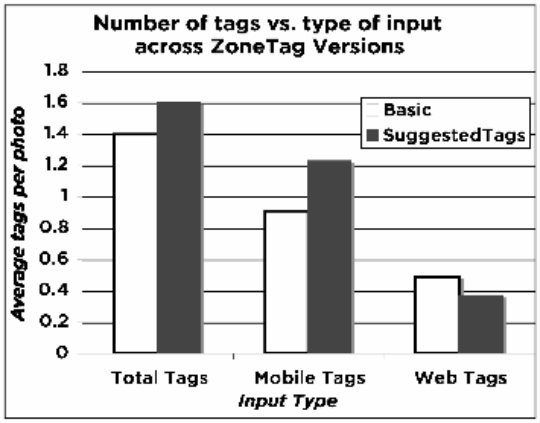

Figure 3. Number of tags added using the various input methods, in each of the ZoneTag versions

The figure highlights several important facts. As shown by the bars in the left-most column, the average total number of tags per image increases from 1.4 (*Basic*) to 1.6 (*SuggestedTags*). More importantly, the increase is entirely due to tags entered on the mobile interface (second column). At the same time the number of tags users added to the photo on the web, after the upload (shown in the figure's third column), dropped. This drop might suggest that mobile tagging with *SuggestedTags* was more complete and satisfactory than with *Basic*, often rendering the addition of tags on the web unnecessary.

5. DISTRIBUTION

ZoneTag is currently available for download for Nokia Series 60 2nd Edition phones, at http://zonetag.research.yahoo.com.

6. REFERENCES

[1] http://meaning.3xi.org/

[2] http://portals.telin.nl/contextwatcher/

[3] http://europe.nokia.com/nokia/0,,71739,00.html

[4] http://www.flickr.com

[5] Davis, M., King, S., Good, N., and Sarvas, R. From Context to Content: Leveraging Context to Infer Media Metadata. In *Proc. MM 2004*. ACM Press (2004), 188-195.

[6] Sarvas, R., Herrarte, E., Wilhelm, A., and Davis, M. Metadata Creation System for Mobile Images. In *Proceedings of the Second International Conference on Mobile Systems, Applications, and Services (MobiSys2004) in Boston, Massachusetts*. ACM Press, 36-48, 2004.