World Scientific Research Journal Volume 7 Issue 3, 2021
ISSN: 2472-3703 DOI: 10.6911/WSRJ.202103_7(3).0001

Database Construction and System Design based on Scene Image

Yihang Li[1], Weifeng Xu[1, *] and Guping Zheng[1]
[1]School of Computer and Control Engineering, North China Electric Power University, Baoding 071003, China

Abstract

With the extensive application of deep learning method in the field of image retrieval, researchers have an increasing demand for scene images containing Chinese cultural sites and scenic spots with geographic information. However, there are few open scene image databases with geographic information. Moreover, some scattered small datasets have the problems of few data samples, inaccurate classification and inconvenient query, which bring a lot of inconvenience to the related research work. In view of these situations, the purpose of this paper is to build an applicable scene image database, and on this basis to develop a scene image retrieval system, to achieve simple retrieval function, to meet the needs of database detection and subsequent function improvement.

Keywords

Scene image; Image retrieval; Geographic information; Database; System design.

1. INTRODUCTION

With the continuous development of multimedia information technology and digital image equipment, people can easily obtain and store the real world scene in the form of image. Low cost mass storage and network hosting make people from passive consumption to active photography, and people's demand for images is also increasing. As a result, more and more image information appears in people's daily life, forming a large variety of image library, and with the passage of time, the number of images is growing exponentially. On the one hand, the rapid expansion of data makes the difficulty of multimedia data management gradually increase. On the other hand, due to the diversification of image retrieval technology requirements in social life, more attention has been paid in academic and industrial fields, which makes image retrieval technology develop vigorously. How to make more effective use of these massive image data information to facilitate people's lives has become one of the hot topics in current research.

In daily life, people pay more attention to the content of the image, that is, the scene information described by the image. Therefore, in order to make the image more convenient for people's life, it is particularly important to obtain the image scene information. People get most of the information from the outside world through the eyes, because the visual information can reflect the real face of things more directly, comprehensively and truly. However, in some cases, human visual observation is not accurate enough, and for some unfamiliar scenes, people are eager to understand the other content behind these scenes and obtain additional information. For example, when traveling, it is difficult for tourists to distinguish the difference between two similar scenes for unfamiliar scenes. At the same time, they are eager to quickly obtain the relevant information (such as name, history, structure, value, etc.) behind the scene by simply taking photos of the current scene. Therefore, people began to use more fast and efficient computer to replace human for image retrieval and recognition [1].

Based on the above, this paper intends to collect and establish a real scene image dataset, and on this basis, develop a scene image retrieval system to realize the retrieval function. Through

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202103_7(3).0001

the system, users input characteristic scene images to accurately retrieve the corresponding categories of images and the relevant information behind the images. Users can browse and download all the scene images in the system for scene image recognition and other related research in the field of deep learning. In addition, the system can also be used in the fields of science popularization, education, commerce and production. The construction of the system will improve the efficiency of data retrieval [2].

2. RELATED WORK

ImageNet image dataset [3] is one of the largest datasets in the world. The database contains more than 20000 categories, such as portrait, plant, food, animal and so on. At present, there are more than 14 million images, and each image has a corresponding label. The emergence of the database has aroused the research interest in the application of deep learning method to computer vision in academia and industry. For many years, it has been the benchmark database for image recognition tasks, which promotes the development of image recognition technology. However, the database contains few scene images of Chinese cultural heritage and scenic spots with geographic information. In addition, Oxford5K database [4] and Paris6K database[5] are widely used in the field of scene image retrieval, such as scenic spots or buildings, but they do not contain geographic information, and it is not as convenient to query and supplement the geographic information of each image as to directly use the scene Atlas of our country. These datasets are used to verify the design of many image retrieval algorithms with different features, so they will include images with different subject content and tag content, such as random screened images from Oxford5K datasets, as shown in Figure. 1, images of buildings, and images taken randomly from datasets. For standard image retrieval purposes, these images can exist in the database, but when used for training models, it is better to delete these images or use them as negative samples.

