Image Based Information Access for Mobile Phones

Xi Chen and Markus Koskela
Adaptive Informatics Research Centre
Aalto University School of Science and Technology
P.O. Box 15400, FI-00076 Aalto, Finland
xi.chen@tkk.fi, markus.koskela@tkk.fi

Jouko Hyväkkä
Media Technologies
VTT Technical Research Centre of Finland
P.O. Box 1000, FI-02044 VTT, Finland
jouko.hyvakka@vtt.fi

Abstract

Mobile phones with integrated digital cameras provide new ways to get access to digital information and services. Images taken by the mobile phone camera can be matched to a database of objects or scenes, which enables linking of digital information to the physical world. In this demonstration, we present a prototype system for image based linking of photos taken from pages of magazine. The system is intended not just for the high-end smartphones but for the current mainstream of camera-equipped mobile phones. The proposed system consists of a full architecture for a practical application developed in close collaboration with a magazine publisher.

1. Introduction

Augmenting the user's perception of her surroundings using a mobile device, i.e. *mobile augmented reality* (MAR) [5] is a relatively new field of research, which has been invigorated by the growth in number of capable mobile computing devices. These devices, while becoming increasingly small and inexpensive, allow us to use various computing facilities while roaming in the real world. Cameras and other sensors have been embedded into a large portion of available mobile phone models, and the 3G network provides a fast communication between the mobile clients and the servers. The advantages of these facilities enable the opportunities to attach more information from sources such as the internet to the objects in the real world.

In particular, ordinary mobile phones with integrated digital cameras are nowadays common, and already they can provide new ways to get access to digital information and services. Images or video captured by the mobile phone can be analyzed to recognize the object [2] or scene [11, 3] appearing in the recording. Compared with searching by text in a browser, taking a picture of an interesting object or scene by the mobile phone and then sending it to the server can be a much more convenient method to get extra information about the objects and scenes.

Consequently, the research on applicable image matching algorithms has recently been very active. A mobile image matching algorithm should be robust against variations in illumination, viewpoint, and scale. Mobile applications should work with stringent bandwidth, memory and computational requirements. This requires the optimisation of the performance and memory usage. For example, if feature extraction or image matching can be done directly on a mobile client the system latency can be reduced. In addition, distributing the computation among the users provides better system scalability.

This paper describes a research prototype system aimed at linking of images taken with a mobile phone to interactive and contextual mobile services. This kind of technology can be used for various purposes and it enables linking of the digital information to the physical world. Possible application areas include outdoor advertising, magazine and newspaper advertising, tourist applications, and shopping. In this paper, we focus on a use case with a magazine publisher as the content provider.

The rest of the paper is organized as follows. We first review briefly some relevant related work and discuss the differences to our proposed method in Section 2. In Section 3, we give an overview of the whole system architecture. In Section 4, we describe our image matching techniques in more detail. The current setup of the prototype system is described in Section 5. Conclusions and plans for future work are discussed in Section 6.

2. Related work

Using mobile phones to retrieve additional information related to the users interests has been studied in a number of research projects. Many systems, such as the Nokia's MARA [7] are based on the camera's various sensors, i.e. GPS receiver, accelerometer, and magnetometer. Recently, applications based on image analysis using the mo-

