Vision-Based Location Positioning using Augmented Reality for Indoor Navigation

JongBae Kim and HeeSung Jun

Abstract — *In this paper, we propose a vision-based location positioning system using augmented reality technique for indoor navigation. The proposed system automatically recognizes a location from image sequences taken of indoor environments, and it realizes augmented reality by seamlessly overlaying the user's view with location information. To recognize a location, we pre-constructed an image database and location model, which consists of locations and paths between locations, of an indoor environment. Location is recognized by using prior knowledge about the layout of the indoor environment. The image sequence is obtained by a wearable mobile PC with camera, which transmits the images to remote PCs for processing. The remote PCs perform marker detection, image sequence matching, and location recognition. The remote PCs transmit the recognized location information to the wearable mobile PC. The system provides the ability to identify similar locations in the image database and display location-related information. Accurate and efficient location positioning is achieved by using several vision-based techniques. The proposed system was tested in an indoor environment and achieved an average location recognition success rate of 89%. The proposed system could be applied to various consumer applications including the door plate system, notice board system, shopping assistance system, and bus service route guide system, among others[1].*

Index Terms — Location positioning, Indoor navigation, Image sequence matching, Location model.

I. INTRODUCTION

Computers, from simple calculators to life-saving equipment, are ubiquitous in modern society. In many areas, computers increase human efficiency and save time. Thus, humans have come to expect rapid access to important information through computers. The rapid development of computer technology has yielded the shrinking of computer size and increasing of processing power. Such progress is well realized by wearable computers [1-3]. Wearable computers are small, worn on the body during use, and intended to provide information not directly detectable through human senses. Wearable computers that seamlessly overlay real-time information regarding the human environment may contribute to a convenient and efficient human lifestyle. The augmented reality (AR) technique has been developed to achieve human lifestyle enhancing results, and it is currently used in several applications [3]. AR is an effective means for utilizing and exploiting the potential of computer-generated information. AR techniques are applied to systems for applications including monitoring, remote intelligence, military applications, and location positioning [3-5]. More recently, AR technology has been actively studied for indoor positioning applications. Indoor positioning systems currently use GPS-based [6-8], sensor-based [9, 10], and RFID-based systems [11-14] to establish user location. Among them GPS-based systems are the most common.

In this paper, we propose a vision-based indoor location positioning system that uses the AR technique and does not require any additional devices. Our system employs the AR technique to provide location information to persons unfamiliar with the layout of an indoor environment. In vision-based methods, locations in indoor environments can be recognized by first characterizing each location with special identifiers. Alternatively, as with previously mentioned systems, location can be established using the signal strength of RF(radio frequency) bands of the IR echo distance [15, 16]. The RFID-based system characterizes location by measuring the strength of a signal received from an RF sensor attached to known locations. In our system, however, specific locations are each identified by a marker and with color information and prior knowledge. Here, each marker has a black and white colored square with a characteristic pattern. And topographical information of the indoor environment is established from the prior knowledge of location, and it is represented in the location model by a hierarchical tree structure. Our system has several distinct advantages over all other location positioning systems. First, it is an economical solution because the marker that identifies each location is a simply printed on paper, and the mobile PC and camera are general devices that can be implemented in software. Second, although our system has inferior performance than other systems such as the GPS or sensor-based location positioning systems, it is not limited by signal propagation and multiple reflections. To increase performance, it adapts via an adaptive thresholding method [3] to detect markers under illumination changes, and it uses the location model to reduce execution time during image sequence matching.

The paper is structured as follows. Section Ⅱ describes researches related to the location positioning in indoor environments. In Section Ⅲ and Ⅳ, we present an overview and a detailed explanation of the proposed system. Section Ⅴ of this paper contains a discussion of our results. In Section Ⅵ, we present the conclusions of this paper.

JongBae Kim is with the School of Computer Engineering, Seoul Digital University, Seoul, S. Korea.(e-mail: jbkim@sdu.ac.kr)
HeeSung Jun is with the School of Computer Engineering and Information Technology, University of Ulsan, Ulsan, S. Korea.(e-mail: hsjun@ulsan.ac.kr).

Manuscript received June 21, 2008

II. Related Works

Indoor positioning systems provide location information of indoor environments using various sensors. Previously developed systems are shown in Table I. As shown in Table 1, previous systems have recognized user location in indoor environments with a sensor. These systems work by measuring the angle of signal arrival, the time difference of signal arrival or the signal strength of the sensor. However, the performance of previous systems, including their accuracy and efficiency, is highly dependent on the number of sensors and the structural characteristics of indoor environments. Below is a brief description of some of the location positioning technologies that have emerged in the past few years to aid in navigation.

TABLE I
EXAMPLE OF LOCATION POSITIONING

System	Technique	Usability
Smart Sight [17]	GPS	outdoor
Cricket [9]	RF + Ultrasonic	indoor
Active Badge [10]	Infrared IR	indoor
RADAR [11]	RF	indoor
Drishti [18]	GPS + Sensors	outdoor, indoor
CyberGuide [19]	GPS + Infrared IR	outdoor, indoor
SHOSLIF [20]	Vision	indoor
H. Aoki's method [21]	Vision	indoor

A. *GPS-based Location Positioning*

Major developments of location positioning systems have been focused on enhancing positioning capability in open outdoor environments. The most obvious way to determine location information is to use GPS receivers, which can determine their positions to within a few meters in outdoor environments. Therefore, the GPS-based location positioning systems are used for precise applications such as aircraft or vehicle navigation systems. Unfortunately, GPS radio signals have difficulty in penetrating building walls. Thus the relative distances between the reference points and the location device cannot be easily calculated. This difficulty arises because the communication paths of the GPS radio signals are long and not always empty. So, the GPS-based location positioning devices do not work well indoors, or in many outdoor areas, because the satellite signals are not strong enough to penetrate building walls, dense vegetation, or other obstructive features.

