RESEARCH ON THE CONSTRUCTION OF REMOTE SENSING AUTOMATIC INTERPRETATION SYMBOL BIG DATA

Gao Yin[1, *], Liu Ruomei[1], Liu Jianjun[1], Cheng Tao[1]

[1] National Geomatics Center of China, 28 Lianhuachi Wet Road, Haidian District, Beijing, China, 100830 – (gaoyin, liuruomei, liujianjun, chengtao)@nsdi.gov.cn

Commission III, WG III/10

KEY WORDS: Remote Sensing, Automatic Interpretation Symbol, Big Data, Geographical Conditions Monitoring, Crowdsourcing Update Mode, Open Evaluation, National Wide Application, Database Construction

ABSTRACT:

Remote sensing automatic interpretation symbol (RSAIS) is an inexpensive and fast method in providing precise in-situ information for image interpretation and accuracy. This study designed a scientific and precise RSAIS data characterization method, as well as a distributed and cloud architecture massive data storage method. Additionally, it introduced an offline and online data update mode and a dynamic data evaluation mechanism, with the aim to create an efficient approach for RSAIS big data construction. Finally, a national RSAIS database with more than 3 million samples covering 86 land types was constructed during 2013-2015 based on the National Geographic Conditions Monitoring Project of China and then annually updated since the 2016 period. The RSAIS big data has proven to be a good method for large scale image interpretation and field validation. It is also notable that it has the potential to solve image automatic interpretation with the assistance of deep learning technology in the remote sensing big data era.

1. INTRODUCTION

With the advent of the earth observation big data era, various fields such as ecological evaluation and land regulation have experienced a real demand for rapid remote sensing interpretation and field validation. Traditional field investigation has poor timeliness and high labor costs, making it difficult to implement wide range application and real time monitoring.

Remote sensing automatic interpretation symbol (RSAIS) is a ground photo with some helpful attributes that is able to provide precise field situations for image interpretation as well as accuracy validation. Recently, Peng Gong et al (2015) proposed a global wetlands validation sample dataset based on the Web of Science database. Additionally, they also created 3682 sample points of the Google Earth image, and called for building global wetland information sharing as well as a volunteer service platform. Jun Chen et al (2015) developed the world's first 30 meters land cover products known as GlobeLand30 and applied worldwide ground verification and accuracy evaluation of crowdsourcing. Those studies have vigorously promoted the process of RSAIS research, but the standardized and generalized theories and techniques for large scale and dynamic application has been lacking.

Based on the national geographical conditions monitoring mission of China, this study aims to propose a standard and streamlined RSAIS representation method. Additionally, it aims to develop a complete technical platform for massive data storage, update and quality optimization, and prepare its implementation for the Chinese RSAIS big data construction.

2. METHODS

2.1 RSAIS data characterization method design

The traditional interpretation symbol usually possesses less attributes and is difficult to be applied in spatial region matching and automatic image interpretation. To meet the requirements of precise location matching, automatic interpretation and large scale application, a characterization system with refined attributes was designed, and the detail attributes are indicated in Table 1. The schematic diagram of ground geometric record and region matching are depicted in Figure 1. This method has met the requirements from two aspects: (1) reflects land cover with accurate geometric identity, (2) can be easily recorded and has potential for large scale automatic acquisition.

Attribute	Content and acquisition method
photo	in-site photo of target region: in-site recording
type	land cover type: in-site or indoor recording and classified, known type or content tag
time	shooting time: automatic recording
longitude	longitude of shooting location: automatic recording from *GPSLongitude* information of photo EXIF file, from -180 to 180 degree
latitude	latitude of shooting location: automatic recording from *GPSLatitude* information of photo EXIF file, from -90 to 90 degree
azimuth	angle from shooting direction to north direction: automatic recording from *GPSImgDirection* information of photo EXIF file, clockwise from 0 to 359.99 degree
focal	35 mm equivalent focal length: automatic recording from *FocalLengthIn35mmFilm*

	information of photo EXIF file,, tablet or mobile phone
distance	distance between shooting location and target region: semi-automatic recording based on online map marking

