Applied Mechanics and Materials *Online: 2012-06-04*
ISSN: 1662-7482, Vols. 182-183, pp 854-859
doi:10.4028/www.scientific.net/AMM.182-183.854

Mobile Tour Planning Using Landmark Photo Matching and Intelligent Character Recognition

Cheng-Ming Huang[1], Wen-Hung Liao[2], Sheng-Chih Chen[1]

[1]Master's Program in Digital Content and Technologies, National Chengchi Univ., Taipei, Taiwan, s941622@gmail.com, scchen222@gmail.com
[2]Department of Computer Science, National Chengchi Univ., Taipei, Taiwan, whliao@gmai.com

Keywords: mobile devices, landmark photo matching, intelligent character recognition.

Abstract. The functionalities of smart phones have extended from basic voice communication to gaming, multimedia entertainment, information retrieval and location-based services. In this paper, we attempt to design a mobile application to assist visitors to have better understandings of popular tourist destinations and related routing information while on tour. The users can obtain descriptions of a specific attraction by simply taking the picture of a landmark photo often shown in the travel booklet using their mobile devices. This is achieved by matching the landmark picture with an image database containing popular tourist spots to locate the interested destination. The location information is further confirmed using techniques in intelligent character recognition. Upon successful identification of the interested location, tourist information regarding this destination, along with the routing details will be delivered using location-based service. We anticipate the proposed mobile application to effectively assist foreign visitors by bringing comprehensive, up-to-date tourist information and promoting better travel experience.

Introduction

With the rapid growth of mobile phones and services, an increasing number of applications are being developed to take advantage of the characteristics of mobile devices, such as multimedia streaming, location-based service, and navigation assistance. Mobile tour planning and guidance is one example that combines the best features a mobile device has to offer.

With the abundance of attractions around Taiwan, however, there seem to be a lack of user-friendly guide tools for non-Mandarin speaking visitors. It is true that brochures and tour books are freely distributed in many public places and tour information offices in several popular languages, such as English and Japanese. But whatever the language is, people seem to be more easily attracted by photos in a document filled with textual descriptions. It is often observed that foreign visitors plan their trip by referring to the photos of landmarks or tourist spots in a guidebook. To do so, they need to cross-reference between the caption under the photo and the content, a difficult practice since the keyword is often directly translated from Chinese, as illustrated in Fig. 1. The task becomes more challenging if we are to make arrangements for multiple locations.

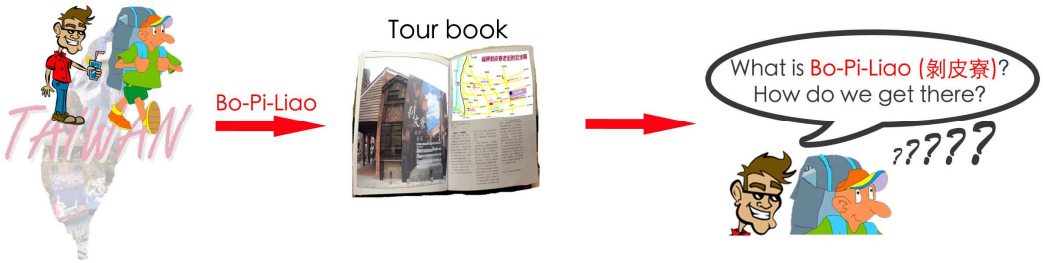

Figure 1. Finding information in a tour book can be challenging.

It is the objective of this paper to devise and develop a mobile application to assist tour planning using recognition of landmark photos and the associated textual information, either printed or handwritten. The proposed solution is built upon three core technologies, namely, location-based

service, large-scale photo matching, and intelligent character recognition. Location-based service is used to determine the current position of the user. Photo matching is employed to recognize the landmarks in a tour book. Intelligent character recognition is required to match texts that appear in a photo, a map or the description. After information regarding the desired destination has been successfully collected, a recommendation of the route, along with materials of interest, will be forwarded to the user by our application.

The rest of this paper is organized as follows. In section 2 we summarize the related work regarding large-scale image search and intelligent character recognition. Section 3 presents the core technologies to be employed in the proposed system, with special emphasis on landmark photo matching. Section 4 discusses the experimental results on landmark photo recognition. We conclude this paper with a brief summary in section 5.

Related Work

The proposed system utilizes landmark photo matching to recognize interested spots depicted in a tour book or brochure. Landmarks are representative images of a particular tourist destination, as illustrated in Fig. 2. Since the picture can be taken from various distances and viewpoints, automatic recognition can still be challenging even if we restrict ourselves to the popular tourist attractions in Taiwan.

