BOOK SEARCH IN STACK ROOMS BY AUGMENTED REALITY AND IMAGE ANALYSIS TECHNIQUES [1]

Shih-Ming Yang (楊士旻)
Institute of Multimedia Engineering
National Chiao Tung University, Hsinchu, Taiwan
Email: kevin810321@hotmail.com

[2] Wen-Hsiang Tsai (蔡文祥)
Dept. of Computer Science
National Chiao Tung University, Hsinchu, Taiwan
Email: whtsai@cs.nctu.edu.tw

***Abstract*—An indoor guidance system based on augmented reality (AR) and image analysis techniques for book search and book-information introduction in a book-related space by the use of a mobile device is proposed. To guide a user to walk along a planned path for book search, a method of attaching a set of so-called signal-rich-art code images at proper locations as landmarks for user localization is proposed. For book-information introduction to the user standing in front of a book stack, a book-spine image recognition method against a pre-constructed book-spine image database is proposed, in which smart book-spine image segmentation and knowledge-based optical character recognition (OCR) techniques are adopted. And book-information introduction is conducted by showing the title and related information of the recognized book on the mobile device for user inspection. Good experimental results are also included to show the feasibility of the proposed system.**

***Index Terms*—Augmented reality, indoor navigation; book recognition, feature matching, knowledge-based OCR.**

I. INTRODUCTION

While visiting a book-related space like a library, a book stack room, a book store, etc., the lack of easily-available information of a desired book is often a problem. For example, if the price label was gone or hidden in a place that was hard to see, it might annoy the customer. An intuitive in-field navigation system is needed to allow people to find the book information they want more quickly. In this study, it is desirable to design an indoor guidance system for book search and book-information introduction based on augmented reality (AR) techniques for users using mobile devices as navigation tools. The concept of the desirable system is shown in Fig. 1.

About related works of navigation or guidance, Barberis et al. [1] proposed an indoor navigation system using a set of visual markers for accurate computation of the indoor location and orientation of a user. Li and Wang [2] proposed a vision-based navigation system focusing on the map-based approach, which generally uses feature-based matching for localization. Liang et al. [3] proposed a three-step pipeline method for image-based localization of mobile devices in indoor environments. Jongbae and Heesung [4] proposed a vision-based indoor navigation system which recognizes the location of a user by applying techniques of marker detection and matching on image sequences captured from a wearable camera, and displays navigation information which the user needs in an AR way. For book search in bookstores and libraries, Lee et al. [5] designed a scanning device by using the Fibonacci lattice sampling technique in color quantization for matching book-spine images to improve the library inventory process. Chen et al. [6] designed a mobile AR system for retrieving book information by recognizing the book-spine image without taking the book off the bookshelf.

(a) (b)

Fig. 1: Examples of AR-based indoor guidance system. (a) A system used in a supermarket. (b) The system implemented in this study.

In this study, we propose an AR-based guidance system for book search and book-information introduction in book-related spaces. The user is assumed to hold a mobile device like a smart phone to take images of books and landmarks in the guidance process. The captured images are augmented by the information of a

---

[1] This research was supported by the NSC, Taiwan under Grant No. 104-3115-E-009-001.

[2] W. H. Tsai is also with the Department of Information Communication, Asia University, Taichung, Taiwan 41354.

planned path to a searched book or augmented by the introduction information of the searched book after the book is reached. A more detailed description of the function of the proposed system is as follows.

For the function of book search, the user is guided by the proposed system to a desired book after a book name is input into the system. Specifically, the system uses the book name to find out the destination shelf on which the desired book is put. Next, the user faces a marker initially, which is a signal-rich-art code image [8], and takes an image of the marker. Then, the proposed system extracts the marker position information embedded in the marker image as the user's location (because the user is standing at the marker site) and then plans an optimal path from the user's location to the destination book shelf. Finally, the system sends the path guidance information to the user's mobile device for the user to inspect and follow to walk forward.

