

APPLICATION OF DATA FUSION IN THE PRODUCTION AND UPDATING OF SPATIAL DATA

Huanxin Chen^{a*}, Qun Sun^a, Li Xu^a, Ziming Xiong^a

^a Institute of Surveying and Mapping, Information Engineering University, 450052 Zhengzhou, China -
chx1557@163.com

KEY WORDS: Spatial Data, Data Fusion, Image, Data Fusion Scheme, Data Integration, Production, Updating

ABSTRACT:

The increasing spatial data provide abundant material for data fusion, and the purpose of the paper is to apply data fusion into the production and updating of spatial data. After outlining the general framework and workflow, the processing contents and methods are specified in sequence. Facing various spatial data from different sources, how to design proper data fusion scheme is the top-priority problem. The method of analyzing and assessing various spatial data is introduced referring to images, which is shown by concrete examples. Then the technical workflow of multi-source data integration is present to eliminate differences and relevant contents are also specified. After building the relationships of homologous entities through spatial data matching, the data fusion which is similar to cartographic generalization in essence can be implemented. Different ways of updating spatial data is introduced to keep the currency of existing data. At last, the spatial data with good quality can be obtained. The efficient and reliability of the methodology in this paper has been proved through practical production.

1. INTRODUCTION

1.1 Motivation

Being the blood of the geography information system (GIS), spatial data is the basement of the digital and information building. So the production of the spatial data never stops, and series database of different scales have been established. The currency of those databases would fade with time lapse if they cannot be updated timely. So the production and updating of the spatial data will be the central work all the time.

With the help of GPS, GNSS and remote sensing (RS) image of high resolution, spatial data with good accuracy and currency can be acquired easily. Facing the increasing of big data, how to use them properly becomes a new problem. The technique of spatial data fusion is just the answer for the scientific use of existing data. Since different kinds of data are the abundant material of data fusion, then better data can be derived from the existing data through data fusion.

1.2 Aims

To apply data fusion technique into the production and updating of spatial data using existing multi-source data, including: scan map, the data of built spatial databases, RS image, relevant statistic and literature, GPS data and so on. And the final goal is to obtain better new data which absorbs the advantages of different multi-source data.

1.3 Related Work

For the production of the spatial data, the primary method can be summarized as vectorization which vectorizes the paper map to get vector data. And hand followed digitising using manual tracking digitiser is just one kind of those methods in the early

stage. With the upgrading of computer, the map is scanned and then vector tracking can be operated semi-automatically with the help of intellectualized man-machine interaction, and relevant techniques include: map element segmentation, mathematical morphology, statistic pattern recognition, artificial neural networks, and syntactic pattern recognition and so on (Zhili Huang, 2005). For spatial data updating, RS image is widely used to update the currency of existing vector data. And in most conditions, data updating just work between vector data and raster data which is mainly used to update the geomorphic data (Ching-Chien Chen, 2006), but the attribute data cannot be guaranteed.

Although some research has been done for the fusion of different vector data or different raster data, applying the data fusion to the production and updating of spatial data at the same time is still a new try.

1.4 Overview

The paper presents a methodology of spatial production and updating using data fusion technique. In order to have a better understanding of this methodology, the paper is arranged as follows. Section 2 draws the general framework and workflow, and then the relevant steps are specified separately. Section 3 explains how to design proper data fusion scheme through analyzing and assessing various multi-source spatial data referring to images. Section 4 introduces the content and technical workflow of data integration in detail. Section 5 specifies the processing of data fusion and shows different ways of updating existing spatial data. At last Section 6 closes the paper with conclusions and recommendations for the further research.

* Huanxin Chen, Ph.D candidate, majors in digital mapping and spatial data fusion.

2. THE GENERAL FRAMEWORK AND WORKFLOW

Before outlining the general framework and workflow, the integrative of map production and spatial data capturing should be introduced briefly. The integrative model produces the spatial data firstly, and then realizes the map making through map visualization (Haiyan Liu, 1999). This model can obtain spatial data and map in one workflow which is the best map production model at present. Based on the integrative model, Figure 1 presents a schematic picture of the general framework and workflow.