Figure 2. Example of landmarks around Taiwan

Images in a tour book may contain depictions of scenery places, person, text and man-made objects. It is therefore necessary to coarsely classify the captured photo to determine what further processing is needed. For example, a picture which mostly contains textual information should be subject to text extraction and recognition to find out the meaning of the text. A landmark photo should be matched against a database of famous tourist destinations to determine the location and search for related information. Picture that contains faces may not provide useful indication of the desired location and should probably be kept as is.

In early stages, image matching usually rely on primitive visual features from image contents, such as color, shape, and texture. Later research moves from single concise feature to multiple image features. In [1], the authors achieved a 99% recognition rate of trademark images by combining color-based and shape-based features. Their proposed approach, however, it is sensitive to image variations, such as scale, rotation, size, color, illumination, or shape distortion. To address these issues, Lowe [2] presented a method for extracting distinctive invariant features known as scale-invariance feature transform (SIFT). SIFT has proved to be invariant to scale, rotation, and invariant to illumination variance partially.

With the wide popularity of digital cameras and social networks, it becomes common for average users to take a vast amount of photos, including buildings, monuments and church, while on tour and share them on the Web. Hence, it is conceivable that image collection is tremendously huge on the

Web. Large-scale image matching is a challenging task. Aside from feature descriptor, there are three main issues of concern, namely, storage, computational cost and recognition performance. It is not surprising to observe that indexing huge databases images has become an active research topic recently [3].

Efficient large-scale image matching calls for compact representation of images. GIST is a global feature descriptor that has been employed to perform web-scale image search [4]. It is designed to 'summarize' the structure of an image by computing the response of partitioned blocks to Gabor filters of different orientations and frequencies, as depicted in Fig 3. We will utilize the gist descriptor, with a slight modification, to represent the image in this research. Specifically, we will first compute the saliency map [5] of the image to detect the regions of interest. The saliency score is averaged in each block to obtain a weighting factor, which is then incorporated in generating the final gist vector for recognition.

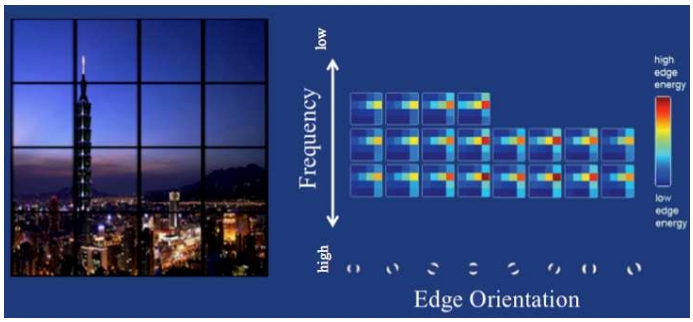

Figure 3. The gist descriptor

The Proposed Architecture

The proposed system aims to provide a mobile service for foreign tourists to plan their trips effectively and effortlessly. We manage to provide comprehensive routing information for tourists while they search their destinations according to the information depicted in the travel guide. Hence, we will integrate location-based services to our system. In addition, we will develop a recognition system which can recognize images or texts, depending on what type of information the travel guide provided. The system architecture and flowchart is depicted in Fig. 4.

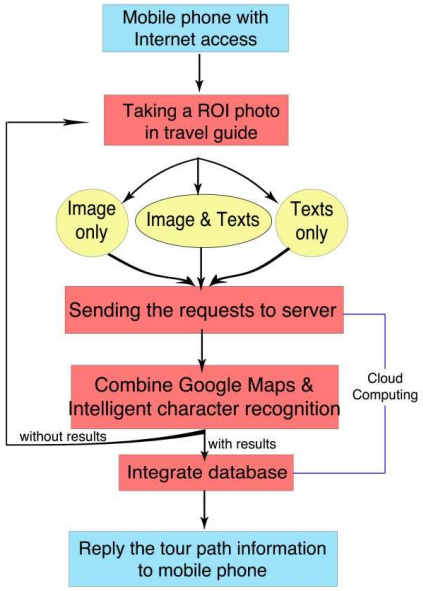

Figure 4. System flowchart

The workflow of a typical user query is as follows (Fig. 4 and 5):

1) The user takes a photo and selects the region of interest (ROI). Three categories are possible, namely, pure image, pure text, or mixture of image and texts.

2) The request is then sent to the remote server. The server is designed to deal with complex computations such as landmark photo/character matching.

