Place Awareness Learned by Mobile Vision-GPS Sensor Data

Chung-Yeon Lee
Graduate School of Brain Science
Seoul National University
Seoul, Korea
cylee@bi.snu.ac.kr

Jung-Woo Ha
School of Computer Sci. & Eng.
Seoul National University
Seoul, Korea
jwha@bi.snu.ac.kr

Beom-Jin Lee
School of Computer Sci. & Eng.
Seoul National University
Seoul, Korea
bjlee@bi.snu.ac.kr

Woo-Sung Kang
DMC R&D Center
Samsung Electronics
Suwon, Korea
ws.kang@samsung.com

Byoung-Tak Zhang
School of Computer Sci. & Eng.
Seoul National University
Seoul, Korea
btzhang@bi.snu.ac.kr

Abstract

Context awareness is necessary for diverse user-oriented services. Especially, place awareness (logical location context awareness) is an essential method for a location-based service (LBS) that is widely provided to smartphone users. However, GPS-based place awareness is only valid when not only a user is located at outdoor positions where a GPS signal is strong enough but also the information on the relation between a physical location and a symbolic place is available. Here we propose a novel place awareness method using visual information as well as GPS data. The proposed method predicts the current place of a user from GPS-tagged scene photographs taken by the digital camera equipped in a smartphone. Our method uses a modified support vector machine (SVM) where scene images and GPS coordinates are given as the instances and the weight parameters of the model, respectively, for classifying the place from the scene. We evaluate our method on the place awareness from approximately 4,000 photographs of four places including hallway, classroom, restaurant, and outdoor. Experimental results show that the proposed method can precisely recognize the place with scene photos only when GPS information is not available and the awareness accuracy is improved in the GPS-available case. Furthermore, we demonstrate a smartphone application using the proposed place awareness method based on vision-GPS data.

1 Introduction

Context awareness is an essential issue for enhancing the performance of user behavior predictions and diverse user-oriented services such as a user-customized recommendation. As ubiquitous computing infrastructure is improved by the increase of smartphone users and the coverage of wireless networks, context awareness has been considered as a more important field [1]-[3]. Especially, context awareness of user location has been actively studied related to the progress of location-based service (LBS) that is a service recommending to users various helpful information based on users' current location such as the recommendation of the nearest famous restaurant [4]-[5].

Location awareness methods are separated to macro, micro, and ad-hoc styles [6]-[8]. A macro location awareness method provides a wide range of the target locations and uses GPS, a mobile communication network, and assistance-GPS (A-GPS) that is their hybrid version [10]. However, it is difficult for macro methods to precisely recognize the specific locations like indoor areas and building complexes because GPS signals become weak and their lower resolution cannot distinguish them in particular. A micro location awareness method deals with the ranges not being covered by macro methods and uses infrared ray, ultrasonic wave, or radio frequency [7]-[8], and hence micro methods require additional equipment for precise localization. Ad-hoc methods are suitable for the environment where sensors are not available to access an installed infrastructure for localization because the location is computed from the connectivity of reference nodes without the specific node providing the absolute coordinates in the ad-hoc method [9]. However, the ad-hoc method can only provide the relative and logical location information instead of the geographical absolute location.

The physical location data must be linked with symbolic representation where users are currently located, such as category of the room; name of the place, to apply the location awareness methods to valuable services such as LBS. Thus, we focus on symbolic location awareness rather than the recognition of physical location and the symbolic location awareness is called place awareness in this study. Most of place awareness methods based on GPS sensors without micro methods assume two necessary conditions: One is that a GPS signal is strong enough to precisely identify the geographical location. The other is that they have information on the relationships between a physical coordinates from GPS sensors and the name of the location like a tagged GIS map database. Therefore, place awareness using GPS-based methods is not available when users are located inside the building even though they spend most of time indoors. Moreover, it is nearly impossible to construct and update the relational information between the physical coordinates and symbolic representations at the level of the inner composition of all the buildings.

Here we propose a novel method for being aware the place where users are located based on a camera sensor as well as a GPS sensor embedded in the smart mobile device. The proposed method uses a modified support vector machines (SVMs) to learn multimodal sensor data which are geotagged scene photographs taken with the smartphone. In the learning, the photographs are represented to the vectors of SIFT (Scale Invariant Feature Transform) descriptors [11], and the SIFT vectors are used as input instance vectors. The current place of a user is classified from a new taken photo based on the learned model. This vision-based place awareness method facilitates to recognize the user’s current place without physical-symbolic mapping of location data. Besides, GPS data tagged with photos are applied to the weight parameter of the SVMs for enhancing the accuracy of place awareness.