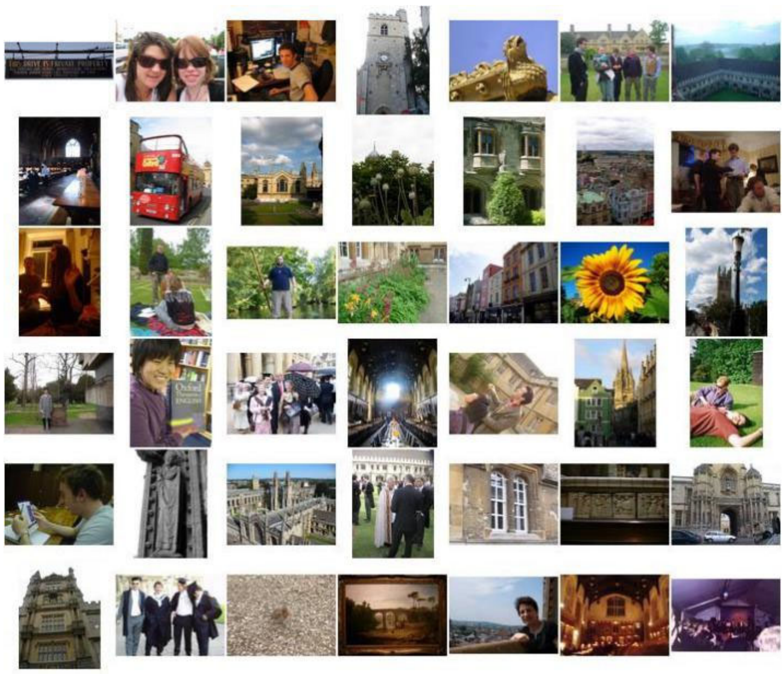

Figure 1. Random selected images from Oxford 5k dataset

Therefore, nearly 2000 scene images with different shooting States and sizes were collected through network and personal photo collection to supplement and improve their geographic information, mainly covering the scene image data of 20 different cultural sites or scenic spots,

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202103_7(3).0001

and the possibility of further expanding the gallery for a wider range of retrieval was reserved. At the same time, the correct annotation of basic query is established, and five retrieval images of different scene images and the corresponding similarity mark file are provided as the standard of evaluating the average accuracy. And part of the wide-angle image query area, in order to cut query, and retain the sample with less interference, if there are tourists, pedestrians occupy a small part of the image, in order to detect the reliability of the retrieval results, as shown in Figure. 2.

