NON-RIGID CONFLATION FOR VECTOR DATA-SETS USING EM ALGORITHM AND MIXTURE OF GAUSSIAN APPROACH

Ganesh Krishnamurthy, ganesh.krishnamurthy@mavs.uta.edu ,
Dept. of Electrical Engineering
Dr.Venkat Devarajan, venkat@uta.edu
Dept. of Electrical Engineering
Dr.Iriniel Dragan, dragan@uta.edu
Dept. of Mathematics
University of Texas, Arlington

1. INTRODUCTION

Conflation technology offers a cost-effective, efficient method of combining the best attributes of separate vector databases. There is an ever-increasing need for reliable and highly automated spatial data integration. This integration may be:

- **spatial**, where two or more datasets depicting complementary information for the same area (e.g. one depicting road networks, and another depicting aerial photo of the same location) are brought together to extend their coverage or to aggregate their content, thus supporting analysis.
- **temporal**, where two or more similar datasets conveying the same thematic information at different time instances (e.g. a sequence of satellite images of the same area over a long time period) are brought together to analyze change.

The essence of the data integration process is the conflation (fusion) of spatial data from two different sources, which results in a single data set that incorporates data and attributes from both sources. Registration technique can be rigid or non-rigid depending on underlying transformation. Key characteristic of rigid transformation is that all distances are preserved. Effective algorithms exist for rigid transformation, some of which are discussed in ref [1], [2], [3].

Despite these efforts, fully automated conflation still remains a major challenge, primarily due to the inability of existing matching algorithms to cope with complex non-linear transformation models required for a complete solution. Our intent is to improve the accuracy of the conflation by implementing a non-rigid transformation to the above solution.

Below we discuss the two main approaches to solving the conflation problem

a. **Deterministic**: in which distance metrics are used as similarity measures between two matching candidates. A common distance metric used in this approach is the Hausdorff distance. Some examples of this approach are Algebraic Structural Analysis (ASA) algorithm [1], [2], Automatic Multi-Source conflation developed by Geosemble [3].
b. **Probabilistic**: in which empirical probabilities are iteratively assigned to potentially matching features. Though not a lot of work using this approach has been done to solve conflation problem in geospatial realm, it is popular in medical imagery analysis.

This paper will discuss in detail the essentials of probabilistic approach to solving correspondence/conflation problem for given vector datasets. We wish to implement the approach on an open-source GIS platform called, OpenJUMP.

2. APPROACH

The current version of the software implements a rigid transformation on datasets, which does not properly align all the nodes and edges. Our idea is to implement a non-rigid transformation on top of the rigid transformation to properly align the data sets and thus reduce errors in the result.

The whole idea is to treat registration as a Maximum Likelihood (ML) estimation problem with a constraint which coherently moves the points. The coherent movement is justified by the fact that a non-rigid transformation is followed by a rigid transformation. An Expectation Maximization (EM) approach along with Deterministic Annealing is used to optimize the ML solution [4]. The E-step basically estimates the correspondence under the given transformation, while the M-step updates the transformation based on the current estimate of correspondence.

As a precursor to implementing the approach on OpenJUMP we tested it on MATLAB and it shows promising results. Some results are attached below. We refer to TEMPLATE as the master data set, which we believe to be the ground truth onto which the MODEL data (inaccurate) has to be conflated.

