Research on the Method of Managing Personal Massive Photos Using Geodatabase

Tao Cheng [1], Xuchao Xing [1], Xinyan Zheng [1],Yang Gao[1]

[1] National Geomatics Center of China, Beijing, 100830, China – (chengtao, xingxuchao, xinyan_zheng,gaoyang)@ngcc.cn

Keywords: Photo, Geodatabase, Exif, Geographic Information.

Abstract

In view of the limitations of the existing methods for personal massive photos management and browsing, a method of using geodatabase to manage and browse massive personal photos is proposed. Firstly, vector geodatabase is created and defined based on geographic information technology. Then, attribute information including the file name, photo date and time, photo point longitude, photo point latitude, location name, photo storage path and so on of each photo in the file system are extracted. Finally, based on the longitude and latitude of the photo point, the geospatial location data of each photo is generated, and written into the vector geodatabase together with the attribute information. Based on the storage path of photos in the file system, the links between the geospatial database and photos are established. Furtherly realize the attribute information and geographical space location management of massive photos, as well as the browsing and viewing of photos. Take personal photos as an example to introduce the data processing and verify the practicability of the method. The result shows that this method uses the idea of geodatabase to manage massive personal photos, which is conducive to reflect the geographical spatial distribution and density characteristics of the photos, and can use the time axis to connect the track of personal history, which is convenient for browsing and analyzing the experience.

1. Introduction

According to the "China Mobile Internet Development Report (2022)" released by People's Daily Online, the number of mobile internet users in China has reached 1.029 billion (Tang, 2022) . With the popularization of smartphones and the continuous improvement of the functional performance, using mobile phones to take photos has become the most common way for people to take photos. The convenience lead to the massive growth of personal photos, and as the resolution of sensors increases, their data volume exhibits characteristics of big data. After taking photos, people usually name the folder for storing photos based on personal habits, such as the date, place where the photos were taken, or the name of the scenic area. Some people store photos in the default folder name on their phones; the most of photos in the folder often use temporary file names, some photos will have their file names modified as users needed; But due to the large number of photos, few people will rename all the photos. Moreover due to the lack of unified standard specifications to follow, the naming of photo file names in different folders is confusing and prone to duplication. In addition, users may forget the storage location and file name keywords of the photos they want to query, which leads to difficulties in searching and utilizing photos.Therefore, the storage, management (Wang and Yao, 2012), directory organization, fast querying, and convenient browsing of massive personal photos have become an urgent need in front of people. In order to improve the convenience of querying and analyzing photo data, it is necessary to scientifically store and manage photos. At present, photo batch management software mainly includes ACDSee, Bkviewer, etc (Meng, 2007) .These types of software have the characteristics of small space occupation and fast running speed, and they have strong practical value in digital photo management. Meanwhile, with the application of Exchangeable Image File Format (Exif) in digital photos(Liu, 2010) .photo attribute information and shooting data can be recorded in digital photos, including camera parameters, photo original date and time, X-axis resolution, Y-axis resolution, GPS longitude, GPS latitude, etc. Exif can be attached to JPG, TIFF, RIFF, and other files, and can be browsed and viewed in general software. These pieces of information provide a foundation for the uniqueness of each photo in photo management.

Currently, there is an increasing amount of research and application on Exif. Tian Fangjie and others pay attention to the current situation that traditional management methods cannot meet the needs of photo retrieval. They proposed an identification based photo organization and management method, which uniquely identifies photos through location codes and time codes based on the characteristics of spatiotemporal uniqueness (Tian, Cheng and Pu, 2018) By analyzing Exif information, establishing information retrieval relationships, editing query interfaces, Bo Jianjie et al. developed a method that enables users to quickly find the desired photos from a massive collection of photos (Bo and Hu, 2018). Zhang Zhanyou developed a photo organization software based on Exif information, which can easily organize photos into different folders according to their shooting dates; he also studied the unique identification code of photos by combining parameters such as file name, file date and time, shooting date and time, Exif information, file size, etc. (Zhang, 2015) . Li Yao discussed how to develop a software system using C # language that can automatically read Exif information and recognize, classify, and save images based on Exif information for users to query (Li, 2012). Ma Ju et al. proposed a mobile phone image search algorithm that integrates region based camera Exif information, and used this algorithm to develop a search system for mobile phone user (Ma , Lu and Guo, 2007) .Significantly, the National Archives Administration issued and implemented the "Digital Photo Archiving and Management Standards" in 2014, providing a basis for digital photo archiving and management, which also utilizes Exif information (National Archives Administration of China, 2014).

