# Automatically and Accurately Conflating Satellite Imagery and Maps

Ching-Chien Chen, Craig A. Knoblock, Cyrus Shahabi, and Snehal Thakkar
University of Southern California
Department of Computer Science and Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292 USA
{chingchc, knoblock, shahabi, snehalth}@usc.edu

## 1 Introduction

There is a wide variety of geo-spatial data available on the Internet, including a number of data sources that provide satellite imagery and maps of various regions. The National Map[1], MapQuest[2], and University of Texas Map Library[3] are good examples of map or satellite imagery repositories. In addition, a wide variety of maps are available from various government agencies, such as property survey maps and maps of oil and natural gas fields. Road vector data covering all of the United States is available from the U.S. Census Bureau.[4] One of the key questions for Geospatial Information Systems researchers is how to accurately and efficiently align imagery, maps and vector data from these various sources. In this paper, we describe our approach to automatically and accurately align satellite imagery with the various online maps that are currently available. The traditional approach to aligning these various geospatial products is to use a technique called conflation [7], which requires identifying a set of control point pairs on the two data sources. The identification of these control points is often performed manually, which is a tedious and time-consuming process that is made even harder by the fact that many of the online sources do not even provide the coordinates of the corner points of the maps. In previous work, we developed an approach to automatically conflating road vector data with satellite imagery [2]. In this paper we describe how we address the even more challenging problem of automatically conflating maps with satellite imagery. Since we build on our previous work, we first review our approach to automatically conflate road vector data with satellite imagery. Then we describe our approach to automatically conflating a map with the satellite imagery by first using the vector data to identify all of the intersections and then utilizing a specialized point matching algorithm to align the two datasets.

## 2 Aligning Vector Data with Imagery

The first step in aligning maps with imagery is to identify the location of all of the intersections in the imagery. Since image processing is both expensive and inaccurate, we find the road intersections in the imagery by first aligning road vector data with imagery and then locating the road network intersection points from the vector data.

In [2], we described several techniques for automatic conflation of road vector data with satellite imagery. The most effective technique we found exploits a combination of the knowledge of the road network with image processing in a technique that we call localized image processing. In this approach, we first find feature points, such as the road intersection points, from the vector dataset. For each intersection point, we then perform image processing in a localized area around the intersection point to find the corresponding point in the satellite image. The running time for this approach is dramatically lower than traditional image processing techniques due to the limited image processing required. Furthermore, exploiting the road direction information improves both the accuracy and efficiency of detecting edges in the image.

---

1 http://seamless.usgs.gov
2 http://www.mapquest.com
3 http://www.lib.utexas.edu/maps/index.html
4 http://www.census.gov/geo/www/tiger/

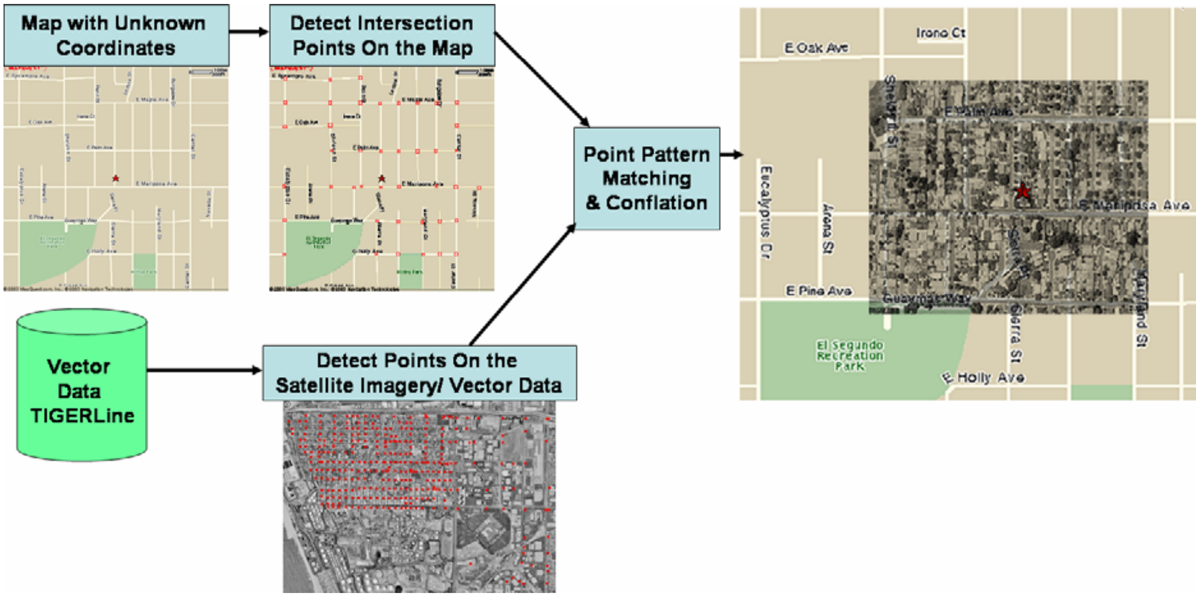


*Figure 1. Align Imagery With Maps*

An issue that arises is that the localized image processing may still identify incorrect intersection points, which introduces noise into the set of control point pairs. To address this issue, we utilized a filtering technique termed Vector-Median Filter[1] to eliminate inaccurate control point pairs. Once the system has identified an accurate set of control point pairs, we utilize the rubber-sheeting techniques described in [7] to align the vector data with the satellite imagery. In our test sets, this approach produced an accurate alignment of the vector data with the imagery.