We evaluate the proposed place awareness method based on vision-GPS multimodal sensor data, with approximately 4,000 geotagged scene photographs including four kinds of places: hallway, classroom, restaurant, and outdoor. Experimental results show that our method provides high accuracy of place awareness over 80% although GPS information is not available. Furthermore, the results present that vision-GPS multimodal data improve the accuracy of the awareness when GPS data are available.

2 Smartphone-based context-aware framework using multimodal sensor data

The context of a user generally means the current situation of the user in terms of location, time, and user-related task such as a schedule [12]. The increase of usable sensors makes context awareness become powerful because the user situation can be more precisely recognized by integrated information of diverse sensors. Thus, a smartphone that equips various sensors such as camera, GPS receiver, accelerometer, and so on is a very effective and suitable device for recognizing the context of users. We are studying a context awareness method using multimodal sensor data based on the smartphone as shown in Figure 1. Our smartphone-based context awareness framework consists of three parts: data integration module, context awareness engine, and context-aware based smartphone applications. Data integration module preprocesses and integrates the multimodal data from various heterogeneous sensors equipped in user’s smartphone. Context awareness engine consisting

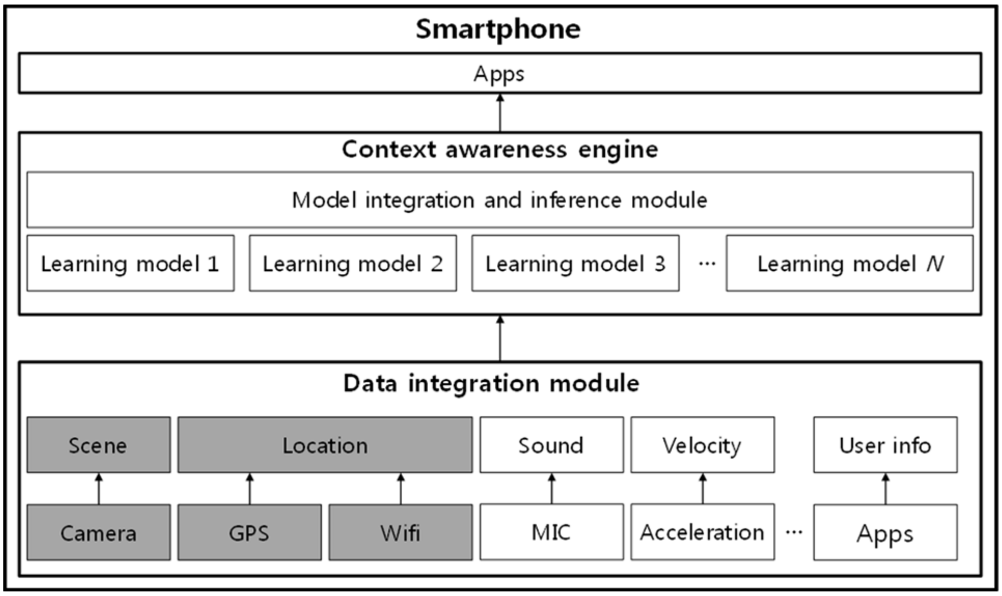


Figure 1: Framework for user context awareness using multimodal sensor data based on smartphones

of learning and inference modules recognizes and predicts the user contexts with respect to place, current time, and tasks. Smartphone applications are then conducted based on the predicted contexts. This paper covers the awareness of the user's current place based on camera and GPS sensors among the overall framework.

3 Place awareness with vision-GPS sensor data

3.1 Data gathering and preprocessing

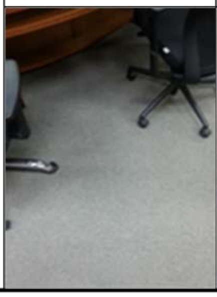
The place awareness framework based on vision-GPS multimodal sensor data recognizes the place where users are located with a learned SVM classifier from the geotagged scene photo taken with user's smartphone. The proposed framework assumes five conditions:

1) User's smartphone is locked by user or time.
2) Whenever the smartphone is unlocked, the embedded camera takes a scene without user's notice.
3) When the user unlocks the smartphone, the direction of the camera lens is frontal or descending.
4) The place where the user is located is light enough to recognize the taken scene.
5) Whenever a photo is taken, the GPS coordinate value of the location is tagged in the image.