B. *Sensor-based Location Positioning*

Sensors, IR, electromagnetic wave and other sensors have been adapted for location positioning systems. Sonar or IR sensors placed at fixed positions within a room-scale space receive sensor signals and, through software analysis, make the sightings available to services and applications. To operate effectively, sensors require the deployment of large arrays of hundreds to thousands of IR beacons on ceiling tiles. The position and orientation of the IR sensor are estimated by sighting the relative angles and positions of the ceiling IR beacons. A single transmitter emits a magnetic signal to track the position and orientation of numerous sensors. The location error of these sensors is typically within a few meters. However, these techniques require extensive wiring, which makes them prohibitively expensive and difficult to deploy. More importantly, these techniques only allow for room-scale location, and are therefore not suitable for wide-scale deployment.

C. *RFID tag-based Location Positioning*

The RFID-based location positioning method uses the RF tags and a reader with an antenna to identify user location. Tags are generally affixed to places or objects such as consumer equipment so that the objects can be located without line of sight. Tags contain circuitry that gains power from radio waves emitted by readers in their vicinity. Tags then use this power to transmit their unique identifier to the reader. The detection range of these tags is approximately 4-6 meters. Therefore, this technique offers room level precision. As result, the use of the RFID tag method requires a great infrastructure in order to be highly precise and effective at location positioning. Additionally, the use of sensing equipment for indoor location positioning results in exorbitant costs. Likewise, the installation of RFID sensors on ceiling tiles or walls can be costly. Thus, there exists a need for a location positioning system that overcomes the shortcomings associated with current indoor positioning methods while still being economical. The development of such a method is discussed herein.

III. Overview of The Proposed System

A. *System Design*

Location positioning systems for indoor environments should have minimal weight and consistent performance. Therefore, our location positioning system comprises a mobile tablet PC, wireless camera, Head Mounted Display (HMD), and desktop PCs with a wireless LAN interface. Fig. 1 shows a front and top view of a user using our system. The mobile PC is used to input the destination and to display the location information and map. The HMD is used to annotate the direction sign on the user's view with location information. The camera, which is mounted on the user's cap, captures image sequences and transmits them to remote PCs via a wireless LAN. The remote PCs estimate location information from the received image sequences and transmit the result to the user's mobile PC.

B. *Overview*

A process diagram of our system is shown in Fig. 2. Given an image sequence taken as input from a cap-mounted wireless camera, the proposed system annotates the recognized location information. Our system mainly consists of two parts: wearable units and remote PC units. The wearable units capture and display images. The remote PC

Fig. 1. Block diagram of the system

units conduct four main processes: marker detection, image sequence matching, location recognition, and location annotation. Once the mark detection process identifies a potential marker in an image sequence, the remote PC units output location information that corresponds to the predefined marker patterns. Then, the image sequence matching process outputs location information by analyzing both the features of the input image sequences and the image sequences stored in the location dictionary. The results of the previous two processes and the location model are used to pinpoint the user's current location. Finally, the location annotation process annotates the direction sign on the HMD with current location information.

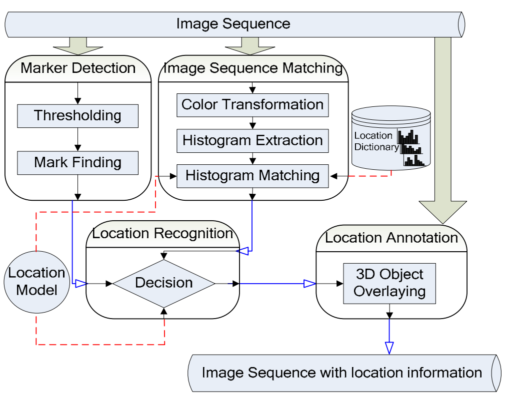


Fig. 2. Processes diagram of the proposed system

IV. Location Positioning in Indoor Environments

A. Marker Detection

This process detects a marker from an input image sequence. The image sequence shows the indoor environment in front of the user, e.g., doors, walls, or hallways. This process primarily distinguishes the candidate marker and non-marker regions (e.g., background) in the image sequence and detects real markers in the candidate marker regions. To characterize each location, a 20×30cm marker that consists of an asymmetric, black and white color square pattern is used. Each marker is affixed to a specified position in an indoor environment such as a wall, door, stairs, corridor, or wall board plate. The markers come with the ARToolKit [22] and ARTag [23].

It is generally difficult to detect the marker from an image sequence. The image sequences obtained by the cap-mounted camera have several artifacts, such as noise from digitization. Moreover, the image sequence records the global motion of the user. Due to such conditions, the marker detection process must operate in real-time. To do this efficiently, the process requires a method to reduce the extent of each image that must be searched for a marker. Therefore, we adapted vision techniques to a process that aids in marker detection. This process consists of image thresholding, labeling, contour detecting, square checking, and pattern matching. First, the input images are turned in binary images using the thresholding method with a specified threshold value. Then, square regions (candidate marker regions) are distinguished from each binary image by the labeling, contour detecting, and square checking processes. The marker detection process finds all the squares in each binary image. For each square region, a pattern inside the square is scanned and compared to pre-trained pattern templates. As a result, location information is detected by matching the detected pattern in a square region to a pre-registered pattern template that designates a specific location in an indoor environment.