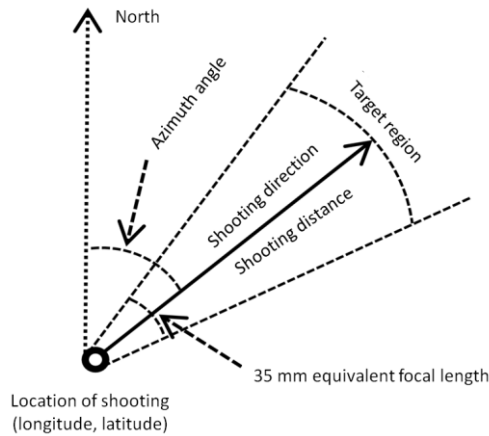Table 1: Attributes designed for RSAIS recording

Figure 1. Schematic diagram of RSAIS geometric parameters and automatic image position matching

Throughout the process of actual data acquisition, the key criterion is to make the RSAIS more typical to represent local land cover characters. All the RSAIS data should be able to represent the characters of its land cover, and comprehensively reflect the general characteristics of the regional land cover types. In a certain area where the distribution of RSAIS data and real land cover are relatively consistent, the total count of acquired data should represent the overall characteristics of the region. This is not to mention that distribution should closely align with the trend of the land cover type distribution.

2.2 Distributed and cloud data storage architecture

For massive data storage, the latest elastic architecture consisting of distributed database and cloud storage method was designed. All RSAIS photos are stored in the distributed cloud storage file system, and attributed information are formatted and stored in the distributed database. The photo and attribute of each RSAIS is associated with a unique identifier. All files are considered as redundant backup storage based on cluster, which could support elastic expansion and concurrent access. The subdivision grid spatial index of GeoSOT is applied for high efficient spatial retrieval.

2.3 An offline and online update mode

An offline and online update mode is implemented for the data flexible update. In some national or regional territories or resources investigation projects, the RSAIS can be updated with offline batch update mode, which is conveniently made for large-scale as well as all other types of productions and updates. In addition, a web based crowdsourcing update method is adopted for global, dynamic, and continuous updates. With sufficient volunteers, the RSAIS data of most regions and all land types can be updated with an acceptable frequency, which will automatically promote RSAIS big data.

2.4 A dynamic evaluation mechanism

A dynamic evaluation mechanism is applied as a means to improve the quality of while also promoting the value of RSAIS big data. Users are allowed to tag RSAIS online and openly score the RSAIS they used from several key aspects, such as correctness, typicality, photo quality, attribute quality and comprehensive evaluation. The evaluation data will build an open RSAIS evaluate library on the server side. The scores or tags of the updated RSAIS will be updated regularly, and RSAIS data with poor quality and accuracy will be reflected in a lower priority.

3. APPLICATION

3.1 Workflow of RSAIS big data construction in China

The National geographic conditions monitoring is an important new mission of China and it is vital work in obtaining access to fundamental natural, ecological and human activities information of land surface. The basic goal is to acquire annual and nationwide land cover datasets from remote sensing images with resolution that is greater than 1 meter. Through this project, a national wide RSAIS database covering 86 land types of China was constructed and updated annually and dynamically. The workflow of RSAIS big data construction in China is depicted in Figure 2.
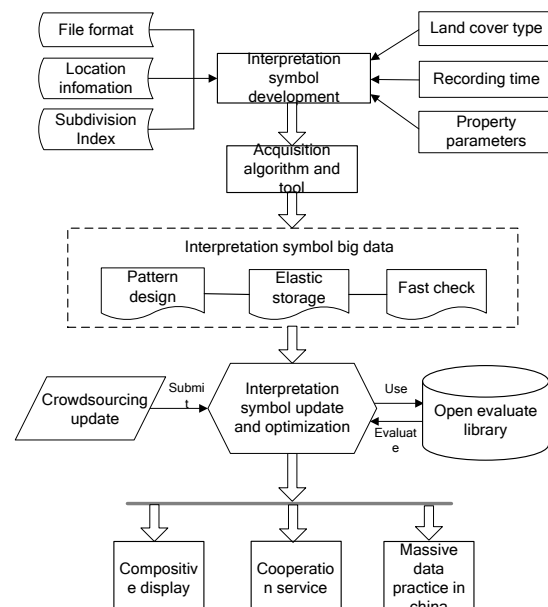

