PHOTO BASED INTERACTION AND QUERY FOR END-USER APPLICATIONS - THE GEOPILOT PROJECT

Volker Paelke

IKG, Institute for Cartography and Geoinformatics
Leibniz Universität Hannover
Appelstr. 9a, D-30167 Hannover, Germany
Volker.Paelke@ikg.uni-hannover.de idowman@ge.ucl.ac.uk

Commission Ⅳ, WG-IV-6

KEY WORDS: Photography, Multimedia, Interface, GIS, Photogrammetry, Recognition

ABSTRACT:

We present GeoPilot, a system that combines photogrammetric techniques with GIS functionality to automate the identification of images. GeoPilot can be used to "tag" photos with keywords and location information to help users manage increasingly large photo collections and to retrieve information related to a photographed object on site, providing intuitive access to location-based information. The system consists of user-interface frontends for the photo management and location-based-service tasks that provide adapted access to the base functionality. Image identification is provided by feature detection and nearest neighbour search algorithms that match images against a reference database. The integration of Web 2.0 services allows to improve coverage of database and tagging information, e.g. by drawing on tagged images in on-line photo repositories like Flickr, while additional location information (e.g. GPS or GSM phone positioning) can be used to accelerate the matching process.

1. INTRODUCTION

1.1 Goal

As digital photography becomes a ubiquitous part of modern life, more and more people face the challenge of organizing their growing repositories of digital images. The annotation of images with keywords and positions – "Tagging" and "Geo-Tagging" – enable more efficient means of retrieval based on key-word search and spatial browsing. While these techniques are becoming increasingly popular in on-line image repositories like Flickr, a significant amount of work is required to tag images and few users are currently willing to invest this work for huge image collections.

Geo-Tagging, in which images are annotated with their spatial location, could be automated: One possible approach is to integrate a GPS receiver into the camera, but it is also possible to match the track of a photographer recorded with a GPS receiver with images in an off-line process according to their time stamps.

However, currently few cameras integrate a GPS receiver and only a minority of casual photographers is prepared to invest the effort required for geo-tagging.

The goal of the GeoPilot system is to derive (geo-)tagging information from existing databases of tagged images to automate the process. The approach is not only applicable to newly taken pictures but can be used on existing image collections as well. In addition, the same functionality can be used on-site to provide an intuitive interface to access location-based information, e.g. in touristic contexts. By incorporating Web 2.0 photo hosting services like Flickr which feature an increasing amount of tagged and geo-tagged images the dynamics of social web-services help to improve the coverage over time. The same functionality can also be used to provide additional information about photographed objects on-line to photographers, providing an intuitive "point-and-shoot" interface to location-based information retrieval.

1.2 Current practice

The effective management of large photo collections requires a systematic approach to their organization and retrieval. The predominant access-scheme to private photo collections is currently browsing, possibly aided by a temporal organization scheme as provided by the camera time-stamps. While suitable for small collections and in cases where the user can specify the date of capture, this approach becomes problematic once the collection grows, in cases where the exact date is not known and if multiple users / cameras become involved. Both on-line photo collections like Flickr and management tools like iPhoto therefore rely on (manual) tagging of images to enable text based queries. Effective support for tagging is therefore desirable.

1.3 Functionality

To help users in their interaction with photos two central usage scenarios for the GeoPilot project where identified and then explored within a student project group:

- Off-line tagging – the automatic annotation of (large) image collections with keywords.
- On-line information retrieval – in which a photograph is used to retrieve location-based information on-site.

The same base functionality is used in both scenarios, combining photogrammetric techniques with GIS functionality to enable efficient retrieval of matching images based. Each scenario is supported by an adapted user-interface frontend. For the photo management task the process is largely automated, while the on-line information retrieval is supported by intuitive "point-and-shoot" interface for location-based information access.