Furthermore, for book-information introduction, the proposed system analyzes the images captured with the mobile-device camera after the user reached the book shelf on which the desired book is located. The system then extracts the features of the image of the spine of the book by an SURF algorithm [7] and matches the features against a pre-constructed database of book features to obtain the book's introduction information if the image matching is successful. Finally, the system sends the obtained book information to the mobile device held by the user. Such information is overlaid in an AR way onto the book item shown in the current image taken of the book shelf by the mobile-device camera.

In the remainder of this paper, the configuration of the proposed system is described in Sec. 2, establishment of a book-spine image database associated with book titles is described in Sec. 3, matching of book-spine images for book recognition is described in Sec. 4, the proposed AR techniques for the system are described in Sec. 5, and experiment results are shown in Sec. 6, followed by some conclusions in the last section.

II. System Design And Processes

A. System Design

As shown in Fig. 2, the proposed system is constructed to be of a client-server structure. The server runs on a centralized computer for conducting complicated works with heavy computations. The server sends the guidance information to the client system running on the user's mobile device. On the other hand, when the user enters the environment, the client program running on the user's mobile device is connected to the server through a network and receives relevant information from the server.

B. Learning Process

The system operations include two stages − a learning process and a guidance process. The goal of the learning process is to establish an environment map, which includes information about the places available for users to visit, the book information, and the magnetic fields in the environment. The learning process includes two phases: learning for user localization and learning for book recognition. In the former phase, the system conducts learning of the position information of the places for visits and the readings of the magnetic field in the book-related space. And in the latter phase, the system constructs book-related data, including book image features and book-related information.

Fig. 2: Network architecture of proposed system.

C. Guidance Process

In the guidance process, while conducting book search, the server analyzes the images of the marks received from the client, and sends back the related path guidance information to the client. And to achieve book-information introduction, the server analyzes the book-spine images received from the client, and sends the corresponding book information to the client. Also, the client program displays the information on the screen of the user's mobile device. More details are described in the following sections.

III. Augmenting Book-Title Database by Learning Book-Spine Images

A. Learning Book-Spine Images

It is assumed that there exists already a book-title database in which the titles of all books in the book-related space are kept. In order to provide a convenient application of the proposed system, we have to augment this database with book-spine images, each image being associated with a book title. The result is called a book-spine image database. For this, we have to take book-shelf images and segment each of such images into separated book-spine parts, which we call *book-spine images*.

For segmentation of book-spine images, at first we hold the mobile device to face a book shelf and take an image of it. An example of such images is shown in Fig 3(a) which contains several book-spine items. Then, we apply canny edge detection [9] to the image to obtain a result like that shown in Fig. 3(b). Next, we detect all contours in the resulting image and keep only those contours longer than a threshold, like that shown in Fig. 3(c). Finally, we apply the Hough line transform to find long straight lines like the blue lines shown in Fig. 3(d). Finally, a semi-automatic process is conducted to separate the book spines, resulting in respective book-spine images like those shown in Fig. 3(e).

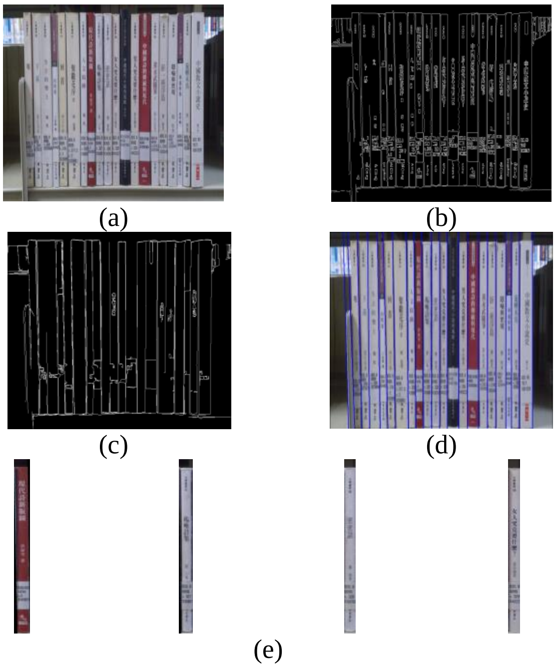

Fig. 3: Result of segmenting book-shelf images. (a) A book-shelf image. (b) Edge detection result. (c) Result of keeping longer contours. (d) Result of finding long lines by Hough transform. (e) Some segmented book spines.