Figure 2. The workflow of RSAIS big data construction in China

3.2 National wide RSAIS big data statistic

The national RSAIS big data of China was established with field photos from the 2013 to 2015 period, which consists of more than 3 million samples and covers 86 land types of the entire mainland. The total data size is about ten TB. All the samples had been widely applied in image interpretation and national wide ground validation. The sample count of the 31 provinces is depicted in Table 2.

Province	Sample Count	Province	Sample Count
Beijing	20987	Hubei	113253

Tianjin	5836	Hunan	81693
Hebei	106456	Guangdong	154180
Shanxi	103974	Guangxi	75446
Neimenggu	223663	Hainan	31785
Liaoning	92977	Chongqing	66734
Jilin	36813	Sichuan	217403
Heilongjiang	136054	Guizhou	70013
Shanghai	5574	Yunnan	256408
Jiangsu	207800	Xizang	36075
Zhejiang	73176	Shanxi	132052
Anhui	52064	Gansu	103783
Fujian	60992	Qinghai	35583
Jiangxi	107275	Ningxia	39640
Shandong	106161	Xinjiang	180475
Henan	99496	**Total**	**3033821**

Table 2: Sample counts of China RSAIS samples across 31 provinces

3.3 Massive RSAIS data display in two-dimensional and three-dimensional system

In order to display an accurate representation of the massive RSAIS data, both two-dimensional and three-dimensional display systems based on distributed parallel scheduling technology were built. The spatial distribution of some RSAIS data and photo samples in the two-dimensional display system are shown in Figure 3, and some other RSAIS data in three-dimensional display system are shown in Figure 3.

Figure 3. The spatial distribution of the RSAIS data and photo sample in two-dimensional display system
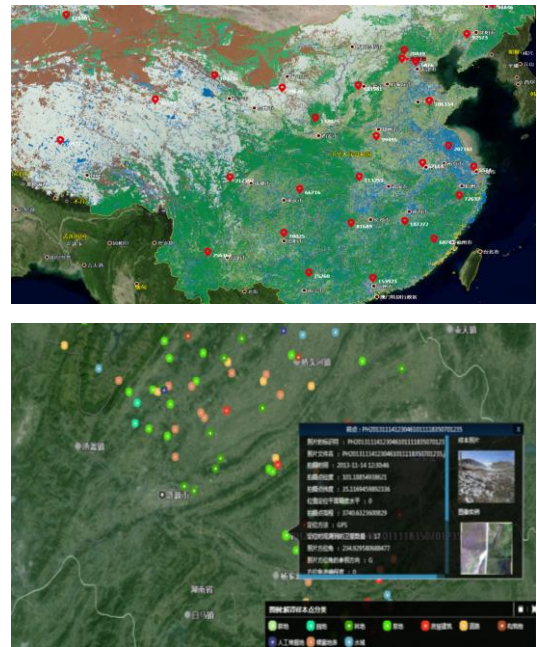

Figure 4. The spatial distribution of the RSAIS data and photo sample in three-dimensional display system

3.4 Crowdsourcing update and open evaluation platform

Since China owns a wide scale of land area, the geographical conditions are quite complicated, and landscape has reflected extremely varying regional differences. Consequently, a web based crowdsourcing update prototype platform was built for the open RSAIS dynamic update. Users are able to submit typical RSAIS data as a means to update and challenge the RSAIS data for national wide discussion, improving the comprehensive and representative aspects of Chinese RSAIS big data. In addition, an open evaluation prototype platform is built for data sharing and evaluation information submitting, which is quite helpful for RSAIS optimization. The interfaces of the crowdsourcing update and open evaluation platform are shown in Figure 5.