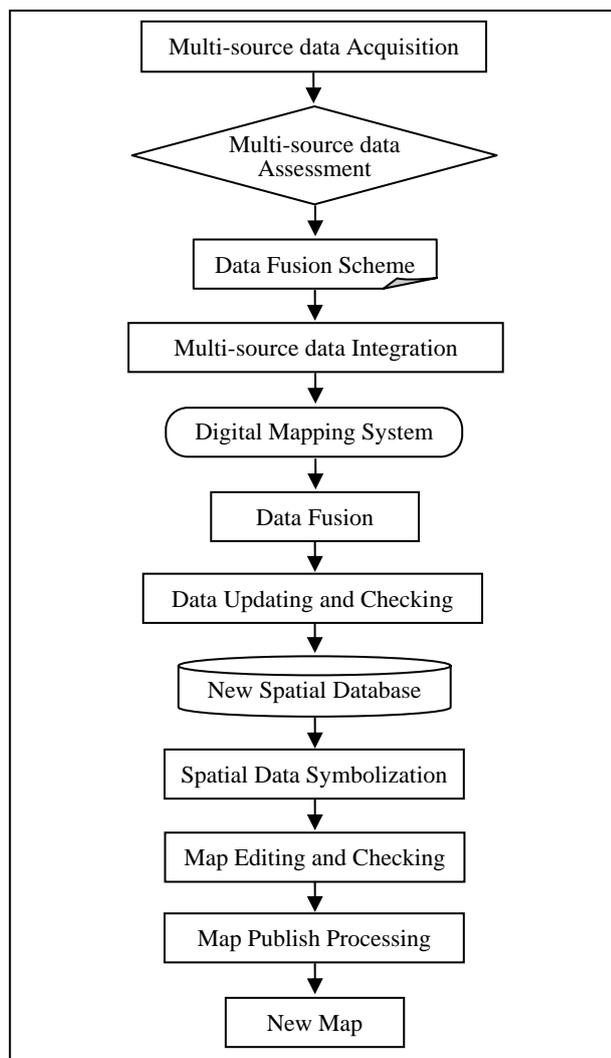


Figure 2. The general framework and workflow

Firstly the source data, statistic and literature should be collected. Secondly those data should be analyzed and assessed in order to design proper and scientific data fusion scheme. Thirdly the selected multi-source data are gathered into the digital mapping system after data integration. Fourthly the data fusion is operated according to the previous designed scheme. After data updating and checking, the derived data are storage into the new databases. All of steps above are to get new spatial data. Based on the new data, a further step can be done to obtain corresponding new map. And the relevant processing including: symbolization, map editing and checking, and map publish. The detailed content and technique of the main steps would be described in the following parts.

3. DATA FUSION SCHEME DESIGN

Since the data fusion scheme would influence the efficiency of following steps and the final data quality. So the scheme design is the key step in the whole workflow. Facing various spatial data from different sources, how to design proper and scientific data fusion scheme is a challenge. After introducing the kinds of multi-source data, the paper would specify the way of scheme design.

3.1 The Kinds and Uses of Multi-sources Data

3.1.1 Published Paper Map: The published paper map is the basic source of multi-source data, including: series scale maps, the geographical map, administrative map, circulation map, thematic map and atlas. The databases are mainly built through vectorizing these paper maps. For the region which has no vector data, the way of raster to vector from paper map is still a feasible method.

3.1.2 The Built Spatial Database: The series scale's basic geo-spatial databases which have been built are the main source of data fusion. At the same time, different kinds of thematic data are also the important material, such as traffic data, administrative data, and vegetation data and so on. Because those databases are built in early time, the accuracy and currency of the data would decrease with time lapse. In order to guarantee the quality of the built databases, real-time updating mechanism should be set up.

3.1.3 The Remote Sensing Images: Modern detecting technology offers abundant data which can be used to improve the currency and reliability of spatial data, and the remote sensing image is one of the means which is widely used in spatial data updating. The Soviet Union uses the RS image of Peace to update the topographic map of 1:50000 -1:100000. Norway uses the SPOT image to revise the 1:50000 traffic map. France uses SPOT data in the mapping of 1:100000 topographic map and America uses Landsat data to revise the medium-scale and small-scale map. China also uses TM image to update the data of 1:250000, and use SPOT image to update the data of 1:50000 and 1:100000. With the start-up of high resolution detecting project, the application of remote sensing image is not only the main way of geospatial information updating, but also feasible for the spatial data production of large scale.

3.1.4 GPS Data: Being the new technique of positioning data acquisition and updating, Global Position System has the advantage of high-precision, low cost and high agility. GPS data can be used to set up geodetic control network, leveling network and other spatial datum, and assistant positioning data for the acquisition of the spatial data. And the vehicle-mounted GPS data is also can be used to the acquisition and updating of traffic data.

