The Road to the blend of Augmented Reality and Intellectual Capital: a Case of Data Management for Outdoor Mobile Augmented Reality

Rehman Ullah khan,
FSCSHD,CoESTAR,
Universiti Malaysia Sarawak
Kota Samarahan, Sarawak, Malaysia.

Muhammad Khalique
FEB, Universiti Malaysia Sarawak
Kota Samarahan, Sarawak, Malaysia.

Dr. Edmund Ng Gaip Weng
FSCSHD,CoESTAR,UNIMAS
Kota Samarahan, Sarawak, Malaysia.

Dr Shahren Ahmad Zaidi Adruce
FSCSHD,CoESTAR,UNIMAS
Kota Samarahan, Sarawak, Malaysia.

Abstract

Augmented reality (AR) presents a particularly powerful user interface (UI) to context-aware computing environments in a Knowledge-based economy. AR systems integrate virtual information/intellectual capital into a person's physical environment so that he or she will perceive that information as existing in their surroundings. In a limited mobile platform we propose a framework which covers the main problems of limited resources in mobile, server dependency for data management, processing and network latency, for outdoor mobile augmented reality. This model will be a gateway to explore and apply augmented reality and intellectual capital in future with full spirit.

Key words: Augmented reality, Image database, Image retrieval, Surf, Intellectual capital

Introduction

In a knowledge-based economy, innovation and technological advancement have great impact on the economy and lifestyle of people. For example 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. They are an attractive platform for see-through augmented reality applications. This type of applications acquires video frames through the camera, processes these images and then displays the live video on the device screen with some overlaid information and graphics. 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 web sources. Result can be achieved by performing an image matching between the image captured by the camera and an image database. The images can be matched with the database by extracting the features from the new image and looking for similar features in the database. Once the system recognizes the user's target it can augment the viewfinder with graphics and hyper-links that provide further information, guidance or services. This paper presents an architecture we have developed for supporting these kinds of point-and-find applications.

Related Work in Image Retrieval

We want to recognize real-world images but we want to do it under a variety of viewing and lighting conditions. We want the algorithms to work in the presence of transient clutter in the foreground, and changes in appearance. Object categorization algorithms (Ferencz, Learned-Miller, & Malik, 2004; Fergus, Perona, & Zisserman, 2003) typically require an expensive training step and are less discriminative among similar looking object categories than the algorithms based on robust local descriptors (Lowe, 2004)on which we base our system.

Research on image search services using mobile devices includes work by Zhou et al (Zhou, Fan, Xie, Gong, & Ma, 2006) 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. (Lowe, 2004), GLOH by Mikolajczyk and Schmid (Mikolajczyk & Schmid, 2005), and SURF by Bay et al. (Bay, Tuytelaars, & Van Gool, 2006). For our application, SURF features have the most favourable computational characteristics. Image retrieval algorithm adapts the framework of Lowe et al. (Lowe, 2004) and Khalique et al (2011) to work with stringent bandwidth, memory, computational requirements and intellectual capital.

Recent work in object-based image retrieval uses a vocabulary of "visual words" to search for similar images in very large image collections (Philbin, Chum, Isard, Sivic, & Zisserman, 2007). In this formulation, a bag of visual words is used as an index during query and retrieval. Device application and algorithms also focus on fast retrieval from a database of features. We focus on minimizing the query time and exploiting the spatial structure of the query to develop techniques that are robust.

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 (Kähäri & Murphy, 2006). 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.(Fritz, Seifert, & Paletta, 2006) 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 (Bay, Fasel, & Van Gool, 2007). Local descriptors have also been used for tracking. Skrypnyk and Lowe (Skrypnyk & Lowe, 2004) 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. (Takacs, Chandrasekhar, Girod, & Grzeszczuk, 2007) track SURF features using video coder motion vectors for mobile augmented reality applications.

Both Layar (Raimo van der Klein, 2006) and Wikitude (GmbH) are examples of currently available mobile augmented reality applications 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 an iPhone.

Both this applications suffers 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 results 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.

