

Roads Data Conflation using Update High Resolution Satellite Images

Abolfazl Abdollahi¹, *, Hamid Reza RiyahiBakhtiari²

¹ Remote Sensing and Geospatial Information Systems, Faculty of Geography, Kharazmi University of
Tehran, Iran

A_Abdollahi69@yahoo.com

² Faculty of Natural Resources and Earth Science, University of Shahrekord, Shahrekord, Iran.
reza_reyahi@yahoo.com

ABSTRACT:

Urbanization, industrialization and modernization are rapidly growing in developing countries. New industrial cities, with all the problems brought on by rapid population growth, need infrastructure to support the growth. This has led to the expansion and development of the road network. A great deal of road network data has made by using traditional methods in the past years. Over time, a large amount of descriptive information has assigned to these map data, but their geometric accuracy and precision is not appropriate to today's need. In this regard, the improvement of the geometric accuracy of road network data by preserving the descriptive data attributed to them and updating of the existing geo databases is necessary. Due to the size and extent of the country, updating the road network maps using traditional methods is time consuming and costly. Conversely, using remote sensing technology and geographic information systems can reduce costs, save time and increase accuracy and speed. With increasing the availability of high resolution satellite imagery and geospatial datasets there is an urgent need to combine geographic information from overlapping sources to retain accurate data, minimize redundancy, and reconcile data conflicts.

In this research, an innovative method for a vector-to-imagery conflation by integrating several image-based and vector-based algorithms presented. The SVM method for image classification and Level Set method used to extract the road the different types of road intersections extracted from imagery using morphological operators. For matching the extracted points and to find the corresponding points, matching function which uses the nearest neighborhood method was applied. Finally, after identifying the matching points rubber-sheeting method used to align two datasets. Two residual and RMSE criteria used to evaluate accuracy. The results demonstrated excellent performance. The average root-mean-square error decreased from 11.8 to 4.1 m.

KEY WORDS: Conflation, Control Points, Matching Control Points, Rubber-Sheeting Algorithm

1. INTRODUCTION

In today's societies, considering the growing population and industries, construction of new road network and improvement of the existing network is of particular importance. Efficient road network infrastructure provides economic and social benefits. Accurate and update information of road network is the fundamental requirement for new road design, construction, maintenance, and management projects. Due to the size and extent of the country, providing these data using traditional methods is time consuming and costly. In addition, the rapid trend of spatial changes of complication especially in developing countries requires that along with the preparation and production of spatial information, updating of existing maps and other spatial information should considered. Significant advances in geographic information systems have opened up new horizon in this domain, as a powerful tool for storing, retrieving, processing, analyzing and displaying spatial information in recent decades. Today, the improvements in spatial and spectral resolution of satellite images have led to obtain geometric data with higher precision to produce and update coverage maps as a practical and optimal way.

Instead of displaying all relevant information in a single framework, users actually need to combine spatial data to provide additional information that is not available in any single information resources. The process of combining the information from two (or more) geodata sets to make a master data set that is superior to either source data set in either spatial or attribute aspect is called conflation (Yuan and Tao, 1999). In fact, the conflation of spatial information is one of the main issues in GIS (Usery et al., 2003). Depending on the aim of the project, conflation may used to increase spatial accuracy and consistency, updating or adding more attributes associated with the spatial features and updating or adding new spatial features to a data set. According to the types of applied geospatial datasets, the conflation can be categorized into vector to vector, vector to raster and raster to raster groups. There have been a number of efforts to automatically or semi-automatically accomplish vector to vector conflation but there are fewer research activities on vector to raster data conflation.

As a result, spatial data products and related data displayed in some conflation methods are required. Conflation of spatial information requires a different set of data, which can

* Corresponding author

be integrated, and then a single combinational data set of integrated elements be created. A critical step is reducing the inconsistency and spatial contradiction among multiple sets of data. This research presents a novel approach for an operational road database updating system and some of our research achievements along this direction.

2. LITERATURE REVIEW

The history of map conflation dates back to the early 1980s which first use of the auto-conflation process was designed during a project between the US Geological Survey and the Bureau of Statistics to integrate a digital map of urban areas in the United States (Saalfeld 1988).

Cobb et al. (1998) presented a method for combining the vector data properties with the datasets produced by the National Mapping and Imaging Agency. They consider attributes matching as a kind of classification problems, which can be applied through the theory of evidence or uncertainty argumentation, such as fuzzy logic.

Different researchers from the GIS and computer vision have shown that at the intersections of road networks an exact set of control points has been provided (Chen et al., 2003).

Xiong and Sperling (2004) provided a semi-automatic matching for integrating the network's database. The algorithm automatically establishes strong tensions for nodes, edges, and segments between the two networks using a cluster-based matching mechanism.