B. Book Title Extraction by Knowledge-based OCR

After segmentation of book-spine images is conducted, it is desired to recognize the characters in each book-spine image so that the image can be associated with the corresponding book in the book-title database, resulting in a book-spine image database augmented with book titles. For this aim, we use the OCR technique. It is assumed, as mentioned previously, in this study that for a book-related space there exists already a book-title database in which the titles of all books are kept. The idea of our method for character recognition is to utilize the known characters in the book-title database to remedy the high difficulty of recognizing the characters on the book spines. This high difficulty of character recognition comes from the fact that the shapes of the characters on the book spine are usually non-regular or even artistic. Thus, we cannot directly use the OCR results. Instead, we find out from the book-title database all titles of the books which are on the shelf, and then compare the OCR result of a book-spine image with all the book titles by the longest common subsequence (LCS) algorithm. The book title, which has the LCS within the OCR result, is taken as the desired result, which then is linked to the book-spine image.

IV. Book Spine Recognition Using Speeded-up Robust Features (SURFs)

After the user has reached the book shelf on which the searched book is put, he/she will take an image of the book shelf, and the system will then try to recognize the book spine of the searched book in the acquire image. If the recognition is successful, corresponding book information will be augmented on the user’s mobile-device screen for inspection. For this to be conducted, the features of book-shelf and book-spine images should be extracted in advance before they can be matched for the purpose of book-spine recognition.

A. Extraction of SURFs

In this study, the type of feature we use is speeded-up robust feature (SURF). We “learn” SURFs of the book-spine images using an SURF extraction algorithm after the book-spine image database is constructed. Specifically, we conduct this feature extraction procedure on each book-spine image and keep the feature descriptors with the image in the database.

In the guidance phase, after the user has reached the searched book on a shelf, he/she holds the mobile device to face the shelf and take an image of it. An example of such images is shown in Fig 4(a). Next, the system extracts the SURF points from the taken image. The extracted feature points of Fig. 4(a) are shown in Fig. 4(b), where the size of each circle specifies the corresponding scale and the line in the circle indicates the orientation of the feature point.

(a) (b)

Fig 4: Feature extraction from an acquired book shelf image. (a) An original client-camera image. (b) The extracted feature points of the client-camera image.

B. Matching of Book-spine and Camera Images

We match the pre-extracted features of the book-spine images with those extracted from the book-shelf image for the purpose of book-spine recognition as mentioned above. The number of feature points used in the matching determines the speed of the recognition work. The size of an originally uploaded book-shelf image is 1632×1224. According to our experimental experience, the number of detected feature points is between 1500 and 2500; and correspondingly, the time for feature extraction and that for matching are about two seconds and a half second on the average, respectively.

After conducting SURF matching, several *matched* feature points are obtained. For the purpose of augmenting information on a user-device screen, after we obtain the matched feature points, we try to find a bounding box to enclose them so that the book spine in the image can be precisely located and observed visually easily by the user. To accomplish this, we have to determine where to display the green bounding box on the mobile-device screen. For this, a scheme is proposed in this study to find a transformation from the book-spine image to the shelf image in terms of a homography matrix. The homographic transformation may be described as an equation as follows:

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1X_1 & -y_1X_1 & -X_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1Y_1 & -y_1Y_1 & -Y_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2X_2 & -y_2X_2 & -X_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2Y_2 & -y_2Y_2 & -Y_2 \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3X_3 & -y_3X_3 & -X_3 \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -x_3Y_3 & -y_3Y_3 & -Y_3 \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -x_4X_4 & -y_4X_4 & -X_4 \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -x_4Y_4 & -y_4Y_4 & -Y_4 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \end{bmatrix} = 0 \qquad (1)$$

where (X_1, Y_1) through (X_4, Y_4) are four feature points in the book-spine image, and (x_1, y_1) through (x_4, y_4) are the corresponding ones in the book-shelf image, and h_1 through h_9 are parameters to be solved. After solving Eq. (1) to get the solutions h_1 through h_9, we can construct a homography matrix as follow:

$$H = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \qquad (2)$$

After applying the homography matrix (2), the system can find the precise location of the book spine in the book-shelf image and draw a green bounding box to enclose the book spine for easier inspection by the user. An experiment result is shown in Fig. 5.