The assumption above does not restrict the practical use of the proposed method because the conditions involved in the assumption are usual cases for smartphone users.

Data for learning the awareness model consists of approximately 4,000 photo images of four places including classroom, hallway, restaurant, and outdoor and they are taken with a camera-equipped smartphone (Samsung Galaxy S3, 8M-pixel front camera) under the conditions above. All the places are located at one building in Seoul National University. Images of classroom are taken at four different classrooms and images of hallway are taken at six different floors of the building. The resolution of all images is fixed as 480×640 pixels with 72 dpi, and flashlight and optional image corrections are not used. The levels of lens

Table 1: The size of image datasets and the example images of 4 target places

	Classroom	Hallway	Restaurant	Outdoor
Frontal Direction	430	497	548	553
				
Descending Direction	544	494	520	518
				

opening and exposure are automatically controlled by the camera. Table 1 illustrates the number of image dataset and the examples of the images of the 4 target places.

Feature vectors are then extracted from each image using SIFT descriptor after the images are resized to 120×160 pixels and transformed to grayscale images.

$$Descriptor(x, y) = [\theta(x, y), m(x, y)] \tag{1}$$

where $\theta(x, y) = \tan^{-1}\left[\dfrac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}\right]$ and

$$m(x, y) = \left[\left(L(x+1, y) - L(x-1, y)\right)^2 + \left(L(x, y+1) - L(x, y-1)\right)^2\right]^{\frac{1}{2}}.$$

As seen in (1), the SIFT descriptor is derived as the vector of 128 dimensions including histograms computed at all the 4×4 grid points and 8 quantized directions for each interest point. However, using all feature vectors increase the cost of learning greatly in proportion to the number of training images. Thus, we reduce the dimension of the feature vectors using k-means clustering. Since the size of feature vector is determined by k, we find the optimal k for effective learning.

GPS coordinates of each place are measured by A-GPS system equipped in the smartphone while the images are taken. To preprocess them, GPS coordinates recorded in EXIF (Exchangeable Image File format) meta data of each image are extracted, and the notation of coordinates consisting of degrees, minutes and seconds ($dd°mm'ss''$) are converted to the degree level (*DD.DDDD*) using (2)

$$DD = dd + \left({}^{mm}\!/_{60} \right) + \left({}^{ss}\!/_{3,600} \right) \quad (2)$$

GPS data which exceed the range of the real coordinates of the building for this experiment (the north latitude: 37.4480~37.4490; the east longitude: 126.9515~126.9530) are excluded from the training data to avoid the efficiency loss due to outliers. The numbers of remaining GPS data after excluding the outliers are 84 of classroom, 494 of hallway, 326 of restaurant, 493 of outdoor for descending direction, and 58 of classroom, 197 of hallway, 287 of restaurant, 322 of outdoor for frontal direction.

3.2 Learning model for place awareness

The place awareness based on the feature vectors extracted from images is evaluated using a SVM classifier. In this experiment, we use LIBSVM [13] which runs Sequential Minimal Optimization (SMO) [14] that is widely using for quadratic programming problems.

SVMs can be learned with adding weights to features regarded as more interested. Thus, we use GPS data acquired simultaneously with images as the weights for SVMs classifier to enhance the accuracy of the place awareness. (3) is the objective function of SVMs using the extracted GPS data as weight vector W.

$$f(x) = \sum_{i=1}^{M} W_i \alpha_i^* K(X_i^*, X) + b^* \quad (3)$$

where $W = DD_E \times DD_N$, X_i^* is i-th vector of M support vectors derived from the learning, and the optimization bias $\alpha^*$and b^* are solutions of the quadratic programming problem while learning. Radial Basis Function (RBF) [15] is used for the kernel function in (4).

$$K(X_i^*, X) = \exp\left\{ -\gamma \left\| X_i^* - X \right\|^2 \right\}. \quad (4)$$

4 Experimental results

Table 2 shows the accuracies of the classifications on four place scenes and the elapsed times for learning when using scene photos only. The values in the Table 2 are averaged after 10

