Analytical Conflation of Spatial Data from Municipal and Federal Government Agencies

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By
Hoseok Kang, B.S. M.S.

****
The Ohio State University
2002

Dissertation Committee:
Professor Alan Saalfeld, Advisor
Professor Harold Moellering
Professor Raul Ramirez

Approved By

Geodetic Science and Surveying
Graduate Program
ABSTRACT

As spatial data resources become more abundant, the potential for conflict among them increases. These conflicts can exist between two or many spatial data sets covering the same areas and categories. Therefore, it becomes increasingly important to be able to effectively relate these spatial data sources with others and then create new spatial datasets with matching geography. One extensive spatial dataset is the Census Bureau's TIGER file, which includes Census tracts, block groups, and blocks. At present, however, Census maps often carry information that conflicts with municipally-maintained detailed spatial information. Therefore, in order to fully utilize Census maps and their valuable demographic and economic information, the locational information of the Census maps must be reconciled with the more accurate municipally-maintained reference maps and imagery. Several algorithms for facilitating map conflation already exist. The appropriateness of a map conflation algorithm depends on the quality of the source and reference maps, the level of required accuracy, the scales of the maps, and the availability of auxiliary spatial or non-spatial information. None of the current methods explicitly address topology modification. Therefore, those methods can only merge maps when the two members of a matched pair of common features are topologically equivalent.
We present two different map models corresponding to two different mathematical approaches to map conflation. The first model is based on the cell model of map in which a map is a cell complex consisting of 0-cells, 1-cells, and 2-cells. The second map model is based on a different set of primitive objects that remain homeomorphic even when generalized. The second model facilitates map conflation by guaranteeing the existence of both local and global homeomorphisms everywhere. Matching operations in map conflation are defined in one of two ways, depending on the map model used. The second model allows for the possibility of non-homeomorphic matching. The first model does not. If corresponding matching features within a subregion happen to have the same dimension and be homeomorphic, then more conventional matching strategies may be applied. If corresponding features, however, are not homeomorphic, then an operation called topological surgery may be implemented. A 0-cell conflation test is proposed to deal with the non-homeomorphic case. A new hierarchical based map conflation is also presented to be applied to physical, logical, and mathematical boundaries and to reduce the complexity and computational load. Map conflation techniques developed explicitly for Census maps are formulated and implemented. These new methods guarantee producing a conflated result that is topologically consistent with reference maps. This dissertation also describes how national maps could be updated effectively by exchange of map information between federal and local governments.

Implementation issues are discussed in terms of the theoretical design of map conflation system and results presented for the case study area, Delaware County, OH. We selected three area types, downtown, residential, and water-feature intensive areas, on
which to test our conflation strategies. We also present three transformation methods two of which offer new approaches to map conflation. The traditional method is Delaunay Triangulation (DT). The two new methods are Weighted Delaunay Triangulation (WDT) and linear feature based transformation (LineMorp). Our study results indicate that transformed map features approximate ground truth as follows, ordered from best to worst: LineMorp, WDT, and DT. New relationships between map transformation and map generalization or map revision are found from the transformation results.
Dedicated to my wife and our family
ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my adviser, Dr. Alan Saalfeld, for his intellectual guidance, enthusiasm, and helpful discussions that made this dissertation possible. His influence will be felt for the rest of my life. I am also grateful to Dr. Harold Moellering for constructive suggestions and his enlightening lectures on analytical cartography. I would also like to acknowledge Dr. Raul Ramirez for his valuable instructions and comments.

My sincere appreciation goes to Ms. Shoreh Elhami, GIS director at Delaware County Auditor’s Office, Ohio for giving me exposure to the GIS world. Without her support, this dissertation would not have been done.

I also appreciate the work of Craig Rollins at the Geodetic Science Department who did some proofreading for English grammar.

And last but definitely not least, I am indebted to my wife, Hyunjung Yu, as well as my mother and parents-in-law, who supported me, and encouraged me to reach my potential over the last seven years. I would like to thank my sons, Andy and Daniel, who add the joy of life to every day. Finally, I would like to honor my late father who would have been proud of me.

This research was supported by the NIMA grant under NMA201-01-1-2012.
VITA

June, 1967 ........................................... Born – Seoul, South Korea

1991 ................................................. B.S. Oceanography,
                                  Inha University, South Korea

1992 – 1995 ..................................... The Korea Ocean Research &
                                  Development Institute (KORDI),
                                  South Korea

1997 ................................................. M.S. Geodetic Science and Surveying
                                  The Ohio State University

1996 – Present ................................. Graduate Research or Teaching Associate,
                                  Center for Mapping or Geodetic Science,
                                  The Ohio State University

2001 – Present ................................. GIS Programmer
                                  Delaware County’s Auditor Office, OH
PUBLICATIONS

Research Publication


FIELDS OF STUDY

Major Field: Geodetic Science and Surveying

Studies in Computer Mapping and Geographic Information System
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Background and Problem Statement</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Dissertation objectives and scope</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>Significance of this research</td>
<td>6</td>
</tr>
<tr>
<td>1.4</td>
<td>Organization of the dissertation</td>
<td>6</td>
</tr>
<tr>
<td>2.</td>
<td>Literature Review of Map Conflation</td>
<td>10</td>
</tr>
<tr>
<td>2.1</td>
<td>Match-and-merge Paradigm</td>
<td>10</td>
</tr>
<tr>
<td>2.2</td>
<td>Methods for Matching Map Features</td>
<td>13</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Graph Match Methods</td>
<td>13</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Computer Vision Methods</td>
<td>15</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Match Applications in Agents</td>
<td>17</td>
</tr>
<tr>
<td>2.2.4</td>
<td>Classification of Matching Map Features</td>
<td>18</td>
</tr>
</tbody>
</table>
2.3 Methods for Merging Matched Features ................................................. 20
  2.3.1 Rubber-Sheeting Transformation ..................................................... 20
    2.3.1.1 Exact Alignment on Matched Points ........................................... 20
    2.3.1.2 Transformation Methods ......................................................... 21
  2.3.2 Statistical Fitting ............................................................................. 28
2.4 Early Activities for Improving TIGER at the Census Bureau of the United
  States ......................................................................................................... 29
2.5 Interactive Map Conflation .................................................................... 33
2.6 Summary ................................................................................................. 36

3. Map Conflation Theory and Methodology .................................................. 38

  3.1 Introduction to Map Conflation ............................................................. 38
  3.2 Mathematical Model of Maps ............................................................... 44
    3.2.1 The Cell Model of Maps ................................................................ 45
    3.2.2 A Homeomorphism Model for Map Conflation ............................. 46
  3.3 Map Conflation Principles ................................................................... 47
  3.4 Hierarchical Conflation ....................................................................... 49
  3.5 Heterogeneous Map Conflation ............................................................ 54
  3.6 Anchor Nodes ....................................................................................... 55
  3.7 Map Feature Matching ......................................................................... 56
    3.7.1 Generalization and Matching ......................................................... 61
      3.7.1.1 Cartographic and Digital Generalization ................................. 61
      3.7.1.2 Graphic and Human Visual Resolution Limitation ................ 64
      3.7.1.3 Generalization and 0 cell Match ............................................. 65
      3.7.1.4 Generalization and 1 cell Match ............................................. 68
    3.7.2 Graph Theory and Match ............................................................... 69
  3.8 Summary .............................................................................................. 75

4. Map Transformations in Map Conflation .................................................... 76

  4.1 Introduction to Map Transformation .................................................... 76
  4.2 Constrained Delaunay Triangulation .................................................. 78
  4.3 Piecewise Rubber-Sheet Map Transformation by Weighted Delaunay
      Triangulation (WDT) ........................................................................... 83
  4.4 Linear Feature Based Transformation (LineMorp) ............................. 87
  4.5 Restore One-to-one Relation of Control Points on the Boundary .......... 92
  4.6 Summary .............................................................................................. 93

5. The Spatial Structure of Census Maps ....................................................... 95

  5.1 Introduction to Census Maps ............................................................... 95
  5.2 GBF/DIME .......................................................................................... 97
  5.3 TIGER ................................................................................................. 99
Appendix A. GIS in Local Government, Delaware County Ohio

A.1 Introduction ......................................................... 183
A.2 Delaware County and Its Administrative Organization .................. 185
A.3 The History of GIS in Delaware County, Ohio .......................... 190
A.4 Spatial Data Creation, Update, and Management in Delaware County, Ohio ......................................................... 192
   A.4.1 The Cost of Spatial Data Creation .................................. 193
   A.4.2 Available Spatial Data ............................................. 194
   A.4.3 GIS staffs and GIS Software and Hardware ..................... 197
   A.4.4 Spatial Data Management and Update ............................ 200
      A.4.4.1 Map Sheet and Parcel Identification Number ............... 202
      A.4.4.2 Subdivision .................................................. 205
      A.4.4.3 Quality Control (Q/C) ...................................... 206
      A.4.4.4 Spatial Data Backup ........................................ 208
A.5 The role of GIS in Local Government ...................................... 208
   A.5.1 Census Project .................................................. 209
   A.5.2 Ohio's Current Agricultural Use Valuation (CAUV) .......... 210
   A.5.3 Internet GIS ..................................................... 212

Appendix B. List of Avenue Codes .............................................. 216

   B.1 Avenue Script that Builds a Polyline Shape File ..................... 216
   B.2 Avenue Script that Builds a Polygon Shape File .................... 218

List of References ....................................................................... 220
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 CSS Image and Its Maxima (Morktarian, 1996)</td>
<td>16</td>
</tr>
<tr>
<td>2.2 Type of Map Feature Match</td>
<td>19</td>
</tr>
<tr>
<td>2.3 Transformation using Simplicial Coordinates</td>
<td>23</td>
</tr>
<tr>
<td>2.4 Influence Regions (Doytser, 1995)</td>
<td>24</td>
</tr>
<tr>
<td>2.5 Type of Map Transformation in Map Conflation</td>
<td>27</td>
</tr>
<tr>
<td>2.6 Conflated and Geo-coded MAF (blue flags) vs. the Geo-coded Voter Registration (red flags) in a Subdivision, (Elhami, 2000)</td>
<td>32</td>
</tr>
<tr>
<td>2.7 Quality Control Interface for Matching Pairs</td>
<td>34</td>
</tr>
<tr>
<td>2.8 Alignment for Unmatched Linear Feature by an Interactive Method</td>
<td>35</td>
</tr>
<tr>
<td>3.1 Mapping relationships between Real World and Hardcopy Map, and between Hardcopy Map and Computer-assisted Vector Map Features</td>
<td>39</td>
</tr>
<tr>
<td>3.2 Mapping Relationship between Real world and Augmented Map</td>
<td>40</td>
</tr>
<tr>
<td>3.3 Spatial Processes for Map Conflation</td>
<td>49</td>
</tr>
<tr>
<td>3.4 Hierarchical Relationship of U.S. Census Geographic Entities (Census Bureau, 2000)</td>
<td>51</td>
</tr>
<tr>
<td>3.5 Hierarchical Subdivision of Census Tract, Block Group, and Block</td>
<td>54</td>
</tr>
<tr>
<td>3.6 Heterogeneous Strings on a Polygon Boundary</td>
<td>55</td>
</tr>
<tr>
<td>3.7 Junction Node</td>
<td>56</td>
</tr>
<tr>
<td>3.8 Neighbor Checking in Node Matching</td>
<td>59</td>
</tr>
</tbody>
</table>
3.9 Spatial Transformations (Buttenfield, 1991)........................................64
3.10 Generalizations of Linear Feature that Change Geometry and Topology........66
3.11 0-cell Conflation Test Illustrated.........................................................67
3.12 Topological Survey Operation...............................................................68
3.13 Finding Matched String After 0-cell Match...........................................71
3.14 Tolerance Buffer from the String A’- B’..................................................72
3.15 Non-path Edges Removed from a Graph...................................................74
4.1 Delaunay Triangulation and Constrained Delaunay Triangulation...............82
4.2 DT and CDT for One Census Tract Polygon.............................................83
4.3 Voronoi Diagram and Delaunay Triangulation........................................84
4.4 Influences of the Second-order Voronoi Regions....................................85
4.5 The Comparison of DT (piecewise rubber-sheeting by Delaunay Triangulation) and WDT (piecewise rubber-sheeting by Weighted Delaunay Triangulation) Transformation.................................................................87
4.6 Transformation for Single Line Match Pair.............................................89
4.7 Transformation for Multiple Line Match Pairs..........................................90
4.8 The Boundary Change from Source to Reference Polygons..........................92
4.9 The Transformation of Inside Points between Source and Reference Polygons.....93
5.1 Sample Map for Topological Geocoding (Saalfeld, 1999)............................99
5.2 GBF/DIME Structure.............................................................................100
5.3 TIGER File Structure (Saalfeld, 1999).......................................................102
5.4 Major Steps to Build and Manage TIGER Files for 1990 Census (Marx, 1984)....104
5.5 Distance Differences between TIGER and GPS Surveyed Match Pairs (Liadis, 2000)..........................................................196

xiv
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.6 The Locations of Control Points</td>
<td>107</td>
</tr>
<tr>
<td>5.7 Distance Difference between Delaware GIS Base and Census Maps</td>
<td>107</td>
</tr>
<tr>
<td>6.1 Logical System Design for Map Conflation</td>
<td>111</td>
</tr>
<tr>
<td>6.2 Information Reality in Map Conflation</td>
<td>113</td>
</tr>
<tr>
<td>6.3 Information Structure</td>
<td>114</td>
</tr>
<tr>
<td>6.4 Canonical Structure</td>
<td>116</td>
</tr>
<tr>
<td>6.5 Data Structure</td>
<td>117</td>
</tr>
<tr>
<td>6.6 ArcView® Screen Design and Customization</td>
<td>120</td>
</tr>
<tr>
<td>6.7 Customized Tools</td>
<td>121</td>
</tr>
<tr>
<td>6.8 Census Tracts</td>
<td>122</td>
</tr>
<tr>
<td>6.9 Census Block Groups</td>
<td>123</td>
</tr>
<tr>
<td>6.10 Census Blocks</td>
<td>124</td>
</tr>
<tr>
<td>6.11 Digital Orthophoto</td>
<td>125</td>
</tr>
<tr>
<td>6.12 Road Center Lines</td>
<td>126</td>
</tr>
<tr>
<td>6.13 Railroads</td>
<td>127</td>
</tr>
<tr>
<td>6.14 Hydrographic Boundary</td>
<td>128</td>
</tr>
<tr>
<td>6.15 Township Boundary</td>
<td>129</td>
</tr>
<tr>
<td>6.16 Municipal Boundary</td>
<td>130</td>
</tr>
<tr>
<td>7.1 Conflict Condition between Census Map and Reference Map in Downtown</td>
<td>135</td>
</tr>
<tr>
<td>Area</td>
<td></td>
</tr>
<tr>
<td>7.2 Conflict Condition between Census Map and Reference Map in Residential Area</td>
<td>136</td>
</tr>
</tbody>
</table>
7.3 Conflict Condition between Census Map and Reference Map in Hydrographic Area

7.4 RMSEx and RMSEy in Downtown, Residential, and Hydrography Area

7.5 The Histogram of Distance Difference between Intersection and Turning Nodes and Their Matching Ground Truth

7.6 Control Points on Tract Boundary in Downtown Area with Inside Block Polygons

7.7 Control Points on Tract Boundary in Residential Area with Inside Block Polygons

7.8 Control Points on Tract Boundary in Hydrographic Area with Inside Block Polygons

7.9 Q/C Process for Matched Pairs by Visualization

7.10 Finding a Target String Search

7.11 The Replacement of the Source Boundary by the Reference That Has Geometric and Topologic Consistent Conflation Results

7.12 Updated Boundary by the Matched String with Digital Orthophoto as Background

7.13 Updated Boundary by Standard Rubber-Sheeting Transformation

7.14 Comparison Before and After Transformation for Three Areas

7.15 The Distribution of Control Nodes on the Boundary and Matched Points inside the Boundary

7.16 The Histogram of X and Y Distance Difference for the Inside Points in the Downtown

7.17 An Example of Partial Match in the Case of Reference Maps Is Not Available

7.18 An Example of Partial Match in the Case Topology Surgery Is Needed

A.1 Delaware County with Municipalities and Townships Boundaries, Ohio Location
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Hierarchical Relationship of Census Tract, Block Group, and Block</td>
<td>53</td>
</tr>
<tr>
<td>5.1 Some Examples of Census Map Products</td>
<td>97</td>
</tr>
<tr>
<td>5.2 The Collected Anchor Points by Update Operation and Tract, Delaware County, OH (Liadis, 2000)</td>
<td>105</td>
</tr>
<tr>
<td>7.1 RMSE, Mean Distance, Range, and SD between Intersection and Turning Nodes and Their Matched Ground Truths</td>
<td>138</td>
</tr>
<tr>
<td>7.2 The Number of Match Pairs according to the Exact and the Partial Matched</td>
<td>145</td>
</tr>
<tr>
<td>7.3 RMSE, Mean Distance, Range, and SD of Control Nodes on the Boundary</td>
<td>154</td>
</tr>
<tr>
<td>7.4 Comparison of the Three Transformation Methods in the Three Areas</td>
<td>156</td>
</tr>
<tr>
<td>7.5 Comparison of RMSE after Translation by Mean X and Y Difference</td>
<td>163</td>
</tr>
<tr>
<td>A.1 Land Uses of Parcel and Their Area, 2002 Year (1st Quarter)</td>
<td>184</td>
</tr>
<tr>
<td>A.2 Activities of DALIS Project in Delaware County, Ohio</td>
<td>190</td>
</tr>
<tr>
<td>A.3 Spatial Data Creation Costs</td>
<td>191</td>
</tr>
<tr>
<td>A.4 Positional Accuracy of Base Map, Digital Orthophoto</td>
<td>192</td>
</tr>
<tr>
<td>A.5 Available Spatial Data in Delaware County, Ohio</td>
<td>193</td>
</tr>
<tr>
<td>A.6 The Attributes of Parcel Coverages</td>
<td>194</td>
</tr>
<tr>
<td>A.7 The Attributes of Road Centerline</td>
<td>195</td>
</tr>
<tr>
<td>A.8 Used Hardware and Software</td>
<td>198</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Background and Problem Statement

As spatial data resources become more abundant, the potential for conflict among them increases. These conflicts can be described as both spatial and non-spatial and can exist between two or many spatial data sets covering the same areas and categories. Using a uniform spatial reference when creating multiple spatial datasets could prevent these conflicts to a great extent. However, using one spatial reference is almost impossible since there are so many independent agencies that create spatial layers based on their own internal needs and purposes. Therefore, it becomes more important to be able to effectively relate these spatial data sources with others and then create spatial datasets with matching geography. For example, spatial data that has more positional accuracy can be related with other spatial data that has less positional accuracy but more valuable attribute information to get a new consistent map that has more positional accuracy than the original and at the same time valuable attributes. One such spatial dataset is the Census Bureau’s TIGER file, which includes Census tracts, block groups, blocks, and also roads, railroads, limited hydrology, and etc. For the remainder of this dissertation,
the term Census Maps will be used for Census tracts, block groups, and blocks. Currently, Census Maps usually do not match the localities’ more precise mapping layers. Therefore, in order to accomplish this task, the geometry and topology of the Census maps must closely match the local and more accurate sources typically obtained from digital photography and/or road centerline coverages. The term ‘geometry’ represents coordinate locations and shapes and ‘topology’ means the relationships between points, line objects, and polygons (Census Bureau, 2000). Map conflation can be defined as combining two or more map data sets from different sources to form a new spatial data set with a unique representation and consistency. Automated map conflation has been implemented since Saalfeld (1993) suggested automated map conflation. Different versions of map conflation can arise in the following cases.

- When combining the same type layers of the same region.
- When combining the different type layers of the same region.
- When combining the same type layers of neighboring regions.

There are multiple reasons for the extensive difference between the geometry and topology of Census maps and local government’s GIS base. The first reason has to do with the scale at which each coverage was originally created. Census maps are created from 7.5’, 15’, and 1:100,000 scale hardcopy maps by heads-down digitizing or scanning (Callahan et al.). However local governments make their own base maps by field surveys or heads-up digitizing of digital orthophotos that are very high resolution to provide
parcel level information. For example, there are 3 scales of orthophotos that are at 1:1,200, 2,400, and 4,800 in Delaware County, OH. Other maps such as road centerlines and parcel boundaries are created at the same scales. When two different scale maps are overlaid at a larger scale, jagged effects cannot be prevented in the smaller scale map, because the smaller-scale map becomes just larger (zoom in) than its own scale without adding any details.