2. RELATED WORK

2.1 Image based interaction

The widespread availability of digital cameras and camera equipped PDAs and smart-phones has led to their use as interaction devices in a number of applications, especially in the domains of virtual and augmented reality. Due to the form-factor of the devices, into which the camera is embedded, these are typically used in an inside-out setup. This means that the camera itself is manipulated in space to effect some interaction. The images or the video-stream captured by the camera are analyzed to derive high-level interaction events that control the application. The additional input mechanism available on the device (e.g. buttons and sometimes the touch screen of a PDA) can be combined with the camera input to create more complex composite interaction techniques. So far, such interaction techniques have mostly been created on an ad-hoc basis by computer vision experts for use in technology demonstrators. A structured overview of the design space of image based interaction techniques is given in XXX. Overall, the exploration of image based interaction techniques and their application is still at an early stage and no standardized approaches have been established so far.

2.2 Photo management

A wide number of applications have appeared in recent years that aim to support users in the management of their growing photo collections. While professional image databases, e.g. those used in the resale of stock photography, are organized according to predefined taxonomies and augmented with keyword ("tags") by professional editors, photo management applications for casual users must operate in a much less structured and defined environment.

Typical examples include Apple's iPhoto, Google's Picasa and Adobe's Photoshop Album. These applications allow users to view, organize, annotate and retrieve images. For the annotation users are not restriced to a predefined list of keywords. Instead they can pick freely chosen labels which simplifies input but can be problematic in retrieval. Because many users do not annotate (tag) their photos in a regular or structured way, the typical mode of retrieval is by the date of capture, possibly associated with a major event.

More recently the public sharing of digital photos has received increasing attention. The most familiar example is probably Flickr, an image hosting website that operates as an online community platform and is a prime example of early Web 2.0 applications. As users who did not take a picture themselves can not possibly associate an image with a given date the annotation on images with "tags" becomes critical to enable effective sharing and retrieval. A key development here is that services like Flickr allow other users to add tags to images, leading to the development of so called folksonomies. Folksonomy is a combination of the words folk and taxonomy and refers to the fact that the annotation is not carried out by experts according to a predefined taxonomy or a controlled vocabulary (as in traditional retrieval systems) but as a collaborative group effort by users (social tagging). In the best case a larger set of descriptive tags is associated with an image to enable easy retrieval. The redundancies of multiple tags helps to alleviate the lack of a structured approach. For popular (and well tagged) images this approach is surprisingly effective.

Another approach to image retrieval from photo collections is content based image retrieval (CBIR). Content based image retrieval aims to identify matching images according to visual featured specified by the user either explicitly or by providing example images (query by example). Imgseek is a well known open source example of this approach. (http://www.imgseek.net/)

Recently researchers have integrated CBIR with existing image databases using image matching and nearest neighbour search (discussed in the following sections) to automate the annotation of images (). Other approaches aim to simplify the physical interaction with image collections through novel interaction techniques like multi-touch or with new 3D navigation techniques in an image space generated by matching large sets of related photographs in a spatial environment ().

2.3 Location based services

Location based services (LBS) provide location specific information on mobile devices. In recent years many prototypes for LBS applications have been investigated in various use cases, e.g. touristic city guides or mobile entertainment.

While GPS seems like an obvious solution for positioning only a small percentage of the deployed PDAs and Smartphones integrate a GPS receiver and few users are willing to carry around an additional receiver. An interesting approach that is the location via wireless networks (using WLAN or cellphone stations), that has been rolled out to a mass market with Google Maps on the GPSless Apple iPhone.

Several projects have aimed to address the need for precise positioning by the use of image information. The basic approach is described by (Johansson and Cipolla [2002]) who developed a system in which the user takes a picture at his current location with the integrated camera. The image is then transferred to a central server where it is matched against a large database of reference images and the position is thus determined. XXX have implemented such an approach in a classic LBS application scenario. Image based positioning is currently receiving a lot of interest (e.g. MobVis XXX and work by Nokia XXX). One interesting application is the possibility to augment the taken images visually with additional information in a mixed-reality setup, e.g. XXX.

