Click4BuildingID@NTU: Click for Building Identification with GPS-enabled Camera Cell Phone.

Chai Kiat Yeo, Liang-Tien Chia, Tat-Jen Cham, Deepu Rajan

Center for Multimedia and Network Technology, School of Computer Engineering
Nanyang Technological University, Singapore 639798
{asckyeo,asltchia,astjcham,asdrajan}@ntu.edu.sg

Abstract
A working prototype of a building identification service which can be used on any camera cell phones equipped with GPS capability has been developed. Users can simply snap photos of architectures and send them, together with the corresponding GPS coordinates, via MMS to a remote server. The server will match the photos with the stored, GPS-tagged images using a combination of scale saliency algorithm for feature matching and earth movers distance measure for scene matching. The estimated location and other information are then sent back to the users via MMS. This prototype will have better accuracy than systems which rely solely on photo recognition given the exploitation of GPS information. Moreover, it is computationally lighter since the recognition engine only needs to compare stored images which lie within the GPS coordinates error range. It is relatively inexpensive as no special phones or subscriptions to telecommunication providers for provision of GPS equivalent location data (i.e. cell location) are needed.

1. Introduction
According to a recent report by Cellular News, Business Monitor International (BMI) forecasts that Asia’s mobile phone penetration rate will be just over 50% by 2010, reaching a hefty 1.75 billion mobile subscribers compared to 820 million at the end of 2005 [1]. This phenomenal growth projection is fuelled by some of the world’s fastest growing mobile markets in India, Vietnam, China, Indonesia and Pakistan. Such is the prevalence of the humble cell phone.

With increased penetration rate, cell phones with built-in cameras constitute 45% of all handset sales in 2005 [2]. This is forecast to surge to 75% by 2009 [2]. In addition, FCC has mandated in its enhanced 911 requirements that by the end of 2005, all cell phone carriers are required to be able to trace the mobile phone calls to a location within 100 meters or less [3] to improve the effectiveness and reliability of the 911 emergency call services. Telecommunication companies have hence been integrating Global Positioning System (GPS) tracking technology into cell phone handsets rather than overhauling the tower networks. Even for non-GPS-enabled handsets, GPS capability can be easily incorporated via Bluetooth GPS receiver module. The combination of GPS and camera technologies on the handset is the main motivation for Click4BuildingID@NTU to provide a more accurate and quicker turnaround time than normal photo positioning services.

Section 2 details the overview and operation of Click4BuildingID@NTU with a brief description of the algorithms used in the photo matching process. Section 3 provides a brief of some related work. Section 4 describes extensions to the current prototype while Section 5 concludes the paper.

2. Overview of Click4BuildingID@NTU
Figure 1 shows the block diagram of Click4BuildingID@NTU. The frontend comprises an O2 Xphone II which communicates with a Holux GPSlim 236 GPS receiver [4] via bluetooth. This combination goes to show that existing non-GPS enabled handsets can also make use of the Click4BuildingID@NTU service. A bluetooth enabled handset will therefore render the GPS receiver redundant. The backend comprises a public MMS server and the private image server. The public MMS server essentially consists of the NowSMS gateway [5] and a GPRS modem. NowSMS is a commercial SMS or MMS gateway which is used to send and receive SMS or MMS messages through the GPRS modem. SendMMS() and ReceiveMMS() respectively manage the outgoing and incoming messages to and from the NowSMS gateway. ReceiveMMS() separates the photos and text from the received message and sends the received photo together with the received GPS data to the image server for matching. SendMMS() collects replies from the image server for onward conveyance to the client. Messages are stored in the message store. The image server contains the stored images and runs the photo matching engine. All data transfers make use of the TCP/IP protocol for reliable transfer with the help of a web service.
2.1. Operation of Click4BuildingID@NTU

A user who wishes to check his location will select the Get Location from the menu. This will bring the user to the screen shot in Figure 2. To capture the GPS data, the user first selects BT Pair to invoke the Bluetooth discovery process so as to discover the GPS receiver. A pairing is done once the GPS receiver is detected and a virtual serial connection is set up for the handset to receive the GPS data. To alleviate the problem that some coordinates may be wrong due to the GPS module receiving bounced signals from nearby objects, the average of five correctly parsed coordinates are computed and the coordinate that is nearest to the average is sent with the MMS message.

The user then chooses Insert New Image. This will call up a wrapper [6] to platform-invoke the camera to snap a photo of the nearby building. The wrapper greatly enhances the user-friendliness of the application by allowing users the convenience of adding a photo when composing a MMS message. As shown in Figure 2, users no longer have to get out of the program to shoot a picture, remember the path and filename of the picture and get back into the program to attach the picture to a MMS message. In order to differentiate a Click4BuildingID@NTU MMS message from a normal MMS message, it is decided that the subject of the Click4BuildingID@NTU MMS message is set to string “Location” and the destination is the user-id of the client. To make it simple for users and also to prevent users from tampering with a location request message, the subject and the receiver are inserted automatically and the message is sent immediately upon pressing of the send button.

2.2. Photo matching engine

As Click4BuildingID@NTU requires an almost real-time response from the image server, and since computational complexity and time increase exponentially with a higher quality of image matching, the GPS data is used to speed up the matching process and to increase the matching accuracy by reducing the number of matches required.