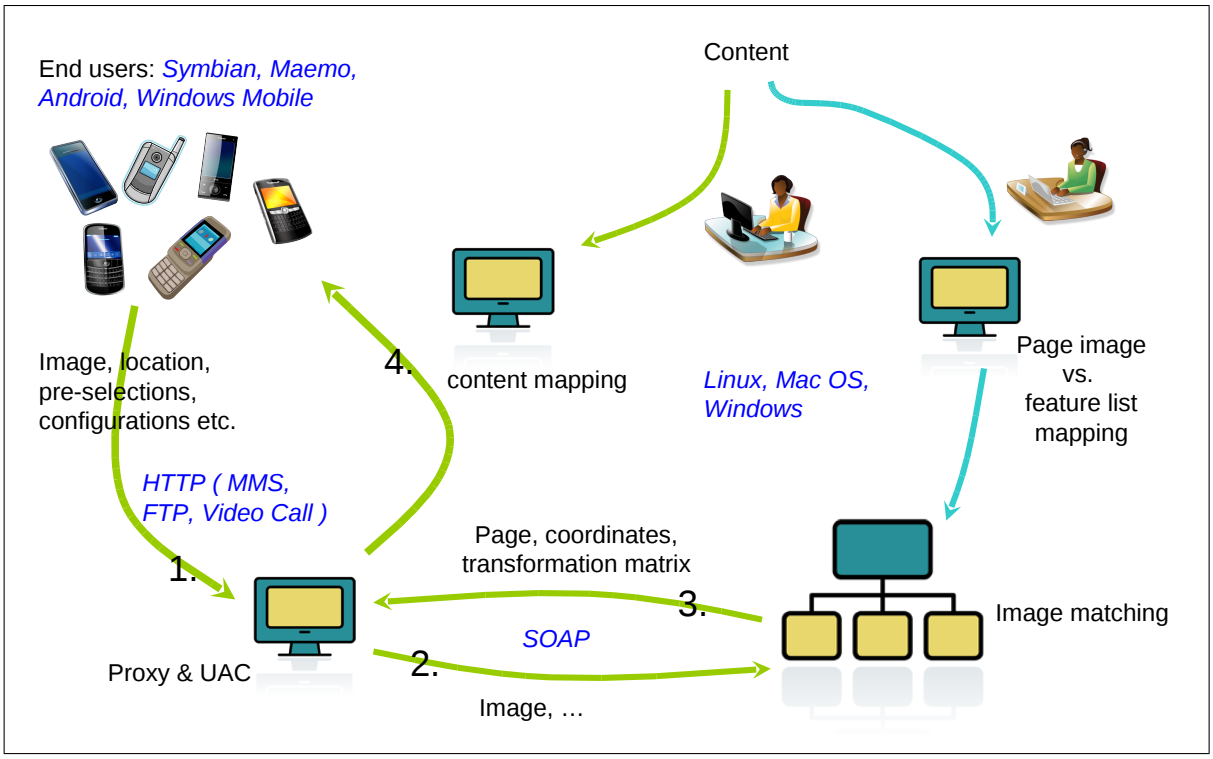

Figure 1. System diagram.

bile phone camera have also been presented. An outdoors MAR system for mobile phones is described in [11], where GPS location data is used to prune the image data prior to the image matching stage. In [3] a similar system using geotagged Wikipedia images in a MAR tourist guide is presented.

The recognition of various objects with mobile phone cameras has also raised considerable research and commercial interest. For example, an application to recognize book and CD covers from live video on mobile phones is presented in [2]. One of the most popular commercial applications with similar functionalities has been launched by Amazon at *A9.com*[1]. After taking a photo with the mobile and sending it to Amazon, corresponding information about the products appearing in the photo or similar products will be sent back to the user if the object is on sale in Amazon. Recently, Google has also launched mobile applications named *Google Goggles*[2] and *Google Shopper*[3], which can recognize book and DVD covers, wine logos, contact info, artwork, and some landmarks. A further example of similar commercial applications is *kooaba*[4].

In comparison to the applications mentioned above, we present in this paper a prototype of a magazine-content retrieval system for mobile phones. The main focus of the application is not on the recognition of magazine covers but on the content pages in the magazine. Therefore, the photos that the users can take are not limited to whole pages. Instead, the photos submitted for recognition can be of small images or details in the magazine articles, or of some advertisements. The system should also work with the mainstream of mobile phones currently in use, i.e. not just with the high-quality cameras included in the high end phones. This can result e.g. in highly blurred input images which may then require some deblurring [4].

Finally, we propose a full system architecture for a practical application developed in close collaboration with the content providers. In this sense, this work is a direct continuation of a study exploring the use of mobile phones for hybrid media in learning [10], where VTT's previous mobile image recognition application was used for linking an ordinary English study book with various digital content related to the photos taken from the book pages.

3. System architecture

From the user point of view the system architecture can be considered as a quite ordinary web service accessible

[1] *http://www.a9.com/*
[2] *http://www.google.com/mobile/goggles/*
[3] *http://www.google.com/mobile/shopper/*
[4] *http://www.kooaba.com/*

with any kind of relatively modern mobile phone equipped with a camera and an internet connection. The system architecture is outlined in Figure 1.

The image recognition and the content retrieval processes start when the user takes a photo of a magazine page with a mobile phone application and sends it for instance as a multimedia message (MMS) to the recognition system. Preferably, the image transfer is handled with a dedicated application which can add more features to the service than just the phone's own camera application. Additional data such as the user preferences and contextual information can also be sent if the data can be utilized and if the application supports it. The user's privacy and legal issues must naturally be taken into consideration when collecting the data from the user.