The image thresholding step has to be performed before the steps above because it can distinguish the pixels of candidate marker regions from whole pixels in an image. The performance of the marker detection process depends highly on the performance of the image thresholding step. Currently, visual markers are widely used in several applications [24-26]. All these systems use a static thresholding method with a fixed value for image thresholding. Therefore, the threshold value of those systems has to be adjusted manually for optimal thresholding. Because lighting conditions in real environments vary often and quickly, the image thresholding step must be adaptable to varying lighting conditions. In our system, the adaptive thresholding method [28] for making the binary image is applied during the image thresholding step. In this method, the threshold value is automatically changed according to the lighting conditions.

Fig. 3(a) shows the original image and, Figs. 3(b)-(d) are the binary images generated by the fixed and adaptive thresholding methods, respectively. As shown in Fig. 3, the adaptive thresholding method under various lighting conditions is performed more efficiently. To calculate the threshold value, adaptive thresholding is conducted every 64 frames, as to allow for adaptation to rapidly changing lighting conditions.

After this step, each pixel with a higher value than the threshold value is labeled, the contour line is extracted from the labeled pixel with a permitted size, and then square regions are selected from the contour regions with 4 vertices by the square checking. Each square region is regarded as a candidate marker region, and if the pattern inside a candidate marker region matches a pre-training pattern template, then the marker is found to correspond to the pattern. As noted above, this process outputs the detected location information (location ID) in the image sequence.

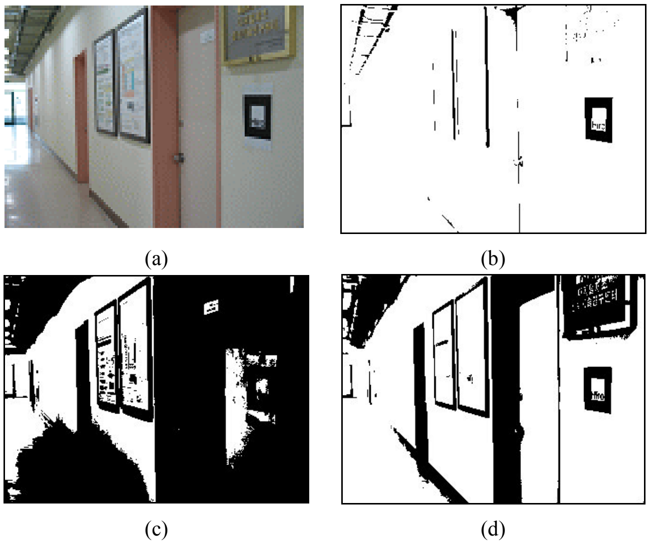


Fig. 3. Results of image thresholding using the fixed and adaptive threshold value. (a) original image, results of threshold image using the threshold values θ=70 (b), 150 (c), and 112 (d)

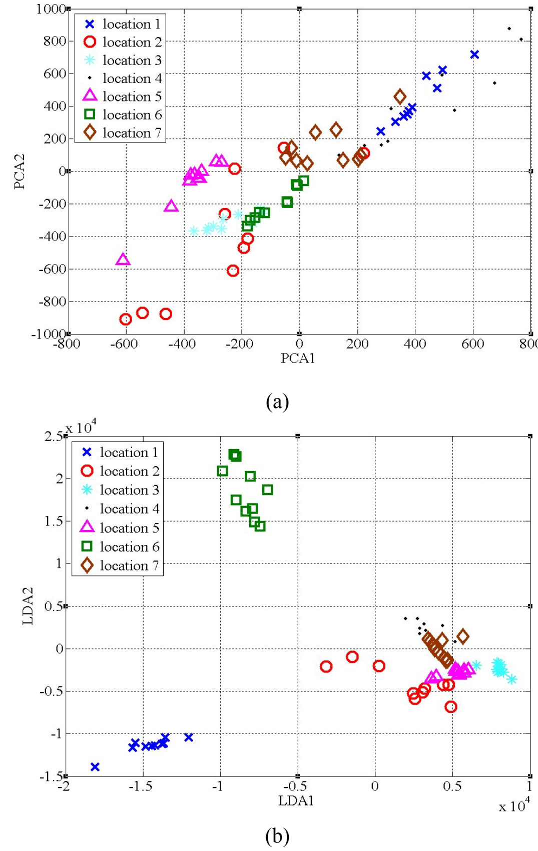


Fig. 4. PCA (a) and LDA (b) distribution of the first-two features of each location

B. Image Sequence Matching

This process estimates user location by analyzing between features of the input and pre-stored image sequences. The location dictionary is constructed from the video sequence captured by a user walking in an indoor environment. The video sequence used for the location dictionary was recorded over a period of 30 minutes at 12 frames per second. The video sequence was then sub-sampled to 8 frames per second to create the location dictionary. Subsequently, the user annotated the sub-sampled frame number and location ID at each location where he stepped into a new area. At each of these locations, features of the sub-sampled frame were calculated and stored in the location dictionary. Each location in the dictionary is represented by a bundle of 64 frames. Therefore, this process compares every 64 frames of the input image sequence to the 8 frame bundle representing each location in the location dictionary. After constructing the location dictionary, features of input image sequence are calculated and matched to features in the location dictionary. When the input image sequence is adequately similar to the pre-stored image sequence in the location dictionary, the image sequence matching process shows the location information. The recognition task is difficult because indoor areas are often of that includes similar features such as hallways, rooms, and lobbies, often have similar features. Therefore, features must be used to characterize each location for matching. Although they are obtained at consistent locations, image sequences may differ widely since they are acquired from different camera viewpoints. Therefore, to operate effectively, the image sequence matching process must be able to characterize whole frames of an image sequence with features of the image. However, more features do not necessarily result in more accurate location recognition. Instead, each feature must have strong location discrimination power [20, 27].