3.1.5 Field data: Although field data is a traditional way, it's also a source of geo-spatial information. The total station instrument, theodolite, distancemeter and other instruments are set on the known points in field work, and then figure out the three-dimensional coordinates of the target points through surveying the direction, distance and altitude difference between target points and known points.

3.1.6 Relevant Statistic and Literature: The data mentioned above are used to get the geometrical information, while the statistic and literature are used to get attribute information. For example, national administrative code and gazetteer can be used to update the attribute data of administrative and settlement feature.

3.2 Data Analysing and Assessment

In order to design proper and scientific data fusion scheme, the work of data analyzing and assessing is rather essential and should be done firstly. Since the spatial data can be divided into two kinds: vector and raster, the methods of analyzing and assessing would be specified respectively.

3.2.1 Raster Data: The imaging time and resolution are two key factors to be considered. For the imaging time, the latest the better. While it is not same for the resolution, high resolution does not feasible any time. Select the images whose resolution is close to the target map resolution. While for the scan map, besides having the scale, map type, production department and other relevant information, select the credible and authoritative map.

3.2.2 Vector Data: The data integrity, accuracy, currency and consistency are the key factors for the vector data. Comparing with raster data, the analyzing and assessing of vector data are more complicated. The current method is visualizing different kinds of vector data in the same system, and put the image with good currency at the bottom lay as reference. Then analyze the geometry accuracy and currency referring to the image, check the consistency, and asses the attribute data by querying the attribute information.

3.3 Data Fusion Scheme Design

The essence of data fusion is deriving better data from the existing data by absorbing the advantages of each data. So in the process of the scheme design, each kind of data should be considered and compared with others. Only based on the thorough comparison and analyzing, can proper and scientific data fusion scheme obtained. At present, this work relies more on human judging and expert decision. In order to expatiate vividly, the design scheme of traffic feature is shown as follows.

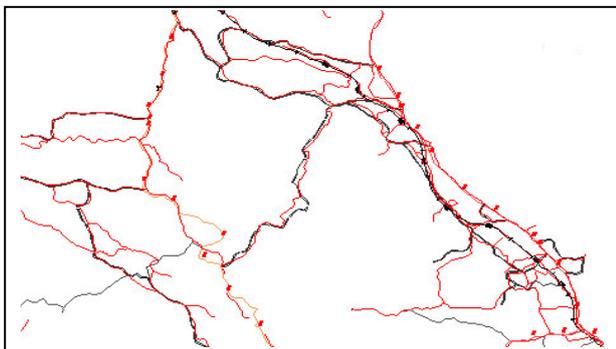


Figure 2. The overlay of three different traffic data

Figure 2 is the overlay of three different traffic data which are symbolized by red, yellow and black respectively. Generally speaking, the red data is comparatively dense which has a better integrity. But it is still early to draw the conclusion. After overlay the latest image at the bottom later, the dark data has better accuracy and currency than the red one which can

judging from the better coherence with the image. So for the geometry data, dark data is used firstly, while for the parts which dark data is absent the red data is selected as complementarities. For the attribute data, the scheme is designed through querying and comparing the attribute information of each kind of data. If the dark data's attribute data is poor, in-depth work should be done to fuse the good attribute data into the dark data. Just in this fusion way the final derived data would have the merits of good integrity, accuracy and currency.

4. MULTI-SOURCE DATA INTEGRATION

The multi-source data are different in source, acquisition approach, data model, data format, coordinate and projection, semantic expression and other aspects. So the multi-source data can't be used to do data fusion directly. The task of multi-source data integration is to eliminate the differences among multi-source spatial data. According to certain standards, those multi-source data are integrated into one uniform system, through which the direct management and structure of multi-source data can be realized.

4.1 Content and Techniques of Multi-source data Integration

4.1.1 Unified Coordinates: Different data use different coordinates and the current coordinates in using includes: old BJ54, new BJ54, CD80, WGS84 and CGCS2000 coordinate. The using of CGCS2000 would influence the geometry precision of data in old coordinates, so unified coordinates is necessary. And the coordinates can be transferred through three parameters or seven parameters models.