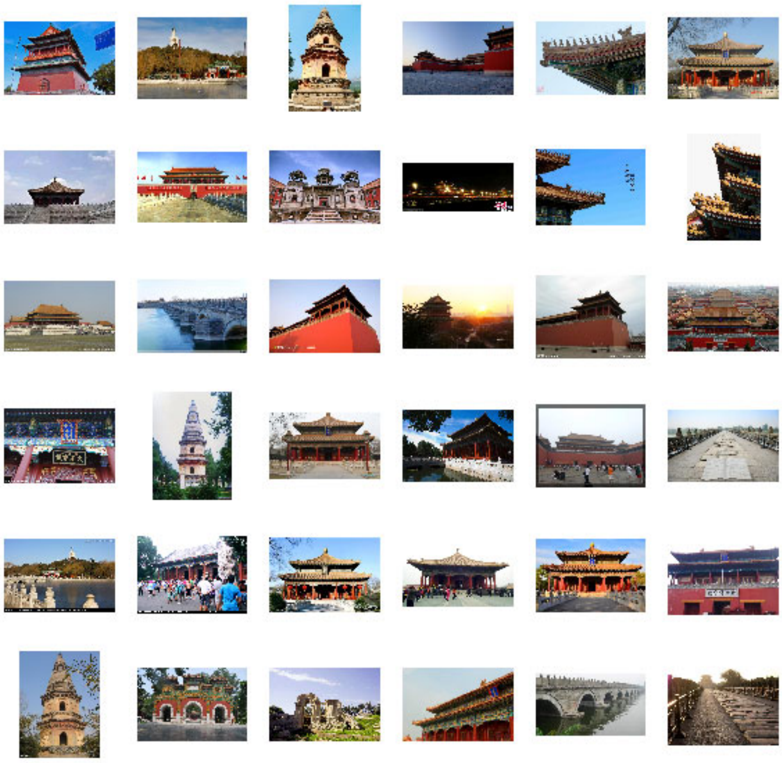

Figure 2. Random selection of images from self built dataset

3. CONSTRUCTION OF SCENE IMAGE DATABASE SYSTEM

The database can be imagined as an electronic filing cabinet, that is, a lot of data are stored together to form a database. Users can query, update, increase or decrease, etc., which is independent of the application program. The database system is a kind of data processing system which can improve the efficiency of data search. It is a collection of storage, maintenance, management data and system. Through the development of the system, the isolated island of data will be broken. Users can retrieve data more easily through the system, and managers can maintain, update and share data more conveniently through the system.

3.1. Requirement Analysis

For image or text correlation analysis, category analysis, output as accurate as possible classification results, the system needs to achieve the following requirements:

(1) When the user enters the scene name in the search bar and clicks the search key, the foreground interface of the web page will show the user the picture of the corresponding scene and its general situation;

(2) When the user enters an unknown scene image, click the search key, the system should output the scene image category, and retrieve the similar scene image. The system should have

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202103_7(3).0001

the ability of feature extraction and efficient identification for the data resources in the scene image database;

(3) The system should have backstage data management and maintenance function, and multiple management accounts can be set to facilitate managers at different levels to log in and use at the same time.

3.2. Function Design

According to the demand analysis of the scene image database system, the system function design is mainly divided into three modules: user module, retrieval module and management module. The structure is shown in the figure below.

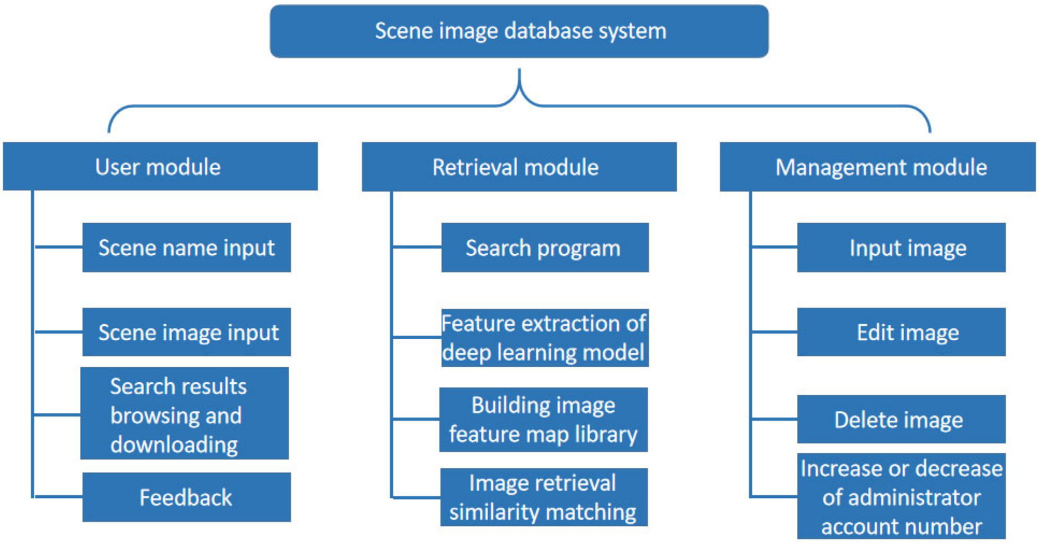

Figure 3. System function structure diagram

3.3. Database Design

The design of scene image database system is generally divided into front-end operating system, data retrieval system and background management system. The front-end operating system is responsible for displaying the retrieval results, the data retrieval system is responsible for processing the retrieval requirements, completing the category matching and image feature matching, extracting the retrieval results from the image database and feature database and displaying them to the front-end users, and the background management system is responsible for managing and maintaining the image data. The basic content of the database is to collect all the scene image data, and the original image data is stored in MySQL database. Some tables in the system database design are as follows.