The second reason relates to generalization. Since the real world is too complex for our immediate and direct understanding and there is a limited space on a hardcopy map, generalization must be applied when a paper map is created. For example, generalization operations include selection, simplification, exaggeration, classification, displacement, symbolization, and so on. It is an ill posed problem to recover the original feature from the generalized one without extra information. This can be described as an inverse generalization problem. This can also be described as a conflict between analog and digital based data. These two main factors make map conflation complex and not easily defined. Since map conflation is an ill posed problem as discussed above, it is extremely difficult to find fully automated solutions.

1.2 Dissertation objectives and scope

The purpose of this dissertation is to make geometrically and topologically consistent source maps according to the reference maps. The selected spatial datasets as an example of map conflation are Census maps and local government’s GIS base, based on the following reasons:
• There are tremendous demands on Census statistics that are closely related to geography but unfortunately, the lack of positional accuracy makes them less use and less effective.

• Census maps are readily available in TIGER and other formats.

• Local governments maintain their own GIS base for various applications such as property taxation in which the positional accuracy is critical. These GIS datasets are a good source for map conflation because they contain up-to-date information, and also contain consistent and complete geometry and topology.

• Local governments’ main interest in Census maps is to compare them with the local maps and find discrepancies that may result in population undercounts. Finding these discrepancies will facilitate placing addresses in their accurate locations and therefore, identifying all residents within a jurisdiction. A variety of demographic information contained in the Census maps also assists local governments in conducting a variety of statistical and analytical functions in their entities.

• There is a nationwide demand to conflate Census maps with more accurate data sources to obtain more accurate positional accuracy and therefore, gain the ability to compare them with more accurate datasets at the local level.

There are three primary goals of this dissertation. The first goal is to formulate a conceptual framework and mathematical model of map conflation. In addition, it intends to develop, design, and implement practical algorithms to conflate source maps with reference maps. As the example of map conflation, Census maps and local governments’
GIS bases are used. This will provide a unique solution of map conflation for local governments. Map conflation has been applied to homogeneous spatial data types mainly. Map conflation of heterogeneous spatial data types has not been well defined or recognized. New approaches for homogenizing Census maps use heterogeneous data and hierarchical conflation will be used for localizing, matching, and transforming map features. New spatial functions for map conflation in irregular domains and heterogeneous spatial data types will also be developed.

The second goal is to develop a new match method. Therefore, a combined use of geometry, topology, and attributes to find matched features is proposed. Additionally, a conflation test that localizes match areas and makes new geometry and topology (free geometry and topology) for map features in the physical map is also developed based on a homeomorphism map model. More over, a new method to match nodes and strings between two maps, which are created at different scales, is developed. This will depend not only on shape analysis, but also graph theory and generalization rules to effectively isolate target nodes and strings.

The final goal of this dissertation is to investigate, explore, and analyze map conflation as a tool of transformation and integration of spatial data in GIS. For the hierarchical map conflation application, transformations through matched map features at each level may affect the result of the next match at the next level of conflation. The success of map conflation when source maps are not available is also dependent upon the type of transformation method. The selection of the method of transformation depends on the number of control points, the distribution of control points, and the properties of
transformation. There is a very close relationship between conflation and integration of spatial data because conflict could exist when spatial data are integrated.

Overall, this dissertation develops and evaluates algorithms of map conflation to provide the foundation research for map conflation in general and Census map conflation in particular.

1.3 **Significance of this research**

This research allows local governments to correlate their in-house detailed parcel data with Census maps that have demographic information at the block level. This new information permits very interesting and intricate statistical, sociological, and spatial analysis of growth and change patterns. This research also provides map conflation as a tool of spatial data integration because map conflict exists everywhere, not only in Census maps. The success of this research for combining spatial information to facilitate analysis and integrate spatial data will lead to more extensive use of these map conflation methods in the future.

1.4 **Organization of the dissertation**

The background, objectives, and significance of the research have been presented in the preceding sections. The organization of the dissertation is as follows:

**Chapter 2:** The overview of the existing map conflation algorithms is presented. The strengths and weakness of them are also discussed. The matching methods required for
finding corresponding features between two maps are classified as spatial or non-spatial and then several matching methods are discussed. The rubber-sheet transformation is thoroughly examined and then classified by local and global methods in terms of the used control points. Finally, the activities of the Census Bureau in the conflation of Census maps are introduced, and projects conducted by local governments using interactive approaches are also discussed.

**Chapter 3:** Two map models are provided for map conflation. The first model is based on the cell model of a map in which a map is a cell complex consisting of 0-cells, 1-cells, and 2-cells. The other model is based on enforcing that a homeomorphism condition is maintained among all maps. This Chapter provides a new approach in map conflation to find matched features that utilizes the homeomorphism between different scale contraction maps. Other match methods that use attributes, geometry, and topology are also discussed. The spatial algorithm to conflate source maps and reference is presented and then each component is thoroughly discussed. Census maps and local government’ GIS base are used as an example. The key terms are hierarchical and heterogeneous conflation, attribute embedding, finding meaningful nodes, conflation test, cartographic 0-cell match, and cartographic 1-cell match.

**Chapter 4:** This Chapter focuses on transformation methods in map conflation. A number of transformation methods with polygon boundary are theoretically examined for their use in hierarchical map conflation.
Chapter 5: This Chapter examines the brief history of Census Maps and their spatial data structure. The spatial data structures discussed in this chapter are GBF/DIME for the 1970 Census and the 1980 Census, and TIGER for the 1990 and 2000 Censuses. The topological geocoding issues of how to relate spatial relations to each other, which is a leading concept and technique in the digital mapping industry, are examined. Finally, the absolute positional accuracy of Census Maps and the relative position accuracy of Census Maps with the local government’ GIS base of Delaware County, OH, are presented. This demonstrates how Census Maps differ positionally from the real world.

Chapter 6: In terms of implementation of map conflation, a number of design activities are discussed. They are logical system design, cartographic system design, screen design, and communication design. The concepts of each design stage are presented and their purpose for map conflation is designed. This chapter also provides the descriptions of spatial data, hardware, software, and program from the implementation standpoint.

Chapter 7: The chapter presents the implementation results of the proposed map conflation and discusses the findings. Three areas such as downtown, residential, and hydrographic are selected because they represent unique characteristics. The transformation performances for the three methods are also discussed for those three areas.
Chapter 8: Research summaries, conclusions, and future works are provided in the final chapter.

Appendix A: This appendix discusses why GIS is needed, how a GIS base is created and managed, and what GIS’s role is in local government. It also illustrates that local government is the source of up-to-date, positionally accurate, and geometrically and topologically complete spatial data. The details of spatial data availability, spatial data history, the cost of creation, update, and backup are discussed for the case study area, Delaware County, OH.

Appendix B: This appendix provides Avenue scripts codes that are used for the interface between Avenue shape file and a user defined data format.
CHAPTER 2

Literature Review of Map Conflation

2.1 Match-and-merge Paradigm

One of main concepts of map conflation is this: If we know some points and their destination locations, other related points can be transformed based on piecewise-linear rubber-sheeting transformations using simplicial coordinates. Many algorithms and their mathematical contexts have been developed at the Census Bureau (White, 1985; Saalfeld, 1985, 1987; Rosen, 1985; Gillman, 1985; Lynch, 1985; Fagan 1986) and are found in Saalfeld (1993). Saalfeld (Saalfeld, 1993) defines map conflation as “the compilation or reconciliation of two variants of a map of the same region”. Another definition is shown in (Lynch, 1985), “combining of two digital maps to produce a third map file which is better than each of the component source maps”. However, the reference maps could be from one to many. Another definition for map conflation could be the combining two or more map data sets from different sources to form a new spatial data set with a unique representation and consistency. Different versions of map conflation can be considered as follows.

- When conflating the same type layers of the same region.
• When conflating different type layers of the same region.
• When conflating the same type layers of neighboring regions.

The first system of map conflation was developed by the Census Bureau to update USGS linework files (no attributes) using their own digital cartographic files (attributes such as name, street type, etc. included). Their goal was to (1) identify and resolve differences, and to (2) transfer name, address range, and other feature attribute information to the USGS files. The algorithm suggested by Dr. Saalfeld (Saalfeld, 1993) employs an iterative approach as follows.

*Standard Conflation Strategy*

1. Identify a few matching pairs of point features.
2. Rubber-sheet one map to bring it into exact alignment on the few matching pairs identified in step 1.
3. Repeat, until no new matches are found:
   A. Compute all nearest neighbor pairs as candidate matches.
   B. Apply configuration measures to confirm, disallow, or defer match classification of candidates found in previous step. If no new candidates are confirmed, relax the configuration match/similarity criteria and reapply them.
   C. Rubber-sheet to align confirmed matches of previous step.
The important part in this approach is the iterative feature-matching routines that are based on nearness or proximity of features. After matching points, map merging and rubber-sheeting routines are followed. In order to align a map area, a map area may be subdivided into non-overlapping triangles. The Delaunay triangulation (Saalfeld, 1987; Sibson, 1980, 1981; Farin, 1990) is unique and locally updateable and it may be built and updated in expected linear time with a suitable data structure. In addition to these, the map conflation system developed at the Census Bureau adopted a video game approach to provide an interactive system for semi-automatic map matching and merging (Lynch, 1975). By using this prototype system, over 90% of map features were successfully identified. However, the Census Bureau did not build a conflation production system because of the excessive cost that field verification of even a small percentage of unresolved differences would require (Saalfeld, 1993).

This algorithm and its underlying match and transformation concepts are used as the guideline of map conflation today and work well if the source and target maps are created at the same scale. This algorithm maintains the same number of map objects after conflation as well as the same topology. This can be a disadvantage if the source and target maps are created at different scales. In that case, making consistent geometry with large scale reference maps is important. In order to make geometry consistent with reference maps, string matches are performed after rubber-sheeting (Elhami, 2000). Candidate strings are selected from the edges of the triangulation. Therefore, string matches are dependent upon the size and distribution of each triangle. However, there is no control to pick candidate strings in match process that cause the limited ability in this
method to update linear features (Elhami, 2000). Other modified versions of map conflation have focused on the methods to effectively relate and match map features. In the next section, other proposed methods are reviewed.

2.2 Methods for Matching Map Features

Two subsections describe match methods based on graph theory and computer vision fields. The third subsection discusses the application of match. Finally, the classification of match method based on spatial and non-spatial categories. These categories are used because the cartographic data could be classified by spatial and non-spatial. The other categories could be used in any other fields.

2.2.1 Graph Match Methods

A graph \( G=(V, E) \) is a mathematical structure consisting of two sets \( V \) and \( E \). The elements of \( V \) are called vertices (or nodes), and the elements of \( E \) are called edges (Gross, 1998). Graphs have been used to analyze a wide range of spatial related problems in which nodes and their connections and the neighborhoods between them have some physical or conceptual interpretation. A map can be modeled by graph theory as a topological object consisting solely of 0-cells, 1-cells, and 2-cells (Barbara and Saalfeld, 1985). It is natural to match map features from smaller to larger, i.e., from 0-cell to 2-cell, because 2-cell match implies 1-cells and 0-cells, and 1-cell match engenders matched of the end point 0-cells of the 1-cells (Barbara and Saalfeld, 1985). In other words, if 0-cell
matches are passed for the map features of two maps, they are same in geometry. The 0-cell and 1-cell matches proposed by Census Bureau are as follows (Saalfeld, 1988).

- **Nearest neighbor pairings**: A pair of points is a nearest neighbor pair if the points are from different maps and each is the closest point from the other map to its paired point. These points are used for candidate matching points and defined by using a nearest point algorithm.

- **Degree of a 0-cell**: For example, the street intersection of one map which has 4 intersections should match to a point with 4 intersections on the other map with approximately the same angles.

- **The spider function of a 0-cell**: Street segments and other linear features may have infinitely many possible different directions. In order to reflect the closeness of direction, group similar directions into a single class.

- **Dependent matching routine**: It applies the relaxed matching criteria to 0-cells which are adjacent to already matched 0-cells and uses the 1-cell network (topological relationship) of the map to precipitate additional matches. After each matching, matched points are moved into perfect alignment and other points of the map are assigned to triangles whose vertices are the matched points.

Linear feature match approach, 1-cell match, is shown in (Filin, 1999, 2000a, 2000b; Gabay, 1994, 1995; Doytsher, 2001). Gabay (Gabay, 1994) uses a matching-tolerance based on the accuracy estimates of location of points on both source and target maps. A matching-tolerance between two points is $T = k \cdot m_{ab}$ where $k$ is the probability factor and
\[ m_{ab} = \sqrt{m_a^2 + m_b^2} \], \( m_a \) and \( m_b \) are the accuracy of each point. Two points can be defined as matching points if the distance between the two points less than \( T \). In the Gabay (Gabay, 1995), the feature correction for unmatched lines are conducted by constrained triangulation based on the matching lines. Matching lines are pre-defined as forced sides in the triangulation and the conformal and affine transformation is used. Two parcel maps that have different positional and topological descriptions were digitized from 1:1250 scale paper maps and were tested.

Filin (Filin, 1999) distinguishes map conflation into point feature based and line feature based conflation. Polyline projection mapping is used to find match polylines and transform them. Cumulative distances and perpendicular rays with breaking points are adopted to find match polylines. This approach will have great performance if there are well-defined linear features between source and reference maps. But finding those well-defined linear features automatically are key issues that are hard to implement in map conflation for real cartographic data. In order to find corresponding objects between two maps, geometric and topological criteria are used to detect counterpart nodes. The Breadth First Algorithm (BFS) is used to find the best fit path between two matched end nodes and the average distance between two paths are evaluated (Filin, 2000a; Doytsher, 2001).

### 2.2.2 Computer Vision Methods

Many attempts to describe the planar shapes or curves have been studied in the field of computer vision or pattern recognition fields. For example, chain coding (Freeman,
Curvature scale space (CSS) (Morktarian, 1986, 1996), and simple geometric border representations such as boundary length, bending energy, and signature (Sonka, 1999) are some of the description methods. Area or length differences between two curves are also reported in (Saalfeld, 1986).

CSS method compares curves by inflection points after increasing the smoothing factor. They were tested using the database of the outlines of fishes to find the similar shape fish if a pattern of the shape is queried (Figure 2.1). Figure 2.1 shows the inflection points according to increasing smoothing factor, $\sigma$. A similar concept can be applied with the Douglas-Peucker algorithm (Douglas, 1973) where the perpendicular distance from the baseline is used instead of Gaussian filter. For example, the more scale is decreased, the more important points are the remaining.

![Figure 2.1 CSS Image and Its Maxima (Morktarian, 1996)]
There have been some effects for map conflation in the field of photogrammetry. Their approaches mainly focus on methods to extract cartographic features from digital aerial photos (image space) and register them to a GIS base. In order to provide control points automatically for map conflation in which a raster map is the only available reference map, extraction of intersection points of roads from the scanned image is shown in (Uebbing, 1999). However, as the current extraction algorithms have limits in a real world situation, the proposed method applies the limited environment, that is, the scanned image of the road network.

2.2.3 Match Applications in Agents

Environmental Systems Research Institute (ESRI) uses simple feature-to-feature distance measurement between coverage for feature matching (Lupien, 1986). The used criteria (Lupien, 1986) are as follows.

1. A point feature in coverage A is matched if only one point in coverage B is within a specified tolerance distance. This tolerance is relaxed in successive runs that follow alignment iterations.

2. Arc features are matched in the same way, except that distance calculations are made from each arc vertex, including its from-and to-nodes, to the nearest point along candidate arcs. If each vertex of an arc in coverage A is within tolerance of an arc in coverage B, and if this is true for only one arc in B, then a match is recorded.
In the viewpoint of practical implementation of an automatic map conflation, Etak (Deretsky, 1993) suggests that the strategy of an automatic map conflation process should be (1) match as many features in both maps as possible, (2) bring unmatched 1-cells, which are topologically consistent in both maps and do not interfere with the existing data, (3) conflate attributes on matched cells, and (4) collect the un-utilized conflicting data to determine whether additional sources and manual work is necessary. Their objective is to remove any manual work, which are Q/C process for map match or alignment for unmatched map feature during map conflation process. However, those manual works can be done after or before process. Their map match methods use the attributes, the geometry and the topology of 0- and 1-cells.

2.2.4 Classification of Matching Map Features

Finding matched map features can be distinguished by spatial or non-spatial based methods (Figure 2.2). A spatial based method, which extensively uses geometry and topology, is divided by nodes, strings, and areas. Geometric matching refers to the similarity of distance, positions, and shapes. For example, in Figure 2.2, proximity, degree (the number of sharing), and spider (a class of directions) are included in Node match. Number (the number of vertex), direction, and length are included in String match. Finally, area, perimeter, and shape descriptions such as compactness \( \left( \frac{\text{perimeter}^2}{\text{area}} \right) \) are included in area, whereas, topological matching refers to the similarity of the relational structure of the points, lines, and areas. For example, in Figure 2.2, Neighbor means that check their neighbor match properties. A non-spatial based
method uses attributes that are associated with map features. For example, rd and road are treated as a match. In that case, knowledge or probability based rules can be also set up beforehand and applied. Matching spatial data sets using statistical approaches is also reported in (Walter, 1999). This method determines the probabilities for each match criteria based on the pre-selected matched pairs. For example, in order to determine the threshold of the angle match, pre-selected matched pairs are examined.

![Map Feature Match Diagram]

Figure 2.2. Type of Map Feature Match

Besides map feature matching, one of the most important factors in the map conflation is map transformation because unmatched map objects should be transformed based on
matched pair points if there is no perfect match for all map objects between the source and reference maps. In the next section, a rubber-sheeting transformation is investigated as a method for the transformation of map objects.

2.3 Methods for Merging Matched Features

This section discusses map-merging methods and classifies them into local and global methods based on the control points used. This classification is one of examples that map transformations are classified.

2.3.1 Rubber-Sheeting Transformation

After finding matched features, especially point features that will be used as control points, a transformation is implemented to merge two maps covering the same area. Rubber-Sheeting Transformation plays an important role in bringing matched points into exact alignment and in transforming unmatched points according to the movement of matched points. First, it divides the whole area into subsections and then it stretches each area defined by matched points to meet its predefined corresponding area. The triangulation is the most popular subdivision method because it has unique characteristics.

2.3.1.1 Exact Alignment on Matched Points

This is the first step of a rubber-sheeting transformation. For matched points, one-to-one alignment is implemented since they are already proved to be the destination locations of
candidate points in the map match. Usually, candidate points are selected as well
distinguished point features such as intersections and critical points so that the chance of
mismatch cases is reduced. These exact aligned match points are used when
transformation parameters are calculated.

