Proceedings of IEEE International Conference on Applied System Innovation 2018
IEEE ICASI 2018- Meen, Prior & Lam (Eds)

A Novel Campus Navigation APP with Augmented Reality and Deep Learning

Cheng-Hung Lin, Yang Chung, Bo-Yung Chou, Hsin-Yi Chen, Chen-Yang Tsai
Department of Electrical Engineering, National Taiwan Normal University
Taipei, Taiwan, R.O.C
Email: brucelin@ntnu.edu.tw, Tel: +886-2-7734-3432
Email: potasdichrome@gmail.com, Tel: +886-978-035-671

Abstract

Augmented reality has been widely used in many applications because of its ability to offer an amazing way to overlay computer-generated images over the user's real-world view, creating a composite view rooted in real and virtual worlds. Augmented Reality is a realistic, direct or indirect view of the physical reality environment whose elements are "enhanced" through computer-generated or sensory input such as sound, video, graphics, tactile, or GPS data. In this paper, we present a novel campus navigation APP that uses augmented reality to provide users with a new and interesting way to meet our campus. With advanced augmented reality technologies such as computer vision and object recognition, the information about the campus environment and its objects is overlaid on the real world and becomes interactive. In order to improve the APP efficiency, this paper presents a virtual terrain modeling interface with deep learning to improve the object recognition ability.

Keywords: augmented reality, deep learning, computer vision, computer vision

I. Introduction

Due to the emergence of powerful mobile devices and bandwidth-improved broadband wireless networks, multimedia mobile applications (such as streaming or check-in places anywhere) can be executed anywhere. Among the multimedia mobile applications, augmented reality (AR) attracts a lot of attention because of its ability to offer an amazing way to overlay computer-generated images over the user's real-world view, creating a composite view rooted in real and virtual worlds. AR provides a realistic, direct or indirect view of the physical reality environment whose elements are "enhanced" through computer-generated or sensory input such as sound, video, graphics, tactile, or GPS data. Since augmented reality technology allows people to interact with the world around them, AR has become another new aspect embedded in modern life.

In this paper, we propose a campus tour APP with AR technology. The campus tour APP allows users to learn about the campus by quickly taking photographs of campus attractions. In order to identify the large number of tourist attractions on campus, the proposed APP combines positioning techniques to reduce search delays. In addition, we adopt machine learning to improve image recognition. Section II proposes the key components of the campus tour APP. Section III discuss the experimental results and Section IV presents the conclusion.

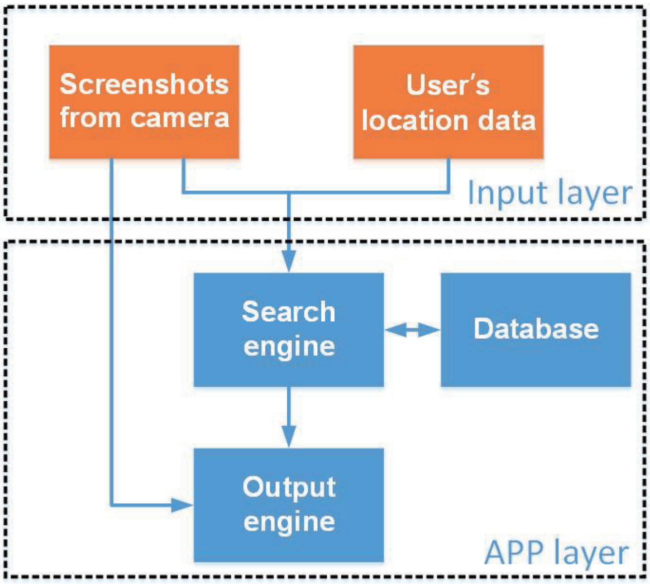

Fig. 1: Architecture of the AR-based campus tour APP

II. Architecture of the AR-based campus tour APP

The key technology of the proposed campus tour APP is based on image recognition. In our project, the features of the campus attractions are extracted by Scale-Invariant Feature Transform (SIFT)[1] and then stored in a database. We also determine the 3D model placement and direction by neural network. Furthermore, we utilize GPS information for assisting recognition performance. In the following, we will discuss the architecture of the proposed campus tour APP.