Chen (2005) described a method for automatically integrating road vector data with geo-referenced images. In this study, an automated method used to find control points over vector data as well as the image. Then vector aligned with the image using an automated algorithm.

Tong et al. (2009) proposed a method for probability-based complications matching by integrating multiple geometric, semantic, and topological measurements. In the proposed method, a general alignment with the weighted average of several measurements, including position measurement, shape measurement, directional measurement and topological measurement, are calculated. The results showed that the proposed method was effective for resolving matching problems in the map conflation.

3. MATERIALS AND METHODS

The main steps in the automated process of vector image conflation are extraction spatial information, automatic intersection extraction from images, control point matching and finally correction of position using the Rubber-sheeting¹method. Figure 1 shows a flowchart of automated vector-image conflation process. For evaluation and implementation of the proposed method, Matlab environment and ArcMap softwares are used in this study.

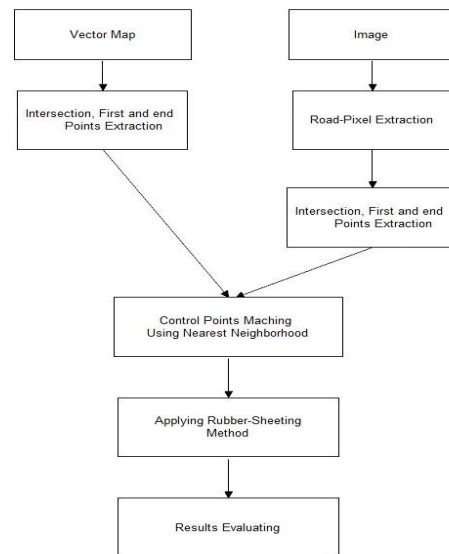


Figure 1 - Flowchart of automatic vector- image conflation process

As shown in Fig. 1, in the first step pixels of road were extracted from image. The SVM² method for image classification and Level Set method used to extract road pixels. In the next step, the intersection, beginning and the end points of road has been determined both on the image and vector map using morphological functions and skeleton methods to find control points. After finding control points, for matching the extracted points and finding the corresponding points, matching function, which uses the nearest neighborhood method, was used. Finally, after identifying the matching points rubber-sheeting method applied to align two datasets.

3.1. Data

The images used in this study are Google Earth Images of Saveh city located in the Markazi Province of Iran. It is located about 100 km southwest of Tehran. Google Earth displays satellite imagery in different resolutions, which enable users to observe objects around the Earth vertically or oblique.

3.2. SVM (Support Vector Machine)

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it used mostly in classification problems. In this algorithm, each data item is plotted as a point in n-dimensional space (where n is number of features) with the value of each feature being the value of a particular coordinate. Then, classification is performed by finding the hyper-plane that differentiates the two classes very well. Support Vectors are simply the coordinates of individual observation. Support Vector Machine is a frontier, which best segregates the two classes.

¹ Piecewise Local Affine Transformation

²Support vector machine

3.3. Level Set Method

The level set method or leveling method is a numerical method to approximate geometric objects and their movement. The advantage of the level-set method is that one can calculate curves and surfaces on a spatially fixed coordinate without having to use parameterization of these objects. In particular, in the level set method, the topology need not be known, and it may change during the computation. This allows easy tracking of the edges of moving objects, such as roads.

3.4. Road intersections and Terminations Extraction

Road intersections and road terminations (endpoints) are generally very reliable and stable across different datasets. They represent robust information that makes them very useful for feature matching. However, few approaches make use of such pertinent information in the large amount literature on automated road extraction from remotely sensed imagery. In this study, morphological functions and skeleton method used to find control points. In general, these functions eliminate noise and pixels in non-roads, fill the gaps and increase the accuracy of the classification. Skeletonization is a process for reducing foreground regions in a binary image to a skeletal remnant that largely preserves the extent and connectivity of the original region while throwing away most of the original foreground pixels. Skeletonization shows a simple skeleton of a binary image. Extracting skeleton of a pattern means narrowing the pattern in such a way that the overall shape of the pattern does not disappear.

3.5. Matching Features

Feature matching is used to identify the correspondence from two different datasets as representations of the same geographic object. This is the most critical step of conflation. In this study, matching function that uses the nearest neighborhood method applied to match the extracted points and find the corresponding points. This method works to find the nearest pixel to the reference point based on the distance. So that it considers the point (reference point) and then for match it with the next point it searches for the reference point around it and finds the nearest neighbor between the two points by defining a threshold. Then, if the distance between them be less than the threshold, the two points will be matched.