2.3.1.2 Transformation Methods
Saalfeld (Saalfeld, 1993) defines Rubber-sheeting (Homeomorphism) as one of map
transformations that allows local distortion or perturbation, but preserves topological
properties and is linear on subsets of the map. There exist several versions of rubber-
sheeting transformations. The traditional Piecewise Linear Rubber-sheeting
(Homeomorphism) or PLH transformation divides the whole area into triangles whose
vertices are node match pairs. Then simplicial coordinates based on the vertices of each
triangle are used to determine the transformation parameters for the points inside each
triangle. PLH preserves the topological neighborhood structure that is very important for
cartographic data. One disadvantage is that it may not preserve straightness at the triangle
boundaries. However, additional match pairs can be added to prevent this defect.
Simplicial coordinates utilize the elementary geometric properties of triangles and are
easy to compute. Linear expression and the area of the triangle are used to compute
simplicial coordinates. Suppose \((x_1, y_1), (x_2, y_2), \text{ and } (x_3, y_3)\) are three vertices of a
triangle (Figure 2.3). Then, area-producing coefficients such as \(A_1, B_1, C_1, A_2, B_2, \text{ and } C_2\)
for any point \((x, y)\) in the triangle can be defined. The actual expressions for \(A_1, B_1, C_1, A_2, B_2, \text{ and } C_2\) are given below:
\[ A_1 = (x_3 - x_2)/T; \]
\[ B_1 = (y_2 - y_3)/T; \]
\[ C_1 = (x_2y_3 - x_3y_2)/T; \]

\[ A_2 = (x_1 - x_3)/T; \]
\[ B_2 = (y_3 - y_1)/T; \]
\[ C_2 = (x_3y_1 - x_1y_3)/T; \]

where \( T = x_1y_2 + x_2y_3 + x_3y_1 - x_3y_2 - x_2y_1 - x_1y_3 \)

Finally, the simplicial coordinates of \((x, y)\) as a convex combination of \((x_1, y_1)\), \((x_2, y_2)\), and \((x_3, y_3)\) are given by:

\[ s_1 = A_1y + B_1x + C_1 \]
\[ s_2 = A_2y + B_2x + C_2 \]
\[ s_3 = 1 - s_1 - s_2 \]

Suppose the corresponding three vertices of \((x_1, y_1)\), \((x_2, y_2)\), and \((x_3, y_3)\) are \((X_1, Y_1)\), \((X_2, Y_2)\), and \((X_3, Y_3)\). The transformed point \((x_i, y_i)\) of \((x, y)\) using simplicial coordinates is given by:

\[(X_i, Y_i) = s_1(X_1, Y_1) + s_2(X_2, Y_2) + s_3(X_3, Y_3)\]
Doytsher (Doytsher, 1995) uses rubber-sheeting for cadastral maps to homogeneously spread distortions that occur at the block borders as a result of junction adjustment, edge matching, and line adjustment. His boundary is not a triangle but a polygon. The influence region is based on geometric relations between the polygon boundary and the points inside the polygon (Figure 2.4).

Point T is the transformed according to the movement of polygon boundaries (inside and outside). Influence regions are defines by the regions “seen” from point T (Doytsher, 1995). Highlight boundaries by thick lines are influence regions in Figure 2.4. The size of influence of region and its distance from the point (here, T) are used for transformation parameters. The proposed transformation equation is shown in (Doytsher, 2001).
\[ \Delta X = \frac{\sum_{j=1}^{m} \sum_{i=1}^{n} w_{j,i} \Delta x_{j,i}}{\sum_{j=1}^{m} \sum_{i=1}^{n} w_{j,i}} \]

\[ \Delta Y = \frac{\sum_{j=1}^{m} \sum_{i=1}^{n} w_{j,i} \Delta y_{j,i}}{\sum_{j=1}^{m} \sum_{i=1}^{n} w_{j,i}} \]

\[ X_{\text{new}} = X_{\text{old}} + \Delta X \]

\[ Y_{\text{new}} = Y_{\text{old}} + \Delta Y \]

where:

\( X, Y \) : coordinates of a point to be transformed.

\( m \): number of polygons (outside plus inside holes).

\( n \): number of vertices in a given counterpart object.

\( \Delta x_{j,i}, \Delta y_{j,i} \): the translation of the vertex to its expected position.
$w_{ij}$: the influence (weight) of that vertex on the transformation of the given point.

The reciprocal of the square distance between a point and reference (before moving) vertex is used.

Doytsher (Doytsher, 1995) lists the characteristics of this rubber-sheeting transformation as follows.

1. Ensuring topological accuracy, so that points that were within the block will remain there (by a proper selection of influence regions).
2. Keeping widths of roads that are parallel to the block constant and spreading corrections of areas proportionally (by a proper selection of weights).
3. Allowing users control and decision both on the weighting of the borders and on the weight function by area.

In both Saalfeld (Saalfeld, 1993) and Doytsher (Doytsher, 1995), barycentric combination of type is used to form the equations. For example, if the $P_i$ are centers of gravity of objects with masses $m_i$, then their barycenter $P_c$ is given by Farin (Farin, 1996):

$$P_c = \frac{\sum_{i} m_i P_i}{\sum_{i} m_i} \text{ where } \sum_{i} m_i \text{ is combined mass.}$$

The combined mass is modified to reflect weights such as area or distance in rubber-sheet transformation. The rubber-sheet transformation has also been used to remove geometric distortions in satellite images in the field of remote sensing (Elliott, 1976; Green, 1975),
to improve old cartographic bases using DOQ (Digital Orthophoto Quadrangles) (Moore, 1999), or to add morphing effects for movie industry related images in the field of computer graphics (Wolberg, 1989, 1990). They work for regular domains in which many uniform properties such as size, shape, orientation, and the number of neighbors are kept.

The types of map transformations for map conflation are shown in Figure 2.5. The classification is based on how the control points that are used for calculating the transformation parameters affect the unmatched points. It can be divided into local and global methods. The whole area of the map is affected by a global method. Typically, a small number of control points are used to calculate the transformation parameters. If there is a systematic error in the map, a global method might be appropriate. However, in the majority of cases, it is not suitable for map conflation because there are so many factors associated each other. Therefore, the map features should be treated independently while maintaining topology.

A local method has some advantages because the map area can be divided into non-overlapping smaller areas so that local control points have local influences. The traditional way is the triangulation of the map area using match pairs. Then, area or distance based transformation parameters using the relationship between a point and triangle vertices can be calculated.
The only disadvantage of this method is that continuity is not guaranteed at the boundary of each triangle because the edges of each triangle are imaginary lines, and not the map feature itself. Polygon based transformations can be considered if a boundary exists. For example, parcel boundaries, building boundaries or, road networks can be used. The advantage of this method is that a critical boundary plays an important role in the transformation.
2.3.2 Statistical Fitting

Statistical or rule-based approach to map conflation is shown in (Wang, 1998; Cobb, 1998). Data registration module, road aggregation module, matching module, fusion module, and validation module are proposed in (Wang, 1998). Data registration brings data from different sources together and makes them commensurate. For example, digital road databases from different sources may be stored in different file formats such as DLG, VPF, etc. Their coordinate systems may refer to different geodetic datums and different map projections. A rule-based aggregation is used to combine several road segments. For examples, the following set of rules is proposed in (Wang, 1998).

1. Road segments can be aggregated only if they are from the same type of roads. For example, interstate roads and US highways cannot be combined.

2. Road segments can be aggregated only if their names are identical. If an actual road name is not available, then a default name “local road” will be used.

3. Aggregation should stop when it encounters a gap.

4. Aggregation should stop when it encounters an intersection at which the number of roads intersected is larger than 2.

The matching module consists of two steps: (1) construction of descriptions for each road network, and (2) probability-based matching algorithms. Road descriptions are centroid, length, orientation, straightness, moments. Using these descriptions of each road segment, continuous relaxation-based and belief network-based algorithms are applied to find match pairs. The fusion module consists of replacement, merging, estimation, and
combination. Finally, the interactive validation module is interactively implemented for quality analysis and control.

Cobb (Cobb, 1998a, 1998b) presents a knowledge-based system that utilizes semantic interrelationships of feature attributes map for VPF data due to its rich information content. "Closeness" and "Similarly" of fuzzy concepts and feature grouping are essential in their approach and both spatial and non-spatial properties are used in feature matching and deconfliction. Classification is used to store attribute data and match features. Other shape similarity tests are also used to help the match process.

2.4 Early Activities for Improving TIGER at the Census Bureau of the United States

As mentioned earlier, the first system of map conflation in Census Bureau used USGS map as the underlying truth. Since then, map revision issues have been raised to update old digital maps such as USGS and Census maps. Census Bureau has recognized two critical deficiencies in TIGER files (Broome, 1998). First, current and past updating has generally failed to improve the positional accuracy and shape fidelity of the features in the database. Second, the TIGER system has not been able to fully utilize the growing number of high quality digital files available from other sources. The Bureau has launched several pilot projects (O'Grady, 1999; Hartung, 1999; Godwin, 1999) that have these four objectives:

1. Develop and test a prototype method to improve the positional accuracy of the coordinates in TIGER/Line files.
2. Develop and test a prototype method to improve the content of TIGER/Line files by adding new road features and/or increasing the shape fidelity of the existing features through densification of the number of coordinates.

3. Develop and compare measures of the relative efficiency, including costs, time and other resources consumed by the DOQ and GPS methods as employed in this test.

4. Develop a data quality classification scheme for currentness, coordinates accuracy and shape fidelity.

The way tested to improve the positional accuracy of TIGER is to bring more accurate coordinate values that can be used as anchor points in rubber-sheeting transformation.

Godwin, O'Grady, and Hartung (O'Grady, 1999; Hartung, 1999; Godwin, 1999) present an economic way to collect anchor points by using DOQ or GPS. In the test site, Newberry County, SC, 3,723 DOQ anchor points are selected by taking 200 hours at the rate of approximately 20 DOQ anchor points per hour. The further components are suggested as potential software developments for fast image display and retrieval, image enhancement tools, and improving analyst interpretation skills (O'Grady, 1999). The advantages of using DOQ are suggested as follows (Godwin, 1999). They are availability, cost efficiency, distribution media, ease of use, overlay, true positions, and resolution and detail. However, The resolution of DOQ could be a disadvantage. The current available high resolution DOQ is 1 meter. In other words, features over 2 meters are clearly identified based on sampling theory. Therefore, the positional accuracy is bounded within that number. A GPS test project was also conducted in the same area
over a two week survey period (Hartung, 1999). Road centerlines and housing unit structures were captured as well as anchor points. The advantage of using GPS is that it provides highly accurate positional data but the disadvantages are that it involves field survey, which could be very expensive and systematic data processing is needed to maintain the same accuracy.

Their projects conclude that it is too preliminary to support any specific approach to help census map conflation, but the TIGER update system must be flexible enough to accommodate quality geographic data based on GPS, DOQ, or any procedure used by those who wish to share in the improvement of TIGER. The Census Bureau continually updates TIGER Line files and feature attributes through both addition and deletion. However, updates are not conducted in any systematic manner throughout the files because the Census Bureau has not formally tracked changes to the files in the past (Weaver, 1999). Weaver (Weaver, 1999) presents qualitative and quantitative measures of the level and nature of changes over time by providing distance network measures and address range measures. This result can be used to determine the importance of the amount of change and to track changes in the TIGER/Line files between versions.

Other approaches from the Census Bureau allow local governments to participate in updating Census data using their own GIS bases. This will provide great opportunities to improve TIGER files for the following reasons:

- The local governments’ GIS bases already contain the desired data. (i.e., digital orthophoto, GPS surveyed point data, large scale maps, etc.)
- The local data are complete and consistent in geometry and topology.
• Duplicate effort in creating reference data by the Census Bureau can be prevented.

• The staffs at local governments’ GIS departments are very familiar with their data and area.

Figure 2.6: Conflated and Geo-coded MAF (dark black flags) vs. the Geo-coded Voter Registration (light black flags) in a Subdivision, (Elhami, 2000)

As part of the Census 2000 activities, LUCA (Local Update of Census Addresses) was launched to provide an opportunity to local governments to review the Census Bureau’s
Master Address File (MAF) (Elhami, 2000; Raza, 1999; Hudson, 1999). In order to efficiently add, modify, and delete the given census file within a certain period, it is necessary to consider implementing map conflation between local governments’ GIS bases and Census maps (Elhami, 2000). For example, conflated 2000 collection block coverage was used in Delaware County, OH for reviewing and making corrections to the MAF (Elhami, 2000). Due to the use of conflated blocks, the block number for every additional address was easily identified. Conflated blocks also helped in comparing multiple geo-coded files such as addresses in the MAF and voter registration file (Figure 2.5). At the conclusion of this program, the County submitted an additional 11,161 new addresses (39% of the total), 235 modifications, and 73 deletions from 28,939 original addresses in the given 3-month time frame.

2.5 Interactive Map Conflation

In order to overcome the uncertainty in map conflation, especially in map feature match, human interaction plays a critical role in implementing a map conflation system. In other words, we cannot fully automate the solution to a not-well-formulated problem. Practically, in that case, the best decision is offered based on human knowledge that current map conflation algorithms are not able to provide in difficult situations. An interactive system also provides an economic and accurate way to pick control points in the map revision process using digital orthophotos. This offers some advantages over independent source control points because it provides high relative geometric accuracy, inexpensive feature problem resolution, and fast processing. Quality processing for match
pairs was suggested in the early conflation system in the Census Bureau and is also implemented by an interactive system in (Elhami, 2000). Its process consists of a control point table, a candidate check dialog, and a visualization window (Figure 2.7).

![Figure 2.7 Quality Control Interface for Matching Pairs](image)

In urban areas, it is relatively easy to recognize correct matches, unlike hydrographic areas where a decision can be difficult because the natural shape is irregular and the map generalization process has been applied excessively (Elhami, 2000). Unmatched features are aligned manually to references, which could be raster or vector maps. If a vector
reference map is available, the interactive system requires the selection of two end nodes in both maps and then replacing one with the other to speed up the process. Avenue program is used to make the interactive system in which matching points are selected by clicking two end points in the source and target maps and an automatic update is performed (Figure 2.8). Manual heads-up digitizing can be implemented if only raster reference maps are available. In summary, the incremental conflation with visual verification is cautious approach.

Figure 2.8 Alignment for Unmatched Linear Feature by an Interactive Method
2.6 Summary

In this chapter, we discussed the matching methods, map transformations, and Census activities to improve TIGER. We reviewed and classified matching methods from graph theory, computer vision, pattern recognition, and so on. Map transformations used for map conflation were reviewed and classified based on local and global. Census Bureau has recognized TIGER is positionally inaccurate and has made some efforts to correct them. But none of them have been selected to implement.

Several algorithms for map conflation have been developed in the past. But, it is hard to determine which is the best approach. The selection of an appropriate map conflation algorithm depends on the type of source and target maps, the level of required accuracy, the scales of the maps, and available spatial or non-spatial information. However, there are many similarities in their approaches. First, they utilize the information (geometry, topology, and attributes) associated with the spatial data by way of graph theory and other mathematical tools.

Second, if there is no information given or utilized, derived functions using statistics or probability theory are developed.

However, none of these methods consider topology correction. In other words, match methods are applied to a generalized topology that results in incorrect matching. The proposed map conflation theory provides topological correction after identifying visual and graph resolution limitation. This is supported by the homoeomorphism map model in which the bijectivity of every map feature is preserved. Since map conflation is a complicated system, it is very important to utilize well-organized and structured tools in
the process and it is only natural to provide a solution in this way. As mentioned earlier, map conflation is an ill posed problem. In other words, the given information is limited. We have to utilize the limited information to its fullest extent.

In the next chapter, the proposed map conflation theory is discussed.
CHAPTER 3

Map Conflation Theory and Methodology

3.1 Introduction to Map Conflation

It is impossible for a cartographer to map everything in the real world. Therefore, cartographic standards determine what gets depicted at any particular scale, and map generalization is adopted to deal with scale change.

In the 1960s, computer-assisted cartography was started, and prototype systems of GIS were built by several pioneers: Tomlinson and others (Tomlinson, 1988) in Canada; Chrisman, White, Franklin, Dutton, Morehouse, Marble, Poiker, and Dangermond (Chrisman, 1988; Goodchild, 1988; Dangermond, 1988) in the US; Rhind and Coppock (Rhind, 1988; Coppock, 1988) in the UK; Abel and others in Australia. Figure 3.1 shows a mapping relation from real world entities to hardcopy map objects, and from the hardcopy map objects to computer-assisted map objects. In the figure, "bijection" means every map object such as a road feature in the hardcopy map corresponds to exactly one object of the computer-assisted map. "Injection" (one-to-one, but not necessarily onto) means that map objects in the hardcopy map correspond to a subset of entities of real
world. This injection is by no means a bijection. There are many real-world mappable entities that DO NOT appear on this, or on any particular hardcopy map.

![Real world](image1.png) ![Hardcopy map](image2.png) ![Computer-assisted map](image3.png)

Real world | Hardcopy map | Computer-assisted map

Figure 3.1 Mapping Relationships between Real World and Hardcopy Map, and between Hardcopy Map and Computer-assisted Vector Map Features.

With the development of digital technologies, more extensive mapping of the real world in the form of digital map is possible by way of additional indirect or direct measurements. Indirect digital maps are created as follows.

1. Capture and store real world data into digital media such as a raster image. Here, assume that sample theory applies to determine a resolution that is sufficient to distinguish the cartographic objects to be mapped. In other words, the width of a road or a hydrographic feature is more than twice the resolution of the media. Therefore, the association of features on the medium and on the map is bijective.

2. Convert linear features on the digital medium to a digital map by digitizing.
Likewise, direct digital maps are created as follows.

1. Store real world location information into a digital database directly by way of GPS-data-capture.

![Diagram showing the mapping relationship between real world, digital orthophoto, GPS-aided survey, and digital map.]

Figure 3.2 Mapping Relationship between Real World and Augmented Map

Figure 3.2 shows each process. Approaches to resolving map differences have the strong mathematical and computational foundations of analytical cartography (Moellerling, 1991, 2000a), which seeks to solve cartographic problems involving spatial data structure, computation, transformation, and visualization. Therefore, it is natural to view conflation as part of analytical cartography. As stated in chapter 1, this dissertation will focus only on conflation of Census maps as an example with other heterogeneous spatial data. To develop a more general framework, however, we will first study broader issues:
the nature of maps and spatial datasets in general and of digital Census maps in particular, and principles for simultaneously reconciling multiple mathematical representations of spatial entities. Second, we will translate our mathematical representations into data structures and our reconciliation principles into algorithms. Finally, we will implement our algorithms to see their performance using Census maps and local government’s GIS base.

Let us begin by looking at available data categories and tools. Many local governments now maintain their own large-scale spatial data in the form of countywide digital orthophotos, parcel maps, topographic maps, road centerline maps, and so on. Agencies within these local governments use Geographic Information Systems to store, manage, analyze, and produce their spatial information. Through experience, these agencies have realized that there are discrepancies among these various spatial data sets themselves and with Census maps, and it is tedious and error-prone work to edit them manually.

The components of Census maps are clearly delineated in TIGER/Line documentation (Census Bureau, 2000). “Census maps such as tract, block group, and blocks usually are small statistical areas bounded on all sides by visible features such as streets, roads, streams, and railroad tracks, and by invisible boundaries such as property lines, legal limits, and short imaginary extensions of streets and roads.” The mission of U.S. Census Bureau is also clearly written in (Census Bureau, 2000), “The U.S. Census Bureau’s mission to count and profile the Nation’s people and institutions does not require very high levels of positional accuracy in its geographic products. Its files and maps are designed to show only the relative positions of elements.” (Even though their goal does
not require high levels of positional accuracy, they are making efforts to improve TIGER accuracy). This is one of reasons why there are discrepancies between them. There are other structural problems for the extensive difference between the geography of Census map boundaries and local government’s GIS base as discussed in chapter 5 and Appendix A. Since Census map boundaries are created by digitizing hardcopy maps, there are two main problems if more accurate maps are overlaid. The first problem is scale dependence. For example, Delaware county officials use 1 foot pixel resolution digital orthophotos as a base map. Census map boundaries, however, were created from 1:100,000 scale hardcopy maps. Uneven effects (coarse representation) cannot be avoided when a map scale factor is merely increased without adding or modifying appropriate feature details. The second relates to generalization. Since the real world is too complex for our immediate and direct understanding and there is a limited space on a hardcopy map, some generalization must always be applied when a hardcopy map is created. According to McMaster, (1992), the digital generalization process consists of ten spatial transformations (simplification, merging, etc.) and two attribute transformations (classification and symbolization). It is not possible to recover an original feature from a generalized one without extra information. This problem is an inverse generalization problem (Can we invert the generalization process? If not, can we generalize a large scale map to small scale in a way that helps to match features? If so, for which features or areas can we facilitate matching through appropriate generalization procedures?). Some feature generalization and graph matching problems are known to be NP (nondeterministic polynomial)-hard or NP-complete problems, also known as provably
intractable problems. In dealing with provably intractable problems and finding acceptable suboptimal solutions, several reasonable approaches are considered standard:

1. Look for special cases of the general problem that apply to a particular situation.
2. Add constraints or make limiting assumptions.
3. Change the problem domain.
4. Change your expectation for a solution by accepting piecewise or sub-optimal solutions.
5. Accept approximations after studying the latest results in approximation theory.

Even if a problem is not intractable, one may still seek to solve a simplified or restricted version of the problem, especially if a solution to the specialized version is all that is required in practice. The proposed map conflation algorithm guarantees geometric and topological consistency with reference maps provided reference maps exist. Based on analyses of the problems and information, suggestion (3), change the problem domain is appropriate to solve map conflation. Since geography feature can have a hierarchical structure, for example, Census tracts contain block groups, and block groups contain blocks, it is useful to conflate within hierarchical boundaries. The two main advantages are:
1) More source data are available at higher or general levels. For example, Census tracts follow most political boundaries such as township boundaries and county boundaries.