(a) (b)

Fig. 5: Results of book spine recognition. (a) Result of matching a book-spine image (appearing as left slim part in figure) with an input image. (b) The green bounding box encloses the correct book spine.

V. AR-Based Guidance for Book Search

We overlay the guidance or book information onto the real image taken of the current scene so that the user can just take the mobile device to follow the guidance for book search provided by the system. To implement AR functions on the mobile-device screen, the client program on the mobile device receives guidance information from the remote server system. The client program will display such information on the mobile device screen. The techniques for carrying out these tasks are described in the following.

A. Finding User Position and Orientation

The user localization method proposed in this study computes both the user's position and his/her orientation. The technique used for computing the user's position is based on Lee and Tsai [8], which is a new message data transfer method via the use of so-called signal-rich-art code images captured by the mobile device. An example of the signal-rich-art code image is shown in Fig. 6 (b) which is a signal-rich-art code image generated from Fig. 6(a).

More specifically, in a book-related space we set up many nodes at different locations in this study. Each node has a marker, which is a printed version of a signal-rich-art code image. In each of these markers, the information of the marker's position is embedded by an information hiding technique proposed by Lee and Tsai [8]. And the information of the marker's position is an 8-bit binary number. Each marker image has a different 8-bit binary number for use as the location number. After the user takes an image of the marker, the system extracts the location number using Lee and Tsai method [8] to find the user's position in the book-related space. It is in this way we accomplish user position computation. In addition, the user will know the next marker location after he/she takes a picture of a marker. Because we have learned the indoor environment in advance, we know all the markers' position. Then, we can get the direction from the current marker's position to the next marker's position.

(a) (b)

Fig 6: Using signal-rich-art code images to conduct human localization. (a) An original image. (b) Generated signal-rich-art code image of (a) using Lee & Tsai's method [8].

Now, we have to compute the user's orientation. For this, we use the mobile phone's built-in orientation sensor, which is an electronic compass, to accomplish the work. The electronic compass measures the azimuth angle of the device by detecting changes in the magnetic field in the surrounding environment. It is noted that the azimuth angle of the mobile device may be taken to be that of the user because the user is holding the mobile device whose heading direction is roughly identical to the user's orientation. After starting the guidance, the client program keeps detecting the heading direction by the electronic compass, and the azimuth angle of the electronic compass can so be used to calculate the direction from the position of the current marker to that of the next marker. The direction can then be drawn as an arrow which the system displays on the mobile-device screen to direct the user to the destination while he/she is being guided by the system.

B. Path Planning by Dijkstra Algorithm

In this study, the user should take an image of a marker located at the initial node at first. Extraction of the marker's location extracted from the acquired image in a way as described previously is then taken to the initial location of the user. In addition, the destination node will be found out after the user key in the book title of the wanted book. Then, the proposed system will use the Dijkstra algorithm to plan an optimal path from the initial node to the destination to guide the user to the book shelf on which the wanted book is put.

C. Rendering for Displaying Guidance Arrow

To allow the user a more convenient use of the system in the guidance process, as mentioned, we display a guidance arrow on the mobile-device screen in an AR way for the user to inspect during his/her walk from one node to another. According to the user position computed in the way as described previously, we can know which node on the planned path has been reached. Then, based on the planned path and the information of the user orientation, we can decide which kind of guidance arrow should be displayed on the mobile-device screen. The three kinds of guidance arrow used by the proposed system are shown in Fig. 7.

Fig. 7: Three kinds of guidance arrows used in this study. (a) Leftward. (b) Forward. (c) Rightward.

