A Multimedia, Augmented Reality Interactive System for the Application of a Guided School Tour

Ko-Chun Lin[1], Sheng-Wen Huang[1], Sheng-Kai Chu[1], Ming-Wei Su[1], Chia-Yen Chen[1], and Chi-Fa Chen[2]

[1] Department of Computer Science and Information Engineering, National University of Kaohsiung, Taiwan
[2] Department of Electrical Engineering, I-Shou University, Taiwan
ayen@nuk.edu.tw

Abstract. The paper describes an implementation of a multimedia, augmented reality system used for a guided school tour. The aim of this work is to improve the level of interactions between a viewer and the system by means of augmented reality. In the implemented system, hand motions are captured via computer vision based approaches and analyzed to extract representative actions which are used to interact with the system. In this manner, tactile peripheral hardware such as keyboard and mouse can be eliminated. In addition, the proposed system also aims to reduce hardware related costs and avoid health risks associated with contaminations by contact in public areas.

Keywords: Augmented reality, computer vision, human computer interaction, multimedia interface.

1 Introduction

The popularity of computers has induced a wide spread usage of computers, as information providers, in public facilities such as museums or other tourist attractions. However, in most locations, the user is required to interact with the system via tactile means, for example, a mouse, a keyboard, or a touch screen. With a large number of users coming into contact with the hardware devices, it is hard to keep the devices free from bacteria and other harmful contaminates which may cause health concerns to subsequent users. In addition, constant handling increases the risk of damage to the devices, incurring higher maintenance cost to the providing party. Thus, it is our aim to design and implement an interactive system using computer vision approaches, such that the above mentioned negative effects may be eliminated. Moreover, we also intend to enhance the efficiency of the interface by increasing the amount of interaction, which can be achieved by means of a multimedia, user augmented reality interface.

In this work, we realized the proposed idea by implementing an interactive system that describes various school locations, both verbally and visually, according to the location pointed to by the user on a school map. The implemented system does not require the keyboard or the mouse for interaction; a camera and a printout of the map

F. Huang and R.-C. Wang (Eds.): ArtsIT 2009, LNICST 30, pp. 231–238, 2010.

are used to provide the necessary input instead. To use the system, the user is first given a printed school map, as often given out to visitors to the school, he/she can then check out the different locations by moving his/her finger across the paper and point to the marked locations on the map. The movement and location of the fingertip are captured by an overhead camera and the images are analyzed to determine the user's intended actions. In this manner, the user does not need to come into contact with anything other than the map that is given to him/her, thus eliminating health risks due to direct contact with harmful substances or contaminated surfaces.

The paper is organized as follows. Section 2 describes the various stages in the implementation of the system; section 3 demonstrates the usage of the implemented system, and finally in section 4, we discuss possible extensions and future applications of the system.

2 Background

Augmented reality (AR) has received a lot of attentions due to its attractive characteristics including real time immersive interactions and freedom from cumbersome hardware [10]. There have been many applications designed using AR technologies in areas such as medical applications, entertainment, military navigation, as well as many other new possibilities.

An AR system usually incorporates technologies from different fields. For example, technologies from computer graphics are required for the projection and embedding of virtual objects; video processing is required to display the virtual objects in real time; and computer vision technologies are required to analyse and interpret actions from input image frames. As such, an AR system is usually realized by a cross disciplinary combination of techniques.

Existing AR systems or applications often use designated markers, such as the AR encyclopedia or other applications written by ARToolkit [11]. The markers are often bi-coloured and without details to facilitate marker recognition. However, for the guidance application, we intend to have a system that is able to recognize colour and meaningful images of objects or buildings as printed on a brochure or guide book and use them for user interactions.

3 System Design and Implementation

The section describes how the system is designed and implemented. Issues that arose during the implementation of the system, as well as the approaches taken to resolve the issues are also discussed in the following.

3.1 System Design

To achieve the goals and ideas set out in the previous section, the system is designed with the following considerations.

Minimum direct contact. The need for a user to come into direct contact with hardware devices such as a keyboard, or a mouse, or a touch screen, should be minimized.

User friendliness. The system should be easy and intuitive to use, with simple interface and concise instructions.

Adaptability. The system should be able to handle other different but similar operations with minimum modifications.