Table 2: Comparison of accuracy and learning time for place awareness as the number of visual features

k	Frontal		Descending	
	Accuracy (%)	Learning time (sec)	Accuracy (%)	Learning time (sec)
50	69.34	4.896	75.64	4.096
100	72.45	5.799	78.75	4.218
200	74.61	9.740	80.03	6.964
500	80.25	29.279	80.87	23.613

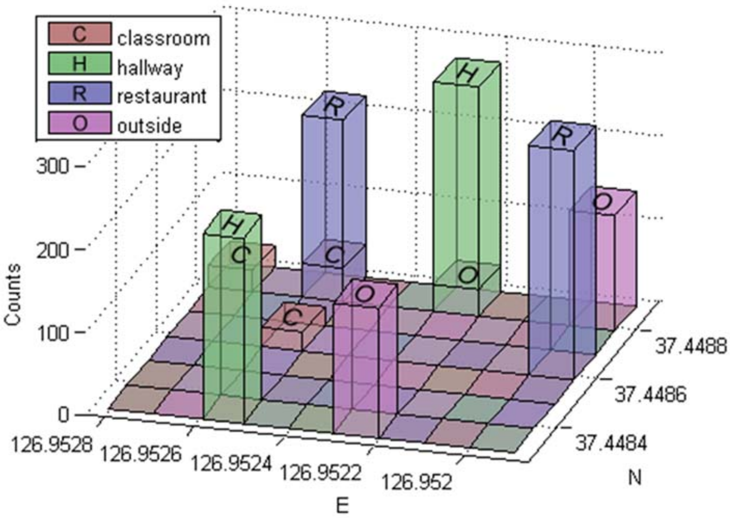


Figure 2: Histogram of GPS coordinates of four places

experiments of 10 fold cross-validation. The accuracies are higher for the classification of descending directional images than frontal directional images, and elapsed times are also shorter for descending directional images. The highest accuracies in both directional cases were shown in the experiments using 500-dimensional feature vectors that is k=500. It is natural that the larger k provides the higher accuracy because more features can characterize the images more precisely. However, it is more effective to use 200-dimensional feature vectors in consideration of the learning time.

Figure 2 shows the geodesic histogram on GPS data of the descending direction after preprocessing. GPS data explicitly represent the geographical location but they are not precise and sparsely dotted due to its limitation of the resolution while feature vectors of image dataset characterize the logical location. Besides, some data of different places are overlapped each other such as classroom-restaurant and hallway-outdoor because GPS data can't recognize the places located at different floors on the same point.

The place classification using only GPS data with the decision tree algorithms (J48 implemented in WEKA [16]) shows largely high accuracy (around 95%), but the problem is that data are not reliable since only 41% of frontal directional data and 68% of descending directional data are used, and specifically 13% and 15% of GPS data are used for classroom. Moreover, GPS data themselves have a limitation to represent logical information of places without tagging, but it is impossible to tag entire places in all the buildings. Nevertheless, GPS data can provide consistency between feature vectors extracted from images of same places because of its discrete property. Table 3 shows the accuracy of place awareness in consideration of comparison between using image data only and image data with GPS data. Using image data with GPS data shows higher accuracies in both frontal directional and descending directional dataset. Consequentially, GPS data have been used as weight vectors

Table 3: Comparison of accuracy of place awareness for different data types

Modality	Accuracy (%)	
	Frontal	Descending
Scene Only	78.63	80.07
Scene with GPS	81.83	81.00

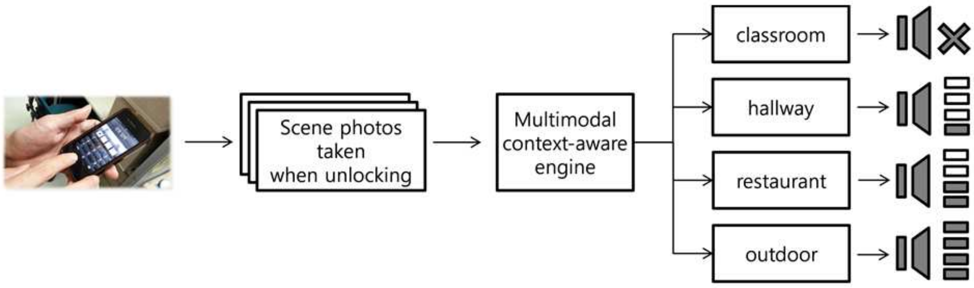


Figure 3: Application scenario: automatic bell volume controller based on user's current place using vision-GPS data-based place awareness engine

for classification of feature vectors of image dataset and improved the accuracy of place awareness.