Color, as a component of visual context, is an important source of information for image matching. Therefore, many researchers have used color information to match images. H. Aoki used a hue histogram with 32-bins to match frames [21]. However, the use of a hue histogram to match features has several limitations. When large variation exists between frames, prominent hue features and the global contrast regions may not be stable. In response to this deficiency, the proposed method identifies features as the variation of color information between frames in an image sequence. However, the dimensionality and number of image sequences is large. Thus, a dimension reduction method is applied in real-time in our system. Linear Discriminant Analysis (LDA) is used to effectively reduce the dimension of the image feature space and to successfully discriminate each location [27].

Principal Component Analysis (PCA) is generally used to reduce the feature dimension. But, PCA extracts the most expressive features in an image due to their having the most significant variation in the feature distribution. Fig. 4 shows the PCA and LDA distributions of the first–two features of 10 frames at each location (total 7 locations). Fig 4(a) shows the PCA features to be scattered, and the features of different locations are mixed. However, the features identified by LDA are grouped, and the features of different locations are not nearly as mixed as those identified by PCA. Therefore, to discern the location, this process applies the Euclidian distance function of the LAD features.

The color information obtained by a camera is influenced by many factors (ambient light, object movement, etc.). The color information of two images taken under the same lighting conditions may be significantly different. Also, the color space is sensitive to illumination color changes. One way to increase resilience to changes in lighting intensity is through the use of chromaticity features. Therefore, the image sequence matching process transforms the input RGB space into an HSI space, and the first-five LDA features are extracted from the hue space. To determine the current location from the input image sequence, it is most simple to measure the Euclidean distance between features. Therefore, this process calculates the similarity between feature vectors (x_i) of the input image sequence (*X*) and feature vectors (y_i) of the image sequences (Y^n) in the location dictionary with *n*th locations. The similarity calculation process is shown in Fig. 5.

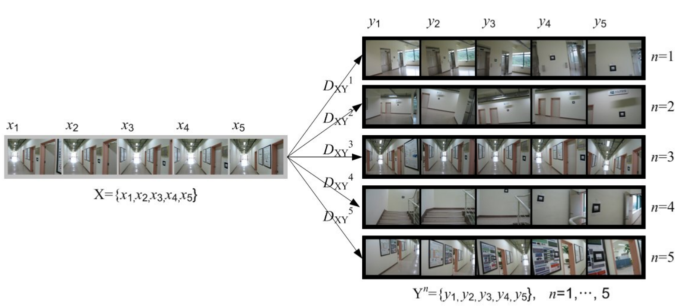


Fig. 5. Similarity calculation process for image sequence matching

In this figure, the first-five LDA feature vectors of the input image sequence and the image sequences in the location dictionary are noted as

$$\begin{aligned} &x_i = y_i = \{LDA_k\},\ k = 1,...,5 \\ &X = \{x_i\}, i = 1,...,64 \\ &Y^n = \{y_i^n\}, i = 1,...,64, n = \textit{the sets of moveable locations} \end{aligned} \tag{1}$$

For matching, the Euclidean distance between two feature vectors is calculated as $d(x_i, y_i^n) = \left|x_i - y_i^n\right|^2$. The distance matrix ($D_{XY^n}$) used to calculate the similarity is shown as

$$D_{XY^n} = \begin{bmatrix} d(x_1, y_1^n) & d(x_2, y_1^n) & \cdots & d(x_i, y_1^n) \\ d(x_1, y_2^n) & \ddots & & \vdots \\ \vdots & & & \vdots \\ d(x_1, y_i^n) & d(x_2, y_i^n) & \cdots & d(x_i, y_i^n) \end{bmatrix} \tag{2}$$

The distance matrix (D_{XY^n}) has the value of the minimum sum of elements between the $d(x_1, y_1^n)$ element and the $d(x_i, y_i^n)$ element. H. Aoki's employed the same method [21]. Two example distance matrices are shown as Eq. (3).

$$D_{XY^1} = \begin{bmatrix} 3 & 8 & 12 & 4 \\ 7 & 2 & 1 & 9 \\ 3 & 9 & 16 & 15 \\ 20 & 16 & 21 & 0 \end{bmatrix},\ D_{XY^2} = \begin{bmatrix} 8 & 2 & 12 & 4 \\ 18 & 2 & 9 & 22 \\ 4 & 10 & 7 & 15 \\ 20 & 2 & 35 & 0 \end{bmatrix} \tag{3}$$

To find location, the shortest path set and minimum sum of the path are calculated from (1,1) element to (4,4) element in each matrix. In Eq. (3), the shortest path sets of each of the distance matrixes are (1,1)-(2,2)-(2,3)-(2,4)-(3,4)-(4,4) and (1,1)-(2,2)-(3,3)-(4,4), and the minimum sums are 30 and 17. In this case, the test image sequence is similar to the 2nd location in the location dictionary because the path and the sum of the distance matrix are short and minimal. After the distance matrix between features is calculated, function *f*() outputs the location ID from the location dictionary that is most similar to the input image sequence (Eq. (4)).

