Study for Multi-Resources Spatial Data Fusion Methods in Big Data Environment

Zhiquan Huang, Yu Fu & Fuchu Dai

To cite this article: Zhiquan Huang, Yu Fu & Fuchu Dai (2017): Study for Multi-Resources Spatial Data Fusion Methods in Big Data Environment, Intelligent Automation & Soft Computing, DOI: 10.1080/10798587.2016.1267237

To link to this article: http://dx.doi.org/10.1080/10798587.2016.1267237

Published online: 25 Feb 2017.
Study for Multi-Resources Spatial Data Fusion Methods in Big Data Environment

Zhiquan Huang\textsuperscript{a,b}, Yu Fu\textsuperscript{a} and Fuchu Dai\textsuperscript{a}

\textsuperscript{a}North China University of Water Resources and Electric Power, Zhengzhou, China; \textsuperscript{b}Xinjiang Institute of Engineering, Urumchi, China

\textbf{ABSTRACT}\textbf{ABSTRACT}\textbf{ABSTRACT}\textbf{ABSTRACT}

The rapid development and extensive application of geographic information system (GIS) and the advent of the age of big data bring about the generation of multi-resources spatial data, which makes data integration and fusion share more difficult due to the differences on data source, data accuracy and data modal. Meanwhile, study for multi-resources spatial data fusion methods has an important practical significance for reducing the production cost of geographic data, accelerating the updating speed of existing geographical information and improving the quality of GIS big data. To expound the formation and developing trends of multi-resources spatial data fusion methods systematically, and on the basis of referring to lots of related technical documents both at home and abroad, this paper makes a conclusion and discussion about multi-resources spatial data fusion methods, and foresees the prospects of data fusion in big data environment, which has certain reference value for the related research work.

\textbf{1. Introduction}\textbf{1. Introduction}\textbf{1. Introduction}\textbf{1. Introduction}

With the advent of the age of big data and the extensive applications of 3S (remote sensing, geographic information system and satellite positioning) technique in all walks of life, there is a growing demand for spatial data. Extracting the richer, more reliable and useful information from data rather than single-source data, needs merge spatial data with different data sources, standards and models. Therefore, data fusion problem has been produced. Essential characteristics of data fusion on the one hand reflected in the input information of multi-source, on the other hand reflected in comprehensive application of a variety of information processing technology. Spatial data includes remote sensing images, topographic maps, GIS data and GSP data, also can be undivided into raster and vector in the format.

Multi source spatial data fusion is a multi-platform, multi-level, multi-sensor, multi-temporal, multi-spectral, multi-angle and multiple resolution spatial data fusion, because multi-resources spatial data exists the differences on multiple semantic natures, multiple-scale features and storage format, and there are differences in data models and storage structures. These data characteristics of the age of big data would bring about great difficulties for data share and data fusion between GIS departments. The study for multi-resources geospatial data fusion methods has vital practical significance for reducing the production cost of geographic data, fastening the updating speed of existing geographical information and improving the quality of GIS big data. The paper mainly discussed remote-sensing image fusion, data fusion between GIS vector data and remote-sensing image data, and the author puts forward his own opinions. Finally, some conclusions and considerations about further work are given in Section 6.


Remote-sensing image fusion technology is the process of pre-processing the remote-sensing image data with same or different phase of the same object from different sensors and then using fusion algorithm to obtain remote-sensing images with more useful information. Data fusion of remote-sensing image can optimize the information about remote-sensing image with different sources, search for useful information and reduce the uncertainty, incompleteness and errors in the aspect of studying object cognition.


The data fusion methods of remote-sensing image can be classified into three types: Pixel-level fusion, feature-level fusion and decision-level fusion (Pau, 1988; Pohl & Genderen, 1998). The fusion algorithm based on pixel-level is generally used at
the moment. There are many pixel-level fusion methods (Jia, 2005; Mao, Li, Yang, & Ohmi, 2011; Zhao, 2009). The following will introduce these common pixel fusion methods.

(1) HIS Transform Fusion Method

While in the remote-sensing image fusion, there is a need to transform RGB space to HIS space in which data with different resolutions are recombined. That is HIS transform fusion (Mei & Peng, 2001; Varshney, 1997). HIS transform fusion model was put forward by Harrison and Jupp in 1990. Because it can accurately and quantitatively describe the characteristics of each component, it applies extensively in the process of processing digital image.

