iPiccer: Automatically retrieving and inferring tagged location information from web repositories

Benjamin Proß
Institute for Geoinformatics
University of Münster
Weseler Str. 253
48151 Münster, Germany
benjamin.pross@wwu.de

Johannes Schöning
German Research Center for
Artificial Intelligence (DFKI)
Stuhlsatzenhausweg 3
66123 Saarbrücken, Germany
johannes.schoening@dfki.de

Antonio Krüger
German Research Center for
Artificial Intelligence (DFKI)
Stuhlsatzenhausweg 3
66123 Saarbrücken, Germany
krueger@dfki.de

ABSTRACT

We present iPiccer, a system that allows mobile camera phones users to interact with meta-tagged image material available from web repositories, such as flickr in two novel ways. Users of the system are able to infer photo tags from their location and orientation, but are also able to infer their location and orientation from their spatial photo tags. An implementation of iPiccerTaker is outlined, and the potential of this new form of interaction with web repository data is discussed.

1. INTRODUCTION & MOTIVATION

Almost every current digital camera and mobile camera phone automatically stores extensive metadata about each photo taken. This metadata includes information about the shutter speed, aperture, flash, and even whether the camera was in "scene mode". While these data may be of use for professional photographers, less experienced or interested photographers do not benefit directly from this type of information. On the other hand, location-related metadata is widely appreciated. Nowhere is this more evident than in the fact that nearly every photo sharing site/album comes with a feature to place taken photos on a map, thus adding the missing location metadata. We introduce iPiccer, a new way to interact and contribute to this important location information for web repositories. Using flickr and Wikipedia, iPiccer both aids in the tagging of photos with location-relevant tags, as well as using these tags to help users infer their location. While the former feature is relatively easy to understand, the latter feature requires a bit more explanation. Let us consider the example of a colleague who always has a hard time orienting her/himself when s/he leaves the subway in downtown Chicago. If this colleague is able to take a photo and is able to suggest a location-relevant tag for that photo, iPiccer would be able to assist this colleague in inferring both his location and his orientation.

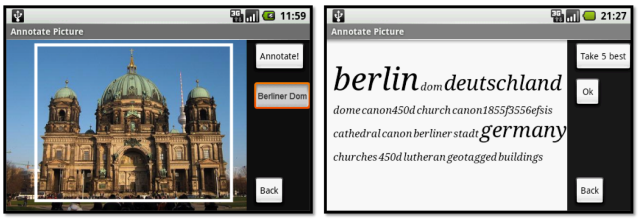

Figure 1: The iPiccer prototype: After the user has taken a picture the system automatically adds the sight name (left) and additional tags to the meta information (right). In addition the location information is verified and the direction the photo was taken, derived from the sight location and the location the photo was taken, is stored.

2. RELATED WORK

Zonetag [1] is similar application to our application presented that allows the users to upload pictures to flickr[1]. It also suggests tags based on the cell location of the phone. There is also a support for external GPS devices. Tags are suggested that the user has used before in the same location and that friends of the user have used in the location. Also tags are suggested that correspond to nearby items from the Yahoo! Local database. The Tag Suggestr [4] uses image recognition to suggest additional tags for a picture. It needs two to three initial tags that are used for further picture collection. Pictures with matching tags will be compared to the input image via the SIFT algorithm [5]. The candidate tags are weighted according to the similarity of the input image to the picture including the tag. Tag Suggestr does not search for exact image matches and is not suited for mobile purposes. In contrast to the application presented here, both approaches just work in one way: they receive tag suggestions. They do not work in the other direction, i.e. they do not add new tags to matching pictures. Still no image matching is performed and so the suggested tags are rather general and solely related to the location. More similar Crandall [2] also deals with tags and image recognition. They focus on analyzing and organizing large photo collections. The SIFT algorithm is used here to extract visual features from pictures. Visual, textual and temporal features are consequently used to estimate the location of a picture and to identify places that people like to photograph.

