

A framework for Outdoor Mobile Augmented Reality

Dr. Edmund Ng Gaip Weng¹, Rehman Ullah khan², Dr Shahren Ahmad Zaidi Aduce³ Oon Yin Bee⁴

¹ FSCSHD, CoESTAR, UNIMAS
Kota Samarahan, Sarawak, Malaysia.

² FSCSHD, CoESTAR, UNIMAS
Kota Samarahan, Sarawak, Malaysia.

³ FSCSHD, CoESTAR, UNIMAS
Kota Samarahan, Sarawak, Malaysia.

FSCSHD, CoESTAR, UNIMAS
Kota Samarahan, Sarawak, Malaysia.

Abstract

Augmented reality (AR) presents a particularly powerful user interface (UI) to context-aware computing environments. AR systems integrate virtual information into a person's physical environment so that he or she can get the latest information about the environment. In a limited mobile platform we propose a framework for outdoor augmented reality which covers the main problems of limited resources in mobile, such as server dependency for data management or processing and network latency.

Keywords: Augmented reality, Context-aware, Virtual information, Image features, Image recognition, Static database.

1. Introduction

In present era, "Smart Mobile" phones offer a unique combination of features: they are easily accessible, can be carried everywhere, have photo/video capabilities and are getting powerful enough to perform complex tasks.

Due to mobile AR we are able to point our device at anything around us and, without prompting, that device will recognize what is there, incorporate our interests, and layer of information about what we're looking at. With AR technology, a consumer simply points at an object to get information. Aim at a house, for example, and find out whether the resident is selling anything on eBay Classifieds. Or point to an apartment building, and find out whether there are vacancies and what the landlord wants for rent. Augmented reality let a user look through the camera at a building and see information about restaurants, stores, and other

businesses inside. The mobile platform is effectively transformed into a looking glass the user can use to explore the world.

Our goal is to have a "tell me what I am looking at" system usable to explore objects through the mobile camera. This system relies on object information captured by mobile camera, location data (GPS), orientation of the device (compass data) and information from a database. The augmented result can be achieved by comparing the features extracted from the camera image and already stored features in database. Image recognition is based on the result of features comparison. Once the system recognizes the user's target it can augment the camera view with information like business ads, promotions, finding best restaurant etc. This paper presents a framework we have developed for supporting these kinds of point-and-find applications.

2. Related Work in Image Retrieval

Image recognition is a challenging task especially for mobile platform. Because several factors like viewing direction, lighting condition and quality of images affect this process. So the proposed framework will recognize real-world images providing solution to all these conditions.

Object categorization algorithms [1, 2] typically require an expensive training step and are less discriminative among similar looking object categories than the algorithms based on robust local descriptors [3] on which we base our system.

Research on image search services using mobile devices includes work by Zhou et al[4], which focuses on identifying matching robust local features in multiple query images. Only features that appear in many query images are used for querying a database. In this work, capturing a sequence of query images is done using a mobile phone, but the processing is done on a server.

Research on robust local descriptors is very active. Recent examples include, but are not limited to, SIFT by Lowe et al. [3], GLOH by Mikolajczyk and Schmid[5], and SURF by Bay et al. [6]. The proposed framework uses SURF algorithm which have the most favourable recognition characteristics. Our image retrieval algorithm adapts the framework of Lowe et al. [3] to work with optimised query for features retrieval, memory and computational requirements.

Recent work in object-based image retrieval uses a vocabulary of “visual words” to search for similar images in very large image collections [7]. In this formulation, a bag of visual words is used as an index during query and retrieval. The proposed framework also focuses on fast retrieval from a database of features. The researchers focus on minimizing the query time and exploiting the spatial structure of the query to develop techniques that are robust.

3. Related Work in Mobile Augmented Reality

The first commercial project to attract some attention was presented by Nokia in 2006, when Mobile Augmented Reality Application - MARA was presented as a sensor based augmented reality system for mobile [8]. The project used a GPS device connected to a Nokia S60 mobile, equipped with a standard camera, to overlay location based information on the video captured by the camera. Not many details were released about this project, now discontinued.