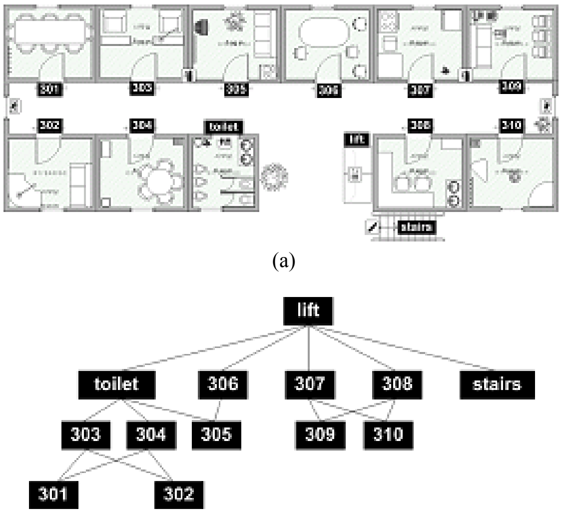
$$S = \min_n f\left(D_{XY^n}\right) \tag{4}$$

In Eq. (4), *n* represents the sets of moveable locations connected to the previous recognized location. The sets of moveable locations are determined by the location model. Therefore, this process does not have to match all of the whole image sequences from the location dictionary.

C. Location Recognition

The location recognition process recognizes the current location using the results of the marker detection and image sequence matching process and the location model. The location model presented in Fig. 6 was designed for an office at our university. Several of the locations that are shown in Fig. 6(b) are in indoor environments. The lines in the figure represent potential paths of a user. It is assumed that movement between nodes that do not have lines is not impossible. The location model is represented by an array of locations. Each node is connected to every node by the potential paths of the user. Therefore, the location model can be easily constructed and efficiently modified by an operator.

In our case, location models were made for every building at our university. The location model for a specific building is selected by inputting the destination of a user. By using the location model, the range of image sequences searched for in the location dictionary is narrower, which allows for low matching error.



(a)

(b)

Fig. 6. Layout of the (a) laboratory area and (b) location model

Three potential location recognition output scenarios must be considered. The first is in the case where the results between the marker detection process and the image sequence matching process are outputted. The second is the case where the difference between results is outputted. The third is the case where only the result of the image sequence matching process is outputted.

To determine the current location in the first case, the location recognition process constantly checks the moveable locations connected to the previous recognized location. The process moves to the next location when the result is found to be one of the moveable locations.

The current location is determined in the second case if the moveable locations connected to the previous recognized location include one of the results of the marker detection or image sequence matching processes. However, if the moveable locations do not include one of the results of the two processes, then the location recognition process outputs the previous recognized location.

In the third case, if only one of the previous two processes has a result, and if the moveable locations include the result, it the location is determined. However, if the moveable locations do not include the result, then the location recognition process outputs the previous recognized location. The location recognition process is performed as detailed in Table II .

TABLE II
RULES GUIDING THE LOCATION RECOGNITION PROCESS

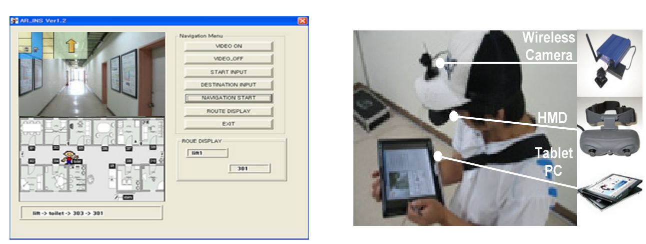
a	– the result of marker detection process	
b	– the result of image sequence matching process	
pl	– the previous recognized location in the location model	
$\{cl\}$	– the set of the moveable locations connected to the pl in the locati on model	
-*if* $a=b$	$a \in \{cl\}$,	output a
	$a \notin \{cl\}$,	output pl
-*if* $a \neq b$	$a,b \in \{cl\}$,	output a location shortest distance from pl
	$a \in \{cl\} \cap b \notin \{cl\}$,	output a
	$a \notin \{cl\} \cap b \in \{cl\}$,	output b
	$a,b \notin \{cl\}$,	output pl
-*if* only output b	$b \in \{cl\}$,	output b
	$b \notin \{cl\}$,	output pl

D. Location Annotation

The location annotation process annotates the location information on the user's view. To annotate the location information on the user's view in a seamless manner, we used the OpenGL graphics library. Location information is provided using virtual 3D direction sign graphics and text such as "go straight", "turn left" and "turn-right". The location information is received by the remote units, and the location information is annotated by the wearable unit. The wearable unit annotates the 3D direction sign on the top-left position of the user's view. The user can see the 3D sign on the input image sequence through the HMD. The distance to the destination, according to the office layout, is also displayed through the HMD.

V. EXPERIMENTAL RESULT

In order to verify the effectiveness of the proposed system, an experiment was performed using indoor image sequences acquired by the cap-mounted camera. Images were captured at a rate of 8 frames per second and were digitized to a size of 320×240 pixels. The experiments were performed with a Pentium Mobile 1.8GHz tablet PC with Windows XP, a wireless camera, desktop PCs (P-3.2GHz, 2GB RAM), HMD, and the algorithm was implemented using a MS Visual C++ development tool. Fig. 7 shows the interface and system setup of the proposed system. The top-left side of the interface shows the input image overlaid with a direction sign, the bottom-left side shows the office map overlaid with the user's current location, and the right side consists of buttons to set the departure and destination locations, as well as text boxes that display the recognized location. This interface is displayed on the mobile PC, and the HMD displays the top-left image of the interface.