What the HIS transform fusion adopts in this paper is frequently-used cylinder transform (Carper, 1990; Yuan, 2005), and the specific formula is as follows:

Positive transformation formula

\[
\begin{bmatrix}
I \\
v_1 \\
v_2
\end{bmatrix} = \begin{bmatrix}
\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\
\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\
\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & 0
\end{bmatrix}
\begin{bmatrix}
R \\
G \\
B
\end{bmatrix}
\]

\( S = \sqrt{v_1^2 + v_2^2} \) (2)

\( H = \tan^{-1} \frac{v_1}{v_2} \) (3)

Inverse transformation formula

\[
\begin{bmatrix}
R_{\text{new}} \\
G_{\text{new}} \\
B_{\text{new}}
\end{bmatrix} = \begin{bmatrix}
\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\
\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\
\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & 0
\end{bmatrix}
\begin{bmatrix}
I \\
v_1 \\
v_2
\end{bmatrix}
\]

Where I is transformed component and \( v_1 \) are intermediate variables introduced for calculation. \( R, G \) and \( B \) are pixel gray values of image before transformation, while \( R_{\text{new}}, G_{\text{new}} \) and \( B_{\text{new}} \) are pixel gray values of image in RGB space after the fusion.

HIS transform fusion is one of the most common methods in image fusion. This fusion method can basically maintain the spectral features of multispectral image before fusion and fuse geometric information with high resolution. It improves the spatial texture characteristics and strengthens the spatial detail performance of multispectral image, but there still exist some shortcomings. When improving geometry resolution, it will lose partial spectral information of origin images and only fuses three wave bands of multispectral image at the same time.

(2) Brovey Transform Fusion Method

Brovey transform fusion is also called color normalized transform fusion. Multispectral image can be resolved into intensity and hue. After the normalization of multispectral image bands, image information can be strengthened by product of bands and high resolution image, and its calculation formula is shown as follows:

\[
R_{\text{new}} = \frac{R}{(R + G + B)} \times I
\]

\( G_{\text{new}} = \frac{G}{(R + G + B)} \times I \)

\( B_{\text{new}} = \frac{B}{(R + G + B)} \times I \)

Where \( R, G \) and \( B \) are pixel gray values of three wave bands of multispectral image respectively. \( I \) is pixel gray value of high resolution image \( R_{\text{new}}, G_{\text{new}} \) and \( B_{\text{new}} \) are fused values of \( R, G \) and \( B \). This algorithm makes both ends of histogram of fusion image have great contrast. Besides, it can enhance the ratio of high intensity value of image to the low one and improve visual effects (Alvarez, 2003). However, after the fusion, the intensity of image is low and the phenomenon of losing spectrum information is serious, and it is greatly influenced by the noise. Only three wave bands in the multispectral data can be used, so it has extremely low utilization ratio about the original data.

(3) Principal Components (PC) Transform Fusion Method

Principal components (PC) Transform is multidimensional orthogonal linear transformation based on the image statistical characteristics, and is called K-L transform at mathematics. Principal components transform can centralize the useful information in multiple bands, which has the effect of concentrating variance information and compressing the data volume. What’s more, it can reduce the correlation between image bands after fusion and increase distinction degree between categories (Feng, 2012). Its transformation formula can be expressed as follows:

\[
Y = TX
\]

Where \( X \) is multispectral image data matrix before transform and \( Y \) is principal component space pixel matrix after transform. \( T \) is principal component transformation matrix. Principal component transform fusion can make operation for the multiple band images, whose degree of spectral distortion is less than that of HIS and Brovery Fusion Method. After the fusion, the spectral characteristic keeps well and definition is relatively high.

This method plays an important role in the aspects of compression, enhancement and change detection of image data, etc. However, during the process of using image with high resolution to replace the first principal component of multispectral image, it causes some loss of spectral characteristics information, which affects spectral resolution of image fusion to a great degree.

(4) Gram-Schmidt Transform Fusion Method

Gram-Schmidt transform fusion, GS transform for short, is a common multidimensional image or matrix orthogonal transformation algorithm, and it is put forward by Schmidt in early 1970s, while the application in the aspect of remote-sensing imaging fusion is proposed by Craig A. Laben in 1998 (Laben, 2004). The basic idea of this algorithm is making multispectral image bands simulate one panchromatic band by adopting re-sampling methods, and making GS transformation for multispectral bands and simulated panchromatic band, then using panchromatic band with high spatial resolution to substitute the first component after GS transformation, finally making GS inverse transformation to substituted data and the final fusion images are gained.