2) Transformations and matches are local in scope. For example, after conflating Census tracts, the transformations and match information for sets of Census block groups and blocks are constrained to fall within the Tract boundary. The computational complexity and computation load are reduced.

So far, the hierarchical boundaries have been defined by physical domain. In addition, mathematical or virtual boundaries can also be defined to focus matching operations. Map conflation can consist of nothing more than making explicit a homeomorphism (a bijective bicontinuous function) whose existence has just been proved. Corresponding features need to be mapped onto each other to establish a location correspondence. Features appearing on only one depiction need to be assigned a location on the other. The Jordan-Schönflies Theorem (for a statement of this theorem, see, for example, Mohar, 2001) implies that every point or feature that is interior to one or more closed simple loops must be matched with a unique point or feature within the corresponding loops. This property greatly reduces the search space for matching features to within a nested collection of subregions. The local interaction between generalization and scale is presented for one sample pair of mathematical boundaries.
3.2 Mathematical Models of Maps

Two distinct mathematical models are provided for maps. The first and more familiar model is based on a mathematical theory of cellular surfaces composed of idealized cell objects of zero, one, and two dimensions. The first model looks at graphs embedded in orientable surfaces. The second model is based on a different newly-formulated mathematical principle that requires that all map features—line, point, or otherwise—have non-zero area. This newly-stated mathematical principle, in turn, derives from the physical reality of perceptual acuity that requires an object to have significant extent in all directions in order for the object to be visible and be seen.

Whereas the first model allows for collapsing of features (forcing areas to become lower-dimensional lines or points), the second model need not. When a lake representation changes from a polygon with interior to a single 0-dimensional point, the transformation is many-to-one and, hence, not invertible. When a lake representation changes from a polygon with interior to a “fat” point (i.e., a small disk), then a local homeomorphism (a bi-continuous bijection) may still be established. This second theory helps to localize feature matching in map conflation by searching for and then establishing a one-to-one bi-continuous relation everywhere.

3.2.1 The Cell Model of Maps

In the traditional mathematical model for a map (Corbett, 1979), a map is a cell complex consisting of 0-cells, 1-cells and 2-cells. Removing the open 2-cells from a rectangular map or a globe leaves behind the 1-skeleton, a plane graph embedded in a compact region
of Euclidean space. This model remains the basis for many systems currently in use by US federal government agencies, including the Census Bureau’s TIGER files, the DLG files of the US Geological Survey, and NIMA’s Vector Product Format (VPF) files.

For conflating datasets that use the cell model, difficulties arise when no cell-structure-preserving homeomorphism exists. For example, one cannot match maps, point for point, when a corresponding feature pair is represented on the two maps by cells of different dimensions. Such cells of different dimension admit no homeomorphisms whatsoever between them. In addition, mapmakers occasionally conceptually change dimension of the object they are depicting (a 2D road has width, a 1D road centerline does not; a 2D manhole cover has area, its 0D point location on a map does not) without changing the drawn representation. Drawn map features are not idealized mathematical objects, although they may purport to represent idealized mathematical objects (e.g., a boundary line is a zero-width line, but it cannot be drawn as such). Finally we note that decreasing a map’s scale requires more than a simple contraction, a uniform scale reduction everywhere. Mapmaking practice requires that the positional adjustment not be uniform (exaggeration may be necessary to capture a feature's character).

3.2.2 A Homeomorphism Model for Map Conflation

One idealized mathematical model for a map is a homeomorphism (a bijective bicontinuous function) or a diffeomorphism (a homeomorphism that is also has partial derivatives everywhere) of the Earth’s surface. Diffeomorphisms admit partial derivatives between the coordinates of one surface and coordinates of the other, which makes sense
for a mathematical surface (sphere or ellipsoid) or a surface satisfying spatial relations that are described in term of potential fields and differential equations, such as the geoid. The carrier topology of a surface with a cell decomposition is the underlying point set topology of the surface. The carrier topology can be derived from any cell decomposition. Nevertheless, the carrier topology is unique, and therefore any two different cell decompositions of the same region must produce the same carrier topology.

By this model, all representations of a single region are homeomorphic (or possibly even diffeomorphic) to a common surface, and, hence, are homeomorphic (or diffeomorphic) to each other. For example, all map projections are diffeomorphic away from “interruptions,” the discontinuities resulting from cuts made to flatten the map.

For two idealized maps, this homeomorphism must exist, and the homeomorphism can be explicitly described. Map conflation in this framework consists of nothing more than making explicit the existing homeomorphism. Corresponding features need to be mapped onto each other to establish a location correspondence. Features appearing on only one depiction need to be assigned an appropriate location on the other. The Jordan-Schönflies Theorem (see, for example, Mohar, 2001, for a statement of the theorem) implies that every point or feature that is interior to one or more closed simple loops must be matched with a unique point or feature within the corresponding loops. This property greatly reduces our search for matching features to within a nested collection of subregions.

Matching operations in map conflation are now defined in two ways, that is, totally homeomorphic and potentially non-homeomorphic matching. If corresponding matching
features within a subregion are of the same dimension, and hence homeomorphic, conventional matching strategy may be applied. If corresponding features are not homeomorphic, then an operation called topological surgery (replacing an entire neighborhood by a more detailed neighborhood that agrees on the boundary) may be implemented. Handling the different subregions will be discussed in Section 3.7.1.

3.3 Map Conflation Principles
A spatial process diagram for map conflation is presented in Figure 3.3. It has the capability to produce geometrically and topologically consistent maps. Each component must use mathematical and analytical theories that deal with irregular domains. Higher or general level boundaries represent outline views of geographic spaces. Details can be examined in lower level boundaries with successively finer resolutions. An attribute-embedding component was developed to assign a spatial key to each line segment, which links to source database. A meaningful node is defined not according to geometry properties, but according to attribute properties. The categories to distinguish meaningful nodes are (1) same type and (2) same junction. Cartographic 0-cell matching uses a 0-cell conflation test to see if the neighborhood of the 0-cell has a legitimate topology, in other words, if the two subregions are homeomorphic. Geometry, topology, and attribute match tests are performed if the regions are homeomorphic. Using two matched 0-cell pairs and its correspondent 1-cell in the source map, cartographic 1-cell match is implemented.
Finally, map transformations are applied to transform the features of lower boundary inside higher boundaries. The results of map transformation are used as the inputs for the next level conflation. These spatial processes may be iterated to match elements within lower boundaries.

3.4 Hierarchical Conflation

Hierarchical structuring has many advantages not only for GIS fields but also for many other fields as well. The main methodology of hierarchical structuring is to subdivide complex reality or to refine broad problems to provide various levels of understanding.
According to Timpf, (1998), there are three types of hierarchies. They are aggregation, generalization, and filter. The aggregation hierarchy is built by aggregating sets of individuals. The generalization hierarchy defines classes in such a way that the higher level is a more generic class and lower level is a more specific class. For example, the generic class road consists of highway, state road, and local road. The filter hierarchy filters a set of individuals according to filter criteria and generates a subset of these individuals on a higher level. Useful hierarchical approaches to solve GIS related problems are abundant in literature. Car (1996) shows that human beings use hierarchies extensively to simplify their conceptual models of reality and to solve spatially reference problems more efficiently. She builds hierarchical road structures such as expressways, highways, and local roads to find optimal paths. Hierarchical spatial reasoning is based on the following cognitive assumptions (Car, 1994a, 1994b).

- **Humans divide a large road network hierarchically**

- **The amount of detail increases from the top level to the lowest level**

- **The path found with hierarchical reasoning is (close to) optimal (if hierarchical levels are formed according to expected travel speed).**

- **Wayfinding occurs in small subnetworks (this reduces complexity)**

By the way of hierarchical structuring, complexity and processing time can be reduced because the effective problem area becomes more manageable and the chance of better performance is increased because an ordered structure is provided. In order to subdivide a spatial raster domain into homogeneous blocks, a quadtree, which always has four subdivisions, may be used (Samet, 1989; Burrough, 1998; Samet, 1990; Laurini, 1995).
This permits variable resolution, in which detail can be represented at the higher resolution. The U.S. Census Bureau uses the following hierarchical geographic entities (Figure 3.4). These geographic entities are defined by legal and statistical criteria.

The determination of the boundary of legal entities is based on laws, treaties, or other administrative or government actions. Legal entities include states, counties, minor civil divisions, incorporated places, American Indian reservations, off-reservation trust lands, and Alaska region corporations (Census Bureau, 2000). A statistical entity is an area for which the U.S. Census Bureau tabulates data. It includes census tracts, block groups, blocks, and urbanized areas (Census Bureau, 2000).

![Diagram of hierarchical relationship of U.S. Census geographic entities](image)

Figure 3.4 Hierarchical Relationship of U.S. Census Geographic Entities (Census Bureau, 2000)
Each abbreviation is as follows.

ALANHH : American Indian area/Alaska Native area/Hawaiian home land
ANRC : Alaska Native Regional Corporation
BG : Block Group
CD : Congressional District
MA : Metropolitan area
SLD : State Legislative District
TAZ : Traffic Analysis Zone
UA : Urban Area
UGA : Urban Growth Area
VTD : Voting District
ZCTA : ZIP Code Tabulation Area

Each description of a census tract, block group, or blocks is well explained in the manuals of the Census Bureau, (2000). The regions form a nested structure in which a census tract includes its block groups and a census block group includes its blocks. In order worlds, the external boundary of all block groups is the boundary of census tracts and the external boundary of all blocks is the boundary of block groups.

Table 3.1 shows the hierarchical relationship of Census tract, block group, and block. Those registered boundaries are changed every 10 years based on population numbers as illustrated in Table 3.1. If there is considerable land development in an area, then there are even more opportunities to redefine boundaries by repartitioning and consolidation. Figure 3.5 shows a step in hierarchical conflation for Census maps.
<table>
<thead>
<tr>
<th>Purpose</th>
<th>Tract</th>
<th>Block Group</th>
<th>Block</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Decennial census data</td>
<td>Decennial census data</td>
<td>Decennial census data</td>
</tr>
<tr>
<td>Area</td>
<td>A small, relatively permanent statistical subdivision</td>
<td>Contiguous area inside census tract</td>
<td>The smallest of the census geographic areas</td>
</tr>
<tr>
<td>Size Range</td>
<td>1,500 – 8,000 inhabitants, optimum size is 4,000</td>
<td>600 – 3,000 inhabitants, optimum size is 1,500</td>
<td>Average 100 inhabitants</td>
</tr>
<tr>
<td>Boundary</td>
<td>Visible feature, legal or governmental boundaries, non-visible in some instance, and always nest within counties</td>
<td>Visible and non-visible feature</td>
<td>All cases, visible features such as streets, roads, streams, and railroad tracks, and invisible features such as city, town, township, county limits, and short imaginary extensions of streets and roads.</td>
</tr>
<tr>
<td>Revision</td>
<td>Seldom, but if there is physical changes such as new highway construction. Sometimes spilt or combined if there is population changes</td>
<td>More than Census Tract</td>
<td>More than Census Block Groups</td>
</tr>
<tr>
<td>Spatial Relation</td>
<td>Parent</td>
<td>Child</td>
<td>Grand child</td>
</tr>
</tbody>
</table>

Table 3.1 Hierarchical Relationship of Census Tract, Block Group, and Block
3.5 Heterogeneous Map Conflation

Heterogeneous reference maps may be used to find corresponding anchor points of source maps. Heterogeneous reference maps are polygon and polyline maps such as township boundary, road centerline, hydrographic, railroad, municipal boundary, and parcel boundary. There are also two kinds of heterogeneity. One refers to the map type itself, and the other refers to the variety of components that bound a polygon.

The boundaries is homogeneous in terms of being a polygon boundary itself, but they are not homogeneous in terms of their attributes within geography feature. For example, for a single Census tract, some line segments are roads and others are political boundaries. The boundary of a tract polygon is usually subdivided according to type and then saved as homogeneous strings (Figure 3.6).
If various types of maps are overlaid, new point features, intersections between the layers, are created and new topology can be also built. As far as map conflation is concerned, a new topology build in the reference map may not be necessary because the new topology build may not necessarily reflect any features in the source map. Intersection points are not pre-calculated because the situation when they are needed is known in the process of matching. Therefore, this process is carried out on the fly. The advanced intersection finding algorithm, a sweep line algorithm \(O(n \log n)\), could be used but a brute force algorithm \(O(n^2)\) may also be acceptable if the number of line segments is small.

![Figure 3.6 Heterogeneous Strings on a Polygon Boundary](image.png)
3.6 Anchor Nodes

In most recently developed map conflation systems, anchor nodes are defined as well distinguished points such as intersection and turning points so that we may automatically and easily distinguish them in the digital map. However, for hierarchical boundary map conflation, there are no necessary conditions for a point to be an anchor node. Since difference types of data constitute Census map boundaries as we discussed in section 3.5, it is useful to match homogeneous strings by type before finding anchor nodes (Figure 3.6). Besides those anchor nodes, in order to make connections between tract, block group, and block, junction nodes among them may be useful (Figure 3.7).

![Diagram of Junction nodes between tract polygons and Junction nodes on a tract boundary](image)

**Figure 3.7 Junction node illustration**

For example, junction nodes between tract polygons play a role in making a consistent topology when they are combined. Junction nodes on a tract boundary are nodes that share with the lower level features, block group and block, inside a tract polygon. These
relations are still valid in the next lower level. The attribute embedding process depends on spatial data system. For example, each line segment of Census map boundary is assigned a unique key (tlid) that will lead a link to the TIGER/Line Database (Broome, 1990) which, in turn, has all attribute information.

### 3.7 Map Feature Matching

Many methods have been developed to find 0-cell matching pairs between two maps, as mentioned in literature review. These methods mainly depend on attribute, geometry, and topology information to find possible candidates and then statistical methods can apply to choose the best-fit pairs if there is no unique exact match case. The proposed attribute, geometry, and topology matching module finds match pairs from multiple candidate points. Attribute matching may be separated into exact and approximate matching tests. Exact matching tests checks if two maps have exact common parts in prefix, name, suffix, and other information. Approximate match adopts generalization operations. For example, for type of road, “rd” and “road” constitute the same category. For the number of lanes, two lane roads can also be represented by a single lane in a small-scale map. Geometry and topology matches have been implemented as follows.

**0-cell**

- Geometry: point – point (proximity)
- Geometry and Topology: node – node (neighbor proximity, degree)
- Heterogeneous point (intersection of different type features)
1-cell
- Geometry: string – string (distance, direction)
- Geometry and Topology: chain – chain (neighborhood, connectivity)

2-cell
- Geometry: G polygon – G polygon (area, perimeter, slope)
- Geometry and Topology: GT polygon – GT polygon (neighborhood, connectivity)

Multiple candidates may be chosen within any given buffer distance but it is not possible to set a universal buffer distance. Instead, the $n$ nearest points from the location where the point of source map is projected onto reference map were chosen for investigation. Since there are several reference maps, those multiple candidates may even include intersection points between reference maps. First, 0-cell matching criteria based on proximity, degree, neighbor proximity, and neighbor degree are applied. Next 1-cell match is evaluated based on distance, direction, neighborhood, and connectivity. Finally, 2-cell matching is made based on area, perimeter, neighborhood, and connectivity, if applicable. Neighbor checking gives further evidence to use to avoid choosing a false match (Figure 3.8).
Without neighbor checking, node 1 may be incorrectly matched with node 2' because they are the closest pairs and their degree (sharing number of lines on a node) and direction are matched. However, this case will be rejected with neighbor checking because neighbors (node 2 and node 3') do not have matching node direction. Therefore, the second-closest pairs (node 1 and node 1') will be chosen.

Since there are multiple match criteria and multiple candidates, several scenarios are possible. The easiest case to resolve is that all match criteria are passed only for the closest match pairs so that the rest of candidates are not even considered. The uncertainty for determining correct match pairs occurs when there are some candidate matching pairs for which only some of the matching tests are passed. In that case, one may try to establish weights for the matching criteria and then try to find a best fit based on those weights. This can be written as follows.
Best fit = \( \text{Max}(P_1 \sum_{i=1}^{n} w_i M_i, P_2 \sum_{i=1}^{n} w_i M_i, P_3 \sum_{i=1}^{n} w_i M_i, P_4 \sum_{i=1}^{n} w_i M_i, \ldots) \)

Where:

\( P_n \) : Candidate match pairs

\( w_n \) : Weight

\( M_i \) : Match test

However, assigning weights is not anything like a straightforward or simple task. For example, distance may be more important in some cases, whereas the degree of a node may be more important in others. It may turn out that no rules exist because the domain is too irregular and/or the relative weighting scheme is too subjective.

The most important thing is that modeling uncertainty is not needed if unique match candidates of the same dimension exist. In other words, we have to find as many exact match cases as possible and then model uncertainty if there is no one best choice. Therefore, if there are no exact match pairs in matching test, determining weight functions to get an approximate match would not be the best choice. Another possibility is to change geometry and topology of the source data to help derive the exact match. This can be done by employing the homeomorphic model of map conflation discussed in Section 3.2.1.

From the point view of completeness in making a map or homeomorphism, map features between maps that have same coverage should be matched each other. However, this is not case in map conflation because one map is at a smaller scale, and generalization
procedures have already been conducted. Therefore, the uncertainty in any matching process already exists. It is impossible to invert the generalization process, but it may possible to recover the original shape before generalization from reference maps. The proposed method provides consistent geometry and topology for a candidate point in the source map after modeling generalization in the reference map. The advantages of this approach are as follows.

1) It can find many exact match cases to reduce uncertainty.
2) It can change geometry and topology in the source map to represent more realistic geometry and topology (rebuild one-to-one relation).
3) It can allow one candidate point to be divided into multiple points to restore the original shape (add details).

3.7.1 Generalization and Matching

This section discusses the relation between generalization and matching strategies and provides 0-cell conflation test.

3.7.1.1 Cartographic and Digital Generalization

Cartographic generalization, an essential component in mapmaking and for successful map communication, is the process of selecting and simplifying the reality according to the scale, the objective, graphic limits, and map users of a target map. This graphic process corresponds to the fundamental human activity of abstracting and reducing complexity (Timpf, 1998). However, it consists of highly subjective operations, which do
not readily decompose entirely into logical rules because the process depends considerably on an individual cartographer's skill (Timpf, 1998). Our classification of these subjective operations follows Hake (1975), Robinson (1995), and Anson, (1993).

- Selection: determination of features mapped
- Simplification: exclusion of unwanted detail and enhancement of desired features.
- Classification: order, scale, and group features.
- Exaggeration: enhance or emphasize important characteristics of features.
- Displacement: relocation of feature to prevent conflict.
- Amalgamation/Merging: combine features.
- Symbolization: determination of graphic symbols to represent various features.
- Induction: logical interference.

Since the inception of digital cartography, automated cartographic generalization has been researched extensively (Brassel, 1988; Buttenfield, 1991; McMaster, 1992) to provide a conceptual or implemental framework. However, none of the known methods promise complete solutions because of the subjective nature of generalization itself and its difficulty in being implemented automatically. Brassel and Weibel (Brassel, 1988) propose a conceptual framework for knowledge-based generalization that has five steps: structure recognition, process recognition, process modeling, process execution, and display. The first step aims at the identification of objects or aggregations, their spatial relations and relative importance. This step is controlled by the objective, data quality, scale, and graphic and perceptual limits. Process recognition defines control parameters, the types of data modification, and the parameters of the target structures. The next step, process modeling, is a compilation of rules and procedures from a process library. Process execution creates generalized target data using predefined rules and procedures. Finally, the display step converts generalized target data into a target map. McMaster and
Shea (Buttenfield, 1991) provide an extended model of Brassel and Weibel (Brassel, 1988). It decomposes generalization into three main components, (1) the intrinsic objectives of why we generalize, (2) an assessment of the situations which indicate when to generalize, and (3) an understanding of how to generalize using spatial and attribute transformation. Why and how to generalize has some overlap with the model of Brassel and Weibel (Brassel, 1988). How to generalize includes the ten spatial transformations (simplification, smoothing, aggregation, amalgamation, merging, collapse, refinement, exaggeration, enhancement, and displacement), and the two attribute transformations (classification and symbolization). From the point view of automated map production, a cartographic language, which treats cartographic objects in the forms of language using its own alphabet and grammar, has also been provided (Ramirez, 1993).