(a) (b)

Fig. 7. Interface (a) and system setup (b) of the proposed system

A. Evaluation of the Marker Detection

The marker detection process is applied to every frame of the input image sequence. To show the robustness of the marker detection process, we performed the detection process for a total 53 locations on image sequences captured at day and night. We followed Eq. (5) to test the performance of the location detection system.

$$R = \frac{1}{n} \times \sum_{i=1}^{n} \left(\frac{c_i}{T_i} \times 100 \right) \tag{5}$$

Here, n is the number of locations included in the captured image sequence, T_i is the number of frames that show the ith location in the image sequence, and C_i is the number of times that the ith is recognized by the system. The robustness of the marker detection process was tested with a motion blur sensitivity test. In our system, the movement of the camera follows the user's view. A blurry image, which may decrease the success rate of marker detection, may result from a user's quick movement. The robustness of our marker detection process under increasing motion blur is shown in Table Ⅲ. For this test, the term "ground-truth marker" is used to denote the ground-truth bounding square marked around a region of each marker. For the test, the images were blurred in one direction (the camera is moving from the left to right) using a 9×9 motion blur filter. The fixed threshold value was set to 100. The marker detection success rate gradually decreased with increasing blurriness. Still, the detection success rate was maintained at over 92% using the adaptive threshold method.

TABLE Ⅲ
MARKER DETECTION RESULTS FOR A MOTION BLURRED IMAGE SEQUENCE ANALYZED USING THE FIXED(FT) AND ADAPTIVE THRESHOLDING(AT) METHODS

Image sequence (286 frames)						
Blurring factor	4	6	8	10	12	14
with FT(=100)	87.3	82.6	73.9	68.9	63.9	57.2
with AT	93.7	92.3	92.0	81.6	75.1	60.8

B. Evaluation of the Image Sequence Matching Process

The image sequence matching process analyzes the difference between features of the input image sequence and features of the location dictionary. The process uses LDA features to match features from both image sequences. As mentioned, LDA performs dimensionality reduction while preserving as much of the class discriminatory information as possible. To evaluate the image sequence matching process, we measure the false positive rate (*FPR*) and false negative rate (*FNR*) using the PCA and LDA features and color and hue histograms as features for the image sequence matching process. The PCA and LDA features are used in the first-two features. Color and hue histograms are the most widely used for color feature representations [21]. The histogram information is partially reliable for matching purposes even in the presence of small variations in the frame's visual appearance of frame. The 32-bin histograms are calculated and recorded for every frame. The *FPR* is the proportion of negative instances that were erroneously reported as positive. The *FNR* is the proportion of positive instances that were erroneously reported as negative. The optimal case is when values of the *FPR* and *FNR* are set to zero. For this test, one video sequence with 64 frames per location was stored in the location dictionary, 10 test image sequences per location were obtained, the number of locations was 53, and the best result was obtained when the Euclidean distance between features was at a minimum. Because the test sequences were obtained during 24 second intervals, the test image sequences had a total of 192 frames per location. The measures of the *FPR* and *FNR* are shown by Eq. (6), where # represents the number of each component.

$$FPR = \frac{\#\,false\ positive}{\#location \times \#test\ image\ sequence} = \frac{\#\,false\ positive}{53 \times 10}$$
$$FNR = \frac{\#\,false\ nagative}{\#location \times \#test\ image\ sequence} = \frac{\#\,false\ nagative}{53 \times 10} \tag{6}$$

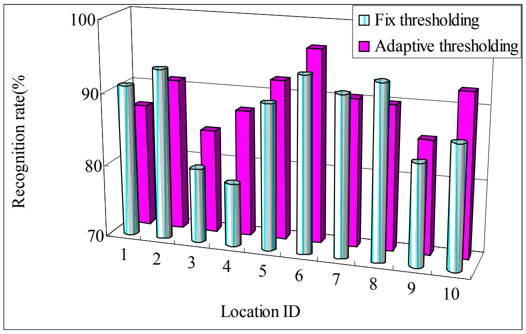
Table Ⅳ shows the *FPR* and *FNR* results obtained using several features for image sequence matching. As shown in Table Ⅳ, LDA is suitable for image sequence matching, and the performance of image sequence matching is higher than PCA.

TABLE Ⅳ
THE RESULTS OF THE FPR AND FNR USING SEVERAL FEATURES FOR IMAGE SEQUENCE MATCHING

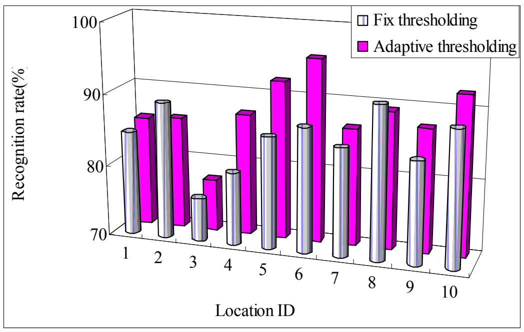
Measures / Features	FPR / #false positive	FNR / #false negative
Color histogram with 32-bin	0.200 / 106	0.115 / 61
Hue histogram with 32-bin	0.136 / 72	0.075 / 40
PCA first-two features	0.092 / 49	0.030 / 16
LDA first-two features	0.045 / 24	0.017 / 9

C. Evaluation of the Location Recognition with the Location Model

The location recognition process could be evaluated by comparing the measured positions after they were calibrated with the content of the location dictionary. For this evaluation, we assume that the user walked in the middle of corridors and followed the nodes of the location model. Fig. 8 shows the location recognition results obtained using the fixed and adaptive thresholding methods with the location model. As previously mentioned, location recognition was performed with a better success rate by using the adaptive thresholding method instead of the fixed thresholding method. Two measures were used to evaluate the accuracy of location recognition: average probability and average execution time. Table Ⅴ shows the results of the location recognition accuracy assessment. Daytime- and nighttime-captured image sequences were used for this test. As shown in Table Ⅴ, the rate of incorrect recognition and execution time were decreased by using the location model.