VI. Experimental Results

The environment for conducting experiments in this study is a lab space with book shelves. A map of it is shown in Fig. 8, which includes two target places (shown as green regions), 30 book items (part of them shown as blue texts) and 5 marker images (shown as red circles). The first experimental result we show is that of segmentation of book-spine images. Figs. 9(a)-(d) show the results of segmentation of book-spine images. Fig. 9(a) is an image of a book shelf with many vertical book spines. The result of applying canny edge detection is shown in Fig. 9(b). Fig. 9(c) is the result of using a threshold to filter those contours with small perimeters appearing in Fig. 9(b). And Fig. 9(d) is the result of applying the Hough line transform to find straight lines in Fig. 9(c), which are supposed to be the border lines of the books. Then, Fig. 10 shows some book-spine images we obtained from the segmentation result of Fig. 9 and the result of OCR of the book-spine images. In addition, Fig. 11 shows the results of book-spine recognition. These results and many others obtained in our experiments show the feasibility of the proposed method.

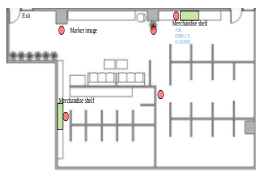

Fig. 8: The map of the experimental environment.

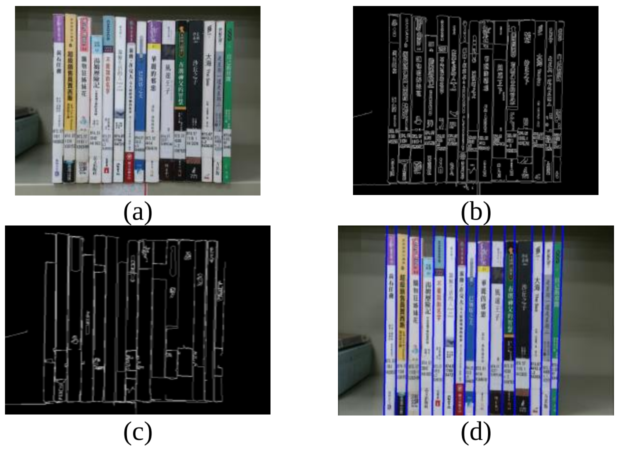

Fig. 9: Result of segmentation of book-spine images. (a) A book shelf image. (b) Result of applying canny edge detection. (c) Result after using a threshold to filter small-perimeter contours. (d) Result of applying Hough line transform.

VII. Conclusions

A system for AR-based guidance system for book search and book-information introduction using a mobile device in book-related spaces has been proposed. To design such a system, several techniques have been proposed and shown to be effective, as summarized in the following. (1) A method for constructing an augmented book-spine image database associated with book titles by skillful techniques of book-spine image segmentation and knowledge-based OCR has been proposed. (2) A method for recognition of book-spine images by SURF extraction and matching has been proposed, by which the system can recognize book-spine images and obtain the precise location of the desired book in a book-shelf image. (3) A method for indoor AR-based guidance in book-related spaces by overlaying guidance arrow on the mobile-device in an AR way screen has been proposed, by which a user can be guided node by node to a destination book shelf. (4) A method for user localization using special markers (i.e., signal-rich-art code images) and electronic compass readings has been proposed, by which the proposed system can compute the user's position and orientations. Good experimental results have revealed the feasibility of the proposed system.