Table 1. Scene information table

Table name	Scene_Information		
Field name	Field type	Field constraints	Field description
id	int(8)	Primary Key	Number
name	varchar(45)	Not Null	Name
description	varchar(1000)	Not Null	Introduction description
longitude	float(53)	Not Null	Longitude
latitude	float(53)	Not Null	Latitude

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202103_7(3).0001

Table 2. Scene image table

Table name	Scene_Information		
Field name	Field type	Field constraints	Field description
id	int(8)	Primary Key	Number
tid	int(8)	Not Null, Foreign Key	Scene information ID
title	varchar(200)	Not Null	Image title
cover	varchar(1000)	Not Null	Image storage path
cont_sign	varchar(1000)	Not Null	Identification of similar Library
create_time	datetime	Not Null	Image adding time
update_time	datetime	Not Null	Image update time

3.4. System Construction

Based on deep learning, this paper combines database technology with network technology to build scene image database retrieval system. Through the design of retrieval program, combined with MySQL database system, the search and search function of the system is realized, and the scene image name entered by the user is searched to find the corresponding scene image and its information; Combined with deep learning feature extraction method, the image recognition function of the system is realized, and the scene image input by the user is recognized; PWA[6] and geographic information are used to realize the image retrieval function of the system, and similar image retrieval is carried out for the scene image input by users.

4. DESIGN OF SCENE IMAGE RETRIEVAL SYSTEM

Feature extraction is the key step of image retrieval system[7], which plays a decisive role in the performance of image retrieval system. With the great breakthrough of deep learning in various fields, deep learning model has made a leading position in feature extraction, so the application of deep learning related methods in image retrieval is gradually widespread. In order to promote the application of deep learning in cultural tourism, this paper develops a scene image retrieval system based on the scene image database, which realizes simple retrieval function and meets the requirements of database detection and subsequent function expansion.

4.1. Algorithm Overview

This paper adopts the image retrieval algorithm based on the combination of PWA and geographic information, which is roughly divided into three stages: "probability suggestion" selection, local detector weighted aggregation and GPS geographic information filtering. The algorithm flow chart is shown in Figure. 4.

4.1.1 The choice of "probability suggestion"

Step 1: Feed forward convolution neural network of database image. After selecting the appropriate convolution layer, the depth feature map is extracted in this layer;

Step 2: Sum the depth convolution features of each feature map of the convolution layer, and then calculate the variance of each feature map on the database image;

Step 3: Rank the variance of the characteristic graph, and take the channel ranking at the top as the "probability suggestion". PWA thinks that the feature graphs with the largest variance have different responses on different targets, so they have strong discrimination ability;

Step 4: Calculate the spatial weight of "probability suggestion".