The query is first processed in a proxy server which can handle both anonymous and authenticated users. The data from the query is filtered and the image is forwarded to an appropriate image matching service using the SOAP protocol. The data filtering removes the unnecessary data and adds all the known meta-data so that image matching service can narrow the search as much as possible.

The image matching block returns the matching results, consisting of the matching magazine issue, page number, x and y coordinates, and the transformation matrix. The results are mapped to the related content, such as additional information, news, and videos, provided by the publisher. Links to the related content are then returned to the user's mobile phone.

Before the image matching block can process the query images, it requires access to all the supported magazine pages from the magazine publisher. The image matching database has to be constantly updated, also by removing outdated magazine pages, for example when a certain magazine campaign period is closed. Moreover, the additional content visible to the end users is even more temporary and volatile. The content provider therefore has to be able to add frequent updates to the content while the magazine page information located at the image matching service remains unchanged.

4. Image matching

In this section, the image matching component is described in more detail. The image matching algorithm is based on pair-wise matching of local features using the SIFT [8] or SURF [1] descriptors. The pair-wise matching is implemented using approximate nearest-neighbor search, and the resulting correspondences are verified using a geometric consistency check. Figure 2 shows an example matching between a query photograph and a magazine page in the database.

a) b)

Figure 2. An example matching: (a) the query photo, (b) the matching magazine page.

4.1. Feature description and database setup

The dedicated application on the user's mobile phone resizes the image and sends it via the proxy server to the image recognition server for feature extraction and matching. In general, using a higher resolution results in a larger number of local features and enables a higher accuracy in the matching, albeit with an increase in the matching time. The query images are currently resized to 512×384 pixels, which we have found to be a suitable trade-off between processing time and accuracy. Then, either SIFT or SURF features are extracted from the query image by the server. An alternative approach that we are considering would be to perform the feature extraction in the phone, as is done e.g. in [11].

We use currently two alternative database structures. First, we treat each magazine page as a single entity and match them separately to the query image. Second, we consider each issue of a magazine as a single entity, and the database consitutes of several entities depending on the number of magazine issues included in the application. Compared to the recognition of book or CD covers, the recognition of magazine content pages is more demanding since the recognition is not based on the whole page but on specific images within the pages. Several target images can exist in the same page, each linked to different extra information. Consequently, these images may be rather small in print. An example of matching such images in the database is shown in Figure 3.

When the target is a small image, the query photograph is often out of focus due to the difficulty of the mobile phone cameras to autofocus on macro distances. As a result, often only some dozens of SIFT or SURF features are extracted from the query. With such a few features in the query photo, the features extracted from the source image of each magazine page should contain enough features in order to reach sufficient accuracy also with small target images. There-

Figure 3. An example of matching a specific image on a magazine page.

fore, the image of each page in the database is stored with 1024 pixels in the longer edge. Compared for example to the system of Chen *et al.* [2] with images of 320×240 pixels, the images in our database are an order of magnitude larger. Especially for the pages full of texts the number of local features is up to 7000 in one page.

4.2. Approximate nearest-neighbor search

The next step in the matching pipeline is finding the corresponding matched features from the image database. We use the Fast Library for Approximate Nearest Neighbors (FLANN) [9] to find the pair-wise matches to the local features. FLANN supports two searching methods: hierarchical k-means trees with a priority search order and multiple randomized k-d trees. It can automatically configure the searching methods and parameters based on the user requirements for the memory usage, searching time and tree building time. In our application, the tree building time and memory usage are not consequential but the searching time is critical.

In the operation mode with each page as a separate entity, we build a set of randomized k-d trees for the descriptors of each magazine page. The query image is matched to each page in turn either exhaustively or until a page exceeding a set threshold of matched descriptors is found. We use the distance ratio criteria of [8] for accepting or rejecting a matched descriptor pair. The matched page is then verified by a geometric consistency check described below. This mode of operation is simple and accurate but does not scale up to large database sizes. Still, it can be utilized with a relatively small number of magazine issues as the average number of pages to match in a query can be reduced by using non-exhaustive processing and utilizing contextual and a priori information.