3.6. Map alignment

After identifying the matched control points, rubber-sheeting method used to align two datasets. The Rubber-Sheeting transformation is based on geometric and mathematical theories. This technique is called "Rubber-Sheeting" because figuratively it considers a deformed map as a sheet made of rubber which is stretched to some fixed nails representing the correcting reference points. In rubber-sheeting adjustments, one layer has been aligned with another that is often in close proximity. The source layer adjusted to the more accurate target layer. During rubber-sheeting, the surface is literally stretched, moving features using a piecewise transformation that preserves straight lines. During this process, links are placed to stretch or warp

the data to align to the underlying datasets. Rubber-sheeting is used to make small geometric adjustments in data usually to align features with more accurate information. In Conflation process rubber-sheeting is used to align layers in preparation for transferring attributes.

4- RESULTS

The images used in this research are Google Earth Images. The road vector data is also relating to outer city roads provided by the National Cartographic Center.

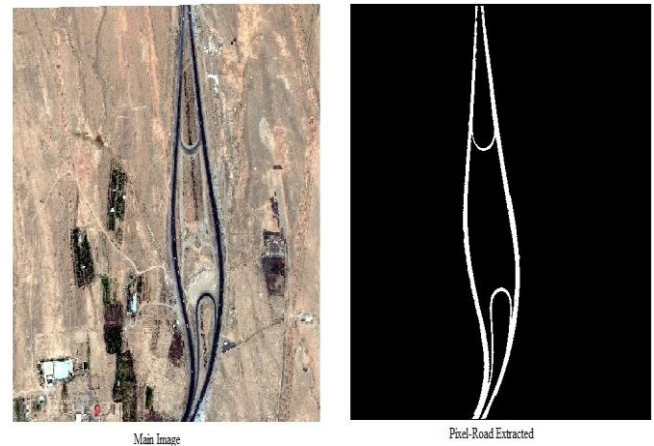


Figure 2. Road extracted from Google Earth image

After extracting road pixels (Fig. 2), vector-image conflation was done automatically. To evaluate accuracy, road was extracted manually from image as a reference (Fig. 3) so that it can be used to evaluate the precision of conflation.

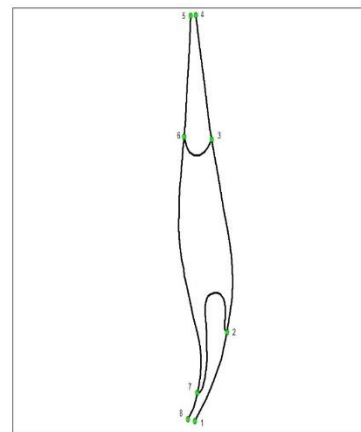


Figure 3- Road and control points extracting from image manually

Then main road, deformed road (vector old) and finally the ultimate conflation has been compared for referencing in order to show how to improve the accuracy. Finally, after selecting the control points and matching them, the conflation of vector-image has been done using the Rubber-Sheeting method which is shown in Fig. 4.

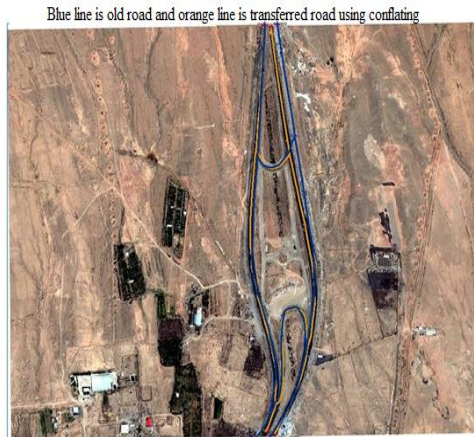


Figure 4: Vector-image conflation using Rubber-Sheeting method

In Fig.4, the blue lines are belonging to the old vector which after the conflation has been taken, this line has been almost transmitted to the center of the image which is marked with orange color. After performing rubber-sheeting method the road has been transmitted to the center of image and as it can be seen, the orange line is also in the middle.

In order to evaluate the accuracy of the results two residual and RMSE¹ criteria were used. The residual error represents the distance between the image point location calculated by the appropriate polynomial equation and the location of the specified point on the GIS data. The RMSE error is the difference between the predicted value of statistical model and actual value. The RMSE value can be calculated by equation (1):

$$\text{Equation RMSE} = \sqrt{\sum \frac{(x_{\text{target}} - x_{\text{reference}})^2 + (y_{\text{target}} - y_{\text{reference}})^2}{n}} \quad (1)$$

Table 1 shows the estimate of residual and RMSE errors before the conflation process as well as after applying the Rubber-Sheeting method. As can be seen, the RMSE error is 11.839 meters before conflation process and then decreased to 4.124 meters after applying rubber-sheeting method. This improvement indicates that the algorithm used for the conflation process is good and efficient.