Contributions and Overview

In this study, the researchers have developed an outdoors augmented reality system 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 (Bay, et al., 2006) to run on the mobile phone. The researchers optimized the performance and the memory usage of this already efficient algorithm. The matching quality of implementation is comparable to that of the original one.

Proposed mobile phone implementation of the SURF algorithm (Bay, et al., 2006) focuses on using memory efficiently and on reducing computation. The algorithm consists of three major steps: interest-point extraction, repeatable angle computation, and descriptor computation. In the first two steps of the algorithm, the researcher used the integral image (Viola & Jones, 2001) for efficient Haar transform computation as described in the original paper.

The researchers intended the image matching to be done directly on a mobile device for several reasons. First, this significantly reduces the system latency. Second, distributing the computation among the users provides for better system scalability. Finally, there is no issue with the changing location of the user.

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 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. 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 utilize their intellectual capital.

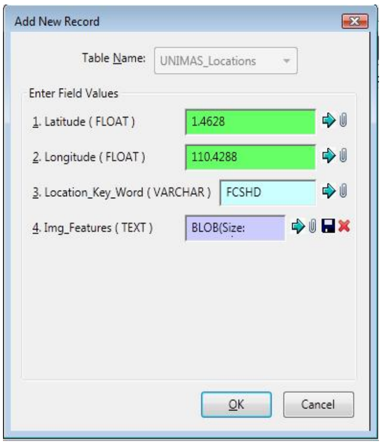

Figure 1. Table for storing image features.

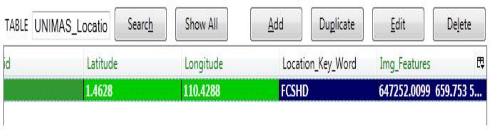

Figure 2: Sample record of faculty of Cognitive Sciences and Human Development UNIMAS Malaysia

To further enhance the performance of SURF algorithm, the researchers use multithreaded implementation. 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 model; a mobile device takes images of the real object. Features are then extracted from the query image. These features are matched 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. After recognition further information will be added to generate a new query for www or semantic web. The proposed model will extract very precise and specific information from web. Then the information will be formatted according to the limited display of mobile device. More interesting findings are that these information will be interactive.