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202103_7(3).0001

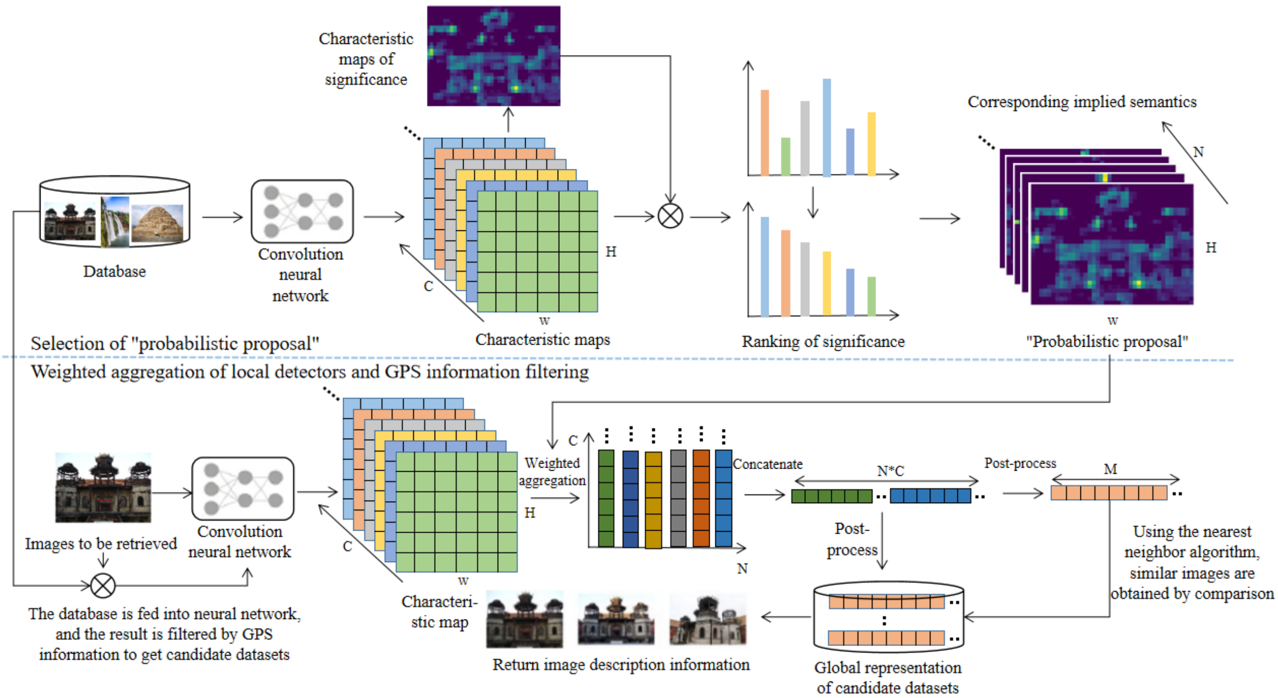

Figure 4. Algorithm flow chart

4.1.2 Weighted aggregation of local detectors

Step 1: Feed forward convolutional neural network is used to extract the depth feature map of the image to be retrieved;

Step 2: The depth feature map is weighted by the spatial weight obtained in the selected "probability suggestion" process;

Step 3: The weighted depth feature maps are aggregated by the method of summation and confluence, and the local features are obtained;

Step 4: After directly connecting the local features corresponding to each local detector and using PCA to reduce the dimension[8], the global representation of the final feature is obtained.

4.1.3 GPS geographic information filtering

The scene images obtained by mobile devices often contain GPS geographic information. The application of GPS information in the field of image retrieval will greatly improve the accuracy and speed of retrieval[9]. By comparing the GPS information of the image to be retrieved and the database image, the candidate datasets within a certain GPS range are filtered. The convolution features of each image in the candidate dataset are processed to get the corresponding global representation. After that, the K-Nearest Neighbor (KNN) algorithm[13] is used to compare the global representation of the image to be retrieved with that of the candidate dataset, and the similar images in the candidate dataset are returned and sorted according to the similarity.

4.2. Building Environment and Platform R & D Technology

One application and data server is used to run the server program. The configuration is shown in the following table. If the scene image database continues to expand, the application server and data server can be split.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202103_7(3).0001

Table 3. Server configuration table

Basic parameters	2U rack server
Processor	CPU type: Intel Zhiqiang e5-2600 CPU model: Xeon E5-2620v4 CPU frequency: 2.1GHz
Memory	DDR4 16GB *1
Storage	SSD 960GB *1

The system environment mainly includes server and client running environment. Server running environment:

Development language: Python 3.6; development environment: Ubuntu 18.04.2; database: MySQL;

Client running environment:

Operating system: Windows 7 or above; Web browser: Google, IE and other classic commonly used browsers.

Based on the cross platform and strong expansibility Python language technology, the search program is designed. Combined with MySQL database, the browser / server mode BS[14] architecture system which can be compatible with a variety of operating systems is selected. This mode simplifies the client and makes the system development and maintenance stable. The administrator can quickly organize the information into the system through the background to realize real-time update.