3. SYSTEM

3.1 Overview

In the GeoPilot system the same base-technology is used for both the automatic annotation of large photo collections and the on-site access to location and object specific information. GeoPilot uses a Geo-Database in which reference images are stored and managed according to their spatial coordinate. The use of a spatial database allows to limit the following analysis and query operations to the immediate surroundings, improving the speed of the operation and thus enabling the use of more refined analysis and matching techniques. An image under consideration is compared to reference images in the vicinity. This operation should be both robust and fast. Several approach were considered in the project. For the final implementation we have selected a custom matching approach that builds on SIFT (Scale Invariant Feature Transform) features as these are independent from image-scale and have been demonstrated to be quite robust against changes in perspective, lens and lighting conditions. As the matching process using SIFT features is quite complex and computationally intensive a close coupling to the underlying Geo-Database is required to limit the number of candidates and make the approach viable in a real-time system. The main approach is shown in Figure 1.: SIFT features are extracted (2) from the image supplied by the user (1). These SIFT features are matched (3) against the SIFT features that were determined for the reference images in the database. If positioning information is available it is used to restrict the search to the immediate surroundings of the capturing position. if a match is determined and verified the corresponding augmentation information is retrieved from the augmentation database (4) and integrated into the image (in the automatic annotation use case) or presented as a multimedia presentation (in the LBS use case).

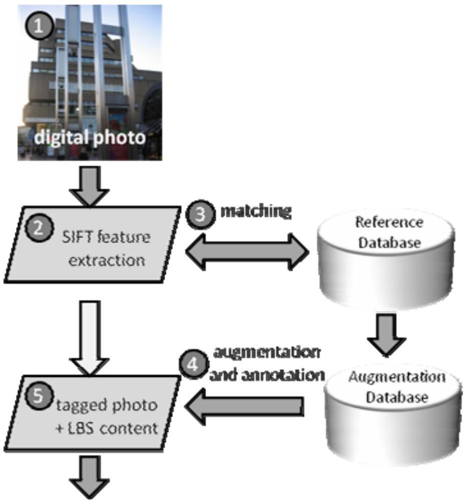


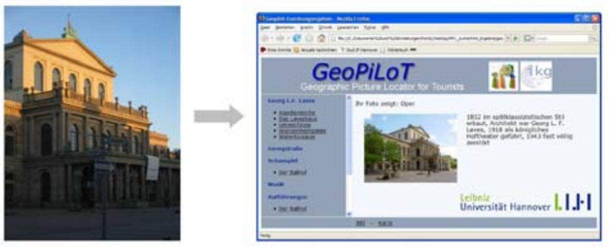
Figure 1. Operation principle of GeoPilot

3.2 Development process

The development of services aimed at end-users can only be successful if the requirements and preferences of the intended user population are carefully considered throughout the development process. In the domain of user-centered design a number of processes have been developed and a best-practice framework for the iterative development of usable interfaces has been codified in ISO standard ISO-13407. For the development of an end-user service like GeoPilot the consideration of users and usability aspects is critical. Unfortunately, current user centered design activities are not well integrated with general software engineering activities. Software engineering processes and frameworks typically consider user centered design as an atomic activity that is conducted somewhere in the larger software engineering process. While this approach is suitable for the development of well-defined standard software projects for it becomes problematic in domains where the target system is not that well defined at the outset of the project. We therefore adapted an exploratory design process that integrates into agile software engineering practices for the GeoPilot project.

3.3 User Interface

The GeoPilot system features two distinct user interfaces for the two use cases. Both are designed to be as transparent as possible to the user. For automated annotation the functionality is embedded into a photo management application, but integration with online photo hosting services like Flickr would also be possible. In this use case the GeoPilot functionality works automatically in the background without active user intervention.


GeoPiLoT
Geographic Picture Locator for Tourists
ikg
Leibniz Universität Hannover

Ihr Foto zeigt: Oper

1852 im spätklassizistischen Stil erbaut, Architekt war Georg L. F. Laves, 1918 als königliches Hoftheater geführt, 1943 fast völlig zerstört

Figure 2. Offline-interface for automatic photo annotation

In the use case of on-site access to data the capturing of a photo by the user serves to trigger the retrieval process for location or object specific information. The photo is transmitted to the GeoPilot server, possibly with the addition of location information from cell-based positioning or GPS (if available), to accelerate the matching process. On the server a multi step process is executed to retrieve the augmentation information. In the first step SIFT features for the newly taken image are extracted. In the second step nearest neighbor search of the extracted SIFT feature vectors against the database of reference images is used to determine possible cadidates for matching. The candidates are then matched with the image to determine if a real correspondence exists. In this case the location and annotation information from the augmentation database is integrated into the EXIF header of the new image) and the appropriate augmentation information is retrieved from the database and transmitted to the user. This can either be traditional location-based information, using only the position information to determine which content is appropriate, or object specific information, which can be embedded in the image, leading to a new kind of snapshot augmented-reality application in which the camera becomes a kind of “magic lens”, returning either the newly taken image with additional integrated information, or e.g. alternate views from the database, 3D renderings or other forms of presentation. Figure 2 shows the classic LBS information in a web-based demonstrator, Figure 3 shows an Snap-Shot Augmented Reality view of a landmark building.