In the context of augmented reality, Fritz et al.[9] use a modified version of the SIFT algorithm for object detection and recognition in a relatively small database of mobile phone imagery of urban environments. The system uses a client-server architecture, where a mobile phone client captures an image of an urban environment and sends it to the server for analysis. The SURF algorithm has been used successfully in a variety of applications, including an interactive museum guide [10]. Local descriptors have also been used for tracking. Skrypyk and Lowe [11] use the SIFT features for recognition, tracking, and virtual object placement. Camera tracking is done by extracting SIFT

features from a video frame, matching them against features in a database, and using the correspondences to compute the camera pose. Takacs et al. [12] track SURF features using video coder motion vectors for mobile augmented reality applications.

Both Layar[13] and Wikitude[14] are examples of currently available mobile augmented reality browsers that rely on the location data to display directions and label buildings. Layar displays small icons and text in the direction of interesting location (which are fetched from a remote database) depending on to the direction, the camera is pointing to according to GPS, electronic compass and accelerometer data. It was first announced in May 2009 and is now available for Android and iPhone platforms.

Wikitude is similar to Layar in displaying icons and text according to orientation data acquired through GPS, compass and accelerometer. It was launched in August 2009 and allows for community created content to be added on the remote database the mobile phones connect to. Wikitude is available on Android and iPhone.

Both these applications suffer from GPS and compass accuracy issues. While they perform well in optimal conditions, in many realistic scenarios the labels are notably offset and, due to drag in compass or accelerometer data, they can be slowly moving around even if the user is perfectly still. Additionally, the only value that discerns the visualisation of a given label is the distance as the crow flies. In case of a high number of points of interest in the nearby area, this will result in labels stacked on top of each other even if other buildings are interposing between the user and the points of interest, creating a cluttered feedback that doesn't reflect the visual perception of the user.

While applications resulting from the location based approach might look similar to the one that we proposed, there is a substantial difference in how we get to the same result. In our approach, the location data is used only to reduce the search space of visual features to match, and the displayed result is ultimately derived from an image matching process, while for these applications the location data is the only input to determine, if the user is pointing the direction of a given building.

4. Contributions and System flow

In this study, the researchers have developed an outdoors augmented reality framework for matching images taken with a GPS-equipped camera-phone against a database of location-tagged

images. The system then provides the user links or services associated with the recognized object. The system is fully implemented on the mobile device, and runs at close to real-time, while maintaining excellent recognition performance.

To ensure that the image matching algorithm is robust against variations in illumination, viewpoint, and scale, the researchers adapted the SURF algorithm [6] to run on the mobile phone. The researchers optimized the performance and the memory usage of this already efficient algorithm. The matching quality of this framework is comparable to that of the original one.

Proposed mobile phone implementation of the SURF algorithm [6] focuses on using less memory consumption and on reducing computation. The algorithm consists of five major steps: interest-point extraction, repeatable angle computation, and descriptor computation, comparison among features and augmentation. In the first two steps of the algorithm, the researcher used the integral image [15] for efficient Haar transform computation as described in the original paper. In third step it creates a descriptor containing image features. In fourth step it compares features from the camera image and already stored features in real time. Finally on the bases of features comparison result it augments the camera view for user guidance.

The researchers intended that the whole process to be done directly on a mobile device for several reasons. First, this significantly reduces the system latency. Second, the system is not dependent on a server. Finally, there is no issue with changing location of the user. These are the most significant contributions of this framework.

Using location information to limit irrelevant data has been critical to our system's performance. The proposed method for indexing the query which reduces the recognition time up to one third of the client server system because there is no network connection, no downloading and uploading of features files, images or loxels. The proposed system will not access and match every record in the database but it will use the index of latitude, longitude and a key world for location as shown in the figure1 and figure 2. This combination of GPS data and the key word also play a critical role in generating a dynamic query for a web or a semantic web for farther information to guide the user.

Fig. 1 Table for storing image features.

id	Latitude	Longitude	Location_Key_Word	Img_Features
	1.4628	110.4288	FCSHD	647252.0099 659.753 5...

Fig. 2 Sample record of faculty of Cognitive Sciences and Human Development UNIMAS Malaysia.

To further enhance the performance of SURF algorithm, the researchers use multithreaded implementation in iPhone. The proposed model based on changes of some methods of open SURF library to become compatible with iPhone file system.