This dissertation does not try to provide a general model or complete solution for the automated cartographic generalization problem. Since linear features such as road, railroad, and hydrographic boundary related to map conflation are our main concern, a model to predict original linear features before cartographic generalization with the help of reference map will be discussed. The McMaster and Shea (Buttenfield, 1991) model also provides that linear feature are generalized by “simplification, smoothing, merging, collapse, refinement, exaggeration, enhancement, displacement, classification, and symbolization operations” among spatial and attribute transformation (Figure 3.9).
Those operations between cartographic and digital generalization have common parts in terms of conceptual and implementational aspect. For example, “simplification, merging, collapse, refinement, exaggeration, enhancement, and displacement” are especially common operations for linear features of Census maps.
3.7.1.2 Graphic and Human Visual Resolution Limitation

As discussed earlier, cartographic generalization is done according to the scale, the objective, graphic limits, and map users of a target map. Among generalization factors, the scale, the objective, and map users of source maps are given in the case of Census map conflation.

Census maps are created by digitizing USGS 1:100,000, USGS 7.5’ (1:24,000), and USGS 15’ paper maps. The objective of the USGS map series is to represent entire Nations in the forms of paper maps. The 1:100,000 series started earlier than 7.5’ and 15’ series. At those scales, most of the major, intermediate, and minor of road, railroad, and hydrographic boundary are represented (Thompson, 1988). There is, however, a high probability that some minor features such as private roads, drawbridges, and exposed wrecks will be missing from the 1:100,000 scale maps (Thompson, 1988). Map users of these maps are general public so that map representation and symbols should be widely understandable. Therefore, one can assume that the boundary information for Census maps is all there. One can also expect that those boundaries have been subjected to a cartographic generalization process, which could have changed geometry and topology of boundary information, because of graphic and human visual resolution limitations. For example, two lanes might have been combined into one lane if the space between two lanes is small enough. Two intersections that are close each other might have been combined (Figure 3.10).
Figure 3.10 Generalizations of Linear Feature that Change Geometry and Topology

There are physical distances of graphic and human visual resolution limitation for linear features. The separation distance between two lines should be at least 0.15 mm and this distance should be even greater if the lines are very fine (Cuenin, 1972; Keates, 1989). The minimum line width that is discernible is 0.06 mm. Finally, the length of a line should not be less than 0.6 mm (Cuenin, 1972; Keates, 1989).

3.7.1.3 Generalization and 0-cell Match

For purposes of finding matching pairs for map conflation, the scale difference between source and reference maps should be taken into consideration before geometry, topology, and attribute matching rules are applied because a scale decrease may result in the maps no longer being homeomorphic. Therefore, the following 0-cell conflation test is proposed.
$k$ is a constant that controls displacement; if there is no displacement, then $k$ is 1. Graphic limit is 0.15 mm and 0.6 mm for the separation and the length, respectively. "Distance (L in Figure 3.11)" refers not to the distance between candidate match pairs in two maps but rather to the distance between a node and its neighbor point in the reference map.

![Diagram](image)

Figure 3.11 0-cell conflation Test Illustrated

In Figure 3.11, the map intersection at $1'$ is the source map, and the solid line map with the multiple intersections at 1 and 2 is a reference map. When node $1'$ looks for its corresponding node in the reference map, the 0-cell conflation test is applied to see whether node $1'$ is a potential generalized node (in other words, to check for local homeomorphism). L, the distance between node 1 and 2 in the reference map, is used to see if the rescaled distance is within graphic limit or human visual limitation. If the
rescaled distance is within graphic limit, then node $1'$ finds its multiple match pairs by relaxing its requirement of maintaining topology. In this case the geometry and topology of the source map should be modified according to that of the reference map.

![Figure 3.12 A Topological Surgery Operation](image)

Figure 3.12 shows a topological surgery operation for the node that has multiple candidate matching pairs. The node $1'$ of the left map of Figure 3.12 finds its multiple match pairs in Figure 3.11. The four regions (A, B, C, and D) in the source map (the middle map) are treated separately and then are replaced with their four-match regions (A', B', C', and D') in the reference map (the right map). Finally, one four-way intersection becomes two three-way intersections to reflect true topology after conflation.

This example is a good illustration of generalization by cartographic simplification. If there is generalization by feature merging, such as a divided highway shown as a single lane, then the two lanes may still be recognized using the 0-cell conflation test. From the point view of Census geography, however, it is not important to separate lanes because
the median zone contains no demographic data. For example, major highways usually have a median area. These roads are nevertheless described single connector using a centerline in the small scale and two distinct connectors, one for each direction in the large scale. Therefore, as long as a globally consistent topology is maintained, either connector may be chosen to match the single connector representation.

In summary, if there is a many-to-one or a one-to-one matching relation between reference and source and their relation does not affect topology outside neighborhoods that have matchable boundaries, then the 0-cell conflation test may be relaxed. For example, displacement should not alter topology after generalization. One can also argue that there should always be corresponding features in the larger scale reference map for all of the generalized linear features in a source map.

3.7.1.4 Generalization and 1-cell Match

In general, when two strings (polylines made up of a sequence of connected line segments) are tested for matching, the shape of the strings and their relative positions and orientations are used as criteria for selection (Hangouet, 1995; Saalfeld, 1986). Sometimes it may be satisfactory to use simple shape, location, and orientation matching rules for 1-cell matching. However, since map conflation includes consolidating maps at different scales, matching features may fail to exhibit even similar geometry. For satisfactory matching of features of map at different scales, we must understand how generalization may alter geometry. Some of the effects of generalization have been well studied. Topper's selection law, for example, describes the relationship between the
number of features and map scales (Topfer, 1966). One classic string simplification algorithm, the Douglas-Peucker algorithm, has been used to reduce the number of vertices of the string (Douglas, 1973). A pre-selected buffer-distance is used to remove a polyline’s vertices if their distance from a simplifying line is less than the buffer-distance. A modified version of the Douglas-Peucker algorithm has also been developed to preserve topology in a neighborhood of the simplified polyline (Saalfeld, 1998).

The vertices of the string in the reference map are given. Therefore, string comparison may still be implemented after the line simplification of a string in the reference map.

3.7.2 Graph Theory and Match

Suppose that two pairs of matched nodes (A↔A’, B↔B’) are found by 0-cell matching and there is a string (A’ to B’) in the reference map. The task is to find the string in the source map that corresponds to the target string (A’-B’) in the reference map. Figure 3.13 shows this situation.

Since the underlying 1-skeleton of a map is a plane graph, there are many graph matching algorithms that deal with aspects of this problem. There are many approaches to select a candidate string from the source map’s graph if start and end nodes are known. One might be tempted to try a greedy algorithm, such as: starting from start node A, first choose a segment that is connected to node A and remains closest to the string A’ - B’. But this approach has several disadvantages.
Figure 3.13. Finding Matched String After 0-cell Match

1. There is no universally agreed-upon procedure for determining closeness. For example, metrics involving length or direction can be used to select the "closest" line segment. But this is case by case because one string is finer and the other is coarse.

2. A good partial string match at the beginning segments may lead to a very poor overall string match.

3. Locally acceptable matches may fail to capture the true overall shape of the string.
Therefore, global comparison between strings should be done to consider generalization effect.

Figure 3.14. Tolerance Buffer from the String A' - B'

However, this approach could force one to examine many candidate strings that start from node A and end at node B in the graph. Most of them are simply not worth considering. A tolerance buffer may be used to restrict the search space since the geometry of the reference string (A'-B') and its two end matching nodes (A-A' and B-B') are known (Figure 3.14).

Using only those edges within the tolerance region, the k shortest paths are found. The k shortest paths algorithm uses the well-known Dijkstra method repeatedly (Katoh, 1978).
Dijkstra’s method finds the shortest path between two vertices in a graph. If the graph \( G = (V, E) \), where \( V \) is a given set of vertices and \( E \) is a given set of edges, then Dijkstra’s method stores the total cost from the source vertex to the current vertex by using temporary and permanent labels. The temporary labels are vertices that have not been reached and the permanent labels are given to those vertices whose cost to the source vertex is known (Gross, 1998). Dijkstra’s algorithm to find the shortest path from vertex \( s \) to \( v \) is as follows (Gross, 1998).

1. Initialize the Dijkstra tree \( T \) as vertex \( s \).
2. Initialize the set of frontier edges for tree \( T \) as empty.
3. \( \text{dist}[s] := 0 \)
4. Write label 0 on vertex \( s \).
5. While Dijkstra tree \( T \) does not yet span \( G \)
   a. For each frontier edge \( e \) for \( T \)
      i. Let \( x \) be the labeled endpoint of edge \( e \).
      ii. Let \( y \) be the unlabeled endpoint of edge \( e \).
      iii. Set \( P(e) = \text{dist}[x] + w(e) \).
   b. Let \( e \) be a frontier edge for \( T \) that has the smallest \( P \)-value.
   c. Let \( x \) be the labeled endpoint of edge \( e \).
   d. Let \( y \) be the unlabeled endpoint of edge \( e \).
   e. Add edge \( e \) (and vertex \( y \)) to tree \( T \).
   f. \( \text{dist}[y] := P(e) \)
   g. Write label \( \text{dist}[y] \) on vertex \( y \).
6. Return Dijkstra tree \( T \) and its vertex labels.

The time complexity of Dijkstra method is \( O(n^2) \), where \( n \) is the number of vertices.

The graph \( G (V, E) \) created from tolerance region may have islands and dead ends. In order to make simple graph in terms of start and end vertex, delete islands and dead ends are performed before \( k \) shortest path implementation (see Figure 3.15).
The k shortest paths algorithm uses the additional constraint that additional paths branches from a specified initial portion of the first shortest path. In other words, if there is branch from the first shortest path, Dijkstra method is repeated but the start node is changed to the next branch vertex. The final algorithm to find k shortest paths for map conflation is as follows.

```
set tolerance region
G(V, E) is initialized
set start and end vertex
delete islands and dead ends
call Dijkstra to find 1st shortest path
while if there exists branch in the n-th path
```
Finally, line generalization on those candidate strings may be implemented to see how a large-scale string can be modified to produce a representative small-scale string. 1-cell match operations based on length, curvature, and separation by area between strings are then used to match candidates with the reference string. String projection, a point-by-point allocation respecting order along each string, is used if a matched string is identified. In that case, a source string may be replaced by a reference string. The use of string projection in map transformation offers several advantages. It establishes geometric consistency with the string on reference map, and hence has detail features. It also guarantees topological consistency in a neighborhood of the matched string features. Although explicit one-to-one mapping inside a string may be temporarily lost, the bijection may be recovered by computing relative offsets inside the strings.

3.8 Summary

In this chapter, map conflation theory is discussed. First, two map models are presented to understand the nature of maps and spatial datasets. The first model is based on a mathematical theory of cellular surfaces composed of idealized cell objects of zero, one, and two dimensions, for which collapsing of map features is allowed. The second model is based on guaranteeing the existence of homeomorphism, for which one-to-one explicit
links are preserved all of the time. Map conflation in this framework consists of nothing more than making explicit the existing homeomorphism. Corresponding features need to be mapped onto each other to establish a location correspondence. 0-cell conflation test is proposed to correct topologic difference if corresponding features are not homeomorphic. A combination of attribute, geometry, and topology criteria are applied if corresponding matching features within a subregion are of the same dimension. String matching is conducted based on matched 0-cells and a reference string used as a model.

A new hierarchical based map conflation is also presented to be incorporated with physical, logical, and mathematical boundary and to reduce the complexity and computational load. Map conflation principles are formulated and Census maps are used as a conflation example. They consist of attribute embedding, find meaning node, cartographic 0-cell match, cartographic 1-cell match, and map transformation. These spatial processes may be iterated to match additional elements using fewer constraints. In the next chapter, map transformations for map conflation are discussed.
CHAPTER 4

MAP TRANSFORMATIONS IN MAP CONFLATION

4.1 Introduction to Map Transformation

Map transformation is conducted on map objects that do not have matched pairs in each hierarchical map conflation level and those transformed map objects are used as the inputs for the next level conflation. In the traditional piecewise rubber-sheet linear map transformation (White, 1985; Saalfeld, 1985, 1987, 1993; Gillman, 1985), the Delaunay Triangulation (DT), which has many unique properties, is used to provide a piecewise subdivision into triangles. Then, map objects in a triangle are transformed to the corresponding triangle. However, in the hierarchical map conflation, there is a constraint, that is, a hierarchical boundary. In order to satisfy a boundary problem and to list other possible methods that result in more complete solutions, it is necessary to analyze several map transformations (Sugihara, 1999; Hiyoshi, 1999, 2000; Gross, 1999; Beier, 1992; Gans, 1969; Okabe, 1999) such as the Constrained Delaunay Triangulation (CDT), which utilizes simplicial coordinates or weighted function by neighbor triangles, and linear feature based transformations.
CDT has been used for Warnitz networks in which some important topographic lines are preserved after triangulation. However, it is not easy to predefine those lines in a map conflation. Therefore, in the triangulation type transformation, the edges of each triangle play an important role in the transformation. One of the characteristics of this transformation is that the neighborhood between edges (imaginary control lines in DT or topographic lines in CDT) of triangles and points inside of a triangle is maintained no matter what the movement of neighbor control points or the closeness of a point transformed with neighbor control points. Another is that the linearity is preserved inside a triangle but there can be the bending of lines on the boundary of a triangle. It will completely follow the meaning of rubber-sheeting if the movement and the closeness of neighboring control points are considered because the transformation is dependent on all the control points (global transformation). However, it is inefficient or impractical to consider all the control points for the transformation, because, there are main weights in control points close to a point needed to transfer. The weighted piecewise rubber-sheet map transformation uses at least four control points or four triangles to determine the coefficients of the transformation. This will give more suitable result than the traditional rubber sheeting where thin and narrow triangles are created in a constrained boundary and used as transformation.

Other linear feature based transformation can be used for hierarchical map conflation because the transformations are bounded by polygon areas, linear lines can be easily defined on those polygon boundaries, and triangulation that result in thin and narrow
triangles does not need to be built. A linear feature based morphing will be closely analyzed.

4.2 Constrained Delaunay Triangulation

This triangulation, which subdivides the planar map into a finite family of triangular regions, provides a key tool for automating the interplay among (1) finite sets, (2) finite combinational topology, and (3) two-dimensional, infinite-set, continuous geometry/topology (Saalfeld, 1999). The triangulation uses the finite number of vertices, edges, and triangles to represent continuous planar map and the infinite number of points of that region are related with the triangulation. Every triangle has 3 vertices, 3 edges, and 3 triangle neighbors. Triangulation on \( n \) vertices is called a maximal (no more edges can be added) connected plane (edges do not cross each other) graph on those vertices and has the following property: If there are \( n \) vertices, and \( m \) of those vertices lie on the boundary of the convex hull formed from those \( n \) vertices, then every triangulation has \((3n-m-3)\) edges and \((2n-m-2)\) triangles (Saalfeld, 1999). Therefore, a triangulated planar graph with \( n \) vertices has maximum \((3n-6)\) edges and \((2n-5)\) faces in the case that the convex hull is a triangle.

There are several triangulation methods such as minimum-total-edge-length and Delaunay triangulation. However, Delaunay triangulation has many unique properties that are suitable for map conflation systems (Saalfeld, 1993; Samet, 1989; Preparata, 1985) and other applications (Gold, 1989, 1991; Sambridge, 1995; Mocozet, 1997).
Some of those properties are briefly reviewed in the following. More details are found in the literature (Saalfeld, 1993; Samet, 1989; Preparata, 1985; O’Rourke, 1993):

1. A Delaunay Triangulation is unique.
2. Circumcircle property for each triangle.
3. Disk property for edges.
4. Maximum-minimum angle property.
5. Best-lexicographic order property.
7. The dual structure of the Voronoi diagram.
8. The external edges of the DT form the convex hull.
9. Edges of the EMST (Euclidean Minimum Spanning Tree) are those of Delaunay. That is, $EMST < DT$.

A Delaunay Triangulation can be built using a divide and conquer algorithm ($O(n \log n)$) or an incremental insertion algorithm ($O(n^2)$), where $n$ is the number of vertices on which to triangulate. The incremental insertion algorithm is simple, elegant, and robust compared to the other algorithm and its expected behavior is much better, $O(n \log n)$ on average. It can be easily modified for the Constrained Delaunay Triangulation. The Delaunay Triangulation for incremental insertion algorithm (Saalfeld, 1999) is shown below.
Delaunay Triangulation for the Incremental Insertion Algorithm

Procedure Add(p, D)
// Note: D is the triangulation, p is a new vertex not in D to be added to D. All the triangles of D are kept in
// clockwise order.
if p is inside the convex hull of D
    then AddInsideHull(p, D)
else AddOutsideHull(p, D)

Procedure AddInsideHull(p, D)
Find the triangle Δabc of D containing p
Replace Δabc by 3 triangles {Δpab, Δpbc, Δpca} in D
SwapTest(ab, p, D)
SwapTest(bc, p, D)
SwapTest(ca, p, D)

Procedure AddOutsideHull(p, D)
Find the sequence {hi, hi+1, ...., hk} of the full clockwise sequence of hull vertices
that are visible from p
Replace the subsequence {hi, hi+1, ...., hk} by {hi, p, hk} to update the clockwise
sequence of hull vertices (and also update hull edge set accordingly)
for each j=1 to k-1
    Add the triangle Δp hj+1hj to D
    SwapTest (hj+1hj, p, D)

Procedure SwapTest(uv, p, D)
if uv is an edge in the convex hull, then return
if Δpuv is not in triangulation D, then return
Let w be the third vertex in the triangle Δuvw to the left of uv in triangulation D
if CircleTest(u, v, w, p) then
    Replace Δuvw and Δpuv with Δpuw and Δpww in D
    SwapTest(uw, p, D)
    SwapTest(ww, p, D)

A Constrained Delaunay Triangulation (CDT) is a triangulation that has predefined
edges. The predefined edges will be preserved in the triangulation (Figure 4.1). Those
edges can be any important linear features in GIS applications such as topographic edges
in elevation mapping and boundaries of each Census maps in Census map conflation system.

![DT](image1) ![CDT](image2)

Figure 4.1 Delaunay Triangulation and Constrained Delaunay Triangulation

There are two ways to construct CDT, the modification of empty circle property and the refinement of constrained edges. The refinement of constrained edges introduces additional points on those boundaries to make small edges so that the triangulation will include them. This method is not suitable for map conflation systems because it is difficult to make explicit relations between the introduced points and the points in the source map. Therefore, the modification of empty circle property is preferred. The empty circle property is modified to treat the predefined edges as opaque so that vertices are not affected from these edges. Therefore, what is modified in the incremental construction of
the DT is the addition of one more condition, the visibility (circumcircle property has an exception for constrained edges), in Procedure SwapTest. In hierarchical map conflation, the predefined edges are external boundaries as shown in Figure 4.2.

![Census Tract Polygon](image)

**A**

Census Tract Polygon

**B**

Delaunay Triangulation

**C**

Constrained Delaunay Triangulation

*Figure 4.2 DT and CDT for a Census Tract Polygon*

In Figure 4.2(A), the Census tract polygon consists of 87 vertices. Figure 4.2(B) shows DT and Figure 4.3(C) shows CDT where tract boundary is used as constrain. The method to calculate simplicial coordinates for the transformation is discussed in Section 2.3. This method has many merits such as being simple and fast. Additionally, the linearity is preserved inside a triangle. However, a line cannot be connected smoothly on the boundary of a triangle. This problem can be solved if Voronoi neighbor weighting is used.
4.3 Piecewise Rubber-Sheet Map Transformation by Weighted Delaunay Triangulation (WDT)

This method is also known as the natural-neighbor interpolation and was introduced by Sibson (Sibson, 1980). The Voronoi diagram and the Delaunay triangulation are dual relations such that Voronoi sites(a) $\leftrightarrow$ DT triangles(its vertices are P1, P5, and P), Voronoi bisector(a-e) $\leftrightarrow$ DT edges(P-P1), and Voronoi polygons(a,b,c,d, and e) $\leftrightarrow$ DT vertices(P). Figure 4.3 shows these relations and the change to the Voronoi diagram and the DT after inserting the point P (left to right).