Figure 5. The interface of the crowdsourcing update and open evaluation platform

3.6 Potential value of RSAIS big data

In the era of remote sensing big data, the RSAIS big data provides prior field knowledge support for image automatic interpretation, and possesses the basic potential for breaking through the worldwide image automatic interpretation problem from the mechanism. In addition, a number of case studies are conducted to provide scientific reference for expanding the application scenario of the RSAIS big data. It is possible to unearth a large amount of valuable information using big data calculation and analysis.

Reflect the links between land cover types and the ground field terrain feature: In the absence of field work, the RSAIS data can help to accurately distinguish some confusing land cover types from image.

Reflect the land cover types of similar geographical environment area: In the adjacent area or similar geographical areas where it is difficult to reach the field work, image interpretation can be implemented based on the similar spectral, texture and geographic correlation analysis.

Reflect the land spatial distribution and diversity characteristics: Based on spatial analysis and statistical calculation, it is possible to obtain the spatial distribution and density characteristics of some land covers between different administrative divisions all over the country.

Reflect the spatial-temporal characteristics of specific vegetation: The RSAIS data is captured from different areas and times, and it is easy to analyze the morphological and image features of a specific land type such as broadleaf arbor in different seasons.

4. CONCLUSION

The RSAIS representation, storage, update and optimization techniques that are adopted through the practices in China have shown an efficient and scientific method for RSAIS big data construction. The dynamic updating RSAIS data of China is considered to hold important knowledge of land covers and supports an automatic image interpretation method that was developed based on multi-scale segmentation and deep learning classification. In our following work, prior national wide knowledge reflects potential in being able to solve image automatic interpretation with the assistance of deep learning technology.

Along with the national wide monitoring project conducted on a yearly basis, the RSAIS big data will gradually evolve into a spatial-temporal field prior knowledge big data, which contains a lot of field land covering knowledge across the country. This will also lay the foundation for the natural resources survey as well as for the geographical environment analysis and information mining.

REFERENCES

Y Zheng, Z Niu, P Gong, J Wang. A database of global wetland validation samples for wetland mapping. *Science bulletin*, 2015, 60(4): 428-434.

J Chen, J Chen, A P Liao, X Cao, L J Chen, X H Chen, C Y He, G Han, S Peng, M Lu, W W Zhang, X H Tong, J Mills. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, 103: 7-27.

D R Li, H G Sui, J Shan. Discussion on key technologies of geographic national conditions monitoring. *Wuhan Daxue Xuebao*, 2012, 37(5): 505-512.

Z H Gao, X Zhou. Cause analysis and accuracy improvement of confusing land cover types in China geography census. *International Conference on Agro-geoinformatics*, 2014: 1-5.

T Jiang, Construction and Application of the visual interpretation bank of remote sensing image. *Geospatial Information*, 2010, 31-33.

T Chen, R J Yuan, Z H Gao, Y Gao, X M Shi, Integrative neatening method of sample data for remote sensing image interpretation. *Geomatics World*. 2014, 21(5): 96-99.

T Chen. Database construction and its application of sample big data collected in remote sensing image. *Computer Systems & Applications*, 2017, 26(5): 43-48.

Maria Antonia Brovelli, Monia Elisa Molinari, Eman Hussein, Jun Chen, Ran Li. The first comprehensive accuracy assessment of globeland30 at a national level: methodology and results. *Remote Sensing*. 2015,7(4): 4191-4212.