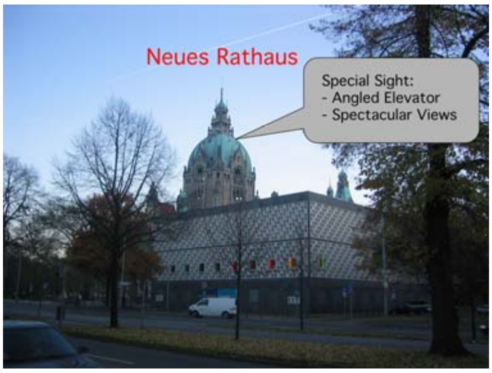


Figure 3. Annotated image in Snap-Shot Augmented Reality mode with location and object specific Information

3.4 Feature detection

For the intended application an effective way is required to determine similarities in (partly) overlapping images, taken under different lighting condition with varying equipment. Standard cross-correlation cannot easily handle the variations in the images, which has led to the development of alternative approaches to image matching that operate by detecting local features that should be invariant under a wide range of conditions. SIFT – Scale Invariant Feature Transform – is an algorithm developed by Lowe (XXX) using local features that are based on the visual characteristics and are invariant to changes in image scale and rotation, as well as robust against variations in lighting and projection. SIFT features are easy to determine and allow for effective matching against a large database of reference images, making them a popular choice for applications in which robust image matching against large databases is required. To determine the best match the image matching must be combined with an effective nearest-neighbor search as described in the following section. Figure 4 shows a visualization of the features detected for a reference image.

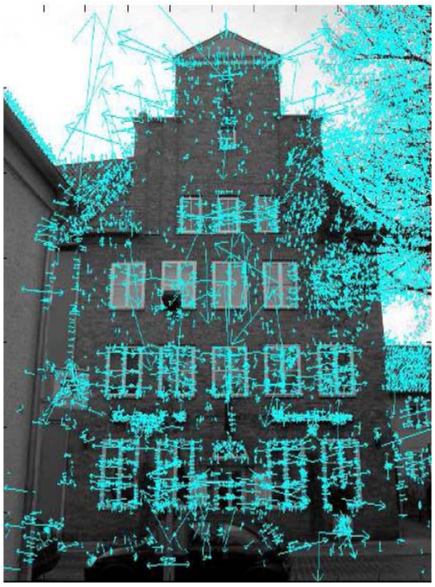


Figure 4. SIFT features extracted in refence image

3.5 Image matching

When using SIFT features as the basis to match images, each feature extracted from an image is characterized by a high dimensional vector. To identify overlapping images, the best matching feature points from two images are identified as candidates for identical points. In a second step the identity is then verified or rejected. Identifying features as possible candidates for identical points requires to determine for each SIFT feature in the current image the closest SIFT feature vector from the set of reference images. This is a nearest neighbor search problem (or proximity search) in a high dimensional vector space. The most simple solution is a linear search of the whole reference set, in which the distance for each SIFT feature in the current image to each SIFT feature in the reference set is determined and the closest stored as the current best match. While linear search is easy to implement the runtime is linear in the number of features in the reference set and thus becomes prohibitive for large reference sets.

More efficient approaches to nearest neighbour search use space partitioning data-structures. A popular example is the kd-tree, in which the search space is iteratively split into two subspaces that each contain half of the points of the parent space. A nearest neighbour search then traverses the resulting tree. Unfortunately, kd-trees are problematic in high dimensional spaces like the one spanned by SIFT vectors, so that approximation techniques and additional acceleration techniques are required to enable adequate search times.