## 3 Aligning Imagery with Maps

The techniques we described for conflating road networks with imagery can be generalized to other geospatial data sources. We can extend our vector-imagery conflation techniques to align imagery with maps whose geo-coordinates are unknown in advance. We assume that the maps we want to integrate show at least a partial road network. We then utilize common vector datasets as “glue” to integrate imagery with maps. In the previous section we described how to find the intersection points in the satellite imagery. The remaining tasks are to find intersection points on the maps and then find the alignment between the intersection points on the maps and the imagery.

Figure 1 shows the overall approach to conflating imagery and maps. First, we automatically conflate the road vector data with the satellite imagery to find the intersections in the image. Next, we find the road intersection points on the map (the example shows a map from MapQuest). Then, we utilize a specialized point pattern matching algorithm to align the two datasets.

One of the frequently extracted features on maps is road intersections because road networks are commonly illustrated on diverse maps. Ideally, intersection points could be extracted by simply detecting road lines. However, due to the varying thickness of lines on diverse maps, accurate extraction of intersection points from maps is difficult [5]. In addition, there is often noisy information, such as symbols and alphanumeric characters on the map, which make it even harder to accurately identifying intersection points. Therefore, we adapted the automatic map processing algorithm described in [5] to skeletonize the maps for extracting intersection points. Although the algorithm can significantly reduce the rate of misidentified intersection points on the maps, it is still possible that some noisy points will be detected as intersection points. However, our point matching algorithm (described next) can tolerate the existence of misidentified intersection points.

Now that we have identified a set of intersections on both the map and the imagery, the remaining problem is to find the mapping between these points in order to generate a set of control point pairs. The basic idea is to find the transformation T between the layout (with relative distances) of the intersection points on the map with the intersection points on the satellite imagery. These intersection points on the image are the intersection points in the vector data since the vector data is aligned with the satellite imagery. The key computation of matching the two sets of points is calculating a proper transformation T consisting of translation and scaling. The geometric point set matching in two dimensions is a well-studied family of problems with application to area such as computer vision, biology, and astronomy [4]. Because it is time-consuming to obtain the mapping, there is a randomized version [4] for this computation on less noisy point datasets. However, this is not appropriate for our datasets because the extracted intersection points from maps could include a number of misidentified intersection points. We developed a new randomized point matching algorithm that exploits information on direction and relative distance available from the vector sets. The information on direction and distance is used as prior knowledge to prune the search space of the possible mapping between maps in the two datasets. The revised algorithm works well in our preliminary experiments even in the presence of very noisy data.

Now that we have a set of control point pairs for the map and imagery, we can use the conflation technique described in [7] to align the map with the satellite imagery. The aligned map and satellite imagery can then be used to make inferences that could not have been made from the map or imagery alone. In addition, the mapping of the vector data to the map can also be used to determine the geo-coordinates of the corner points of the maps, which may have been previously unknown.

## 4 Related Work

While the conflation technique was described in [7] in 1993, there has been relatively little work on automatically conflating maps with satellite imagery. In [8], the authors describe how an edge detection process can be used to determine a set of features that can be used to conflate two image data sets. However, their work requires that the coordinates of both image data sets be known in advance. Our work does not assume that coordinates for the maps are known in advance, although we do assume that we know the general region. There has been a considerable amount of work on conflating vector data with satellite imagery or maps [3, 6]. Our work significantly differs from the previous work in terms of our approach to conflate vector data with satellite imagery. These differences are described in detail in [2].

## 5 Discussion

Given the huge amount of geospatial data now available, our ultimate goal is to be able to automatically integrate this information using the limited information available about each of the data sources. An interesting direction with respect to integrating maps is to be able to take arbitrary maps with unknown geocoordinates and determine their location anywhere within a city, state, country, or even the world. We already have road vector data covering most of the world, so the real challenge is developing a hierarchical approach to the point matching to make such a search tractable.

## References


[1]. Astola, J., P. Haavisto, and Y. Neuvo. *Vector Median Filter. In Proceedings of IEEE*, 1990

[2]. Chen, C.-C., S. Thakkar, C.A. Knoblock, and C. Shahabi. *Automatically Annotating and Integrating Spatial Datasets. In the Proceedings of International Symposium on Spatial and Temporal Databases*. Santorini Island, Greece, 2003

[3]. Hild, H. and D. Fritsch, *Integration of vector data and satellite imagery for geocoding.* IAPRS. **32**, 1998

[4]. Irani, S. and P. Raghavan, *Combinatorial and experimental results for randomized point matching algorithms.* Computational Geometry. **12**(1-2): p. 17-31, 1999

[5]. Musavi, M.T., M.V. Shirvaikar, E. Ramanathan, and A.R. Nekovei, *A Vision Based Method to Automate Map Processing.* Pattern Recognition. **21**(4): p. 319-326, 1988

[6]. Price, K., *Road Grid Extraction and Verification.* IAPRS. **32 Part 3-2W5**: p. 101-106, 1999

[7]. Saalfeld, A., *Conflation: Automated Map Compilation*, in *Computer Vision Laboratory, Center for Automation Research*, University of Maryland, 1993

[8]. Sato, T., Y. Sadahiro, and A. Okabe, *A Computational Procedure for Making Seamless Map Sheets*, Center for Spatial Information Sciences, University of Tokyo, 2001