![Figure 4.3 Voronoi Diagram and Delaunay Triangulation](image)

Any point in the plane can be expressed by the convex combinations of neighboring sites. In traditional rubber-sheet transformation, those three vertices (P2, P4, and P5) are used to make convex combinations providing P is a point that needs to be transformed. In the Weighted Delaunay Triangulation (WDT), the second-order Voronoi regions are used.
(Figure 4.4). The second-order Voronoi regions represent the Voronoi diagram after inserting point P and the first-order Voronoi regions means the Voronoi diagram before inserting point P in the Figure 4.4.

![Voronoi Diagram](image)

Figure 4.4 Influences of the Second-order Voronoi Regions

Voronoi polygon, its vertices are a, b, c, d, and e, is the influence regions by point P if it is inserted. The polygon area can be subdivided by the first-order Voronoi diagram (dashed line in Figure 4.4) and those subdivision areas are influenced by point P’s neighbor sites. Therefore, it can be written as follows.

\[ \omega_i = \frac{Area(P_i)}{Area(P)} \]
Where Area \((P)\) is the polygon area (its vertices are \(a, b, c, d, \) and \(e\)).

Area \((P_i)\) is one of the subdivision areas influenced by point \(P_i\)

Then, all fractions \(\omega_i\) satisfy

\[
\sum_{i=1}^n \omega_i = 1
\]

Finally, the Sibson coordinates of \(P\) have the following relation

\[
P = \sum_{i=1}^n \omega_i P_i
\]

Where \(P\) is an interpolated value at that location and \(P_i\) is the sampled value at each point location.

In WDT, \(\omega_i\) is calculated using the matched points, set them as \(P_i\), and a transformed point, set it as \(P\) in the reference map. Then, a new location for \(P\) in the source map is found by inserting the matched pairs of the source map into (1).

There have been reports about the properties on this method in the interpolation areas (Farin, 1990; Sugihara, 1999; Gross, 1999; Sambridge, 1995).

1. Local procedure: apply only to points that are close to the transformed point by the second-order Voronoi regions.

2. There is a similar relation between the length of a polygon edge and its bisector distance. For example, ratio between length \(p-p1\) and length \(a-b\) can replace area weight.

3. It is a linear and continuous function.
Figure 4.5 The Comparison of DT (piecewise rubber-sheeting by Delaunay Triangulation) and WDT (piecewise rubber-sheeting by Weighted Delaunay Triangulation) Transformation
The expected result of DT (Piecewise Rubber-sheeting Transformation by Delaunay Triangulation) and WDT (Piecewise Rubber-sheeting Transformation by Weighted Delaunay Triangulation) is explained in Figure 4.5.

In Figure 4.5, triangles are shown in dark black line and control based map is shown in light black line. The upper figure shows the condition when a triangle (vertex A, B, and C) that is used for the transformation of point P is found. As shown in Figure 4.5 (top), the triangle (vertex A, B, and C) is thin and its vertices A, B, and C are not closer points to the point P. This situation is improved in the Figure 4.5 (bottom) when WDT is applied using the second-order Voronoi diagram. The 44 selected triangles as shown in dark black line are used for the transformation of point P. Since more control points that are closer to the point P, more accurate results that follow the movement of control points are expected.

### 4.4 Linear Feature Based Transformation (LineMorp)

This new method is introduced to generate intermediate frames for animation to convert one image to another image smoothly (Beier, 1992). Before this method, point based transformations were mainly used like triangulations. Since linear features are easier to recognize than point features, this method is preferred, which is the same situation for cartographic objects. The transformation is based on lines and those line movements and the distance from them are counted for transformation. The influence of the different lines to a transformed point is adjusted using the weight constants.
A linear feature based transformation can be used for Census map conflation because linear features are defined in the process of conflation. The transformation mechanism is as follows.

![Source Map](image1)

![Reference Map](image2)

**Figure 4.6 Transformation for Single Line Match Pair**

If matched line pairs exist in the source and reference map, those lines define a mapping from one map to the other (Figure 4.6). Therefore, the coordinate of a point $X$ in the source map can be defined in the reference map by the relation between a line $PQ$ in the source map and $P'Q'$ in the reference map.

\[
u = \frac{(X - P) \cdot (Q - P)}{{\|Q - P\|}^2}
\]

\[
v = \frac{(X - P) \cdot \perp (Q - P)}{{\|Q - P\|}}
\]
\[ X' = P' + u \cdot (Q' - P') + \frac{v \cdot \perp (Q' - P')}{\|Q' - P'\|} \]

Any point on the line between two points, \( P_1 \) and \( P_2 \), can be written as \( P = P_1 + t(P_2 - P_1) \) where value \( t \) represents the position of \( P \) along the line and goes from 0 to 1 (less than 0 or greater than 1 if \( P \) is outside the range). The value \( u \) corresponds to the value \( t \) and is the position along line \( PQ \) obtained by the perpendicular projection of point \( X \). The value \( v \) is the perpendicular distance from point \( X \) to the line \( PQ \). Using those two parameters, \( u \) and \( v \), a new point \( X' \) is defined in the reference in Figure 4.6. This transforms accounts for a rotation, translation, and scale. If there are multiple line pairs, weighting of the coordinates’ transformation for each line is implemented such that the displacements between the original point location and the transformed location for each line are averaged according to the closeness (Figure 4.7).

![Diagram](Image)

Figure 4.7 Transformation for Multiple Line Match Pairs
Therefore, the weight assigned to each line should be strongest when the point is exactly on the line, and weaker the farther the point is from it (Beier, 1992). The weight function can be defined as

\[ \text{weight} = \left( \frac{\text{length}^p}{(a + \text{dist})} \right)^b \]

Where \( \text{length} \) is the length of a line, \( \text{dist} \) is the distance from the point to the line, and \( a \), \( b \), and \( p \) are constants that can be used to change the relative effect of the lines. The constant \( a \) controls the closeness effect. In other words, closer lines have more influence than others. The constant \( b \) controls the weight of different lines. It is reported that the range of \( b \) between 0.5 to 2 is the most useful (Beier, 1992). The length of a line is also weighted by the constant \( p \). If \( p \) increases, then longer lines have greater relative weights than shorter lines. Finally, the algorithm for multiple line pairs is presented as follows (Beier, 1992).

For each pixel \( X \) in the destination
- \( \text{DSUM} = (0, 0) \)
- \( \text{weightsum} = 0 \)

For each line \( \text{Pi Qi} \)
- calculate \( u, v \) based on \( \text{Pi Qi} \)
- calculate \( \text{Xi'} \) based on \( u, v \) and \( \text{Pi'} \) \( \text{Qi'} \)
- calculate displacement \( \text{Di}=\text{Xi'}-\text{Xi} \) for this line
- \( \text{dist} = \text{shortest distance from X to Pi Qi} \)

\[ \text{weight} = \left( \frac{\text{length}^p}{(a + \text{dist})} \right)^b \]

- \( \text{DSUM} += \text{Di} \times \text{weight} \)
- \( \text{weightsum} += \text{weight} \)
- \( X' = X + \text{DSUM/weightsum} \)
- \( \text{DestinationImage}(X) = \text{sourceImage}(X') \)
Note that because these lines are directed line segments, the distance from a line to a point is \(\text{abs}(v)\) if \(0<u<1\), the distance from \(P\) to the point if \(u<0\), and the distance from \(Q\) to the point if \(u>1\). Figure 4.8 and 4.9 show an example of a linear feature based transformation.

![Source polygon](image1.png) ![Reference polygon](image2.png)

Figure 4.8 The Boundary Change from Source to Reference Polygons

The source polygon will be changed to the reference polygon. Equally spaced points are created inside the source polygon to see their movements after changing to the reference polygon. Each vertex of the reference polygon is found to match a vertex in the source polygon. They are represented as vector lines (Figure 4.9).
Figure 4.9 The Transformation of Inside Points between Source and Reference Polygons

The weight constants used are $a = 0.5$, $b = 2$, and $p = 0.2$. These constants are determined experimental results. Figure 4.9 shows that the transformation of each point is affected by the movement of boundary segments and is blended smoothly with all others. The advantage of these methods is that there is an explicit relation between lines, no control points are needed in the polygon, and the interpolation using thin and narrow triangles in CDT method can be avoided. The computation speed can be slow because each point depends on all the line pairs and if the shape of polygon is complicated or the line pairs are not similar, there will be lots of distortion. However, in census map conflation, each census boundary should remain simple for ease of management.
4.5 Restore One-to-One Relation of Control Points on the Boundary

String match is based on the match pairs of end nodes as discussed in the section 3.7. The explicit (one-to-one) relations are not defined for points inside the string in the string match because it is based on shape similarity, so explicit relations are not important. However, those relations are needed when the transformation parameters are calculated. There are several ways to restore one-to-one relations for points in the strings such as distance ratio, turning points, and a combination of the two methods. The combination of distance ratio and turning points are suitable because one string (A) has more detail shape (more vertices) than the other (B). Therefore, two-way direction matching is implemented. First, for every point in string A, find its matching point in string B. These points are turning points in string B. Next, for the unmatched points in the string B, find their match points in string A, where there are no points physically in the string A because their match points are already found in the previous step. Therefore, their match points can be assigned by the distance ratio on the straight line between turning points. The pseudo-code is presented as follows.

**String A(all points) to String B(turning points)**

for each points in String A

  find match pairs in String B and assign as turning point

end

**String B(points between turning points) to String A(straight line)**

for each string between turning points in String B

  find its match line in String A
divide the line based on the distance ratio of points of the string

calculate those divided points and assign match pairs

end

4.6 Summary

Map transformation is applied to exactly align matched map features and transform unmatched map features in map conflation. In this chapter, the methods of map transformation are discussed. They are piecewise rubber-sheeting transformation by Delaunay Triangulation (DT), piecewise rubber-sheeting transformation by Weighted Delaunay Triangulation (WDT), and linear feature based morphing (LineMorp). The narrow and thin triangles (it means the closer control points do not have any affects to their nearby points) in triangulation are expected if there is a boundary and control points are selected on the boundary. Therefore, WDT is expected to have more accurate results than DT as discussed in section 4.3. In WDT, more closer-triangles are added by using the second-order Voronoi diagram. WDT has been used in interpolation areas but this is a new approach in map conflation area. Another new approach is liner feature based morphing (LineMorp) that has been used in computer animation field. Since a boundary is used in map conflation and linear features can be defined along the boundary, LineMorp is appropriate in the map conflation.

In the next chapter, since Census maps are used as a conflation example, the spatial structure, data history, and accuracy of Census maps are discussed.
CHAPTER 5

The Spatial Structure of Census Maps

5.1 Introduction to Census Maps

The Census Bureau provides basic statistics about the people and the economy to the congress, the executive branch, and the general public every 10 years (Marx, 1984). Some examples of map products of Census 2000 to represent those statistics are shown in Table 5.1. Reference maps show boundaries and names of geographic areas for which the Census Bureau offers statistical data. Thematic maps display the selected kinds of information for a specific purpose, such as displaying the population centers through time. In addition to those maps, field maps are also created to aid staff in taking a census. In order to accomplish this mission, it is important to not only collect geographic data but also link those data to geographic locations. There is no better way to represent those links on the map, which is a graphical representation of the real world. This helps the staff process to collect, identify, record, and analyze geographic data.
<table>
<thead>
<tr>
<th>Reference Maps</th>
<th>Thematic Maps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census Block Maps</td>
<td>Mapping Census 2000: The Geography of U.S. Diversity</td>
</tr>
<tr>
<td>Census Tract Outline Maps</td>
<td>Centers of Population</td>
</tr>
<tr>
<td>Voting District/State Legislative District Maps</td>
<td>• Position of the Geographic Center of Area, Mean and Median Centers of Population 2000</td>
</tr>
<tr>
<td>Census 2000 Population Distribution of the United States</td>
<td>• Mean Center of Population for the United States: 1790 to 2000</td>
</tr>
<tr>
<td>1999 Metropolitan Area Maps</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1 Some of Examples of Census Map Products

These kinds of situations force the Census Bureau to become a mapmaker because a large quantity and diversity of reference maps are needed in a short production time frame (Tomasi, 1990). Geocoding is assigning location information to attribute data. For example, house unit 104 is located on Main Street and in block 308 with specified coordinates. However, it is important to assign topological information as well because it provides lots of benefits such as reducing redundancy and ability to check consistency. Topological geocoding has been researched since the 1960’s by the Census Bureau to provide efficient geocoding methods for their address files. GBF/DIME is the first spatial structure and was applied to the 1970 and 1980 Censuses. TIGER is the next generation of GBF/DIME, which adopts the concepts of database, topology, graph theory, and other related topics. It is used for the 1990 and 2000 Censuses. The positional accuracy of TIGER did not receive much attention until other reference data became available to it. At the time of building TIGER, USGS’s 1:100,000 scale maps were the only available
source containing road, railroad, hydrography, and other miscellaneous transportation features for the entire nation and the Bureau had to convert those features to digital format in a tight time frame. It is often mentioned that using this data source was the best choice to satisfy the needs of the Bureau at the time. However, these days, more accurate spatial data are abundant with the development of newer technologies. Incorporating these local and more accurate datasets into TIGER will be a key issue in the future.

Section 5.2 and 5.3 discuss the transitions of the spatial structure of Census maps. Finally, section 5.4 provides studies of the positional accuracy of TIGER, specially, for the study area of local government, Delaware County, OH, in this dissertation.

5.2 GBF/DIME

The process of assigning geographical location information to Census data is referred as geocoding. For example, assigning addresses of housing units to a Census block is geocoding. Special geocoding methods are critical to support the decennial Census survey by generating street-level address maps for the use of the thousands of Census enumerators and to check the consistency of the survey result (Tomasi, 1990; Clarke, 1995). The first computer readable geocoding file Census Bureau developed is GBF/DIME, which was used for the 1970 and 1980 censuses. GBF/DIME stands for Geographic Base File/Dual Independent Map Encoding. Here, graph theory was first used to implement topological geocoding and each node, line segment, and enclosed area is uniquely identifiable in the map (Clarke, 1995). Spatial information (coordinates) is
also embedded. Figure 5.1 shows a portion of the map to describe the spatial structure of the map and Figure 5.2 shows its GBF/DIME structure.

![Sample Map for Topological Geocoding (Saalfeld, 1999)](image)

Street name and its address range are assigned so that a geocoding function, which is previously done by enumerator, can be done by computer. An additional Master Reference File (MRF) is created to classify each geographic location according to the tabulation units recognized in a particular census or survey (Marx, 1984). These two files have some duplication to describe the same part resulting in confusion and inefficiency for storage and for update events such as insertion, deletion, and modification. For example, the code number of Delaware County is inserted as 043 into the MRF instead of 045 which is the value in the GBF/DIME. This typical mistake causes a cascade of
problems in all subsequent geographic products related to Delaware County and results in much discontent among field staff and data users (Marx, 1984).

<table>
<thead>
<tr>
<th>Line ID</th>
<th>From node</th>
<th>To node</th>
<th>Left Block</th>
<th>Right Block</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>b</td>
<td>f</td>
<td>A</td>
<td>F</td>
</tr>
<tr>
<td>2</td>
<td>c</td>
<td>g</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>b</td>
<td>c</td>
<td>F</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>a</td>
<td>c</td>
<td>B</td>
<td>F</td>
</tr>
<tr>
<td>5</td>
<td>c</td>
<td>d</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>6</td>
<td>d</td>
<td>k</td>
<td>F</td>
<td>C</td>
</tr>
<tr>
<td>7</td>
<td>h</td>
<td>i</td>
<td>E</td>
<td>C</td>
</tr>
<tr>
<td>8</td>
<td>e</td>
<td>h</td>
<td>E</td>
<td>C</td>
</tr>
<tr>
<td>9</td>
<td>a</td>
<td>d</td>
<td>F</td>
<td>B</td>
</tr>
<tr>
<td>10</td>
<td>e</td>
<td>i</td>
<td>C</td>
<td>E</td>
</tr>
<tr>
<td>11</td>
<td>g</td>
<td>k</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>12</td>
<td>j</td>
<td>k</td>
<td>D</td>
<td>F</td>
</tr>
<tr>
<td>13</td>
<td>f</td>
<td>j</td>
<td>D</td>
<td>F</td>
</tr>
<tr>
<td>14</td>
<td>f</td>
<td>g</td>
<td>A</td>
<td>D</td>
</tr>
</tbody>
</table>

Figure 5.2 GBF/DIME Structure

5.3 TIGER

In order to solve the problems of GBF/DIME and other recognized problems from the past, a new spatial structure called TIGER (Topologically Integrated Geographic Encoding and Reference) was developed and used for the 1990 and 2000 Censuses. It
permits the computer to assign residential and business addresses to the correct geographic locations, to produce maps for field operations and publication, and to perform the data tabulation operations for any geographic unit whose boundaries have been recorded in the file (Liadis, 2000). The structure of the TIGER database may be viewed at two levels, the conceptual level and the implementation level. The conceptual perspective illuminates the various entity types, their attributes, and the relationships among them. The implementation perspective concentrates on the internal data structures and the means for accessing and manipulating them (Broome, 1990). Many theories are adopted from database, topology, graph, and other associated fields of mathematics and computer science. Instead of maintaining all data in one file like GBF/DIME in which block or street segment are the basic cartographic objects, TIGER uses as many files as possible which are cross-referenced by a specific network-oriented file structure (Broome, 1990) and divides cartographic objects as different dimensions. They are the 0-cell (nodes), the 1-cell (line segments), and the 2-cell (blocks, census tracts, etc.). The spatial relationships of 0-cells, 1-cells, and 2-cells to each other are described in Figure 5.3 for the sample map of Figure 5.1. Other details of record content can be found in the literature (Broome, 1990; Census Bureau, 2000). This results in a substantial increase in analytical flexibility and automated error checking (Clarke, 1995).
### Figure 5.3 TIGER File Structure (Saalfeld, 1999)

Topological consistency and completeness checks can be implemented by the following.

- Verify every 1-cell has two 0-cells
- Verify every 1-cell has two neighboring 2-cells
- Verify every 2-cell is surrounded by a cycle of 1-cells
- Verify there are no intersections that are not 0-cells
- Verify 1-cells and 2-cells around 0-cell
Building a TIGER file covering entire nations while having an accurate and consistent cartographic base needs the cooperation of other agencies such as USGS. The overview of the major steps to build a TIGER file is shown in Figure 5.4.

Location information strongly depends on the maps of USGS whose scales are 1:100,000, 7.5’, and 15’ where USGS paper maps are converted into digital format by scanning a layer and then vectorizing or digitizing (Callahan; Tomasi, 1990). However, the majority base map sources of TIGER for the 48 conterminous states and the District of Columbia are from USGS’s 1:100,000-scale quadrangle maps which include roads, hydrography, railroads, and miscellaneous transportation features (Tomasi, 1990; Trainor, 1990). In terms of management, major advantages of TIGER are improved map production and consistent updates. In other words, it provides maps and tabulation data in different forms for different groups whenever there is a request.

In addition to the use of TIGER by the Bureau, the usefulness of TIGER is widespread. For example, a GIS analyst can examine all geographically distributed data sets such as agricultural population and retail trade for managing a particular area. The associated attributes in the TIGER database such as street name, address range, ZIP code, and landmark can be linked to local data sets so that inter-relational analysis can be done (Marx, 1990).
Figure 5.4 Major Steps to Build and Manage TIGER Files for 1990 Census (Marx, 1984)
5.4 The Absolute and Relative Positional Accuracy of TIGER

The Census bureau conducts research on the absolute positional accuracy of TIGER (Liadis, 2000). This research, called GPS TIGER Accuracy Tool or GTAAT, compares the distance different between GPS collected points and the equivalent TIGER 0-cell. One of their test areas is Delaware County, OH, which is also the study area of this dissertation. Delaware County was selected due to their participation in the Local Update of Census Addresses (LUCA) program and the fact that under that program, significant updates were made to the county as discussed in Chapter 1. They collected anchor points based on that update operation (Table 5.2).