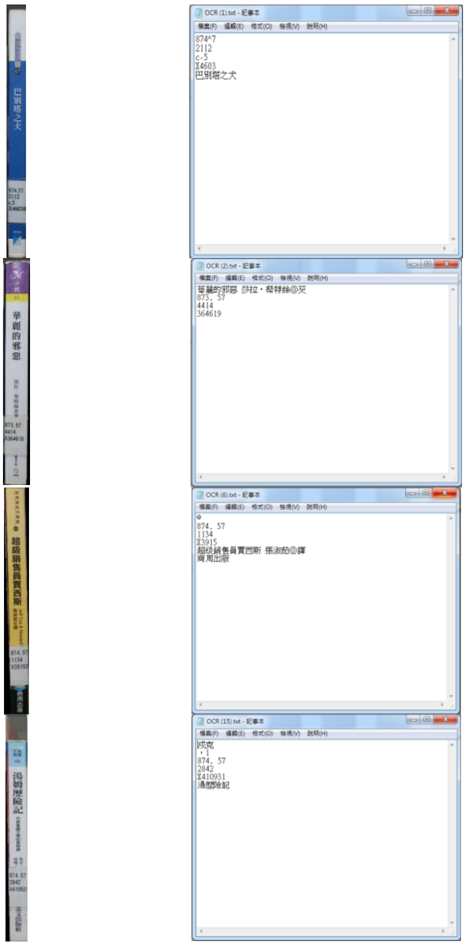

Fig. 10: Result of OCR of a book-spine image where the left-side images are some book-spine images, and the right-side images are results of OCR of the left-side images.

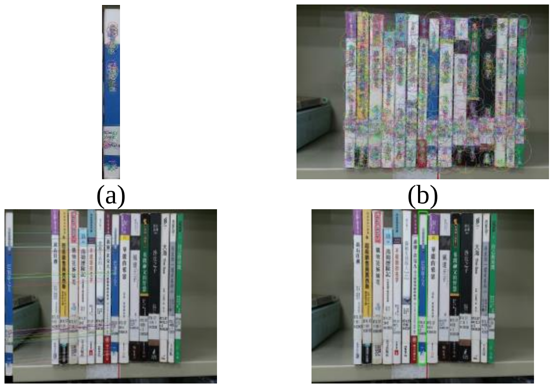

Fig. 11: Results of book-spine recognition. (a) Book-spine image with extracted SURFs. (b) Book-shelf image with extracted SURFs. (c) Result of image matching of (a) and (b). (d) A green box enclosing recognized book spine.

REFERENCES

[1] C. Barberis, A. Bottino, G. Malnati and P. Montuschi, "Experiencing Indoor Navigation on Mobile Devices," *IT Professional*, vol. 16, Issue 1, pp. 50-57, Feb. 2014.

[2] X. Li and J. Wang, "Image Matching Techniques for Vision-based Indoor Navigation Systems: Performance Analysis for 3D Map Based Approach," *Proc. of Int'l Conf. on Indoor Positioning & Indoor Navigation (IPIN)*, Sydney, Australia, pp. 94-95, 2012.

[3] J. Liang, N. Corso and E. Turner, "Reduced-Complexity Data Acquisition System for Image Based Localization in Indoor Environments," *Proc. of Int'l Conf. on Indoor Positioning & Indoor Navigation (IPIN)*, Montbéliard, France, pp. 222-230, 2013.

[4] K. Jongbae and J. Heesung, "Vision-Based Location Positioning Using Augmented Reality for Indoor Navigation," *IEEE Trans. on Consumer Electronics*, vol. 54, no. 3 pp. 954-962, Aug. 2008.

[5] D. J. Lee, Y. Chang, J. K. Archibald and C. Pitzak, "Matching Book-Spine Images for Library Shelf-Reading Process Automation," *Proc. of IEEE Int'l Conf. on Automation Science & Engineering (CASE)*, Washington, DC, USA, pp. 738-743, Aug. 2008.

[6] D. Chen, S. Tsai, C. H. Hsu, J. P. Singh and B. Girod, "Mobile Augmented Reality for Books on a Shelf," *Proc. of IEEE Int'l Conf. on Multimedia & Expo*, Barcelona, Spain, pp 1-6, July 2011.

[7] H. Bay, A. Ess, T. Tuytelaars and L. van Gool, "SURF: Speeded Up Robust Features," Features," *Computer Vision & Image Understanding (CVIU)*, vol. 110, no. 3, pp. 346-359, 2008.

[8] Y. L. Lee and W. H. Tsai, "A New Data Transfer Method via Signal-rich-art Code Images Captured by Mobile Devices," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 25, no. 4, pp. 688-700, Apr. 2015.

[9] J. Canny, "A computational approach to edge detection," *IEEE Trans. on Pattern Anal. Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679-698, Nov. 1986.