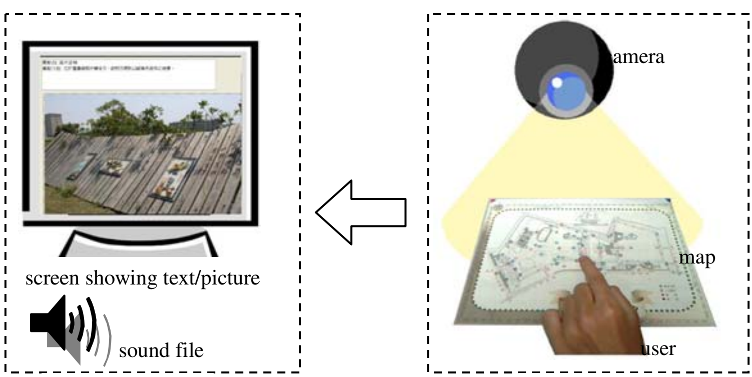
Low cost. We wish to implement the system using readily available hardware, to demonstrate that the integration of simple hardware can have fascinating performance.

Simple and robust setup. Our goal is to have the system installed at various locations throughout the school, or other public facilities. By having a simple and robust setup, we reduce the chances of a system failure.

3.2 System Setup

In accordance to the considerations listed above, the system is designed to have the input and out interfaces as shown in Fig. 1. The system obtains input via a camera, located above and overlooking the map. The camera captures images of the user's hand and the map. The images are processed and analyzed to extract the motion and the location of the fingertip. The extracted information is used to determine the multimedia data, including text, pictures, sound files, and/or movie clips, to be displayed for the selected location on the map.

Based on the designed interface, the system's operations are implemented according to the following procedures. In preparation for the system's execution, the maps are scanned first, and image maps containing the feature points are constructed and stored in a database.



Output **Input**

Fig. 1. Designed interface of the system

Upon the execution of the system, the camera is first adjusted according to the conditions of the operating environment. When a valid map is placed under the camera, the camera will acquire an image of the map and use it as a template for comparison. The system continues to monitor the map via the camera and checks for the presence and motion of the user's fingertip. Once the user's fingertip has stayed pointed to at a particular location for a given time interval, the system acknowledges that the user wishes to learn about the location. The system will then extract the coordinates of the fingertip and perform the required transformation to match the coordinates of the point with the image map. If the location is indeed a feature point, multimedia information associated with the location will be displayed on the screen and played through the speakers.

The flowchart in Fig. 2 describes the main stages in the system's operation.

In addition to the basic operations, the system is also able to perform in different modes and provide multi-lingual interactions. To demonstrate the ability, the system is implemented to recognize two different maps containing different feature points. There are also marked regions on the maps which allow the user to switch the supported language. Currently, we have implemented descriptions of the school in Mandarin, Taiwanese and English to provide a multi-lingual experience for school visitors.

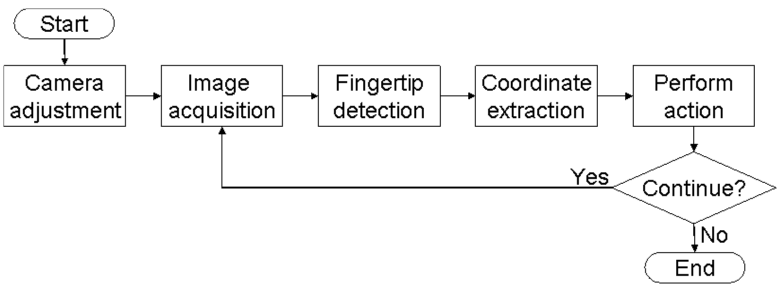


Fig. 2. Flowchart of the system's basic operations

3.3 Camera Adjustment

When the system is first setup, the camera needs to be adjusted for the surrounding environment. Parameters such as the distance between the camera and user's map, the direction of viewing, contrast and colours, need to be selected to obtain valid input images.

In our experiments, we used an USB camera placed vertically above and looking straight down at the map. The camera's resolution is set to 320 by 240 for image capture. It has been found from experiments that such a resolution is sufficient for a distance of 40 to 50 centimeters between the map and the camera. If necessary, the resolution of the captured images can be increased in other applications. The camera is also set to adjusted contrast and brightness automatically to avoid taking pictures that are too dull or too dark for processing.