The proposed place-aware method can be applied to diverse services for smartphone users. In this study, we implemented a simple android-based mobile application which autonomously controls the bell volume of the smartphone based on the user's place using our framework as shown in Figure 3. The application sets up the volume level of the bell or changes the calling alarm from sound to vibration mode vice versa depending on the current place recognized with the scene photo taken when a user unlocks the smartphone. This application can reduce a mistake that a user disturbs the class due to a loud ringing bell.

5 Concluding remarks

We have introduced a smartphone-based context awareness framework using multimodal sensor data and proposed a novel method for place awareness based on a camera sensor as well as a GPS sensor embodied in the smartphone among the framework in this study. We use a modified SVM as a classification model for place awareness. Scene images represented to the vectors of SIFT features are used as instances and GPS coordinates tagged with the images are applied to the weight parameters in our model. The learned model classifies a geotagged scene photo taken with user's smartphone as the current place of the user. The proposed method is evaluated for the place awareness with approximately 4,000 photographs of four places including hallway, classroom, restaurant, and outdoor. Experimental results show that it can precisely recognize each place even with only vision data with awareness accuracy of over 80%, and the accuracy becomes higher when vision-GPS multimodal data is used. Finally, we demonstrate a simple mobile application based on user's place recognized by the proposed vision-GPS based place awareness method.

Future work plans to add more sensors such as sound and accelerometer and learning models into our context-aware framework for enhancing the accuracy of the awareness. Moreover, we will implement the useful smartphone applications based on the proposed framework.

Acknowledgments

This research was supported by the National Research Foundation of Korea (NRF 2012-0005643, Videome), the BK21-IT Program funded by MEST, and Samsung Electronics Co., Ltd.

References

[1] Chen, G. & Kotz D. (2000) A survey of context-aware mobile computing research. *Dartmouth Computer Science Technical Report*. Hanover, NH: Dartmouth College.

[2] Kanter, T. G. (2003) Attaching context-aware services to moving locations. *IEEE Internet Computing* **7**(2):43-51.

[3] Hong, J., Suh, E. & Kim, S. J. (2009) Context-aware systems: a literature review and classification. *Expert Systems with Applications* **36**(4):8509-8522.

[4] Dhar S. & Varshney U. (2011) Challenges and business models for mobile location-based services and advertising. *Communications of the ACM* **54**(5):121-128.

[5] Chon, J. & Cha, H. (2011) LifeMap: a smartphone-based context provider for location-based services. *IEEE Pervasive Computing* **10**(2):58-67.

[6] Whitehouse, K. & Culler, D. (2003) Macro-calibration in sensor/actuator networks. *Mobile Networks and Applications* **8**(4):463-472.

[7] Hightower, J. & Borriello, G. (2001) Location systems for ubiquitous computing. *Computer* **34**(8):57-66.

[8] Koyuncu, H. & Yang, S. H. (2010) A survey of indoor positioning and object locating systems. *International Journal of Computer Science and Network Security* **10**(5):121-128.

[9] Blazevic, L., Le Boudec, J. Y. & Giordano, S. (2005) A location-based routing method for mobile ad hoc networks. *IEEE Transactions on Mobile Computing* **4**(2):97-110.

[10] LaMance, J., DeSalas, J. & Jarvinen, J. (2002) Assisted GPS: a low-infrastructure approach. *GPS World* **13**(3):46-51.

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

[12] Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M. & Steggles, P. (1999) Towards a better understanding of context and context awareness. *Lecture Notes in Computer Science* **1709**:304-307.

[13] Chang, C. C. & Lin, C. J. (2011) LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*. **2**(3):27:1-27:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

[14] Platt, J. C. (1998) Fast training of support vector machines using sequential minimal optimization. In B. Schölkopf, C. J. C. Burges and A. J. Smola (eds.), *Advances in Kernel Methods-Support Vector Learning*, pp. 185-208. Cambridge, MA: MIT Press.

[15] Suykens, J. A. K. & Vandewalle, J. (1999) Least squares support vector machine classifiers. *Neural Processing Letters* **9**(3):293-300.

[16] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I. H. (2009) The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter* **11**(1):10-18.