4.1.2 Projection Transformation: According to the mapping purposes and locations, the existing multi-source data use different projections. It is necessary to set up the corresponding relationship of the points from different plane through projection transformation. For vector data, the projection can be transferred through certain equation. But for raster data, the new image can't be built through point-by-point conversion, because point-by-point conversion may lead to slot and overlap in the new image. So the projection transformation of image is through reversed calculation (Shaomei Li, 2004).

4.1.3 Data Format Exchange: Different software and system use different data models to describe the world, which leads to different data formats. The common used data models are: Arc/Info, MapInfo, MapGIS, AutoCAD, SuperMap and so on. Each data model has its own data format and the exchange methods are ripe.

4.1.4 Data Compression: For the GPS data, the sampling interval is short, so large numbers of redundant points exist in GPS data, which makes the lines not smoothing after symbolization and hard to edit. So the redundant points need to be compressed, simplified and generalized according to the standards of certain scale. The common compression algorithms includes: Ramer-Douglas-Peucker and vertical distance.

4.1.5 Image Enhancement: For the remote sensing images, image enhancement is to reduce or give prominence to the feature of some elements. Image enhancement includes: gray level transformation, histogram transformation, principal component analysis, spatial filtering, texture analysis, image correlation and so on. Different method should be chose according to the practical situation. For the scan map, some macula, holes, break lines, burr and other noise exist after scanning. These noises can be processed through different templates in order to the easy implement of vectorization and map tracing.

Besides the processing listed above, there are still other contents, such as semantic transformation, map matching,

geometry rectification and so on. Based on those integrative processing, the multi-source data integration can be developed.

4.2 Technical Workflow of Multi-source data Integration

In the processing of multi-source data integration, the geometry data and attribute data are organized separately: The former is stored in the design document, and the later is managed by the database. The geometry data and attribute data of the same feature are the one-to-one relationship. And the attribute data are divided into two parts: new database and additional database, and they are connected through index table. The technical workflow of multi-source data integration is shown by Figure 3.

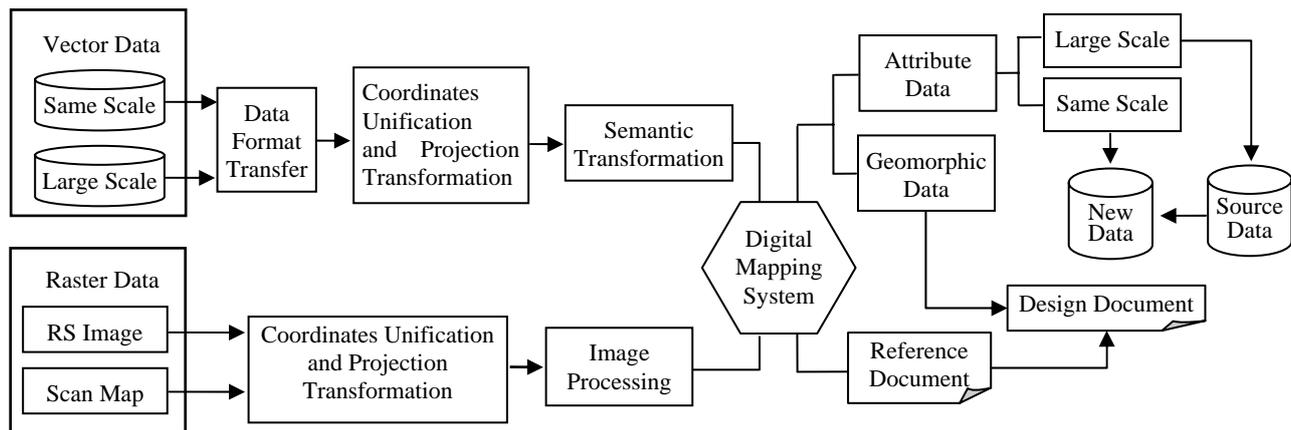


Figure 3. Technical workflow chart of multi-source data integration

For the vector data, both data of the same or the large scale are processed through the procedures of data format transfer, coordinates unification and projection transformation, semantic transformation, and then the processed vector data are integrated into the uniform digital mapping system and symbolized. The geometry data are written into the design document and built the corresponding relationship with the record in the database. The attribute data of the same scale are stored in the new database directly. While the attribute data of the large scale are stored in the additional database as the alternative, which can be extract into the new database according to the data fusion scheme.

For the raster data, after processing of coordinates unification, projection transformation and image processing, the processed raster data are integrated into the uniform digital mapping system and overlaid with the symbolized geometry data in the form of reference document.