3.4 Image Acquisition

The system starts by acquiring an image of a valid map that has been placed in the camera's field of view. The image is used as a template for background subtraction as long as the map is not moved too much. As the camera captures new images, each new image is compared with the background image to determine if the user has placed his/her finger onto the map. When there is a large difference between the background and the current captured image, the system checks for two conditions; user's finger over map, and movement of the map.

When the user's finger points to a location on the map and stays there for a given period of time (e.g. 500ms), the system acknowledges that the user wishes to know about the particular location and acquires the image for further processing.

In addition to checking for finger movements, the system also checks to see if the map is still in the same location. If the map has been moved beyond a certain threshold, the system will display a warning to the user and re-capture a background image. The current implementation requires that the map remains mostly in the same location during the execution of the program. Nevertheless, a more adaptable approach able to deal with map movements is considered for future versions of the system.

At this stage, we can also calculate the angle between the map in the input image and the horizontal axis, θ, as shown in Fig. 3. The value will be used in later process for coordinate transformation.

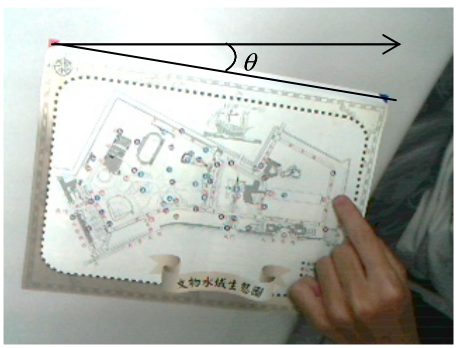
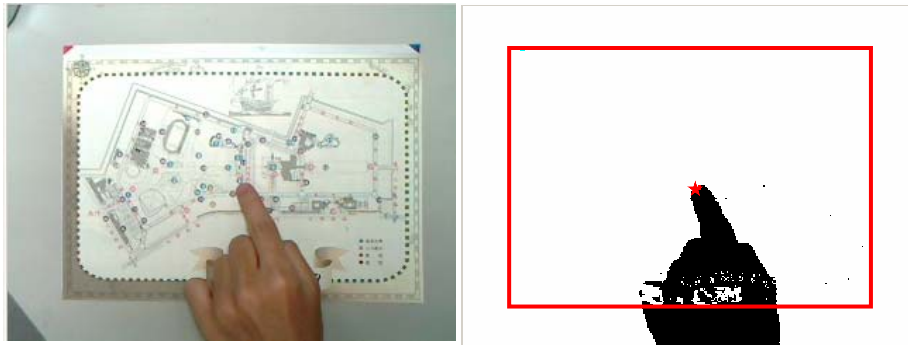


Fig. 3. An example input image and its rotation angle θ

3.5 Fingertip Detection

In the previous stage where the system detects that the user wishes to know about a particular location, a difference image between the background and the captured image would have been produced. The difference image is thresholded to produce a binary image. Morphological operations are performed on the binary image to determine location of the fingertip, (x_0, y_0) with respect to the map in the input image. The input image and the segmented image are shown in Fig. 4, note that the location of the detected fingertip is marked by a star, the boundaries of the detected map is shown as a rectangle.



(a) (b)

Fig. 4. (a) The input image with the user's hand and (b) the located fingertip marked with a star after background subtraction and segmentation

3.6 Coordinate Extraction

The coordinates of the fingertip, (x_0, y_0), with respect to the map, is transformed to the coordinates on the map stored in the database, (x, y) which are used to determine the information associated with the selected location . The coordinates (x, y) are obtained from (x_0, y_0) according to the following equation, in which θ is the map's angle of rotation from the input image.

$$\begin{aligned} x &= x_0 * \cos\theta - y_0 * \sin\theta \\ y &= x_0 * \sin\theta + y_0 \cos\theta \end{aligned} \quad . \tag{1}$$

Once the coordinates (x, y) are obtained, they are used to find a certain block number on the image map. An example of blocks on an image map is shown in Fig. 5, where the blocks have dimensions *k* by *l* and are numbered by integers. Eqn. 2 is used to find the block number of a set of coordinates.