Alternatively, a hierarchical k-means tree with 32 branches is built with all the descriptors from one issue of a magazine, so the number of trees equals the number of issues in the database. The n nearest neighbors of each descriptor of the query image are returned from each tree. Then, for each page image, the total number of its descriptors in this set is calculated, and the N best-scoring pages are selected as candidates for further processing. For the candidate pages, we do a full pair-wise matching of descriptors separately to find the overall best matching pages. This approach provides much better scalability to large databases, as the complexity of the query does not grow linearly with the size of the database. The different issues of magazines do, however, have separate trees due to the dynamic nature of the application setting.

4.3. Geometric consistency

The approximate nearest neighbor algorithm typically results in a substantial number of wrong pairwise correspondences. In the studied application domain, especially the body text on the magazines produces incorrect matches.

To exclude the wrong matches from further analysis, we estimate a homography between the point correspondences using RANSAC [6] and remove the outliers. The white and red quadrilaterals in Figure 2b and Figure 3, respectively, show the estimated projections of the query photos over the matched magazine pages.

5. Demo setup

The presented demo is based on the system architecture described in Section 3 and depicted in Figure 1.

Several phones with the application software installed are available at the scene. Users can take photos of the provided magazines with our phones and submit them for recognition to the image matching server. We will also investigate the possibility of providing the application for download so users could use their own phones, if the phone model is supported.

For the demonstration, the servers will be implemented in a laptop. The proxy server will communicate with the image matching server by a localhost SOAP connection. The image matching server performs the steps described in Section 4 and sends the results back to the proxy server. The server will also visualize the matching results on the laptop screen.

The matching result is then mapped to example content provided by the magazine publisher. In the demo, we will use a static database of links as the augmented content. There will be a few dozen issues of a number of different magazines in the database.

6. Conclusions and future work

The mobile phone based magazine content retrieval system described in this paper provides the users information about advertisements and other images in magazines. The query photos do not need to be taken from whole pages. Rather, the user can zoom in and point the camera directly to the interesting content. The system is intended as a practical application developed with the content providers, and we are aiming to perform a real-user pilot study with the system in the near future.

The required resolution for the database images poses a serious challenge. Currently, the system can handle a small number of magazine issues with a tolerable latency of a few seconds. This is already enough for many practical applications, as the readers typically focus on the latest issues, and additional information is often available to narrow down the search. Still, we are investigating techniques to scale-up the database size e.g. by studying template-based techniques to speed up the matching. We are also trying to minimize the number of low-quality query images by giving feedback to the user about the captured images and rejecting distorted images.

References

[1] H. Bay, T. Tuytelaars, and L. V. Gool. SURF: Speeded up robust features. In *Proc. ECCV 2006*, May 2006.

[2] D. Chen, S. Tsai, R. Vedantham, R. Grzeszczuk, and B. Girod. Streaming mobile augmented reality on mobile phones. In *Proceedings of International Symposium on Mixed and Augmented Reality (ISMAR 2009)*, Orlando, Florida, October 2009.

[3] M. El Choubassi, O. Nestares, Y. Wu, I. Kozintsev, and H. Haussecker. An augmented reality tourist guide on your mobile devices. In *Proceedings of 16th International Multimedia Modeling Conference (MMM 2010)*, pages 588–602, Chongqing, China, January 2010.

[4] S. Esedoglu. Blind deconvolution of bar code signals. *Inverse Problems*, 20(1):121–135, February 2004.

[5] S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. *Personal and Ubiquitous Computing*, 1(4):208–217, December 1997.

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

[7] M. Kähäri and D. Murphy. MARA – Sensor based augmented reality system for mobile imaging. In *Proceedings of International Symposium on Mixed and Augmented Reality (ISMAR 2006)*, Santa Barbara, CA, October 2006.

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

[9] M. Muja and D. G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP'09)*, Lisboa, Portugal, February 2009.

[10] A. Seisto, M. Federley, and T. Kuula. Involving the end users in the development of language learning material. In *Proceeding of IADIS Mobile Learning Conference*, Porto, Portugal, March 2010.

[11] G. Takacs, V. Chandrasekhar, N. Gelfand, Y. Xiong, W.-C. Chen, T. Bismpigiannis, R. Grzeszczuk, K. Pulli, and B. Girod. Outdoors augmented reality on mobile phone using loxel-based visual feature organization. In *Proceeding of the 1st ACM international conference on Multimedia information retrieval (MIR '08)*, pages 427–434, Vancouver, British Columbia, Canada, 2008.