Figure 4. The chart of multi-source data integration

The chart of multi-source data integration for some region is shown in Figure 4. The yellow one is the vector data of large scale, and the colorized one is the vector data of the same scale. While the raster data at the bottom is the image.

5. DATA FUSION AND UPDATING

The multi-source data can be organized and managed congruously after integration. Then the next workflow turns to data fusion and updating. A successful matching of the same feature from different datasets is an important step for data fusion and updating. So this section would expatiate in three parts: data matching, data fusion and data updating.

5.1 Data Matching

Different data has different ways of expression, and data matching is just to identify the same feature in different data sets through similarity measurement of geometry, semantics and topology and build the corresponding relationships (Feng Xu, 2009). The semantic similarity measurement is the best way ideally, but it is unfeasible since different datasets use different data models and expressions which lack unique label attribute information. So the general used factor is geometry similarity in the first step, and then the topology similarity is usually referred as assistant to detect and find the same feature in different datasets.

For different kinds of geographical entities, the corresponding matching algorithms are different. The point feature is comparative simple which is through location, structure and semantic similarity. For linear feature, the factors of location,

shape, dimension, direction, and topology are considered synthetically (Xiaoya An, 2011). While the areal feature seldom uses topology comparing to the linear feature. But the accurate rate of the matching algorithms at present is still has a long way to go. Once the data matching is finished, the work of data fusion is relatively simple to develop.

5.2 Data Fusion

Multi-source data integration eliminates differences and data matching builds the corresponding relationship of the same feature in different data sets, these processing don't change the existing data qualitatively and still keep their own characters as before. The data fusion is just the qualitative change step, which gathers the advantages of existing data extremely to obtain better new data through certain processing. Different motives lead to different content and processing of data fusion. For the spatial data production, the essence of data fusion can be regard as cartographic generalization, since the relevant processing of data fusion can be found in cartographic generalization.

5.2.1 Data Extraction: In the data fusion, the data of different feature layers comes from different datasets. The job of data extraction is to extract the corresponding data from the existing data sets into the new database according to the data fusion scheme. The processing of data extraction is similar to selection in the cartographic generalization.

5.2.2 Data Simplification and Generalization: For the extracted data, the side line of habitation, water area, vegetation and other features, especially for the data of the large scale, need to be simplified according to the mapping standards of final data. Beside that the attribute data also needs generalization to describe quantitative and qualitative characters of the feature properly. The content mentioned above corresponds to the simplification and generalization of cartographic generalization.

5.2.3 Relation Coordination: Since the data comes from different datasets, the logical inconsistency may occur in the data fusion. The purpose of relation coordination is to regulate the relationships of different kinds of features to produce correct topology. Some processing of relation coordination is displacement of cartographic generalization.

Most processing of data fusion can find corresponding content in the cartographic generalization, and they even follows the basic theory and methods of cartographic generalization, only the operation environment changes. It seems that the processing mentioned above aims at geometry data, but it also includes attribute processing in data fusion, such as the classification conformity, combination of attribute items, unification of expression modes and so on.

5.3 Data Updating

With the accomplishment of building multi-scale spatial databases, the updating of these databases will replace the task of spatial production, becoming the next main content of the spatial study. The bulletin of building and updating of national basic database reports that: the task of updating national basic geo-spatial databases is comprehensive utilizing different sources of current information to measure and confirm the positional and attribute information changes of basic geographic elements, recording the changes and releasing the new edition data (Jun Chen, 2004). Although some processing can be

operated semi-automatically, such as map-image registration, change detection, feature extraction and so on, the methods mainly used are visual interpretation, manual editing and other kinds of man-machine interaction.

5.3.1 Updating based on RS Images: RS images are widely used to update spatial data. Overlay the symbolized vector data on RS images and then revise the positional and attribute information of the built spatial databases.

Beside the resolution and imaging time, different ways of image fusion needs to be considered according to the feature types. For the multi-band TM images, the common used ways of fusion between different bands includes: the fusion of 7-4-3 bands resembles true colour image and usually used to update habitation and water system; while the fusion of 4-3-2 bands is the false-colour image which shows gradation well. And the vegetation is expressed in red so it is used to the investigation of land using and resource environment; the fusion of 4-5-3 bands emphasizes the water features which shows the boundary of water and land distinctly, and even the configuration of block can be interpreted easily. So it is usually used to update coast line, ditch, and habitation.