The advantage of Gram-Schmidt transform is no limitation for the number of fused multispectral image bands. The fused images cannot only eliminate the redundant information in the process of elimination and fusion, but also maintain the spectral information of original multispectral image. However, the anti-jamming capability is relatively poor.
(5) Pan Sharp Transform Fusion Method

Pan Sharp transform fusion method is put forward by Doctor Yun Zhang in 2004. Based on the statistics principle, this algorithm uses the minimum variance technique to achieve the optimal matching for the gray value of integrated wave band, and utilizes the principle to adjust the grayscale distribution of single waveband for decreasing the color distortion of fusion results. Furthermore, a series of statistical calculations for all input bands have been made so as to eliminate the dependence of fusion results to data-set and enhance the degree of automation in the process of fusion.

The advantage of Pan Sharp transform fusion method is that due to none limitation on the number of wave bands, it not only makes fusion images obtain high spatial resolution, but also has well spectral fidelity. So it is one of currently and admittedly better algorithms in the fusion of high resolution remote-sensing image (Zhang, 2000).

(6) Wavelet Transform Fusion Method

Wavelet transform fusion method is also called Mallat algorithm. It was proposed by Mallat (1989). Based on the wavelet transform, it can break the images up into algorithm and reconstruction algorithm with different frequency channels. The transform steps are shown in figure 1.

The fused image has two advantages that it not only reserves the spatial information of remote-sensing image with high spatial resolution, but also fuses the abundant spectral information of multispectral image, but the shortcoming is that it is easy to occur the blockiness when enhancing the result due to the substitution of low-frequency information of multispectral image for low-frequency information of panchromatic image. Meanwhile, it leads to the loss of detail information on panchromatic image to some extent, and the extent is related to the order of wavelet decomposition. The lower the order is the more detail information it loses, but more spectral information is reserved and vice versa (Wei & Li, 2003).

3. Data Fusion of GIS is Vector Data and Remote-Sensing Image

The rapid development and extensive application of geographic information system (GIS) lead to the generation of multi-dimensional data. In terms of data structure, vector and grid are two major spatial data structures of GIS. Vector data is based on Euclid geometry to express spatial distribution of geographical entity. Raster data is one kind of data, which divides the space into regular grids and provides corresponding attribute values for each grid to express geographical entity. And raster data includes all kinds of remote-sensing images and gridding map image data. As two different data structures, there are big differences between vector structure and raster structure. The vector data has a simple data structure but large data storage capacity, while vector structure has small data storage capacity, but complex data structure. Grid structure can easily describe boundary complex and obscure objects, which is convenient to process three-dimensional continuous surfaces, but has low spatial positional accuracy. Yet, vector structure is expressed by coordinate, so it has high spatial positional accuracy.

With the continuous development of GIS, in the process of using data, remote-sensing image data has been the main information source in the aspect of updating GIS and evaluating geographical environment, etc. The combination of vector data in GIS and raster image of remote-sensing analytic system becomes the inexorable trend of the developments. In the past, it was widely acknowledged that these two structures were incompatible. The reason is that raster data structure needs a large amount of computer memory for storage and processing to reach or be close to the same spatial resolution of vector data structure. Nevertheless, in the processing with some certain forms, vector structure is very difficult to solve some technical problems. At present, the fusion of GIS vector data and remote-sensing image data has become possible and has been widely applied.

3.1. Superposition and Fusion of Remote-sensing Image and Vector Data

This is simple superposition of two kinds of data and the lowest level of fusion of GIS vector data and remote-sensing image such as the superposition of remote-sensing image and line vector graphic, the superposition of remote-sensing image and DEM data, etc. Selecting vector and remote-sensing image as the underlying data, there is a need to manage jointly. The utilization of these two kinds of data involves the problem of data matching, which needs to unify data with different sources into the common ground. The unified spatial orientation provides a common position basis for vector and remote-sensing image, and the position basis includes projection, plotting scale and coordinate (Xie, 2010). So there is a need to adopt spatial data registration method.

The registration for spatial data can be described by the following mathematical formula (Zhang, She, & Liu, 2006):

$$P_1 = (x_1, y_1, \ldots)$$
$$P_2 = (x_2, y_2, \ldots)$$

In general situations, the coordinates can be understood as geometric parameter vectors for description. If $P_1$ and $P_2$ describe the same spatial characteristic, it can establish the following mapping function in one-to-one correspondence $F$ to satisfy:

$$P_2 = F(P_1, T)$$

This so-called spatial data registration is to determine the above mapping function $F$ that is to determine parameter vector $T$. The frequently-used registration method is that through homonymy points with the certain number (the number of parameter determines the least number of observations), list equation set according to mapping function $F$ and solve parameter $T$.