(a)



(b)

Fig. 8. Location recognition results obtained using the fixed and adaptive thresholding methods with the location model, recognition results of (a) daytime- and (b) nighttime-captured image sequences

Fig. 9 shows typical indoor navigation situations: passing through a corridor, moving to another floor using an elevator and going up or downstairs. The system was found to have an average recognition rate of 89%, and the average execution time for location recognition was 2.3 sec. Incorrect location recognition occurred in corridor locations connected to an outside environment because of rapid changes in illumination. The proposed system can be applied to various consumer applications, such as the door plate system, notice board system, shopping assistance system, and bus service route guide system. The main purpose of our system is to insert computer-generated graphical content into a real scene in real-time. Such a system could be beneficial for shopping, for example, as it would be possible to check prices of goods with out spending time visiting shops and walking around every single aisls. This is beneficial for both consumers and shop owners because less time and money must be devoted to investigating and advertising which products are carried. The processed system could be applied in a shopping environment by putting a marker on each product in a store.

TABLE V

AVERAGE LOCATION RECOGNITION RATE AND PROCESSING TIME ACCORDING TO USING OR NOT USING THE LOCATION MODEL FOR THE DAYTIME-CAPTURED IMAGE SEQUENCE

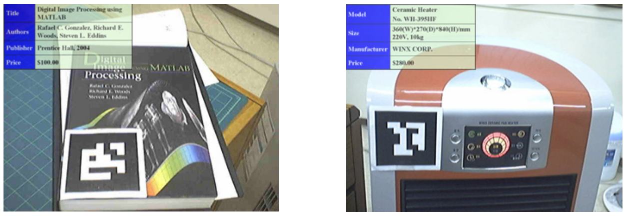
Location ID	Average recognition rate (%) and processing time (sec.)			
	Non-using	time	Using	time
1	85.1	0.31	87	0.19
2	87	0.30	91	0.21
3	78	0.35	84.3	0.21
4	85.8	0.29	87.5	0.19
5	92	0.29	92.1	0.16
6	92	0.27	96.6	0.15
7	91.5	0.31	90.4	0.16
8	89.7	0.30	90	0.20
9	84	0.29	85.7	0.19
10	90	0.27	92.5	0.16
total	87.5	0.29	89.7	0.18

Fig. 9. Results of indoor location positioning

The proposed system could also be beneficial as a bus service route guide. At each bus stop, the system user could see all of the destinations on each bus route at the bus stop. The proposed system could inform the user of the proper bus route and number to reach a specified destination. Generally, due to space limitations, bus service route tables shown at bus stops only show the destinations of each bus line. Visitors unfamiliar with a city may not know the destinations well enough to select a bus line. However, with the proposed system, the user would only have to input the final destination and the bus service route and bus number would be annotated on the user's view. To be employed in an economical manner, our system could be installed at each bus stop. Transit authorities would also save money due to not having to reprint maps when bus routes are modified. Various examples of consumer applications of our system are shown in Fig. 10. To characterize each appliance, a 9×9cm marker is used. Fig. 10 (a) shows the price along with the publication company and the author of a book. And, Fig. 10(b) shows the price, manufacturer and data about a manufacture of an electric heater.

VI. CONCLUSION

A vision-based location positioning system that uses the augmented reality technique for indoor navigation has been proposed herein. This system automatically recognizes a location from image sequences taken of indoor environments.



(a) (b)

Fig. 10. Examples of various consumer applications using our system

Location-related information is displayed by identifying similar locations from an image database. Several vision-based techniques are used to increase location positioning accuracy and efficiency. To increase performance, our system uses adaptive thresholding to detect markers under changing illumination, and it uses a location model to reduce execution time during image sequence matching. The proposed system tested in indoor environments, showed successful location recognition rates of 89%.

REFERENCES

[1] A. Jacobsen, "Wearable computers: developments and possibilities", Norwegian University of Science and Tech., PhD. Thesis, 2000.

[2] J. Lehikoinen and R. Suomela, "Accessing context in wearable computers", Personal and Ubiquitous Computing, Vol. 6, No. 1, pp. 64-74, 2002.

[3] J. B. Kim, "A Personal Identity Annotation Overlay System using a Wearable Computer for Augmented Reality", IEEE Trans. on Consumer Electronics, Vol. 49, No. 4, pp. 1457-1467, 2003.

[4] R. T. Azuma, "A survey of augmented reality", Presence: Teleoperators and Virtual Environments, Vol. 6, No. 4, pp. 355-385, 1997.

[5] N. Navab, "Developing killer apps for industrial augmented reality", IEEE Computer Graphics and Applications, Vol. 24, No. 3, pp. 16-20, 2004.