However, existing management methods still have some limitations. On the one hand, this management method is not conducive to overall understanding of the spatial conditions of personal on-site photography and the number of visits to each

spatial area from a geographical perspective; On the other hand, this management method is not conducive to managing the timeline of personal visits to areas, which can make it difficult for users to sort out their past experiences in chronological order according to the photos.

Aiming at this limitation, a method of using geospatial databases for management is proposed. Using a file system to manage photo files, using a structured two-dimensional table to manage photo attribute information, and using vector data of spatial structure to manage photo spatial location, thereby achieving the management of attribute information and geographic location of massive photos, as well as the browsing and analysis of photos.

2. Method

2.1 Photo preprocessing and information extraction

The prerequisite for spatialized photo management is that photos have coordinate information, namely the longitude and latitude of the shooting point. Meanwhile, consider incorporating digital photos without recorded coordinate information, as well as early film imaging photos, into geospatial database management. Therefore, it is necessary to preprocess personal photos to make them unified photos with coordinate information.

For digital photos that record coordinate information, directly extract their coordinate information and other information; For digital photos without recorded coordinate information, rough coordinate information is first written into the photo, and then its coordinate information and other information are extracted; For film imaging photos, digitize them first, write rough coordinates, rough shooting date and time information into the digitized photos, and then extract their coordinate information and other information. Among them, rough coordinate information can be obtained by querying relevant geographic information data or online channels through photo location names.

The date and time of taking the photo, the longitude and latitude of the shooting point are recorded in the Exif information of the photo, respectively in EXIF format_ DateTimeOriginal, EXIF_ GPSLongitude, EXIF_ The GPSLatitude string is a label, and Exif information is readable, writable, and editable. When extracting and writing, the ExifTool command-line tool can be used (download website: http://www.sno.phy.queensu.ca/~Phil/exiftool/), or free integration tools, or independently developed software implementation. The file name and storage path information of the photo are recorded in the file information of the photo, labeled with FileName and Directory strings, and can be extracted together using the above tools. The location name of the photo shooting point can be obtained from the folder name, file name, or other means where the photo is stored.

Based on the above processing, all the main attribute information of the photo can be extracted. After certain editing and processing, various attribute information is organized into attribute data formats defined in geospatial databases and stored in table data.

2.2 Creation of Geospatial Database

Based on geographic information technology, create and define a vector geospatial database (Kang, Ren and Cai, 2021), and define the geometric types and spatial coordinate systems of spatial data. Geospatial databases can store attribute data, vector data, image data, spatial relationship data, and even video object spatial data (Han, Kong and Qin, 2015).

According to the characteristics of photo management, a point vector geospatial database can be created, with the spatial coordinate system defined as WGS 84 (Cheng, Wen and Cheng, 2009), which is consistent with general-purpose cameras. The definition of attribute data mainly includes photo file name, shooting date and time, shooting point longitude, shooting point latitude, location name, photo storage path, etc. The data types of each attribute data are shown in Table 1, and other attribute items can be extended as needed during use.

attribute item	data type	length	Attribute Description	Example
file name	TEXT	255	The file name of the photo	IMG_2442.JPG
Shooting date and time	TEXT	255	The date and time of the photo shoot, accurate to seconds	20220405165457
Longitude of shooting point	DOUBLE		Longitude of the photo shooting point	116.207125
Latitude of shooting point	DOUBLE		The latitude of the photo shooting point	39.999094
Location Name	TEXT	255	The place name or scenic area name of the photo shooting site	beijing botanical garden
Photo storage path	TEXT	255	The storage path of photos in the file system	D:/photo/20220405Beijingbotanical garden/IMG_2442.JPG

Table 1 Data Types of Attribute Data

2.3 Spatial data generation and updating

Using the longitude and latitude information of the shooting point in the photo, generate point vector spatial data, create a record for each photo, and write all corresponding attribute information into the attribute data of the geospatial database. Based on the storage path information of the photo, realize the connection between the geospatial database and the photo. Thus achieving attribute information and geographic location management of massive photos, as well as browsing and analyzing photos.

For newly taken photos, similarly, after photo preprocessing, information extraction, and spatial data generation, the incremental data is merged into the geospatial database to achieve updates (Pan, Zhong and Zhao, 2004), maintaining consistency between the geospatial database and photo file data, and achieving the integrity of photo management.

3. Experiments and results

Taking 1000 actual personal photos as an example, the photos include three types: digital photos with recorded coordinate information, digital photos without recorded coordinate information, and film imaging photos.