Fig. 1 shows the architecture of the AR-based campus tour APP that contains three major components, including

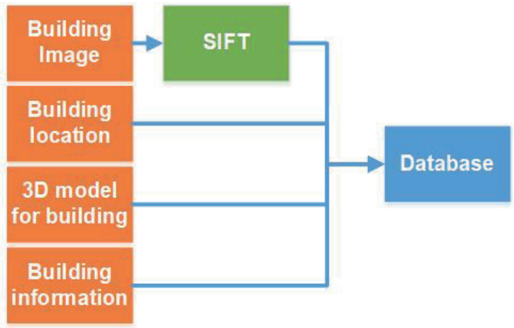

Fig. 2: Database structure

search engine, database, and output engine. The APP imports screenshots and user's location data into the search engine to compare with the database, and then the output engine outputs the match results, including the augmented reality elements, 3D model, and attraction information, on the screen of mobile devices.

A. Database construction

As shown in Fig.2, the database stores the following attraction data: the features of campus attractions, GPS coordinates information, attraction 3D models, and their corresponding information. The features of the campus attractions are extracted by SIFT and saved as a one-dimensional vector. With GPS coordinates information, we can use these as a basis for attraction recognition. When an attraction is recognized, its 3D model and information will be output through the output engine.

SIFT uses the edge and corner positions of an image as its features to identify images in different situations, including rotated or distorted images. Fig. 3 shows an example where SIFT can successfully distinguish the building from three different perspectives. Even if these perspectives are very different, SIFT can still distinguish that they are the same building.

B. Search engine

As shown in Fig. 4, after taking a picture of a target point, the image and the user's GPS coordinates information are sent to the search engine to compare with the database. The search engine first uses SIFT to extract the features of the input image and then compares the SIFT features and GPS data with the database to find the most similar match.

We also use GPS coordinates information as an assistance for data search. Fig. 5 shows that the APP only loads the blue cube database near the user through the user's location. Therefore, an input image does not have to be compared with every attraction in the database. The search speed can be significantly increased.

Fig. 3: Building recognition with SIFT

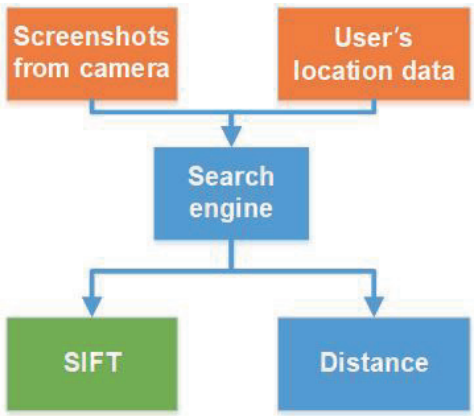

Fig. 4: Search engine structure

C. Output engine with augmented reality

Fig. 6 shows the Output engine structure. According to match results, the Output engine will show the attraction data using augmented reality. The APP achieves augmented reality by placing a 3D model at a specific location in front of the attraction and facing a specific direction. When the user interacts with the 3D model, the attraction's information is also displayed.

The Output engine places a 3D model in a specific place base on a CNN model. We utilize CNN[2] to determine