[6] M. E. Barnard, "The global positioning system", IEE Review, Vol. 38, No. 3, pp. 99-102, 1992.

[7] S. Panzieri, F. Pascicci, G. Ulivi, "An outdoor navigation system using GPS and inertial platform", IEEE Trans. on Mechatronics, Vol. 7, No. 2, pp. 134-142, 2002.

[8] E. Abbott and D. Powell, "Land-vehicle navigation using GPS", Proceedings of the IEEE, Vol. 87, No. 1, pp. 145-162, 1999.

[9] N. B. Priyantha, A. Chakraborth, H. Balakrishnan , "The cricket location-support system", Proc. ACM Mobicom, pp. 32-43, 2000.

[10] R. Want, A. Hopper, V. Falcao, J. Gibbons, "The active badge location system", ACM Trans. on Information Systems, Vol. 10, No. 1, pp. 91-102, 1992.

[11] P. Bahl, and V. Padmanabhan, "RADAR: an inBuilding RF-based user location and tracking system", Proc. in IEEE INFOCOM, Vol. 2, pp. 26-30, 2000.

[12] M. philipose, *et al.*, "Mapping and localization with RFID technology", Intel Research, IRS-TR-03-014, 2003.

[13] H. D. Chon, "Using RFID for accurate positioning", Int. Syst. on GNSS, 2004.

[14] V. Kulyukin, *et al.*, "RFID in robot-assisted indoor navigation for the visually impaired", Proc. in Int. Conf. on Intelligent Robots and Systems, pp. 1979-1984, 2004.

[15] K. Finkenzeller, RFID Handbook: Fundamentals and Applications in Contactless Smart cards and Identification, John Wiley & Sons (2nd edition), 2003.

[16] Z. Xiang, S. Song, J. Chen, H. Wang, J. Huang, X. Gao, "A wireless LAN-based indoor positioning technology" IBM Journals Research and Development, Vol. 48, No. 5/6, pp. 617-626, 2004.

[17] J. Yang, W. Yang, M. Denecke, A. Waibel, "Smart Sight: A Tourist Assistant System", Pro. in Int. Sym. on Wearable Computers, pp. 73-78, 1999.

[18] L. Ran, S. Helal, S. Moore, "Drishti: An integrated indoor/outdoor blind navigation system and service", Pro. in the second IEEE annual conf. Pervasive Computing and Communications, pp. 23-30, 2004.

[19] G. D. Abowd, C. G. Atkeson, J. Hong, S. Long, R. Kooper, M. Pinkerton, "Cyberguide: a mobile context-aware tour guide", Wireless Networks, Kluwer Academic Publishers, Vol. 3, No. 5, pp. 421-433,1997.

[20] J. J. Weng and S. Chen, "Vision-guided navigation using SHOSLIF", Neural Networks, Vol. 1, No. 2, pp. 1511-1529, 1998.

[21] H. Aoki, B. Schiele, A. Pentland, "Realtime personal positioning system for a wearable computers", Pro. in Int. Sym. on Wearable Computers, pp. 37-43, 1999.

[22] http://www.hitl.washington.edu/artoolkit/

[23] M. Fiala, "ARTag revision 1, A fiducial marker system using digital techniques", NRC/ERB-1117, Technical Report, Nation Research Council Canada (NRC) publication: NRC 47419, 2004.

[24] M. Kalkusch, T. Lidy, N. Knapp, G. Reitmayr, H. Kaufmann, D. Schmalstieg, "Structured Visual Markers for indoor pathfinding", Pro. in IEEE Int. Workshop Augmented Reality, pp. 1-8, 2002.

[25] J. Rekimoto, K. Nagao, "The world through the computer: Computer augmented interaction with real world environments", Pro. in ACM Sym. on User Interface Software and Technology, pp. 29-36, 1995.

[26] J. Rekimoto, Y. Ayatsuka, "Cybercode: designing augmented reality environments with visual tags", Proc. in ACM on Designing augmented reality environments, pp. 1-10, 2000.

[27] A. M. Martinez, A. C. Kak, "PCA versus LDA", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 23, No. 2, pp. 228-223, 2001.

[28] J. B. Kim and H. J. Kim, "Efficient region-based motion segmentation for a video monitoring system", Pattern Recognition Letter, Vol. 24, No. 1, pp. 113-128, 2003.

Jong-Bae Kim received the B.E. degree in computer engineering from Miryang National University, Miryang, Korea in 2000, and received the M.S. and Ph.D degrees in computer engineering from Kyungpook National University in 2002 and 2004, Deagu, Korea. He joined a visiting researcher of University of Münster, Münster, Germany in 2003 and the School of Electrical and Computer Engineering of the Georgia Institute of Technology, Atlanta, GA, USA in 2004, respectively. He served a full-time researcher at the Digital Manufacturing Information Technology Research Center in the University of Ulsan, Ulsan, Korea in 2005. His is currently working a professor at the School of Computer Engineering in the Seoul Digital University, Seoul, Korea. His research interests include image/video processing, computer vision, pattern recognition, wearable computing and augmented reality.

Hee-Sung Jun received the B.S. and M.S. degree in electrical engineering from Seoul National University(Seoul, Korea) in 1981 and 1983 respectively, and received the Ph.D. from Rutgers – The State University of New Jersey (USA) in 1992. He is currently a professor at the School of Computer Engineering & Information Technology in University of Ulsan(Ulsan, Korea). His research interests include image processing, computer vision, computer graphics and augmented reality.