The general mathematical model formula of coordinate system transformation is as follows: Suppose that the ground coordinate $(x, y)$ of point $P$ converts to the plane coordinate $(x, y)$ of pixel point.

$$x = a_{00} + a_{10}X + a_{20}Y + a_{01}X^2 + a_{11}XY + a_{21}Y^2 + \ldots$$
$$y = b_{00} + b_{10}X + b_{01}Y + b_{20}X^2 + b_{11}XY + b_{21}Y^2 + \ldots$$

For linear transformation, its mathematical essence is the transformation between two plane-coordinate systems. The complicated transformation requires polynomial with higher power as well as larger quantity of known points (control points during the registration), which are in need when solving polynomial.
Z. HUANG ET AL.

is human-machine interactive interpretive method based on screen digitalization. It can be achieved by field investigation, the establishment of interpretation signs and total factor human-machine interactive interpretation of remote-sensing digital image. It can make the most of various useful auxiliary information and analytic means for remote sensing applied by GIS so as to greatly improve the accuracy and efficiency of remote-sensing information recognition. It makes a big progress comparing with artificial visual interpretation and it is currently a practical technology of remote-sensing digital image interpretation, but still exists the defects such as long time consuming and low efficiency. The third is technique of automatically classifying and identifying remote-sensing images on the basis of the further development of computer techniques and a remote-sensing image digital processing system, which has characteristics of timeliness and speediness. But from present research, due to the limited remote-sensing data spatial resolution and complicated geographical spatial phenomenon, it is difficult to automatically recognize and update the total factor and high accuracy of geographical spatial database, and there has a certain distance from the degree of practicability (Jiang & Ma, 2002).

4. Fusion of GIS Vector Data

At present, one major problem that many GIS application projects faced is how the vector data from different sources are integrated together. Integrating the information from different sources is a basic characteristic of GIS, but it does not mean that simply piling up data from different sources can solve the problem. As one technology of data integration and processing, vector data fusion covers the integration and application of many traditional disciplines and new techniques.

4.1. Fusion of Spatial Vector Data Model

The spatial vector data model has the characteristic of diversity. And achieving the unity between data models is the basic requirement of the fusion of spatial vector data. The common processing method is to design one new data model, which can include all data models in the process of fusion and achieve the uniformity of model in the process of handling by completing the mapping from different spatial vector data model to the new one. In the process of handling, the design of new spatial vector data model is the key of data fusion quality, however, with the increase of fused spatial vector data model kind, this new spatial model exists very strong transverse expansion, and

Figure 1. Wavelet Transform Fusion Steps.

Figure 2. Basic Component of the New Data Model.
on this basis the working of data fusion will increase exponen-
tially (Jiang & Yang, 2011)

The traditional data model is a data model of topological
data model and entity oriented data model, the two has its
advantages and disadvantages. The topology data model can
reduce the data redundancy, and entity oriented data model
has the advantage of the entity management and modifying
convenient. When we design a new data model, the limitations
of traditional data model can be broken through completely,
which has the characteristics of entity oriented, but also con-
tains the topological relationship. As shown in Figure 2, the
new data model can be designed as; basic elements include the
full point, line, plane coordinate data, maintain the integrity
of the elements; elements layer contains topological relation,
have independent topology information that can reflect the
correct spatial relationship, which make spatial analysis been
proceed smoothly.

4.2. Classification Standards of Geographic Factors,
Fusion of Coding

Spatial vector data, especially geographic data, for different
classification systems exist various expression methods cor-
responding to the different coding systems. Achieving the
transformation and unity of classification standards and cod-
ing systems is an important indicator of the fusion quality of
spatial vector data.

Fusion for the classification and grading standards of spatial
objects; in the common fusion methods, it is generally to estab-
lish the standard of third party and the standard conversion
relations with original data format for achieving the unity.

Fusion for the coding systems; the common method is to
establish third-party encoding, which needs to take account
of the coding scheme of original data format, then achieve the
unity of coding by establishing coding transform mechanism.

4.3. Fusion of Spatial Relation

General attributes and data item of graphic entity can achieve
fusion by direct transform. Fusion of geometric position and
processing of topological relations are difficult links which need
special handling (Ji & Liu, 2004).

Fusion of geometric position solves the differences such as
different plotting scale, data acquisition errors and vari-
ous refresh time, etc., between different data sources. It needs
multidisciplinary thoughts and methods like pattern recog-
nition, statistics, graph theory and artificial intelligence, etc.;
achieving the fusion of geometric position usually includes two
steps: Entity matching and matching the data merging of entity.

Most of spatial vector data contain topological information
and are closely related to geometric object of data.