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

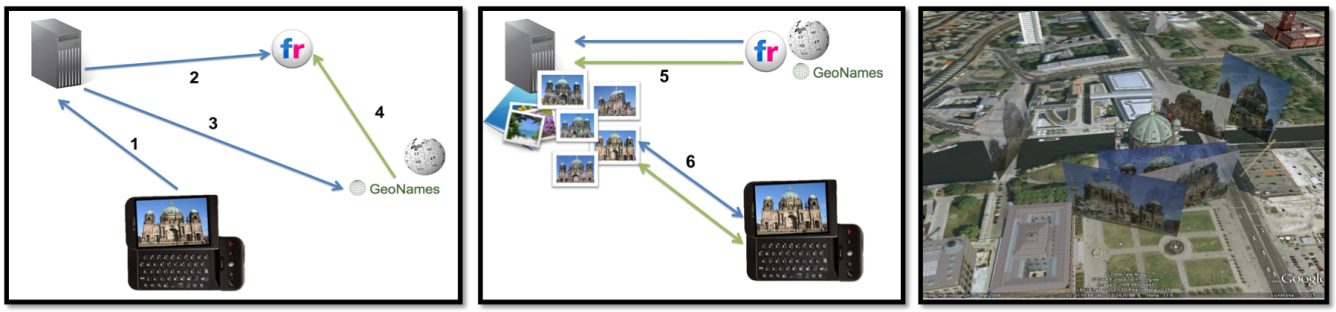

Figure 2: iPiccer System Overview: (1) User sends a photo with location information to the central server. (2,3) Server sends request to flickr and geonames.org. (4) Server sends 2^{nd} request to flickr using the sight name derived from geonames.org (5,6) matching the images with the original photo. System is returning label and tag suggestions (see figure 1) (left, middle). Photos using the direction feature are automatically aligned in Google Earth (right).

3. INTERACTION CONCEPT AND IMPLEMENTATION

The interaction concept is quite simple: A user takes a picture with his GPS equipped mobile phone (in our case the G1 dev phone running the Android operating system[2]. The location information and the picture are sent to a central server via XML-RPC. The server creates a buffer around the current location of the user (buffer size approx. 50m) and requests images within this distance from flickr. The GeoNames Service[3] is queried to find nearby spatial articles in Wikipedia. From the returned list of Wikipedia articles, those in the landmark category are chosen. The titles of these articles are then used in a second flickr photo request. Both image collections are then compared to the input image using SIFT[5]. Using the combination of both image collections high matching rates are achieved, because just a small subset of images is taken and the whole processing time is moderate (The entire processing time is about 4-10 sec. depending on how many photos are taken in the area). Matching is performed via RANSAC [2]. The parameters are set tightly to ensure that only the best matches will pass. If the images match, all relevant information like title, tags, coordinates and id are stored. This information allows an automatic labeling of all the sights that the users can see in the taken picture with a minimal bounding box (compare figure 1). In addition, all images used in the SIFT comparison can be labeled as well. The tags of all matching pictures are composed to a tag cloud and returned to the user. The user can now select tags manually or choose the five best tags (compare figure 1). Users can also add new tags to the picture. This is an important feature given that many times the tags returned by flickr are not obvious given the location only (i.e. the country name,) A slightly different scenario is given, when the user knows a sight's name already. By manually entering the site, this can also be verified and this knowledge can be used to inform the flickr search. Using the location information of a Wikipedia article it is possible to verify the user position and then to calculate the direction in which the photo has been taken (in case of the Android G1, we can use this information to calibrate the in-build compass.).

[2]`http://code.google.com/android/`
[3]`http://www.geonames.org/`

4. CONCLUSION AND FUTURE WORK

We have presented an application called iPiccer for automatic tag suggestion and location/orientation identification. The first prototype has been fully implemented and first user tests are currently carried out. By providing an easy and intuitive way to interact with the web repositories using a camera-phone, we open a new area for further applications. Future work will more fully explore this area. Obviously, one such project could be helping users to place photos in the right orientation in which they have been taken (compare figure 2). Another idea is to use these correctly oriented photos for pedestrian navigation services [3]. In addition, one could think of applications that allow users to see the "hot direction" from which people usually take pictures of a point of interest. Finally, we plan to release the application to the Android user community and analyze overall user experience on a wider scale.

5. REFERENCES

[1] S. Ahern, M. Davis, D. Eckles, S. King, M. Naaman, R. Nair, M. Spasojevic, and J. Yang. Zonetag: Designing context-aware mobile media capture to increase participation. In *PICS 2006 in conjunction with Ubicomp 2006*, 2006.

[2] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.

[3] H. Hile, R. Grzeszczuk, A. Liu, R. Vedantham, J. Košecka, and G. Borriello. Landmark-based pedestrian navigation with enhanced spatial reasoning. *Pervasive Computing*, pages 59–76, 2009///.

[4] O. Kucuktunc, S. Sevil, A. Tosun, H. Zitouni, P. Duygulu, F. Can, and T. Ankara. Tag Suggestr: Automatic Photo Tag Expansion Using Visual Information for Photo Sharing Websites. In *Proc. of Semantic Multimedia*, pages 61–73, 2008.

[5] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.