4.3. System Function Application

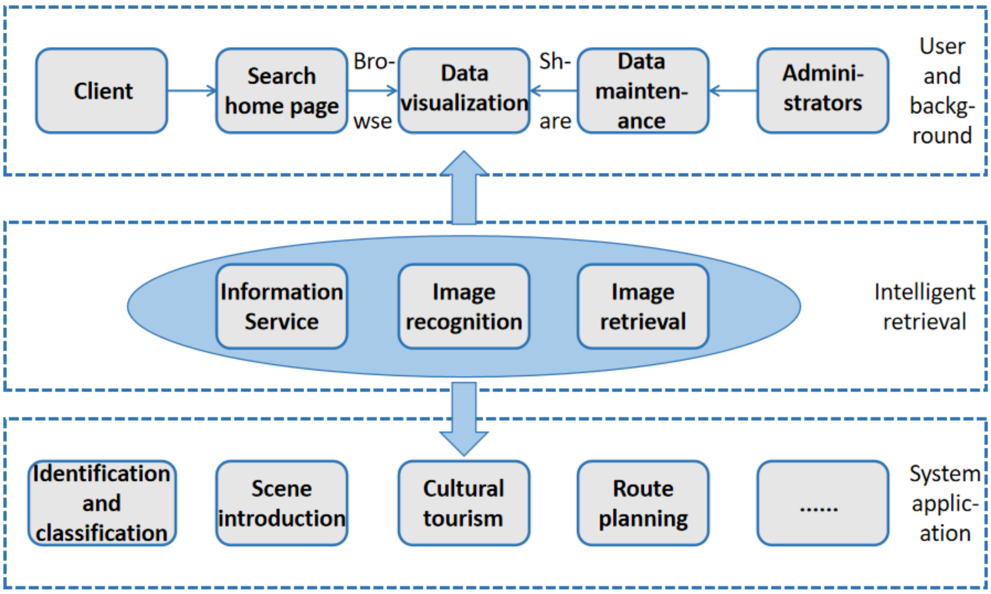

Figure 5. System function application diagram

The database system chooses mysql, a storage database management system based on Linux, as the software platform to design and research a data retrieval system that is convenient for users to operate and use. It focuses on strengthening the processing function of intelligent data retrieval, paying attention to the speed of retrieval and the stability of data. By using the retrieval function of the system, image data visualization can be realized. Users can apply the system to the recognition and classification of scene images, scene introduction, cultural tourism, route planning and other aspects. As shown in Figure. 5.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202103_7(3).0001

5. EXPERIMENTS

5.1. Experimental Setup

The convolutional network model used for feature extraction is VGG16 network model [15] trained by ImageNet dataset. The features of all the scene images in the database in the pool5 layer are extracted, and the number of channels in this layer is 512. Most of the images keep their original size, and some of them are too large to be reduced to about 1000 * 1000. The evaluation method is to calculate the mean average precision (mAP) [16], and the experimental results will show the results of the top 10 sorting queries. Tensorflow is used as the deep learning library of Keras, and the local host is used as the platform of feature extraction process.

The PWA algorithm is directly used in the established database. Ten images to be retrieved, such as Yanxigong, Huangguoshu waterfall and Xixia royal mausoleum, are selected. The preliminary retrieval results are shown in Figure. 6. The mAP of the overall retrieval is 89.5% (at this time, M = 1024, N = 20), and basically all the retrieval items can reach the average accuracy of more than 60%.