<table>
<thead>
<tr>
<th>Update Operation</th>
<th>Tract (114,20)</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990 Enumerator Update</td>
<td>40</td>
<td>8.3%</td>
</tr>
<tr>
<td>1998 Boundary and Annexation Survey</td>
<td>10</td>
<td>2.1%</td>
</tr>
<tr>
<td>Census 2000 Block Canvassing</td>
<td>25</td>
<td>5.2%</td>
</tr>
<tr>
<td>Census Map Preview</td>
<td>36</td>
<td>7.5%</td>
</tr>
<tr>
<td>GBF-DIME Contractor</td>
<td>147</td>
<td>30.4%</td>
</tr>
<tr>
<td>LUCA Updates</td>
<td>26</td>
<td>5.4%</td>
</tr>
<tr>
<td>LUCA Verification Updates</td>
<td>3</td>
<td>0.6%</td>
</tr>
<tr>
<td>Master Address File Office Resolution</td>
<td>2</td>
<td>0.4%</td>
</tr>
<tr>
<td>Restructure 3 Clean-Up</td>
<td>6</td>
<td>2.3%</td>
</tr>
<tr>
<td>TIGER Improvement Program</td>
<td>188</td>
<td>38.1%</td>
</tr>
<tr>
<td>TOTAL Points</td>
<td>483</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 5.2 The Collected Anchor Points by Update Operation and Tract, Delaware County, OH (Liadis, 2000)
Figure 5.5 Distance Differences between TIGER and GPS Surveyed Match Pairs (Liadis, 2000)

Figure 5.5 shows the distribution of distances between collected GPS points and their match points in TIGER. The largest peak is at the 70-150 foot interval, which corresponds to the GBF-DIME Contractor. The average distance is 347.68 foot, which is enough to make conflict with more accurate maps.

The relative accuracy between Delaware GIS base and 1990 Census Maps are surveyed in this research. First, meaningful points or control points, which are intersection and high curvature points, are selected from Census Maps. The number of selected control points is 1315, and visual proof was used to find their match pairs in the Delaware GIS base. The distribution of those points is shown in Figure 5.6. They are well and evenly distributed all over the County except city areas where more density exists because of lots of intersections.
Figure 5.6 The Locations of Control Points

Mean: 162.30
Range: 1185.78
Standard Deviation: 150.81
Upper 95% Mean: 169.83
Lower 95% Mean: 141.99
Unit: feet

Distance Difference

Figure 5.7 Distance Difference between Delaware GIS Base and Census Maps
The histogram and statistics of distance differences of match pairs between the Delaware County GIS base and the 1990 Census Maps are also shown in Figure 5.7. The distance difference is almost \( \frac{1}{2} \) smaller than that obtained by the Census Bureau. However, this difference (mean is 162.30 foot) is still enough to show that there is conflict between Delaware GIS base and Census Maps.

5.5 Summary

The chapter discusses how DIME and TIGER are made by Census Bureau and they are spatially structured. Especially, the detail of creation and management of TIGER are studied in order to provide the data history for map conflation. The one important study is that TIGER is made from 1:100,000 USGS hardcopy maps. The absolute and relative positional accuracy of TIGER are also studied. The absolute accuracy of TIGER is that mean distance is 347.68 foot from 483 survey anchor points. The relative accuracy of TIGER with Delaware County’s GIS bases is that mean distance is 162.30 foot from 1315 anchor points. These numbers tell how Census Maps are geometrically different from the real world.

In the next chapter, the map conflation system design is discussed based on Nyerges’s spatial data design model. Spatial data, software, and hardware used are also presented.
CHAPTER 6

Map Conflation System Design, Spatial Data, and Implementation

6.1 Introduction

To find solutions to GIS based problems, it is necessary to define the relationship between machine based data structure and geographic data, and human and computer interaction. A theoretical basis of spatial data structure or design allows us to understand geographic information and spatial data organization. It is also helpful in understanding the map conflation procedure. This process consists of several activities: logical system design, cartographic system design, screen design, and communication design. In the logical system design, the workflow of the proposed application is designed. In other words, the algorithms for each component are specified and they are linked efficiently. Spatial data system design is very similar to data base design that uses the entity-relationships (E-R) modeling techniques. The difference between them is that spatial data need additional information such as geometry and topology that are considered in model design. Since the visualization of spatial data allows us to understand easy problems and
then provide feedback in the spatial data system, screen design that controls human-computer interaction by a Graphical User Interface (GUI) is also critical. Finally, communication design provides commands, tools, and menus that help in the decision making and process as a whole. It is always true that the outcome of effective communication is better results.

Spatial data used in map conflation are presented from the implementation standpoint. The spatial area covered by this dissertation are Census tract, block groups, and blocks for Census maps and digital orthophoto, township, municipal, hydrography, railroad, and road for Delaware County, OH. The data format for all spatial data except the digital orthophotos is ArcView’s Shape file. The software and hardware used to create and run the map conflation system is also discussed.

6.2 Logical System Design

Figure 6.1 shows the logical design for a map conflation system. The logical flow of the map conflation application consists of procedural and recursive steps. The procedure step is to find matched features from 0 cells to 1 cells, update those matched features, and transform the remaining features. The recursive step is to run the procedure process at each level. The 0 cell match algorithms used are the topological surgery for the conflation test objects and attribute, geometry, and topology match for the remaining objects. The extended ArcView’ shape format in which topological information is embedded is used throughout the processes. The best-fit string match finds the correspondence of the reference string. The piecewise linear transformation performs the transformation.
Figure 6.1 Logical System Design for Map Conflation
Real and virtual transformation (Moellering, 1984) is used to solve the uncertainty of spatial data. After one cycle is run for the top level, the next cycle is implemented for the same processes but for a smaller area.

6.3 Spatial Data System Design

The levels of spatial design consist of information reality, information structure, canonical structure, data structure, storage structure, and machine structure (Nyerges, 1981). These items are defined as follows (Nyerges, 1981).

1) Information Reality: Observations that exists as ideas about geographic entities and their relationships which knowledgeable persons would communicate with each other using any medium for communication.

2) Information Structure: A formal model that specifies the information organization of phenomenon in reality. This structure acts as an abstraction of reality and skeleton to the canonical structure. It includes entity sets plus the types of relationships which exist between those entity sets.

3) Canonical Structure: A model of data which represents the inherent structure of that data and hence is independent of individual applications of the data and also of the software and hardware mechanisms which are employed in representing and using the data.

4) Data Structure: A description elucidating the logical structure of data accessibility in the canonical structure. There are access paths which are dependent on explicit links, i.e. resolved through pointers, and other forms of reference. Those access paths dependent on links would be based on tree or link structures as in network models. Those access paths independent of links would be based on tables as in relational models.

5) Storage Structure: An explicit statement of the nature of links expressed in terms of diagrams which represent cells, linked and contiguous lists, levels of storage medium, etc. It includes
indexing how stored fields are represented and in what physical sequence the collected records are stored.

6) Machine Encoding: A machine representation of data including the specification of addressing (absolute, relative or symbolic), data compression and machine code.

Information reality in map conflation consists of the geographic based data that is the portion of reality and is used as tools to fully understand the topic. Figure 6.2 shows the reality in surface structure. Layers used in this process are a digital orthophoto, a road centerline map, hydrography, railroad, political boundary, and Census Maps.

Figure 6.2 Information Reality in Map Conflation
Information structure is described by entity, attribute, and relationships. An entity is a “thing” in the real world with an independent (Elmasari, 1994). An attribute is a characteristic of an entity selected for representation. Relationships, definable for single entities or classes, are the associations between phenomena, such as the land parcel having a house and the car having an owner (Elmasari, 1994). Figure 6.3 provides the information structure.

![Diagram of Information Structure]

Figure 6.3 Information Structure

In the information structure diagram, an entity is shown in a rectangle and its attributes are shown as ovals. There are five main entities such as digital orthophoto,
point, string, polygon, and layers (river, road, township, and so on). In the real world, there are relationships between the real world image and others. But, for the proposed map conflation, digital orthophoto is only used in Q/C-ing process so there is no explicit relationship between the digital orthophoto and the others. A relationship between entities is represented by a diamond. Point, string, polygon, and map entities are inter-related with each other.

The spatial objects used for canonical structure are the pixel for digital orthophoto, the point for control points, the string for road centerlines, railroads, and hydrographic entities, and the GT-polygon for Census maps and political boundaries (Figure 6.4). Since Census maps can have holes, a self-loop is used for the indication of inside GT-polygons.
Data Structure provides the logical structure of data accessibility of the Canonical structure. The accessibilities are represented by explicit links, i.e., resolved through pointers and implicit links, i.e., resolved through other forms of reference. Figure 6.5 shows the data structure.
Figure 6.5 Data Structure
There are two characteristics in the data structure: explicit links are used to indicate an island polygon in Census Maps because they include holes. Parent and child pointers are used to trace topological information in node and edge level. For example, topological information such as the neighbor polygon’s id, edge direction, area, etc. are available on any node. Using this information, the wanted map features can be found with minimizing on-the-fly operations.

6.4 Screen Design

The map conflation application is mainly divided into two parts. One part is the automated map conflation procedures. Those procedures do not need human-computer interaction (it is a one way process). The other part is related to visualization and analyses in a Graphic User Interface (GUI) environment. For this part, one piece of GIS software, ArcView, is used for the following reasons.

- It provides customization tools for GUI environment.
- It allows us to save time for developing GUI environment.
- Data conversion from ArcView’s shape file to any customized format or vice versa is possible.

It is time-consuming to develop a GUI environment and rather difficult for GIS professionals that are not programmers. ArcView’s customized tools allow us to design our own graphic user interface. The biggest problem in using commercial software is related to data format. If the customized data format is not supported in the target software, there is no way to use those data in that software unless it is converted to its
format. In terms of data conversion, ArcView also provides flexibility because shape file format is open to public (ESRI, 1995). Therefore, any customized data can be converted into shape files rather painlessly. However, one main disadvantage of shape format is that it does not support topological information. The topological information has to be rebuilt in the customized data format. The avenue scripts, which convert the customized data to shape polyline and polygon, are provided in Appendix B.

GUI consists of visualization for display and menus, buttons, and tools bar for communication. The custom tools bar for quality control and edit are added. ArcView’s screen design enables us to view a selected map, change window size, run command, and get result graphically and efficiently. Figure 6.6 shows the customization of ArcView’s screen design. The overlay of each layer is controlled by the order of placement. Each layer can be turned on or off when needed. Zoom related functions such as Zoom in and out, Zoom to Select, Pan, etc. are provided in the form of tools, buttons, and menus.

6.5 Communication Design

A GUI enables us to exchange information efficiently. Users do not need to memorize commands that they have to use. ArcView’s menu, button, tool, and customized tools are used for communication operations. Decisions should be made when match pairs or strings are not clear.
Turk (Turk, 1990) distinguishes GIS-based decision as three types; 100% cognitive responsibility on the human side, 0% cognitive responsibility on the human side, and shared cognitive responsibility between human and computer. The decision to determine matched pairs is applicable to the last case because a digital map provides some of information that helps the human decision process. This is also true when automatic and manual editing is need. For example, the background digital orthophoto provides location
information for the human. The customized tools are shown in Figure 6.7. Other commands and menus that are provided by ArcView are categorized by the same topic and they are meaningful and indicative. For example, the file menu consists of file read, save, etc., which are related to file operations so that users can trace the menus that are looking for. Button Bars provide a faster method for executing the same commands available through the menu items that are used frequently (Razavi, 1995). The tool bar is associated the special procedures shown in the customized tools in Figure 6.7. A tool bar is mutually exclusive unless the other tool bar is selected.

Figure 6.7 Customized Tools
6.6 Spatial Data Set for Census Maps

This section describes the spatial data set for Census maps in terms of implementation standpoints.

6.6.1 Census 2000 Tract Coverage

Figure 6.8 shows the Census 2000 Tract Boundaries. There are a total of 29 polygons. The original TIGER format is changed to ArcView’s Shape file to maintain consistent data format for all spatial data sets. The size of this file is 44,238 bytes.

Figure 6.8 Census Tracts
6.6.2 Census 2000 Block Groups Coverage

This coverage is more detailed than the Tract coverage as shown in Figure 6.9. The Block Group coverage consists of 71 polygons. Each tract polygon is subdivided by the average of 2.4 polygons to create the block group coverage. The size of this file is 95,025 bytes.

Figure 6.9 Census Block Groups
6.6.3 Census 2000 Blocks Coverage

Figure 6.10 shows the most detailed level of the Census Blocks. There are 2234 polygons. The size of this file is 1.015,212 bytes.
6.7 Spatial Data Set for Reference Maps

This section describes the spatial data set for reference maps, i.e., Delaware County's GIS bases. Reference maps should have more geometrically and topologically accurate than source maps.

6.7.1 Digital Orthophoto

Figure 6.11 shows a county-wide Digital Orthophoto at the resolution of 1 foot per pixel. The image capture year is 1997. The image is saved in MrSid format and the size of this file is 643,962,230 bytes.

Figure 6.11 Digital Orthophoto
6.7.2 Road Center Line

This is the largest vector coverage and consists of 8145 polylines. Figure 6.12 shows this coverage and the size of this file is 4,976,581 bytes.

Figure 6.12 Road Center Lines
6.7.3 Railroad

Railroad is not one of the major map layers in Delaware County, OH. Three railroads pass through as north and south bound routes as shown in Figure 6.13. Twenty seven polylines are used to save these railroads and the size of this file is 11,831 bytes.

Figure 6.13 Railroads
6.7.4 Hydrology

There are four major hydro features in Delaware County (Scioto River, Delaware Reservoir, Alum Creek, and Hoover Reservoir) amounting to 21 polygons as shown in Figure 6.14. The size of this file is 940,943 bytes.

Figure 6.14 Hydrographic Boundary
6.7.5 Township Boundary

This is one of the political boundaries and is shown in Figure 6.15. This layer consists of 19 township polygons and the size of this file is 408,551 bytes.

Figure 6.15 Township Boundary
6.7.6 Municipal Boundary

Municipal Boundary layer consists of 10 polygons and the size of this file is 136,751 bytes. Figure 6.16 shows this layer.

Figure 6.16 Municipal Boundary
6.8 The Used Software and Hardware

This section describes the software and hardware in map conflation system.

6.8.1 Software

In order to code the proposed map conflation’s logical design, an object-oriented program (C++) is used under the environment of Visual C++ 6.0 where the operating system is Windows 2000. For visualization, ArcView 3.2 for PC is used. Avenue (ArcView’s Script) is also used to provide the interface between ArcView’s Shape file format and user defined data format as discussed in the section 6.4.

6.8.2 Hardware

The generic desktop IBM compatible PC is used. Detailed specification includes: Intel Pentium III processor, memory at 128 MB, and hard disk space of 5GB. The screen sizes are 14” and 19”. A 19” screen is especially useful when visualization proof is needed.

6.9 Summary

In this chapter, map conflation system design is discussed based on Nyerges’s spatial data system design. First, logical design of map conflation system is presented. And then information structure, canonical structure, data structure are discussed. Finally, screen design and communication design that utilize ArcView’s customization are provided. Spatial data set for Census maps and Delaware County’s GIS bases are described in terms of the viewpoint of implementation. Data history and completeness for Delaware
County’s GIS base are provides in Appendix A. Software and hardware used in coding, visualization, and implementation for map conflation system are also provided.

In the next chapter, the results and findings of the proposed map conflation are discussed.
CHAPTER 7

RESEARCH RESULTS AND FINDINGS

In this chapter, the results and finding are presented. The results of each spatial algorithm of the proposed map conflation are provided and the findings are discussed. Since the proposed map conflation conflates census map hierarchically, the first level is shown because the next level procedures are exactly same as the previous one. The result of each map transformation is also analyzed by the comparison of Root Mean Square Error (RMSE), mean distance, range, and Standard Deviation (SD) of distance difference before and after the transformation. The transformation strongly depends on the distribution and the movement of control nodes. The transformed results are used as the input for the next level conflation.

7.1 Research Results

The results are discussed by the order of the spatial algorithm of the proposed map conflation. The analysis of map transformations is also discussed for the three methods in the three difference areas.
7.1.1 The Spatial Algorithm for the Proposed Map Conflation

Figures 7.1, 7.2, and 7.3 show map conflict conditions in downtown, residential, and water-prevalent areas. Red lines represent the Census maps and the blue lines represent the reference maps in each figure. These three areas are selected because each represents a unique characteristic. The first one, which is the downtown area, represents a regular grid pattern in its coverage and has easily recognizable map features such as intersections. The second area depicts areas of new development and construction in residential areas (this is where growth is occurring). Finally, hydrographic features have irregular natural shapes. In the case of hydrographic features, the dramatic differences are mainly due to the map generalization, but also may be due to physical changes in the shorelines as a result of erosion and weathering. Taken together, these characteristics make map conflation for hydrography the most difficult of the three areas. The separation, indicated by arrows in each figure, gives some examples how far apart the two versions of the same feature are from each other.

Some intersections and some turning nodes were surveyed in each tract areas in order to compare the location before and after transformation. The matched nodes to the surveyed nodes were manually selected by visual verification using ArcView software. The control nodes on the tract boundary were automatically selected in the procedure of the control node selection. These nodes and the associated points that consist of strings are exactly matched with the reference strings. Therefore, there will be no distance difference after transformation.
Figure 7.1 Conflict Condition between Census Map and Reference Map in Downtown Area
Figure 7.11 Conflict Condition between Census Map and Reference Map in Residential Area
Figure 7.3 Conflict Condition between Census Map and Reference Map in Hydrographic Area
<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>RMSE$x$</th>
<th>RMSE$y$</th>
<th>Mean Distance</th>
<th>Range</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Downtown</strong></td>
<td>87</td>
<td>51.76</td>
<td>50.95</td>
<td>70.51</td>
<td>104.99</td>
<td>17.53</td>
</tr>
<tr>
<td><strong>Residential</strong></td>
<td>43</td>
<td>220.30</td>
<td>219.32</td>
<td>261.45</td>
<td>618.05</td>
<td>170.14</td>
</tr>
<tr>
<td><strong>Hydrographic</strong></td>
<td>86</td>
<td>356.27</td>
<td>312.69</td>
<td>388.41</td>
<td>1483.56</td>
<td>271.90</td>
</tr>
</tbody>
</table>

All units are in feet.

Table 7.1 RMSE, Mean Distance, Range, and SD between Intersection and Turning Nodes and Their Matched Ground Truth.

Table 7.1 shows RMSE$x$, RMSE$y$, mean distance, range (maximum difference), and SD (Standard Deviation) between the surveyed points before transformation. Accuracy can be evaluated by RMSE calculated as follows (Jensen, 1996; Longley, 1999; Harley, 1975).

\[
RMSE_x = \sqrt{\frac{(x' - x_{orig})^2}{n}}
\]

\[
RMSE_y = \sqrt{\frac{(y' - y_{orig})^2}{n}}
\]

where \(x_{orig}\) and \(y_{orig}\) are the original location and \(x'\) and \(y'\) are the matching location of the original point. Ideally, \(x'\) would equal \(x_{orig}\) and \(y'\) would equal \(y_{orig}\). \(n\) is the number
of observation. Standard deviation (SD) is also used to evaluate the closeness of measurement (precision).

\[
SD = \sqrt{\frac{\sum (d - \bar{d})^2}{n-1}}
\]

where \( n \) is the number of observation, \( d \) is distance and \( \bar{d} \) is the mean of the distances.

The RMSEx (51.76) and RMSEy (50.96) of downtown are the smallest among three areas. Since downtown area is a regular grid pattern as shown in Figure 7.1 and control nodes represent intersections that are well recognizable in the process of mapping, the smallest RMSE is expected. Mean distance is 70.51, range is 104.99, and SD is 17.53. The smallest SD, 17.53, among three areas suggests that the least random errors are occurred during the process of mapping. RMSEx is 220.30, RMSEy is 2119.32, the mean distance is 261.45, range is 618.05, and SD is 170.14 in the residential area.

Figure 7.4 RMSEx and RMSEy in Downtown, Residential, and Hydrography Area.
Finally, they are RMSE\(_x\)(356.27), RMSE\(_y\)(312.69), the mean distance(388.41), range(1483.56), and SD(271.90) in the hydrographic area. The both RMSE of hydrographic is higher than that of residential as well as mean distance, range and SD. These suggest that the difference between Census maps and reference increases if there are the more changes and more irregular shapes in the real world.