5.3.2 Updating based on GPS Data: Generally speaking, the roadway data of vehicle-mounted GPS is better than recorded data, so GPS data is used to update roadway data. Figure 2 is the example of roadway updating: the single line is the old roadway; while the crewel is the GPS data, and the point symbols are the toll stations or other accessorial facilities.

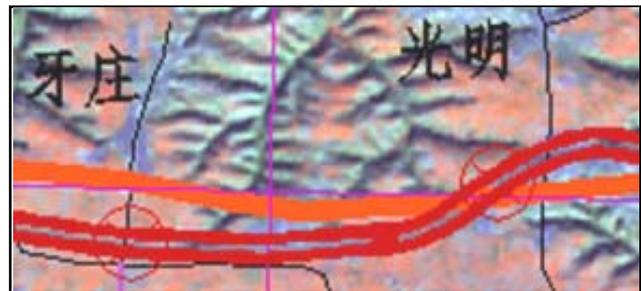
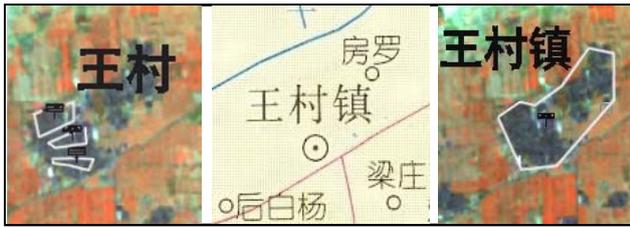


Figure 5. Roadway updating with GPS data

5.3.3 Updating based on comprehensive utilization of multi-source data: Since the updating includes both geometry and attribute, multi-source data are used comprehensively in practice. Only by the way of adducing and analyzing mutually can guarantee the data quality. Figure 3 is the example of habitation updating: the left figure is the overlay of old vector data and RS image, which shows the geometrical contour of the habitation needs to be modified; the middle figure is the administrative map which shows the name of the village changes. And then though referring to relevant statistic and literature it can be found that: with the development of the town, not only the region size extends, but also the population, administrative grade and other attributes change. So both the geometry and attribute data need to be updated; and the right figure is the final updating with the comprehensive utilization of multi-source data.



Huang Zhili, Fan Yangyu, Hao Chongyang, 2005. Interactive vectorization for chromatic scan map. *Computer Application*, 25 (3), pp. 577-579.

Figure 6. Habitation updating with comprehensive utilization of multi-source data

6. CONCLUSION

The production to satisfy different kinds of demands and the updating of the built databases are the two long-term tasks of spatial data. Based on abundant multi-source data, the paper applies data fusion into the production and updating of spatial data, which can gather the advantages of existing data to get better new data. This method has been used in the spatial data production and updating of medium-scale and small-scale. The practice proves that this method is more efficient and reliable than the traditional ways. While there is still a long way to go to perfect this method, such as intellectualizing the assessment and analysing of multi-source data, improving the accurate rate of data matching, enhancing the automation of data updating and so on. With the development of relevant techniques and spatial data increase, the way of applying data fusion in the production and updating of spatial data would be used widely.

7. ACKNOWLEDGEMENTS

The authors acknowledge funding support for this research from National "863" Plan (2012AA12A404) and the National Natural Science Fund (41201391, 41071297).

8. REFERENCES

- An Xiaoya, Sun Qun, Xiao Qiang, 2011. A shape multilevel description method and application in measuring geometry similarity of multi-scale spatial data. *Acta Geodaetica et Cartographica Sinica*, 40(4), pp. 495-501.
- CHEN Jun, LI Zhilin, JIANG Jie, ZHAO Ren-liang, 2004. Key issues of continuous updating of geo-spatial databases. *Geomatics World*, 2 (5), pp.1-5.
- Ching-Chien Chen, Craig A. Knoblock, Cyrus Shahabi, 2006. Automatically conflating road vector data with orthoimagery. *GeoInformatica*, 10, pp. 495-530.
- LI Shaomei, 2004. Study on digital relief shading theories and technologies, Information Engineering University, CHINA
- LIU Haiyan, 2002. The theory and technology of integration of map production and spatial data capturing. Information Engineering University, CHINA.
- Xu Feng, Deng Min, Zhao Binbin, Chen Jian-jun, 2009. A detailed investigation on the methods of object matching. *Journal of Geo-Information Science*, 11(5), pp. 657-663.