Figure 3 gives an overview of the proposed framework; a mobile device takes images of the real object. Features are then extracted from the query image. These features are compared against a local database using composite key of latitude, longitude and key word to access very limited range of records in the database and skipping the irrelevant records in the database. This will speed up the recognition process.

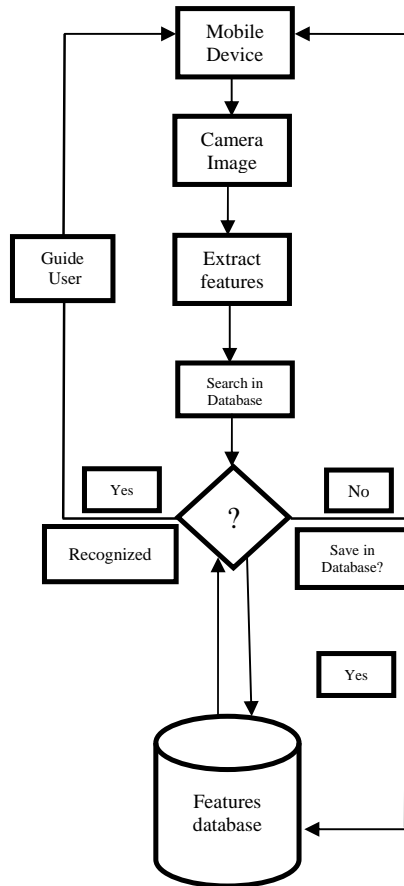


Fig. 3 System block diagram.

5. Conclusion

This study proposed a framework for building augmented reality applications based on local static database of image natural features in mobile device. This framework simplifies the development of context specific applications. Only by changing the database of features, the same framework can be used in education, business, medicine, and tourism etc. Such applications fill a vacuum between the digital and physical worlds. There is no standard data format for augmented reality; therefore it can be enhanced. In further the researchers would like to suggest future researchers to improve the processing speed and data management in static database inside mobile in more depth. Moreover, such frameworks for augmented reality can contribute in knowledge of recognition by natural features and data management.

6. References

- [1] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," *Proc. CVPR*, vol. vol. 2, pp. 264-271, 2003.
- [2] A. Ferencz, E. Learned-Miller, and J. Malik, "Learning hyper-features for visual identification," in *Neural Information Processing Systems*, vol. 18, 2004.
- [3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, pp. 91-110, 2004.
- [4] Y. Zhou, X. Fan, X. Xie, Y. Gong, and W. Y. Ma, "Inquiring of the Sights from the Web via Camera Mobiles," presented at the IEEE International Conference on Multimedia and Expo, 2006.
- [5] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE transactions on pattern analysis and machine intelligence*, vol. vol. 27, pp. 1615-1630, 2005.
- [6] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," *Computer Vision–ECCV 2006*, pp. 404-417, 2006.
- [7] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1-8.
- [8] M. Kähäri and D. J. Murphy, "Mara: Sensor based augmented reality system for mobile imaging device," in *Proceedings of ISMAR'06*, 2006.
- [9] G. Fritz, C. Seifert, and L. Paletta, "A mobile vision system for urban detection with informative local descriptors," in *ICVS '06: Proceedings of the Fourth IEEE*

- International Conference on Computer Vision Systems*, 2006, pp. 30-30.
- [10] H. Bay, B. Fasel, and L. Van Gool, "Interactive museum guide: Fast and robust recognition of museum objects," in *AMR'06 Proceedings of the 4th international conference on Adaptive multimedia retrieval: user, context, and feedback* Springer-Verlag Berlin, Heidelberg, 2007.
- [11] I. Skrypnyk and D. G. Lowe, "Scene modelling, recognition and tracking with invariant image features," in *ISMAR '04: Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'04)*, 2004, pp. 110–119.
- [12] G. Takacs, V. Chandrasekhar, B. Girod, and R. Grzeszczuk, "Feature tracking for mobile augmented reality using video coder motion vectors," in *Proceedings of the Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'07)*, 2007.
- [13] M. L.-F. a. C. B. Raimo van der Klein. (2006, 09/08/2011). *Layer*. Available: <http://www.layar.com/>
- [14] M. GmbH. (09/08/2011). *Wikitude*. Available: <http://www.wikitude.org>
- [15] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," presented at the Conference on Computer Vision and Pattern Recognition (CVPR). 2001.