In the process of data transform, it often produces the oper-
ation needs for transform and option of geometric objects. At
this point, topological information has changed and could not
express topological relations correctly. Therefore, data fusion
software with complete functions is indispensable to inspect the
topological relations, handle and support the reconstruction
for the spatial vector data format with topological relations.

5. Discussion

In the three fusion methods of Remote-sensing image, only
pixel-level fusion is proceed on the basis of original image data.
Therefore, it can reserve more original image information and
provide subtle information what other fusion levels cannot
provide (Zhao, 2009). But there are also some disadvantages,
which are embodied in; high data-processing capacity and
poor real-time performance; relatively high registration accu-
raciness of image data with one pixel; poor analytical capability,
which cannot achieve effective understanding and analysis for
image; relatively high error-correcting ability in the process of
fusion and weak anti-interference ability. However, the main
purpose of pixel-level fusion is to make a service for image
enhancement, image segmentation and image classification,
so it is extensively applied in the fusion.

The fusion of GIS vector data and remote-sensing image
data obtains more abundant amount of information, has
the characteristic of perceptual intuition and readability,
and improves the degree of visual representation directly
perceived by human visual and the accuracy of assistant
decision-making.

Currently, the fusion of these two kinds of formatted data
has become possible and has been widely applied. In the pro-
ject of GIS, many GIS systems have been integrated and could
make unified management for spatial data of vector and raster
structures. While in the field of digital mapping, the fusion of
these two data structures has also been applied extensively.
Furthermore, display the superimposed image and search the
changes through eyes by operator. Use the method of manual
editing to update GIS vector data. Extract new vector data and
store it by adopting the same data structure in the original
database, as a kind of new data source which assists the estab-
lishment, maintenance and renewal of GIS database.

Geospatial Vector Data fusion often focuses on a specific
need. The specific methods solve specific problems, these
methods does not have universality. Map conflation discussed
geometric data fusion, not involved data fusion. Geocoding
technology can realize the integration of spatial data and non
spatial data and enrich the attribute of spatial data, but did
not relate to the geometric data fusion. The tools and methods
using attribute data fusion are hard to express their geometric
features by using ontology.

6. Conclusions

In the big data environment, the rapid development and exten-
sive application of the geographic information system leads to
the generation of multiple sources of spatial data, and the dif-
ference of multisource spatial data brings about great difficul-
ties for comprehensive utilization of data and data sharing. To
reduce the cost of data production and improve the efficiency
and quality of data production, studying the method of multi-
source spatial data fusion plays a vital role for the develop-
ment of all geomatics industry.

The development of multisource spatial data fusion is up
to now, and it is no longer a single technology but an inter-
disciplinary comprehensive theory and method. Although
there exist some problems with the standardized promotion
of data sharing and the improvements of related technologies
such as sensor technology, data processing technique, network
communication technology, artificial intelligence technology,
parallel computing software and hardware technologies, etc.,
the fusion algorithm with robustness and correctness will be
perfected and realized. And the professional multisource spa-
tial data fusion system will produce and become an important
technology of intelligent monitoring and data processing in
complex industrial system, so it can totally satisfy the needs from each application level.

Acknowledgment

We would also like to thank the Dr. Zhao Dongbao, for constructive comments, which helped improve this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was financially supported by the Plan For Scientific Innovation Talent of Henan Province [grant number 154100510006]; Science and Technology Project of Xinjiang Uygur Autonomous Region [grant number 201491105]; Doctoral Student Innovation Fund of Technology Project of Xinjiang Uygur Autonomous Region [grant number 152102210111]; the Key Science and Technology Project of Henan Province [grant number 152102210111]; Doctoral Student Innovation Fund of North China University of Water Resources and Electric Power and High Level Talents Introduction Project of Xinjiang Uygur Autonomous Region.

Notes on contributors

Zhiquan Huang is a professor in the School of Resources and Environment, North China University of Water Resources and Electric Power. He received his Ph.D. degree from Chinese Academy of Sciences in 1999. His current research interests are geographical information systems and geological engineering.

Yu Fu is a Ph.D. student in the School of Resources and Environment, North China University of Water Resources and Electric Power. Her current research is focused in data fusion method and application of geological information. Fu received a Master’s Degree in Science from Nanjing Normal University.

Fuchu Dai is a professor in the School of Resources and Environment, North China University of Water Resources and Electric Power. His current research interests are landslide disaster, remote sensing, and geographic information systems. Dai has a Ph.D. in Engineering from Institute of Geology and Geophysics, CAS in 1997.

References