3.1 information extraction

(1) For digital photos and film imaging photos without recorded coordinate information, following the photo preprocessing method and process, using existing GeoSetter software tools, rough coordinates and other information are written to form a unified photo with coordinate information; After sorting, it is uniformly stored on the hard drive. The folders are mainly named by combining date and place names, and some folders are named more casually, storing photos taken from different dates and locations.

(2) Using the ExifTool command line tool, batch extract all JPG format photo file names, shooting dates and times, shooting point longitude, shooting point latitude, and photo storage path information from the photo storage directory and store them in the photoinfo.txt file. Separate each attribute item with a tab. The command line code is as follows:
exiftool -R -T D:/photo -FileName -exif:DateTimeOriginal -exif:GPSLongitude# -exif:GPSLatitude# -Directory .*.jpg > D:/photoinfo.txt

The results are shown in Figure 1.

Fig.1 Extraction Result of Photo Information

(3) Using general text editing software as a tool, organize the results obtained from the above processing into attribute data formats defined in the geospatial database, and store them in table data, as shown in Table 2. Among them, the location name of the photo shooting point is obtained from the photo storage path information.

file name	Shooting date and time	Longitude of shooting point	Latitude of shooting point	Location Name	Photo storage path
IMG_2497.JPG	20090502183722	116.280833	39.997833	the Summer Palace	D:/photo/20090502 the Summer Palace /IMG_2497.JPG
IMG_2415.JPG	20140531142550	115.599320	39.638190	Beijing Shidu	D:/photo/20140531 Beijing Shidu /IMG_2415.JPG
IMG_2442.JPG	20220405165457	116.207125	39.999094	beijing botanical garden	D:/photo/20220405beijing botanical garden/IMG_2442.JPG
IMG_2500.JPG	20220405184605	116.208875	39.999103	beijing botanical garden	D:/photo/20220405 beijing botanical garden /IMG_2500.JPG

Tab.2 Normalized Photo Information

3.2 Database Construction

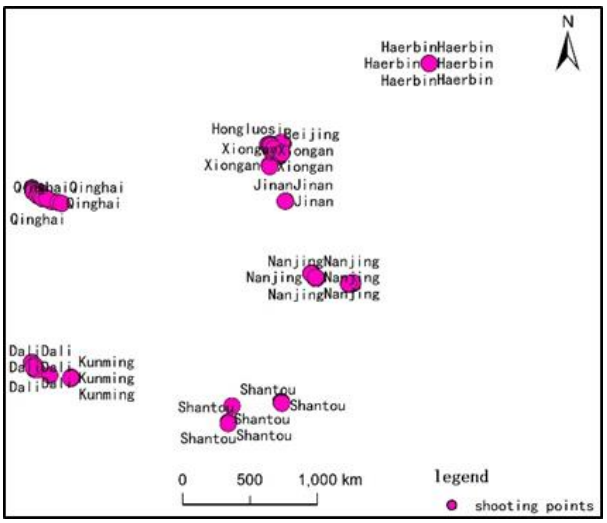
(1) Using the general ArcGIS software as a tool, right-click a folder in ArcCatalog, select New, and then select File Geodatabase. Enter the name of the file database and create an empty database named photo.gdb.

(2) In ArcMap, import table data by adding XY data, and generate point vector spatial data for each photo based on the longitude and latitude of the shooting point in the table data.

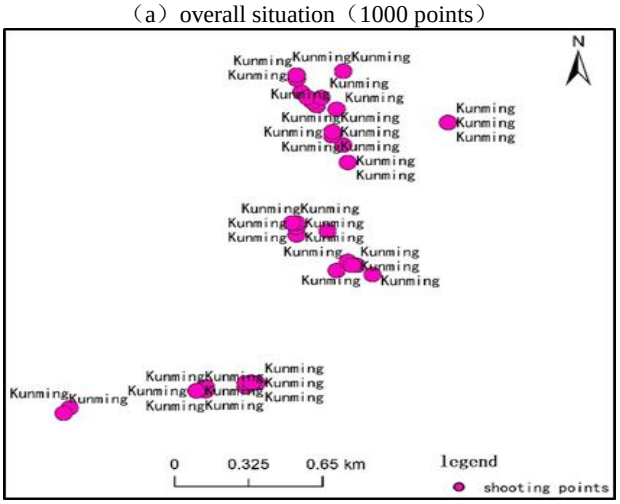
(3) Import the generated spatial data into photo. gdb and set the photo storage path to support hyperlink fields, thereby generating a geospatial database of massive personal photos. In this database, each photo corresponds to a record, a point feature, and its attribute information is stored in the attribute data. The content and type of the attribute data comply with the definition in Table 1. Click the hyperlink icon to open the photo for browsing.