The photo matching engine involves three phases. Phase I basically retrieves a set of stored images within a radius of 100 meters from the client’s GPS location. This greatly helps to narrow the number of matching images and therefore reduces the matching time. Should there be no images from the database, a new MMS message would be sent to inform the user that the location cannot be identified. In another scenario where the user does not send any image for matching although there are stored images retrieved, the minimum Euclidean distance between the GPS coordinates of the retrieved images and the given GPS coordinates is used to estimate the user location. The stored image with the closest coordinate and other information of the location are then sent to the user as an estimate of his exact location. In the normal scenario where user sends in a photo for matching and there are potential candidates which have been retrieved for matching, Phases II and III will ensue.

Phase II makes use of the scale saliency algorithm from Timor Kadir [7] to identify features in the image. This involved a clustering of points using a greedy clustering algorithm. As the algorithm is intended for greyscale images only, the colour images used in the prototype are first converted into greyscale images before applying the algorithm. The scale saliency is chosen for use in the prototype as it is made invariant to rotation, translation, non-uniform scaling, and uniform intensity variations and is also robust to small changes in viewpoint. The limitation in the algorithm is that different types of images will require different parameter settings in the viewpoint. For the purpose of Click4BuildingID@NTU, the parameters are set primarily for the recognition of architectures.

Figure 3 shows a greyscale version of an originally colour image with the features identified by the scale saliency algorithm. These detected features are shown as a collection of red circles, where the radii are indication of the detected scales from the algorithm. A histogram where the bins are scales is created and stored into a file. Invariably, there are some noises (i.e. areas around trees)
being detected as features in the image. However, as long as most features are detected, the noise has little impact on the overall performance.

Fig. 3. Image with features identified by the scale saliency algorithm

Phase III involves the application of the EMD (Earth Movers Distance) algorithm [8,9] on the file storing the histogram of the features extracted by the scale saliency algorithm to compare the similarity between two images. The feature space of the image is a histogram of the scale invariant, maximal entropy features collected in bins that correspond with the detected scales. The EMD typically compares how much energy is used to transform one feature space to the second feature space. The closer the energy needed is to zero, the better is the matching of the feature spaces.

Fig. 4. Location Information Returned by Click4BuildingID@NTU

Figure 4 shows the result returned by Click4BuildingID@NTU after the three phases of processing in the photo matching engine.

3. Related Work

Related work to Click4BuildingID@NTU is the recognition software to be prototyped by Cipolla and Robertson [10]. It is based on the same concept of using cell phones to snap photos and then these photos are sent to a remote server for processing. The database contains a three-dimensional representation of the real environment and it is claimed that the software can actually work out the exact location of the user to within one metre and provide user with directions. Similar to Click4BuildingID@NTU, the algorithm looks for useful features, such as the corners of windows and doors. It then extracts the colours and intensities of the pixels around them and searches the image database for matching data, using the base station the cellphone's signal came from as a guide. Finally, it uses the differences between the two images to calculate the client's position. In Click4BuildingID@NTU, we make use of GPS data to expedite the matching process. Our approach has the advantage that GPS receivers and GPS enabled cell phones are readily available and we do not therefore require the support of the telecommunication companies for the base station information. Such additional reliance on cell phone service providers will render the former to be less cost effective than the GPS approach of Click4BuildingID@NTU.

Another interesting project is World-Wide Media eXchange (WWMX) [11] by Microsoft Research. It provides a centralized index of digital photos tagged by the geographical locations where they are shot. This is an experimental project to explore various commercial and research possibilities pertaining to digital photos and geographical locations. This database will certainly be a great enriching repertoire for services such as Click4BuildingID@NTU.

Other location based services using cell phones are similar to in-car GPS navigation system where GPS-enabled Nextel phones let drivers view and listen to driving directions, complete with street names and mileage. Examples of such services are Televigation's TeleNav [12] and Motorola's ViaMoto [13]. They also provide other location-based services such as downloading business listings within 5 to 10 meters of the client’s location. The only disadvantage is that non-Nextel phones cannot be used for such services.

4. Extension Work

We have developed a working prototype and the next step will be to further enhance the scene matching techniques. There are shortcomings in the current prototype and the scale saliency method that we have employed has its limitations. Identifying distinct landmarks within the university campus is plausible but when we get down to details of each individual hall of residence (which can be quite similar when we view the
close-up architectural structure, like windows and doors), the scene matching will have a low success rate.

Therefore, we have extended our initial work to look into methods of direct one-to-one feature matching and Nearest Neighbor matching of feature distribution. We have also investigated the use of PCA-SIFT as a scale saliency method and tf-idf query vector for matching scenes [14] as shown in Figs. 5 and 6. Another approach is to investigate the use of a 3-D graphic model of the University campus to compare with the real images captured on site and this work is currently in progress.

5. Conclusion

A working prototype (Click4BuildingID@NTU) of a photo positioning service which can be used on any camera cell phones equipped with GPS capability has been developed. The prototype has been tested in the vicinity of the Nanyang Technological University campus. Clients send photos of architectures via MMS messaging to the server and Click4BuildingID@NTU is able to provide their real-time location information.

Acknowledgement – This work would not be possible without the help of two dedicated undergraduates, Lim Hock Soon and Lee Chun Sing, from the School of Computer Engineering, NTU.

6. References