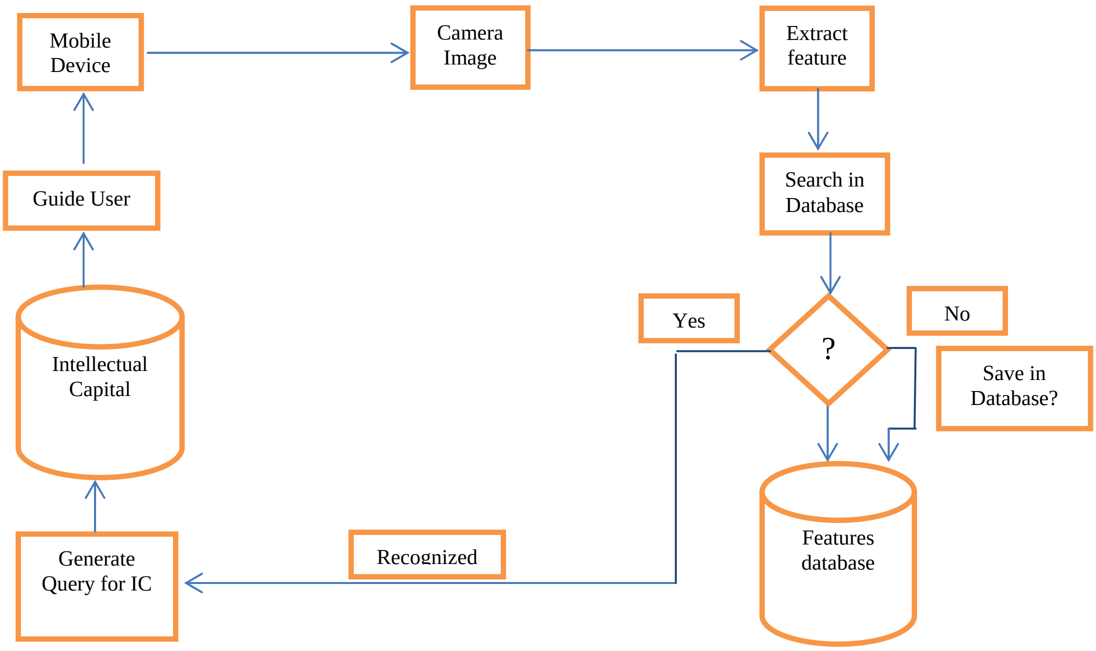

Figure 3: System Block diagram.

Conclusions

This study proposed a framework for building data intensive augmented reality applications based on local static database in mobile device. This framework simplifies the development of context specific applications. Such applications fill a vacuum between the digital and physical worlds. There is no standard data format for augmented reality; therefore, a number of interesting research questions could be identified. The proposed model invited the potential contributors to conduct empirical studies in augmented reality. In further the researchers would like to suggest future researchers to make a blend of augment reality and intellectual capital in more depth. Moreover, augmented reality and intellectual capital can contribute in a knowledge based economy in various dimensions such as technological advancement, social uplifting and economic benefits.

References

Bay, H., Fasel, B., & Van Gool, L. (2007). *Interactive museum guide: Fast and robust recognition of museum objects.* Paper presented at the AMR'06 Proceedings of the 4th international conference on Adaptive multimedia retrieval: user, context, and feedback Springer-Verlag Berlin, Heidelberg.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). Surf: Speeded up robust features. *Computer Vision–ECCV 2006*, 404-417.

Ferencz, A., Learned-Miller, E., & Malik, J. (2004). Learning hyper-features for visual identification. *in Neural Information Processing Systems, 18*.

Fergus, R., Perona, P., & Zisserman, A. (2003). Object class recognition by unsupervised scale-invariant learning. *Proc. CVPR, vol. 2,* 264-271.

Fritz, G., Seifert, C., & Paletta, L. (2006). *A mobile vision system for urban detection with informative local descriptors.* Paper presented at the ICVS '06: Proceedings of the Fourth IEEE International Conference on Computer Vision Systems.

GmbH, M. Wikitude. Retrieved 09/08/2011, 2011, from http://www.wikitude.org

Kähäri, M., & Murphy, D. J. (2006). *Mara: Sensor based augmented reality system for mobile imaging device.* Paper presented at the Proceedings of ISMAR'06.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision, 60*(2), 91-110.

Khalique, M., Shaari, J.A.N., Isa, A.H.B.M. (2011). Intellectual Capital And Its Major Components. *International Journal of Current Research, 33 (6). 343-347.*

Mikolajczyk, K., & Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence, vol. 27*(no. 10), 1615-1630.

Philbin, J., Chum, O., Isard, M., Sivic, J., & Zisserman, A. (2007). *Object retrieval with large vocabularies and fast spatial matching.* Paper presented at the Proc. of Conference on Computer Vision and Pattern Recognition (CVPR).

Raimo van der Klein, M. L.-F. a. C. B. (2006). Layer. *Layer* Retrieved 09/08/2011, 2011, from http://www.layar.com/

Skrypnyk, I., & Lowe, D. G. (2004). *Scene modelling, recognition and tracking with invariant image features.* Paper presented at the ISMAR '04: Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'04).

Takacs, G., Chandrasekhar, V., Girod, B., & Grzeszczuk, R. (2007). *Feature tracking for mobile augmented reality using video coder motion vectors.* Paper presented at the Proceedings of the Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'07).

Viola, P., & Jones, M. (2001). *Rapid object detection using a boosted cascade of simple features*. Paper presented at the Conference on Computer Vision and Pattern Recognition (CVPR).

Zhou, Y., Fan, X., Xie, X., Gong, Y., & Ma, W. Y. (2006). *Inquiring of the Sights from the Web via Camera Mobiles.* Paper presented at the IEEE International Conference on Multimedia and Expo.