Points	Residual error before conflation process	Residual error after conflation process
1	9.681	2.118
2	10.473	3.758
3	19.010	8.523
4	4.805	4.399
5	5.161	3.118
6	3.983	1.910
7	17.772	3.240
8	13.243	1.253
RMSE Error	11.839	4.124

Table 1: Estimating the residual and RMSE errors before and after the conflation process

5- DISCUSSION AND CONCLUSION

In this study, automatic and accurate conflation of vector data maps with high spatial resolution images presented. This method uses automatic algorithms to extract road pixels from image, identify control points, match these points and eventually integrate them together. The basic idea of the proposed approach for road junction detection is based on SVM and level set interactive methods. Finding control points is very important step in map conflation. In this study, control points (intersections, beginning and end points) were found based on the morphology operators and matching those points together and finding the corresponding points was done using the matching function by the neighboring technique. Finally, the Rubber-Sheeting method used to transfer vector data points to the new control points found on the image. With regarding the conflation process, the following studies discussed and their results compared with the results obtained in this study.

Compared to existing approaches and commercial products the proposed method utilize the metadata and attributes of road vector data, maps, satellite imagery and relevant datasets to achieve the automatic feature detection. Unfortunately, until now very few existing approaches for road extraction from remotely-sensed imagery have used these features. Zhang and Isabelle (2004) conducted a study to detect road network changes using wavelet-based approach for road extraction in order to update the Canadian Topographical Database (NTDB).

Song et al. (2009) extends a modified snake algorithm, and introduce spatial contextual measures to extract the different types of road intersections and terminations from imagery. They tested the integrated system in several areas such as rural, suburban and urban. comparing to the our proposed

¹Root-Mean-Square Error

method, snakes algorithm is very sensitive to initial position and also noise in the digital number along the road lines which may lead to an inaccurate result.

Chen et al. (2006) used a new method to automatically annotate and conflate satellite imagery with vector datasets. Their approach utilized the knowledge, such as online data sources or road segment direction and road intersections provided by vector data. Therefore, in comparison with our method the problem of finding control points from satellite images eliminated from the conflation process.

The method used in this research as well as the methods mentioned in the above studies has a good accuracy in automatic vector-image conflation. The average value of residual error for the 8 points before the conflation process is 10.514 m and it reduced to 3.539 m after applying the conflation process. In addition, by examining the RMSE error it is possible to determine the basic role of Rubber-Sheeting method in reducing the error rate. So that, as previously mentioned the RMSE error is 11.839 m before conflation process and then decreased to 4.124 m after conflation process. This improvement indicates that the algorithm used for the conflation process is good and efficient.

REFERENCES

Cobb, M.A., Chung, M.J., Foley, H., 1998. A rule-based approach for the conflation of attributed vector data, *Geoinformatica journal*, 2(1), 7–35.

Chen, C., Thakkar, S., Knoblock, C., 2003. Automatically annotating and integrating spatial datasets. Proceedings of international symposium on spatial and temporal databases, 24–27 July, Santorini Island, Greece.

Chen, Y., Jianya, G., Jianping, P., 2005. Conflation technology using in spatial data integration on the internet: Society of Photo-Optical Instrumentation Engineers, Bellingham. Image processing and pattern recognition in remote sensing. Honolulu HI, US, 56–63.

Chen, C.C., Knoblock, C.A., & Shahabi, C., 2006. Automatically Conflating Road Vector Data with Orthoimagery, *GeoInformatica* 10(4), 495–530.

Saalfeld, A., 1988. Conflation, Automated map compilation, *International Journal Geography Information System*, 2(3), 217–228.

Song, W., Keller, J.M., Haithcoat, T.L., & Davis, C.H., 2009. Automated Geospatial Conflation of Vector Road Maps to High Resolution Imagery, *IEEE Transactions On Image Processing*, 18 (2), 388-400.

Tong, X., Shi, W., & Deng, S., 2009. A probability-based multi-measure feature matching method in map conflation. *International Journal of Remote Sensing*, 30 (20), 5453–5472.

Usery, E.L., Finn, M.P., & Starbuck, M., 2003. Data integration of layers and features for the national map. in

Proceedings of American Congress on Surveying and Mapping, Phoenix, Arizona.

Xiong, D., & Sperling, J., 2004. Semiautomatic matching for network database integration, *ISPRS Journal Photogrammetry and Remote Sensing*, 59(1–2), 35–46.

Yuan S., & Tao, C., 1999. Development of conflation components. In: Proceedings of geoinformatics. Ann Harbour, pp 1–13

Zhang, Q., & Isabelle, C., 2004. Automatic road change detection and GIS updating from high spatial remotely-sensed imagery, *Geo-spatial Information Science*, 7(2), 89-95.