3.3 Result analysis

The generated point vector space data results are shown in Figure 2.



（a）overall situation（1000 points）



（b）enlarge（174 points in Kunming）

Figure 2 Point Vector Spatial Data (Photo Points)

Through, it is possible to visually see the places where individuals have visited and taken photos, as well as the locations where photos were taken. Click on the hyperlink icon on the photo property page to open the photo for browsing, as shown in Figure 3.

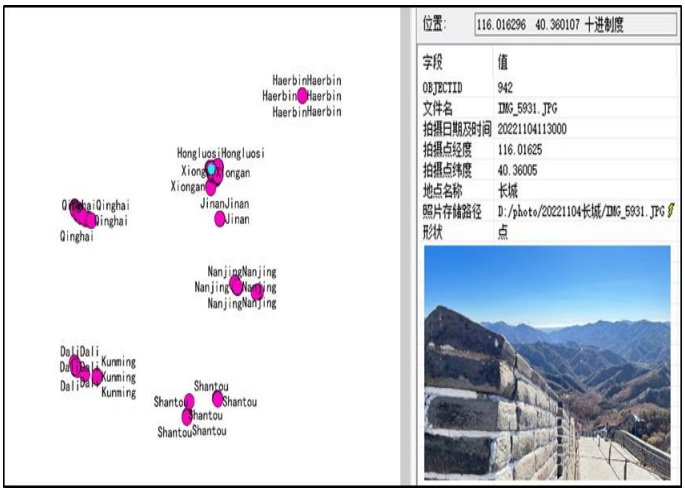


Fig.3 Photo Browsing

By analyzing the density characteristics of the shooting points, the level of activity of individuals in various areas visited can be analyzed, which also indirectly reflects the level of interest in the area. For example, there are 174 shooting points in Kunming, indicating that there are many points that attract individuals to this area. By analyzing the time characteristics of the shooting points, the number and time of visits to this place can be calculated. For example, they visited Kunming on November 8, 2017 and August 15, 2020 respectively, for a total of 2 visits.

By utilizing the time characteristics of the shooting points, the shooting date and time attributes can be sorted in order. Based on this, all shooting points can be connected to generate linear vector spatial data, which expresses the time trajectory of personal visits and photos, facilitating the recall of past experiences, as shown in Figure 4.

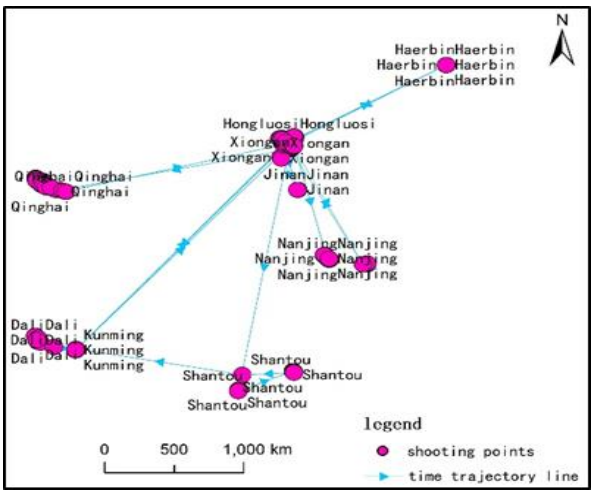


Fig.4 Line Vector Spatial Data (Timeline of Photo Points)

If an individual wishes to publish this information online, they can load the results into a network map. To load onto the National Geographic Information Public Service Platform Tianmap (website: https://www.tianditu.gov.cn/ ） For example, the data to be published online can be filtered out, exported in shp format, and compressed into a. zip compressed file. It can be loaded into the sky map through offline file upload, thereby achieving the overlay display of personal spatial data, sky map vector spatial data, and remote sensing images, and more intuitively browsing and analyzing personal past experiences, as shown in Figure 5.