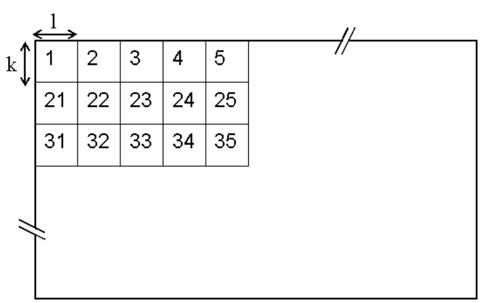


Fig. 5. Example of blocks on an image map

$$\text{block number} = W / l * (x \bmod k) + (y \bmod l) + 1 \ . \tag{2}$$

There are two reasons for dividing the image map into blocks. Firstly, the user's fingertip may not be very precise at pointing to a certain location on the map, therefore, there has to be a tolerance for selection, which is incorporated in the idea of a k by l block. Secondly, by converting a set of coordinates into integers, indexing can be performed more efficiently, since the feature points can each be given an associated block number for referencing.

3.7 Presentation of Selected Information

Each feature point on the map has its associated information, presented in the forms of text, images, and audio data. The data are displayed when the system determines that the user has selected the feature point. In addition, the system is also able to present the information in different languages. The default language is Mandarin, the official language in Taiwan. By pointing to marked regions on the side of the map, the user is able to interact with the system in English or Taiwanese, thus providing multilingual support for different users.

4 Discussion

The implemented system is operational and has received recognition in the departmental project competition. Nevertheless, the system is still in its prototype stage and there are many functions yet to be designed and implemented. For example, currently the system does not tolerate large displacement of the map during execution, since it will be unable to obtain a difference image for segmentation and fingertip detection. In the future, we plan to solve the problem by using a tracking algorithm for map and fingertip detection. In addition, we also wish to expand the system's database, such that it will be able to recognize more maps (and other informative materials) to provide more interactions. There are also plans to incorporate augmented reality functions into the system for more immersive interactions. Overall, the implemented system has shown promising potentials, as well as being fun to use. The system has also provided many ideas for future projects.

5 Conclusion

A multimedia, augmented reality interactive system has been designed and implemented in this work. In particular, the system is implemented to demonstrate its application in providing visitor information for the school.

The implemented system does not require the user to operate hardware devices such as the keyboard, mouse, or touch screen. Instead, computer vision approaches are used to obtain input information from the user via an overhead camera. As the user points to certain locations on the map with a finger, motions are detected and analysed by the system, and relevant information regarding the selected location are displayed. Hence, the user is able to operate the system without contacting any hardware device except for the printout of the maps.

The implementation of the system is hoped to reduce the cost of providing and maintaining peripheral hardware devices at information terminals. At the same time, eliminating health risks associated with contaminations by contact in public areas.

Work to enhance the system is ongoing and it is hoped that the system will be used widely in the future.

References

1. Liu, C.-H.: Hand Posture Recognition, Master thesis, Dept. of Computer Science & Eng., Yuan Ze University, Taiwan (2006)
2. Chen, C.-Y.: Virtual Mouse:Vision-Based Gesture Recognition, Master thesis, Dept. of Computer Science & Eng., National Sun Yat-sen University, Taiwan (2003)
3. Lai, J.C.: Research and Development of Interactive Physical Games Based on Computer Vision, Master thesis, Department of Information Communication, Yuan Ze University, Taiwa (2005)
4. Yeh, H.-C.: An Investigation of Web Interface Modal on Interaction Design - Based on the Project of Burg Ziesar in Germany and the Web of National Palace Museum in Taiwan, Master thesis, Dept. of Industrical Design Graduate Institute of Innovation & Design, National Taipei University of Technology, Taiwan (2007)
5. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. Prentice Hall, Englewood Cliffs (2002)
6. Jain, R., Kasturi, R., Schunck, B.G.: Machine Vision. McGraw-Hill, New York (1995)
7. Klette, R., Schluns, K., Koschan, A.: Computer vision: three-dimensional data from images. Springer, Singapore (1998)
8. Brown, T., Thomas, R.C.: Finger tracking for the digital desk. In: First Australasian User Interface Conference, vol. 22(5), pp. 11–16 (2000)
9. Wellner, P.: Interacting with papers on the DigitalDesk. Communications of the ACM, 28–35 (1993)
10. Azuma, R.T.: A Survey of Augmented Reality. Presence: Teleoperators and Virtual Environments 6, 355–385 (1997)
11. Augmented Reality Network (2008), `http://augmentedreality.ning.com`