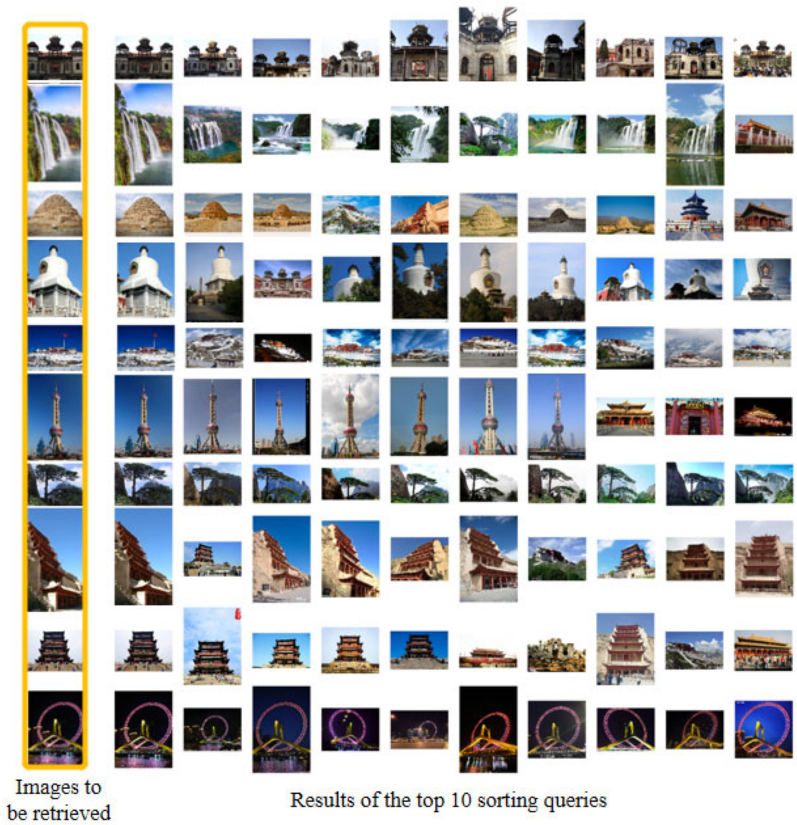

Figure 6. Preliminary retrieval results of ten groups of scene images to be examined

5.2. Performance Testing

Software platform performance test mainly includes server response time index and system capacity index. Generally, the response time of software platform is judged by 2 / 5 / 10 standard, that is, the response within 2 seconds will make users feel "fast response", the response within 5 seconds will make users feel "slightly slow response, acceptable", and the response within 10 seconds will be unacceptable. The capacity index of the system mainly considers the load capacity of the software platform, that is, how many service requests it can bear in a period of time. There are not many users in the initial stage of the software platform in this paper. It mainly verifies the availability of the database system, and only considers the identification retrieval requests with 10 concurrency. Therefore, the requirements for the

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202103_7(3).0001

capacity index are not high. The next version of the software upgrade will focus on improving the capacity index. The specific situation of this performance test is shown in the table below. According to the results in the table, this version of software platform meets the requirements of initial use, and the response is faster when there are a few concurrent requests, and the response is acceptable when there are 10 concurrent requests.

Table 4. Software platform performance test

Test content	Concurrent requests		Remarks
	1	10	
Software platform response time	<5s	<5s	Acceptable response

5.3. Image Recognition

Users input a local scene image through the system, and click search to identify the scene image related information, as well as other similar images. As shown in Figure. 7.