3. RESULTS

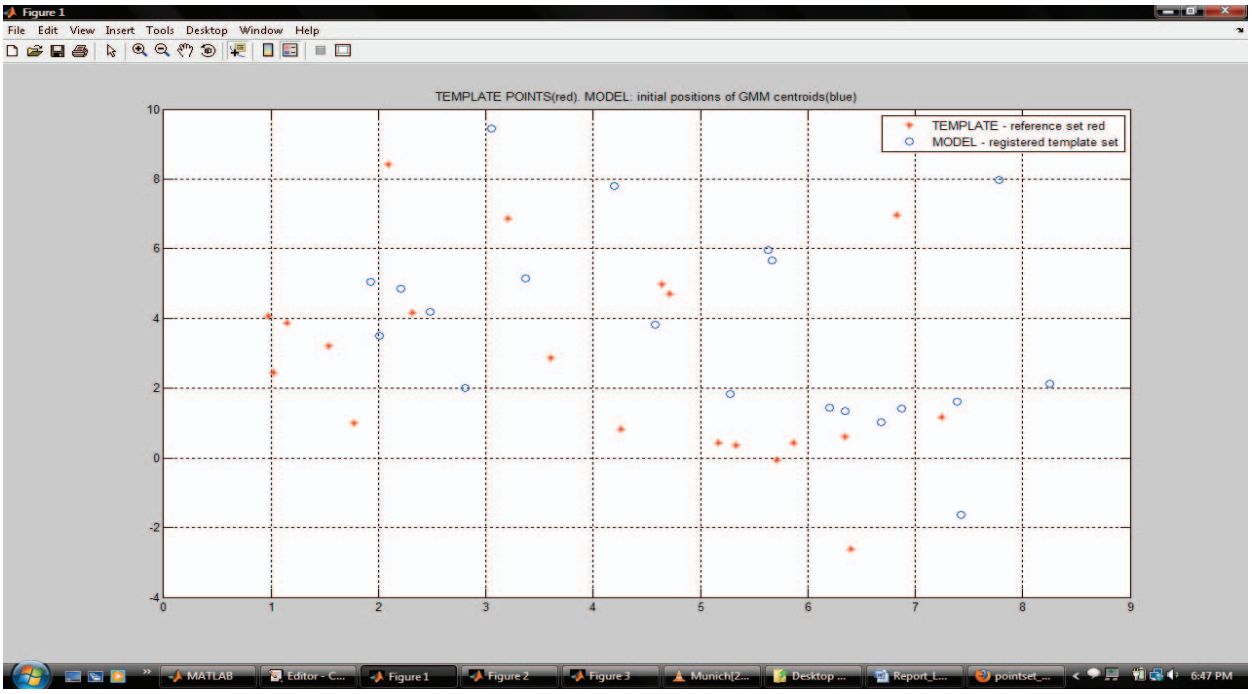

Figure 1: Template: Red dots, MODEL: Blue circles. We have to conflate MODEL to TEMPLATE.

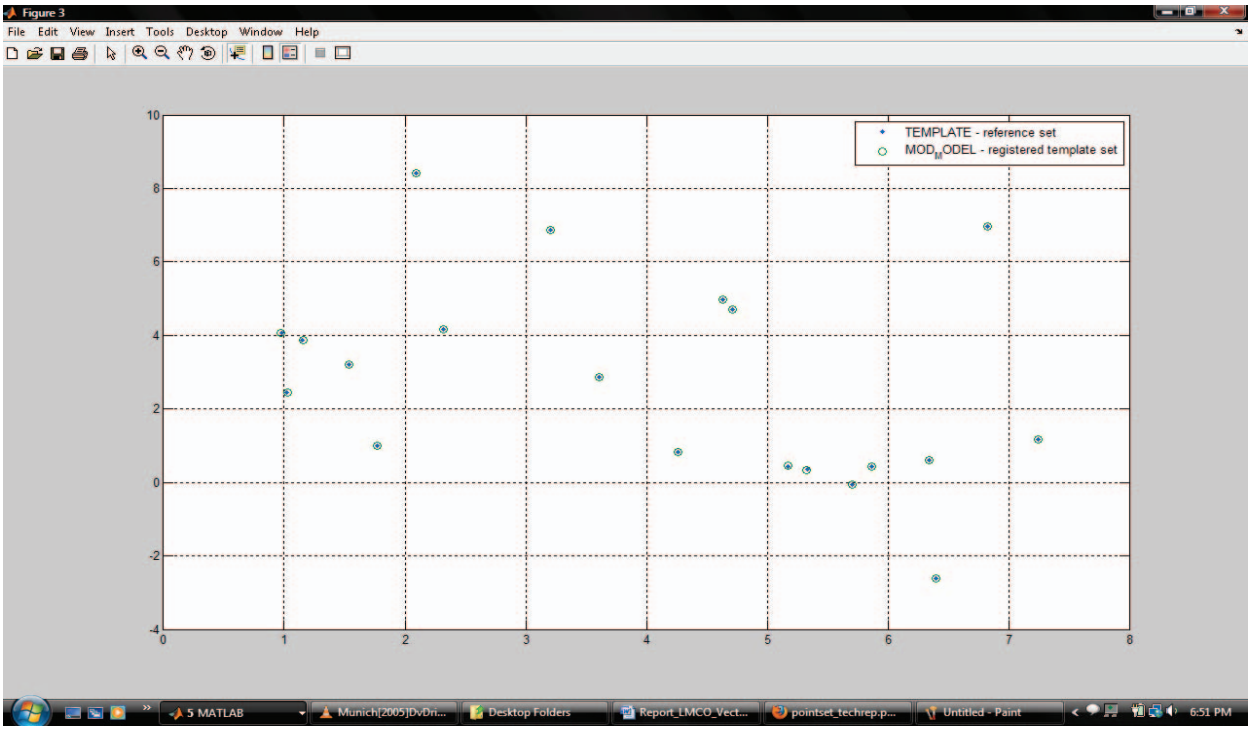

Figure 2: Conflated Data se.Here we see Model data set right on top of Template data set

REFERENCES

[1] Boris Kovalerchuk and James Schwing , "Visual and Spatial Analysis: Advances in Data Mining, Reasoning and Problem Solving", *Eds. , Springer*, February 2005.

[2] B. Kovalerchuk and R. Chase , "Image Analysis Method for Overcoming Source Distortion Using Algebraic Invariant Methods" , *SPIE Defense and Security Symposium*, Orlando, FL, 28 March - 1 April 2005.

[3] Ching-Chien Chen , Craig A. Knoblock , Cyrus Shahabi, "Automatically and Accurately Conflating Raster Maps with Orthoimagery" , *Geoinformatica, v.12 n.3, p.377-410*, September 2008.

[4] Anand Rangarajan, Baba C. Vemuri, Alan L. Yuille: "Energy Minimization Methods in Computer Vision and Pattern Recognition", *5th International Workshop, EMMCVPR 2005*, St. Augustine, FL, USA, November 9-11, 2005, Proceedings Springer 2005.