3) Server starts executing image or character recognition, using corresponding recognition techniques depending on the type of input.

4) There are two possible outcomes after the recognition process. If the recognition result is correct, related tour planning materials will be returned. On the other hand, if there is an unsatisfactory result, go back to step 2) and ask the user to take a photo again.

5) Finally, return the retrieved data (tourist guide or routing information) to user's mobile phone.

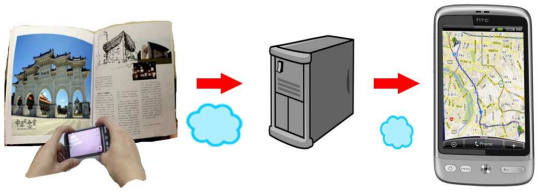

Figure 5. Simulation of typical usage.

This paper focuses on the third step of the process, i.e., landmark photo matching and intelligent character recognition. We will describe the key techniques we adopt in handling these two important tasks in the following.

A. Large Scale Image Matching

Landmark photos refer to the gallery of famous sights which are representative of the tourist spot and easily identified by average users. In [6], a collection of 5312 landmarks around the world has been extracted using images shared on the web. For our application, we restrict the area of interest to Taiwan. As a result, fewer categories (50) of landmarks will be analyzed in this paper.

In contrast to similar image search where only approximate result is required, our application calls for more precise recognition, as only the top-k (e.g., k=3) matches will be utilized to retrieve related tourist information. This usually demands the incorporation of local descriptors such as SIFT, which is expensive to compute and store. Another possibility is to follow a two-stage process, in which global features are first employed to filter out incompatible matches and local features are then compared to arrive at the best matches among the remaining candidates.

To strike a balance between efficiency and precision, we propose to combine two global features, namely, saliency map and gist, to perform the image matching task in this paper. First, the saliency map of the input image is computed using the method described in [5]. The input photo is then partitioned into 4x4 sub-images to prepare for the extraction of gist descriptor. The saliency score, which is the mean of the saliency measure in each sub-region, is utilized to weigh the contribution of the corresponding gist descriptor. Unlike many previous works where the distance between two images are computed using the sum-of-squared-difference (SSD) between gist descriptors, we fed the weighted gist descriptor to a modified support vector machine (SVM) to generate the list of possible matches.

Fig. 6 summarizes the key steps of the proposed landmark photo matching algorithm. To begin with, the query image is scaled to 512x512 since both saliency map and gist descriptor depends on the layout (Fig. 6a). A graph-based visual saliency algorithm is applied to construct the saliency map and locate regions of interest (Fig. 6b). The image is then divided into 4x4 blocks. Each block is therefore of size 128x128. The saliency measure in each block is averaged to arrive at a single weight factor for that particular block (Fig. 6c). Next, we compute the gist descriptor for each image block. We use 8 orientation channels at two different frequencies and 4 orientation channels at another frequency, totaling 20 coefficients for each block. The gist coefficients are concatenated to form a 320 dimensional feature vector. For color images, the dimension is increased to 960 as we will extract the gist features from R,G and B color channels, respectively (Fig. 6d). The concatenated gist vector is weighted according to the saliency score calculated in Fig. 6c to produce a weighted gist feature, which is then forwarded to the classifier based on support vector machine to perform the recognition (Fig. 6e).

B. Intelligent Character Recognition

Intelligent character recognition refers to the processing and classification of non-printed texts. In [7], we have surveyed and experimented with many character recognition algorithms and achieved certain level of success under some constraints. The experience is readily applicable to the proposed route planning service. We will describe the approach employed for robust character recognition in the following.

For feature extraction, we first divide the input image into 4x4 overlapped sub-regions. The percentage of overlapping is set to 50%. Each sub-region is further divided into four concentric rectangles. Then 16 orientations of gradient features are extracted from these four concentric rectangles. A vector consisting of 16-orientation gradient features is created by a weighted combination of the features from each of the concentric rectangle. The 16-orientation feature from each sub-region is concatenated to form a feature vector of 256 dimensions. This can increase the robustness in feature extraction if the input character is deformed or skewed. For feature classification, we employ the well-known support SVM with certain modifications to return a list of possible candidates, each with a probability measure. More details regarding the intelligent character recognition process can be found in [7].