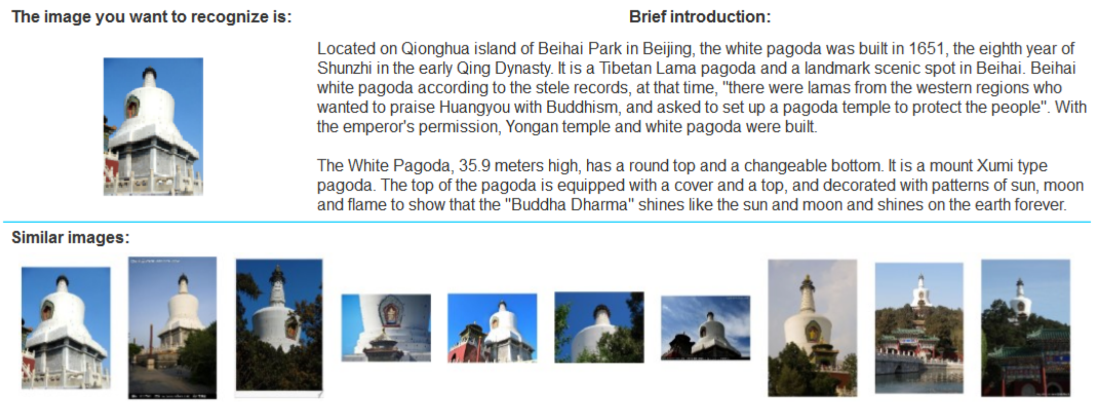

Figure 7. Image recognition test

6. SUMMARY

In this paper, aiming at the problems of scattered resources and inconvenient search in the scene image dataset with geographic information, data processing is carried out. On the basis of the research methods that have achieved certain results in the early stage, the system architecture is designed, the software and hardware environment is built, and the scene image database is developed. It can search image by name, search name by image and search similar image by image. Finally, through the function and performance test of the system, the system runs stably and the test results are good. It provides a test basis for the relevant image retrieval algorithm, so that it can be fully applied in the future scene image retrieval related work. In the follow-up experiments, we plan to further expand the number of scene images, improve the information introduction of relevant images, and further optimize the whole retrieval process to make it more efficient and convenient. We hope that it can be applied to specific scenes in the future to achieve its social and economic benefits.

ACKNOWLEDGEMENTS

This paper supported by Computer basic education teaching research project of AFCEC (2019-AFCEC-125).

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202103_7(3).0001

REFERENCES

[1] F. Radenović, G. Tolias and O. Chum, "Fine-tuning CNN image retrieval with no human annotation", IEEE transactions on pattern analysis and machine intelligence, 2018, Vol. 41 (7), p1655-1668.

[2] K. Ahmad, M. Sahu, M. Shrivastava, and et al, "An efficient image retrieval tool: query based image management system", International Journal of Information Technology, 2020, Vol. 12 (1), p103-111.

[3] O. Russakovsky, J. Deng, H. Su, and et al, "ImageNet Large Scale Visual Recognition Challenge", International Journal of Computer Vision, 2015, Vol. 115 (3), p211-252.

[4] J. Philbin, O. Chum, M. Isard, and et al, "Object retrieval with large vocabularies and fast spatial matching", Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC: IEEE Computer Society, 2007, p1-8.

[5] J. Philbin, O. Chum, M. Isard, and et al, "Lost in quantization: improving particular object retrieval in large scale image databases", Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC: IEEE Computer Society, 2008, p.1-8.

[6] J. Xu, C. Shi, C. Qi, and et al, "Unsupervised part-based weighting aggregation of deep convolutional features for image retrieval", Proc of the 32nd AAAI Conf on Artificial Intelligence, Palo Alto: AAAI, 2018, p7436-7443.

[7] G. Raghuwanshi and V. Tyagi, "Impact of Feature Extraction Techniques on a CBIR System", Advances in Computing and Data Sciences, 2019, p338-348.

[8] J. Hervé and C. Ondřej, "Negative evidences and cooccurences in image retrieval: the benefit of PCA and whitening", Proceedings of the 2012 European Conference on Computer Vision, Berlin: Springer, 2012, p774-787.

[9] D.M. Chen, G. Baatz, K. Koser, and et al, "City-scale landmark identification on mobile devices", Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 2011, p737-744.

[10] Z. Li and K.H. Yap, "Content and Context Boosting for Mobile Landmark Recognition", IEEE Signal Processing Letters, 2012, Vol. 19 (8), p459-462.

[11] G. Schroth, A. Al-Nuaimi, R. Huitl, and et al, "Rapid Image Retrieval for Mobile Location Recognition", IEEE International Conference on Acoustics, Speech and Signal Processing, 2011, p2320-2323.

[12] Y. Zheng, "Tour the World: Building a Web-scale Landmark Recognition Engine", IEEE Conference on Computer Vision and Pattern Recognition, 2009, p1085-1092.

[13] L.E. Peterson, "K-nearest neighbor", Scholarpedia, 2009, Vol. 4 (2), p883.

[14] H. Lu, S. Gao and S. Li, "Design of Mobile Integrated Development Platform for Business System Based on B/S Architecture", Journal of Physics: Conference Series, 2019, Vol. 1237 (4), p033-042.

[15] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition" [EB/OL]. (2015-04-10)[2017-07-20]. https://arxiv.org/abs/1409.1556.

[16] M. Everingham, L.V. Gool, C.K.I. Williams, and et al, "The Pascal Visual Object Classes (VOC) challenge", International Journal of Computer Vision, 2010, Vol. 88 (2), p303-338.