Figure 5（a）

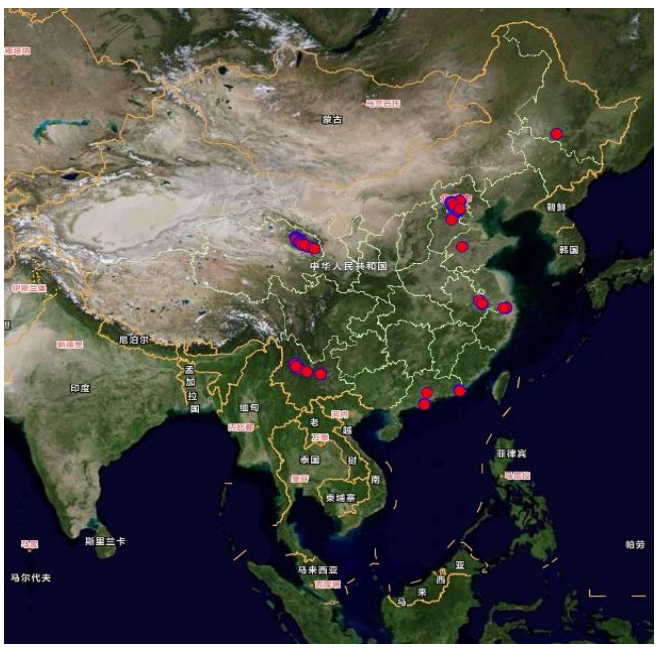


Figure 5（b）

Fig.5 Personal Photo Information Showed Online (Pink Points Are Photo Points, The Base Map Is The Online Map of TIANDITU)

Conclusion

With the development of smartphones and chip technology, mobile photography has become very convenient and easy, and the demand for photo management has gradually emerged. The method proposed in this article based on geospatial databases fully utilizes the positioning function of smartphones, extracts the location information of each photo based on Exif information, and generates spatial data to manage a massive amount of personal photos from a spatial perspective, Being able to clearly display places where individuals have visited and taken photos, and supporting a certain level of spatial analysis, is a relatively novel method.

The geospatial database established by this method is managed offline for personal photos, which is conducive to ensuring the security of photo data and personal privacy; at the same time, the constructed data resources can provide a data foundation for personal needs such as online publishing. The method takes into account both digital photos without recorded coordinate information and early film imaging photos, and proposes a solution that can integrate the spatial management of personal photos, which has promotion and reference significance. The entire technical process involved in the method can be implemented using existing general software, or independently developed software can be integrated for implementation, which is technically easy to implement and has operability; And the method can provide technical support for software development.

Acknowledgements

This work was supported by the National Key R&D Program of China(2023YFB3907600).

References

Bo Jianjie, Hu Guirang. Design and implementation of photo inquiry system [J]. Shanxi Architecture, 2018,44(14):256-258.

Cheng Pengfei, Wen Hanjiang, Cheng Yingyan, et al.,2009. Parameters of the CGCS2000Ellipsoid and Comparisons with GRS80and WGS84. *Acta Geodaetica et Cartographica Sinica*,38(3):189-194.

Han Zhigang, Kong Yunfeng, Qin Yaochen, et al.,2015 A Geospatial Representation Framework for Geo-tagged Video Objects . *Journal of Geo-information Science*, (9):1014-1021.

Kang Mengjun, Ren Fu, Cai Zhongliang, et al.,2021.Application of open source systems in GIS programming course. *Bulletin of surveying and mapping*, 0(5):145-150.

Liu Cai. Examination of Digital Photos using EXIT Information [J]. Chinese Journal of Forensic Sciences, 2010,0(5): 49-51.

Li Yao. Design and Implementation of Picture Information Acquisition Management System Based on Exif [J]. *Computer and Modernization*, 2012,0(9): 134-136.

Ma Jie, Lu Hong, Guo Yuefei. Region Based Cameraphone Image Retrieval Incorporated with Metadata [J]. *Journal of Image and Graphics*, 2007,12(10):1766-1769.

Meng Qingna. Talking about the management software Bkviewer of digital photos [J]. *Camera*,2007,0(4):58-59.

National Archives Administration of China. DA/T 50-2014, Specification for digital photos filing and management [S]. Beijing, 2014:1-6.

Pan Yuchun, Zhong Ershun, Zhao Chunjiang. , 2004. A study on update technologies for spatial database [J]. G*eo-information Science*,6(1):36-40.

Tang Weihong. Annual Report on China’s Mobile Internet Developmeng(2022) [M].*Beijing: Social Sciences Academic Press (China)*,2022:1-25.

Tian Fangjie, Cheng Chengqi, Pu Guoliang. A method for organizing and retrieving masses of photographs based on the space-time code designed on GeoSOT [J]. *Science of Surveying and Mapping*, 2018,43(2):77-82.

Wang Shanshui, Yao Hongling. Explore the modern management of photo archives [J]. *Digital & Micrographic Imaging, 2012,0(4):45-48.*

Zhang Zhanyou. Photo sorting software design [J]. *Computer Programming Skills & Maintenance*, 2015,0(19): 82-83+93.