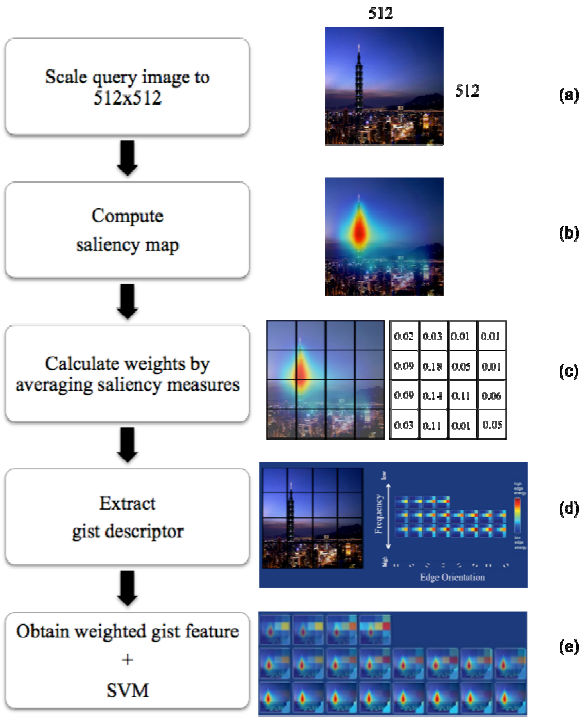

Figure 6. Landmark photo recognition using saliency-map-weighted gist feature

Experimental Results

At present, we restrict ourselves to the recognition of popular tourist destinations in Taiwan. We collect landmark images from tour books, travel brochures, official websites and photo-sharing networks. In this experiment, a total of 9530 images from 50 tourist spots have been gathered. On average, each tourist destination contains 190 images. We randomly select 20 images from each category to form the test set. The remaining 8530 images are used to train the support vector machine using the weighted gist feature described in the previous section. To investigate the role of saliency score, we also perform an experiment using the original gist feature. Table 1 summarizes the recognition results using these two feature vectors.

Table 1. Landmark photo recognition using gist and weighted gist features

Method	Top 1	Top 3	Top 5
Original gist feature	49.5%	68.7%	76.3%
Weighted gist feature (Saliency map+ gist)	64.5%	79.6%	85%

Without incorporating the saliency map, the accuracy is around 50% if we retain only the top match. The accuracy rate rises to 68.7% if the correct match appears in the top 3 candidates. On the other hand, if we integrate the saliency score to emphasize the regions of interest in an image, the performance is improved consistently, from 15% for top match to 8.7% for top 5 matches. These results are quite promising considering the fact that only global features are employed for the recognition task.

Conclusion

We have proposed an innovative idea of integrating location-based service, landmark matching and intelligent character recognition to provide a convenient mobile tool for route planning. The tool can assist users who are unfamiliar with the local area to search for interested destinations and provide suggestions for planning the trip. In this paper, we focus on landmark photo matching and propose a weighted gist feature to cope with the recognition problem. The preliminary results are promising, achieving 85% recognition rate for top 5 matches. We are finalizing the design of other components of the proposed service. We anticipate our proposed application to effectively assist tourists by bringing comprehensive information other than just tourism. Therefore, we plan to incorporate more contents such as education, food or other local culture information in the future.

References

[1]Anil K. Jain and Aditya Vailaya, "Image Retrieval Using Color and Shape", Pattern Recognition, Vol. 29, No. 8, pp. 1233-1244, 1996.

[2]David G.Lowe. "Distinctive Image Features from Scale-invariant Keypoints", International Journal of Computer Vision, 60(2), pp. 91-110, 2004.

[3] Mohamed Aly, Mario Munich, and Pietro Perona, “Indexing in Large Scale Image Collections: Scaling Properties and Benchmark”, IEEE Workshop on Applications of Computer Vision (WACV), Hawaii, January 2011.

[4]Torralba, A.; Fergus, R.; Weiss, Y., "Small Codes and Large Image Databases for Recognition", CVPR 2008, pp.1-8.

[5]Jonathan Harel , Christof Koch, and Pietro Perona, “Graph-based Visual Saliency,” Advances in Neural Information Processing Systems 19, 2007.

[6]Yan-Tao Zheng, Ming Zhao, Yang Song, Hartwig Adam, Urich Buddemeier, Alessandro Bissacco, Fernando Brucher, Tat-Seng Chua and Hartmut Neven, "Tour the World: Building a Web-scale Landmark Recognition Engine", CVPR 2009, pp. 1085-1092.

[7]Jen-Ho Kuo, Cheng-Ming Huang, Wen-Hung Liao and Chun-Chieh Huang, “HuayuNavi: A Mobile Chinese Learning Application Based on Intelligent Character Recognition”, Proceeding of the Sixth International Conference on E-Learning and Games, September 2011. (to appear)