ISBN 978-1-5386-4342-6

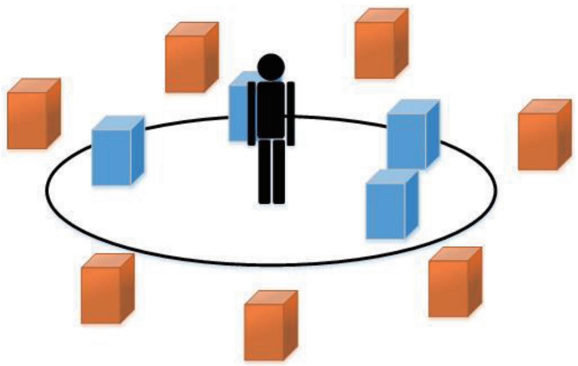

Fig. 5: GPS assistance diagrammatic sketch

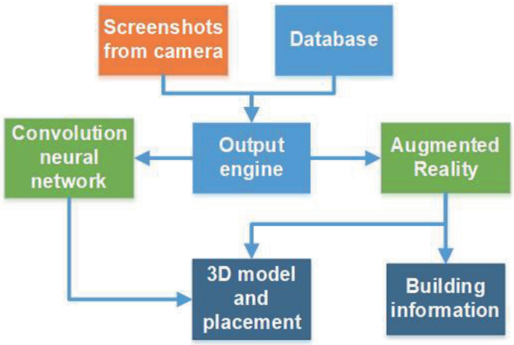

Fig. 6: Output engine structure

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \frac{1}{256} \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}$$

Fig. 7: Convolution kernel

the specific placement of the 3D models in each angle of attractions. After enhancing specific attraction's features with convolution, we make them as inputs of neural network. The output will be the placement of the 3D model and its facing direction. Fig. 7 is the convolution kernel that we used to find the feature of every attraction. Fig. 8 shows the proposed Neural Network which has 324 input nodes, 2 output nodes, and 2 hidden layers. After performing the convolution, we reduce the image to an 18x18 matrix as input to the neural network, and the output will be the position of the 3D model, including the coordinates X and Y and their rotation.

III. Experimental results

We implement the proposed APP on an ASUS_Z012DA mobile phone, which is equipped with a Qualcomm® Snapdragon™ 625 2.0GHz CPU, an Adreno™ 506 GPU, and a 16 million resolution camera. The APP achieves attraction recognition accuracy rate of 90%. The attractions can be easily identified by their appearance and characteristics. However, the use of 3D models with CNN was significantly delayed when running on mobile devices. The convolution process takes too much time, even if we have removed the training process from the phone.

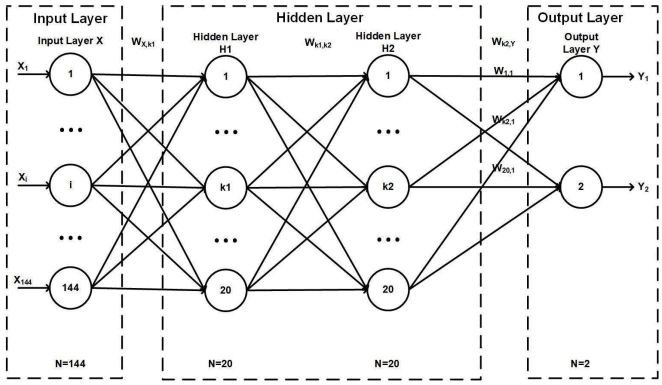

Fig. 8: Back-propagation neural network concept diagram

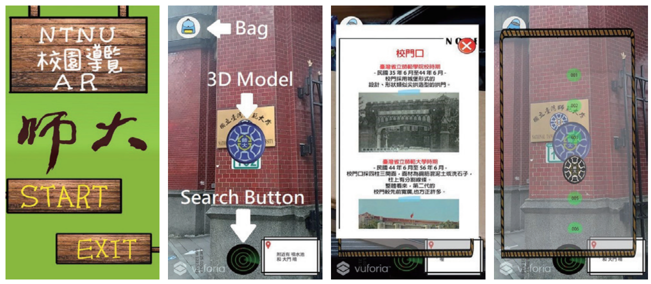

Fig. 9: Application screenshots

IV. Conclusion

We have proposed a novel campus navigation APP that uses augmented reality to provide users with a new and interesting way to meet our campus. The experimental results show that the APP successfully identified the attractions we stored in the database in advance. We will study how to improve the effectiveness of CNN and make it run efficiently on mobile devices.

References

[1] D.G. Lowe. Object recognition from local scale-invariant features, 1999.

[2] Michael Taylor. Make Your Own Neural Network: An In-depth Visual Introduction For Beginners, 2017.

[3] Aurlien Gron. Hands-On Machine Learning with Scikit-Learn and TensorFlow, 2017.

[4] Jin-De Wang. Neural Network and Fuzzy Control Theory, 2007.

ISBN 978-1-5386-4342-6