An interesting alternative are locality sensitive hashing (LSH) techniques in which a special grouping technique is used to hash points into “buckets”. An hash function is then selected, so there is a high probability that points with a small distance in vector space are assigned to the same bucket []. In GeoPilot we have experimented with several approaches to image matching and selected an approach that uses a modified kd-tree and combines it with location information (when available) to restrict search the search spatially to possible candidate images.

Figure 5. illustrates the matches established between a reference image and a newly supplied user images (taken with a different camera and lens).

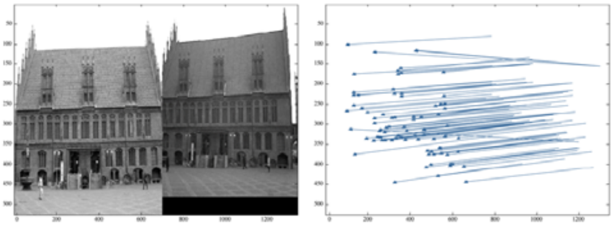


Figure 5. Established correspondences between test image and best match in the reference database.

3.6 (Geo-)Tagging and Augmentation

If a matching image from the reference set is identified and verified, the corresponding meta information (e.g. position, tags and possibly additional application specific content) is retrieved. The relevant meta-information is then integrated into the EXIF header of the image files (in all cases) and in the location based information access use case the corresponding media-elements are rendered and transmitted to the user (e.g. multimedia toursi or routing information, or augmented reality views of the location).

4. RESULTS AND CONCLUSION

The GeoPilot project used a user centered design process to drive the development of a system in which complex technologies are made accessible in an intuitive and transparent fashion to everyday users. The GeoPilot system combines current state-of-the art technologies from GIS and photogrammetry to provide efficient services to non-expert users. It shows that specialized technologies from these domains have interesting and far reaching applications in non-traditional application areas and illustrates how a user centered approach can help to identify such novel design solutions and thus lead to new application domains. The concept of snap shot augmented reality is of specific interest in cases where no continous real-time tracking of the camera is required, as it provides simple access to a wide range of augmented reality presentation techniques with commonly available hardware like smart phones, PDAs and digital cameras. Possible uses range from touristic LBS applications, over object annotation in museums to the use in the visualization of invisible infrastructure (e.g. to enable construction workers to easily check for underground pipes and cables with a mobile device).

REFERENCES

ARYA, S., MOUNT, D. M., NETANYAHU, N. S., SILVERMAN, R., AND WU, A. Y. 1998. An optimal algorithm for approximate nearest neighbor searching ┐ xed dimensions.Journal of the ACM 45, 6, 891–923.

HARTLEY, R. I., AND ZISSERMAN, A. 2004. Multiple View Geometry. Cambridge University Press, Cambridge, UK, March.

LOWE, D. 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, 2, 91–110.

NAAMAN, M., SONG, Y. J., PAEPCKE, A., AND GARCIA-MOLINA, H. 2004. Automatic organization for digital photographs with geographic coordinates. In JCDL '04: Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries, ACM Press, New York, NY, USA, 53–62.

RODDEN, K., AND WOOD, K. R. 2003. How do people manage their digital photographs? In CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press, New York, NY, USA, 409–416.

SCHMID, C., AND ZISSERMAN, A. 1997. Automatic line matching across views. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 666–671.

TOYAMA, K., LOGAN, R., AND ROSEWAY, A. 2003. Geographic location tags on digital images. In MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia, ACM Press, New York, NY, USA, 156–166.

Jia Li and James Z. Wang, ``Real-Time Computerized Annotation of Pictures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 985-1002, 2008.

Noah Snavely, Steven M. Seitz, and Richard Szeliski, Photo Tourism: Exploring photo collections in 3D," *ACM Transactions on Graphics*, 25(3), August 2006.
http://www.microsoft.com/surface/index.html

ACKNOWLEDGEMENTS

GeoPilot was a joint project conducted by the Institute of Photogrammetry and Geoinformation and the Institute of Cartography of Cartography and Geoinformatics of the Leibniz Universität Hannover. The project was co-led by Uwe Sörgel, Kian Pakzad and Guido von Gösseln. The main work on development, implementation and validation was carried out by Ulrich Dieckmann, Niels Hartmann, Behroze Ichaporia, Moritz Menze,.Raika Ost, Alexander Schunert and André Warneke.