Figure 7.4 shows the bar graph for RMSE\(_x\) and RMSE\(_y\) in three areas. The smallest one is in downtown and the largest one is in hydrography as discussed on the above. The difference between RMSE\(_x\) and RMSE\(_y\) is 0.8 in downtown, 0.98 in residential, and 43.58 in hydrography. These differences are the same order as the difference of RMSE for each area.

Figure 7.5 also shows the histogram of distance difference between intersection and turning nodes and their matched ground truth in the three areas. In downtown area, the distance difference is very stable as shown in standard deviation and histogram. The downtown area has not been changed very much since it was built. The old buildings and roads still exist in the downtown area. We could derive systematic shifts are more prevalent in this area than in the other areas when converting to a digital map or making hardcopy map because downtown has the least random errors. In the residential area, the distribution of distance difference is stable below 100 feet, but over 100 feet, the distribution is very irregular. Residential area represents lots of development, especially, in Delaware County, OH where new houses and roads are built rapidly. Census maps did not capture these changes, i.e., they were not updated very well. In the area with water feature, the differences are bigger than in the other areas, and the distribution also shows
irregularity. The RMSE\textsubscript{x} is 356.26 and RMSE\textsubscript{y} is 312.69, which are about 6.5 times bigger than downtown and about 1.5 times bigger than residential, as shown in Figure 7.4. The map generalization was done in this area when USGS 1:100,000 maps were made and river lines keep changing because of erosion and weathering. Therefore, we can conclude that a larger difference is expected.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{images/downtown_residential_hydrography.png}
\caption{The Histogram of Distance Difference between Intersection and Turning Nodes and Their Matching Ground Truth.}
\end{figure}
Figure 7.6 Control Points on Tract Boundary in Downtown Area with Inside Block Polygons
Figure 7.7 Control Points on Tract Boundary in Residential Area with Inside Block Polygons
Figure 7.8 Control Points on Tract Boundary in Hydrographic Area with Inside Block Polygons
7.1.2 The Selection of Control Nodes on the Boundary

Figures 7.6, 7.7, and 7.8 show control points on the three tract boundaries and blocks are shown in each boundary. These control points are distinguished by different types such as road, railroad, hydrographic, political boundary, and junction nodes between neighbor tract polygons or between inside block polygons.

In the downtown area, 45 control nodes are selected. Most of them are intersection nodes as shown in Figure 7.6. 38 control nodes and 48 control nodes are selected in the residential area and in the hydrographic area, respectively.

7.1.3 Node and String Match

Matched points are found based on conflation match strategies that include conflation test, and the combination of attribute, geometry, and topology. The matching tests are applied to attributes such as road names and type of lines, proximity, direction of lines, and sharing lines of nodes. The tolerance of angle is given by 30 degree and the proximity distance is given in 2500 ft to select multiple candidates. Table 7.2 shows the number of match pairs according to match types.

<table>
<thead>
<tr>
<th></th>
<th>Total Matches</th>
<th>Exact Matches</th>
<th>Partial Matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downtown</td>
<td>45</td>
<td>41(91.1%)</td>
<td>4(8.9%)</td>
</tr>
<tr>
<td>Residential</td>
<td>38</td>
<td>31(81.5%)</td>
<td>7(18.5%)</td>
</tr>
<tr>
<td>Hydrography</td>
<td>48</td>
<td>43(89.6%)</td>
<td>5(10.4%)</td>
</tr>
</tbody>
</table>

Table 7.2 The Number of Match Pairs according to the Exact and the Partial Matches
The exact match means all matching tests are passed for the closest potential matching features, and the partial match means that no potential matching features satisfy all matching tests. In that case, best-fit pairs are selected according to the order of the proximity. In all three areas, the exact match cases are 81.5% (residential), 89.6% (hydrography), and 91.1% (downtown). Based on these match pairs, a quality control (Q/C) process is conducted by a visual proof to verify the match pair results. This process is implemented in an ArcView environment in which visualization tools are provided and can be customized as illustrated in Figure 7.9. A Q/C process is an interactive process in which graphic user interface guides the matched pairs in click-and-go environment.

Figure 7.9 Q/C Process for Matched Pairs by Visualization
The deep structure (spatial structure) of matched nodes is transformed to surface structure (display of spatial structure or image) to overlay with digital orthophoto. There are 16 real and virtual transformations (Moellering, 1984). In Figure 7.9, if a user clicks on a table record, then the table record, a control node and its candidate matching record and nodes are highlighted in yellow on their location. The edit box provides all information about matching conditions. For example, tract id, control nodes id, their coordinates, type, (in this case, number “1” represents the node is associated with road layer, and match type “uc”, and (in this case, “110” means all match tests are passed on that control node. The partial matching nodes are thoroughly examined to determine the right node. For example, in the downtown area, among 4 partial matching nodes, three partial matched occurred because there are no reference features on the reference maps, i.e., the detail river lines are not available in the reference maps. Topology surgery (replacing an entire neighborhood by a more detailed neighborhood that agrees on the boundary) is needed in one case. In the residential, all seven partial matches result from no reference maps. Finally, two cases are from no reference data and three cases are from topology in the hydrography.

Strings between each control points are updated after Q/C-ing is done. As discussed in the chapter 3, a matching string is found in the reference map with the help of a guide string and its two end matching nodes. Figure 7.10 shows the steps of finding a matching string. In Figure 7.10(A), the reference string, which is from Census map, is specified based on match pairs between Census maps and reference maps. A buffer region is applied to find candidate line segments from the reference maps. The string that has dots
is candidate line segments in Figure 7.10(A). A candidate string is made by adding those line segments. And if there are multiple candidates like Figure 7.10(A), those multiple candidate strings are compared with the reference string. Finally, after the string that has a loop is ignored because it has a longer length than the reference string and it has a sharp turning point that is not in the reference string, the target string is shown in Figure 7.10(B).

![Diagram](image)

Figure 7.10 Finding a Target String Search.
7.1.4 Conflation Map Transformation

Conflation map transformation consists of the replacement for matched boundary and the rubber-sheeting for map objects inside the boundary.

7.1.4.1 Replacement of boundary

Figure 7.11 shows the replacement of the tract boundary by the matched strings. Since Figure 7.11 is at a small scale, the sub-portion of area is enlarged to provide a better view (Figure 7.12). In Figure 7.11, the red line represents the original tract boundary and the blue line shows the matched strings by the map conflation algorithm. The more magnified image is shown in Figure 7.12 to see how the matched strings are replaced on the digital orthophoto. In Figure 7.12, the red lines represent the original tract boundary before update and the blue lines are the updated boundary based on the matched control nodes and strings discussed on the above. 63 line segments are updated with 252 line segments from reference maps and RMSEx and RMSEy equal 0 because all control nodes are exactly aligned. The better result of this proposed map conflation compared to the standard rubber-sheeting is shown in Figure 7.13. In Figure 7.13, the same area as Figure 7.12 is shown. The green lines represent block lines and red dots represent control nodes used for rubber-sheeting in Figure 7.13. We can compare the string that connects control node A and B in Figure 7.13 and the same string (I – II) in Figure 7.12. As we see, the control nodes A and B are exactly aligned to the ground truth in Figure 7.13, but the string is not aligned to the ground truth because the rubber-sheeting keeps the original linear shape. However, that original string is updated with string (I – II) in Figure 7.12.
Figure 7.11 The Replacement of the Source Boundary By the Reference Boundary That Has Geometric and Topologic Consistent Conflation Results
Figure 7.12 Updated Boundary by the Matched Strings with Digital Orthophoto as Background
Green lines: conflated lines
Red points: control nodes

Figure 7.13 Updated Boundary by Standard Rubber-Sheeting Transformation.
7.1.4.2 Rubber-sheeting for inside map objects of the boundary

Rubber-sheeting is implemented for the map objects bounded by the tract boundary. Three methods are applied. They are DT (piecewise rubber-sheeting transformation by Delaunay Triangulation), WDT (piecewise rubber-sheeting transformation by Weighted Delaunay Triangulation), and Linear feature based transformation (LineMorp) as discussed in the chapter 4. The DT is constructed based on the control nodes on the tract boundary. However, when control nodes are selected on an elongated boundary, narrow and thin triangles are created in the DT. Therefore, the nearby control points do not always have a significant effect in this transformation. In order to provide more control points, neighboring triangles selected by the Voronoi diagram, which is the dual of the Delaunay triangulation, are added in the transformation in WDT. The result of this method is expected to be better than the DT because it uses more control points and less elongated triangles, i.e., it produced a more regular distribution of triangle vertices. Finally, LineMorp is implemented in which all control points have some effect in transformation. However, the nearest control points for a transformed point have more effect than the furthest ones. The weights are \( a = 0.5, \ b = 2, \) and \( p = 0.2 \) (see section 4.4).

7.1.4.3 Rubber-sheeting Analysis

The rubber-sheeting results are dependent upon the movement and distribution of the control nodes. In order to compare each transformation method, RMSEx, RMSEy, mean distance, range, and SD are implemented. First, those statistic values are made for control nodes that are found on the boundary (Table 7.3.A). The RMSEx and RMSEy, mean
distance, range, and standard deviation increase from downtown to residential to hydrography. This result is identical to the results discussed in the section 7.1.1.

<table>
<thead>
<tr>
<th></th>
<th>Control Nodes</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>RMSEx</td>
<td>RMSEy</td>
<td>Mean Distance</td>
<td>Range</td>
<td>SD</td>
</tr>
<tr>
<td>Downtown</td>
<td>45</td>
<td>74.27</td>
<td>59.72</td>
<td>84.90</td>
<td>258.37</td>
<td>43.53</td>
</tr>
<tr>
<td>Residential</td>
<td>38</td>
<td>134.39</td>
<td>218.20</td>
<td>170.44</td>
<td>900.35</td>
<td>192.64</td>
</tr>
<tr>
<td>Hydrography</td>
<td>48</td>
<td>142.78</td>
<td>342.46</td>
<td>256.62</td>
<td>1391.60</td>
<td>268.38</td>
</tr>
</tbody>
</table>

A

<table>
<thead>
<tr>
<th></th>
<th>Control Nodes after String Update</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>RMSEx</td>
<td>RMSEy</td>
<td>Mean Distance</td>
<td>Range</td>
<td>SD</td>
</tr>
<tr>
<td>Downtown</td>
<td>283</td>
<td>68.70</td>
<td>51.62</td>
<td>80.93</td>
<td>285.12</td>
<td>28.96</td>
</tr>
<tr>
<td>Residential</td>
<td>487</td>
<td>141.57</td>
<td>135.32</td>
<td>164.59</td>
<td>905.69</td>
<td>106.24</td>
</tr>
<tr>
<td>Hydrography</td>
<td>639</td>
<td>118.29</td>
<td>305.75</td>
<td>244.81</td>
<td>1401.98</td>
<td>218.22</td>
</tr>
</tbody>
</table>

B

All units are in feet.

Table 7.3 RMSE, Mean Distance, Range, and SD of Control Nodes on the Boundary.

These relations are maintained for control nodes after string update (Table 7.3.B) except RMSEx in hydrography. However, the combination of RMSEx and RMSEy are bigger
than that of residential. If one of matching string is longer than the other, the increases of RMSE occur when adding more nodes on a line to restore one-to-one relation in the matching string (section 4.5). More control nodes are added on to the tract boundary to restore one-to-one relation in transformation. For example, 45 control nodes increase to 253 control nodes in downtown in Table 7.3B. Second, some easily recognizable points such as intersection and turning points are selected inside the tract boundary. Then those matched pairs are selected manually on the digital orthophoto as a background. Therefore, these match points are assumed as ground truth. The RMSEx, RMSEy, mean distance, range, and SD between those selected matched pairs are used as the measure of comparison. They are shown in Table 7.1. Finally, after rubber-sheeting, the transformed matched points are traced and the accuracies are calculated with ground truths for the three transformation methods. Table 7.4 shows these results.

<table>
<thead>
<tr>
<th></th>
<th>Delaunay Triangulation(DT)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSEx</td>
</tr>
<tr>
<td><strong>Downtown</strong></td>
<td>102.62</td>
</tr>
<tr>
<td><strong>Residential</strong></td>
<td>119.72</td>
</tr>
<tr>
<td><strong>Hydrographic</strong></td>
<td>142.82</td>
</tr>
</tbody>
</table>
### Weighted Delaunay Triangulation (WDT)

<table>
<thead>
<tr>
<th></th>
<th>RMSEx</th>
<th>RMSEy</th>
<th>Mean Distance</th>
<th>Range</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Downtown</strong></td>
<td>83.28</td>
<td>48.87</td>
<td>73.36</td>
<td>62.27</td>
<td>12.94</td>
</tr>
<tr>
<td><strong>Residential</strong></td>
<td>91.31</td>
<td>73.35</td>
<td>100.88</td>
<td>258.19</td>
<td>60.29</td>
</tr>
<tr>
<td><strong>Hydrographic</strong></td>
<td>120.47</td>
<td>156.52</td>
<td>188.06</td>
<td>268.74</td>
<td>60.72</td>
</tr>
</tbody>
</table>

### Linear Feature Based Transformation (LineMorp)

<table>
<thead>
<tr>
<th></th>
<th>RMSEx</th>
<th>RMSEy</th>
<th>Mean Distance</th>
<th>Range</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Downtown</strong></td>
<td>85.62</td>
<td>34.76</td>
<td>88.68</td>
<td>65.51</td>
<td>13.95</td>
</tr>
<tr>
<td><strong>Residential</strong></td>
<td>78.97</td>
<td>95.19</td>
<td>96.13</td>
<td>370.28</td>
<td>78.74</td>
</tr>
<tr>
<td><strong>Hydrographic</strong></td>
<td>112.76</td>
<td>162.32</td>
<td>172.49</td>
<td>534.92</td>
<td>97.07</td>
</tr>
</tbody>
</table>

All units are in feet.

Table 7.4 Comparison of the Three Transformation Methods in the Three Areas.

In residential area, the mean distance after transformation is decreased in the order of DT (130.66), WDT (100.88), and LineMorp (96.13). In water-prevalent areas, the mean distance after transformation is also decreased in the order of DT (213.00), WDT (188.06), and LineMorp (172.49). It is expected that WDT produces better results than DT because it uses more control points selected by the second order Voronoi diagram as discussed in the section 4.3. The others such as RMSEx, RMSEy, range, and SD are also decreased in WDT than in DT. However, it is interesting that the mean distance of LineMorp produces the best result in those two areas. They are 96.13 in residential and
Figure 7.14 Comparison Before and After Transformation for Three Areas
172.49 in hydrographic. Since LineMorp is a global transformation, i.e., it considers all control points on the boundary according to weight constants, the transformation reflects all movement of control points. Figure 7.14 plots the change of all statistical values for each of three methods in each of three areas. Further analyses for the three methods in the residential and hydrography show the following.

- The transformation results of WDT are better than DT in every case.
- The range of LineMorp is always bigger than WDT.
- The range of all the three methods after transformation is reduced from 856 to 276.
- After transformation by DT, RMSEx, RMSEy, mean distance, range, and SD are dropped by the half of the original.
- The mean distance of LineMorp is smaller than WDT.

In the downtown area, none of the methods decrease the both of RMSE before transformation. However, DT and WDT show stable results even though they increase the mean distance to 3.34 ft and 2.85 ft, respectively. LineMorp produces the biggest increase in mean distance (19.17 ft). In WDT and LineMorp, RMSEy is decreased to 48.87 and 34.76 from 50.95, respectively, but RMSEx is increased to 83.28 and 85.62 from 51.76. This situation shows the difficulty of map conflation in an irregular domain, i.e., there might be no general rule for every case. However, we can tract such a case if we can compare RMSEs between control nodes and matched nodes. The RMSEx and RMSEy of control nodes are 68.70 and 51.62 and those of matched nodes are 51.76 and 50.95. If both of RMSEs of matched nodes is smaller that those of control nodes, the
result of transformation could increase RMSEs of matched nodes because RMSEs of control nodes is larger. Figure 7.15 shows also a visual proof by visualizing the matched pairs before transformation inside the tract boundary. Black points are the selected points before transformation. Yellow points are their corresponding points selected by visual verification on the digital orthophoto. The main pattern of the movement of the match pairs is seen to be southwesterly. Figure 7.15 also shows the matched pairs used as control nodes in transformation on the tract boundary. Blue points are the selected control nodes on tract boundary and red points are their matched points selected in the reference maps. The displacement of control nodes is to the southwest direction in the upper left, and right boundaries. However, on the bottom boundary, the displacement of them is to the east direction. The transformation results reflect the relations that result in some increase in distances after transformation.

This unexpected situation can be detected by comparing RMSEs and the mean distance between control nodes (Table 7.3) and the transformed point pairs (Table 7.4). The transformed point pairs can be automatically selected in the program (i.e., intersections: points that shared by more than three lines). If RMSEs and the mean distance do not decrease for transformed pairs, there may be many changes for which the movement of control nodes does not correspond to the movement of map feature inside the boundary and RMSEs of control nodes are larger than transformed pairs. In such areas, there are lots of chances that uniform systematic errors occurred when making hardcopy maps or converting to virtual maps.
Figure 7.15 The Distribution of Control Nodes on the Boundary and Matched Points inside the Boundary
X and Y-axis differences for inside matched points are calculated and their histogram is shown in Figure 7.16.

![Histograms of X and Y differences](image)

**X difference (feet)**
- Mean: 49.93 ft
- Standard deviation: 13.71

**Y difference (feet)**
- Mean: 46.50 ft
- Standard deviation: 20.91

Figure 7.16 The Histogram of X and Y Distance Difference for the Inside Points in the Downtown.

Figure 7.16 shows a nearly normal distribution for X and Y distance difference. The mean distances are 49.93 ft and 46.50 feet for X and Y-axis, respectively. If there are uniform systematic errors, simple translation will decrease the distance to the ground truths rather than rubber-sheeting. The simple translation by 49.93 in x and 46.50 in y is applied. A new RMSEx(13.63) and RMSEy(20.82) are calculated, which is better result than original RMSEx(51.76) and RMSEy(50.95). This proves that the downtown area has uniform systematic errors. The other two areas are also shifted to see if there are uniform systematic errors. Table 7.5 shows the results.
<table>
<thead>
<tr>
<th></th>
<th>RMSE&lt;sub&gt;x&lt;/sub&gt;</th>
<th>RMSE&lt;sub&gt;y&lt;/sub&gt;</th>
<th>X Shift by mean</th>
<th>Y Shift by mean</th>
<th>New RMSE&lt;sub&gt;x&lt;/sub&gt;</th>
<th>New RMSE&lt;sub&gt;y&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downtown</td>
<td>51.76</td>
<td>50.95</td>
<td>49.93</td>
<td>46.50</td>
<td>13.63</td>
<td>20.82</td>
</tr>
<tr>
<td>Residential</td>
<td>220.30</td>
<td>219.32</td>
<td>27.15</td>
<td>-101.68</td>
<td>218.62</td>
<td>194.32</td>
</tr>
<tr>
<td>Hydrography</td>
<td>356.27</td>
<td>312.69</td>
<td>120.93</td>
<td>-151.50</td>
<td>335.12</td>
<td>273.54</td>
</tr>
</tbody>
</table>

All units are in feet.

Table 7.5 Comparison of RMSE after Translation by Mean X and Y Difference.

Table 7.5 shows both of the RMSEs reduced by a translation in downtown. However, In residential and hydrography areas, the simple translation does not reduce RMSE as much as DT, WDT, and LineMorp. For example, RMSE<sub>x</sub>(220.30) and RMSE<sub>y</sub>(219.32) in residential is reduced to 218.61 and 194.32 by the translation in Table 7.5, but they are reduced to 91.31 and 73.35 by WDT in Table 7.4. This proves that map transformation by three methods is helpful if the area that has lots of changes and map generalization, i.e., several errors are combined.

In summary, the transformation depends on the movement and distribution of control point and the result of the transformation decrease the distances of the transformed map features to the ground truths. However, it would be possible that the movement of control points might not reflect true relationships for other map features because of the irregular property of problem domain. This case is could be detected by the comparison of RMSE difference between the control nodes and the transformed points inside the boundary. The
transformed results that reduce the distances from the ground truth are used as inputs for the next level conflation.

7.2 Discussion of Findings.

This section discusses the findings from the results of this dissertation.

- Conflict condition between Census maps and Reference maps.

The three areas surveyed are downtown, residential, and hydrography. The mean distance differences to the ground truths are 70.51 ft (downtown), 261.45 ft (residential), and 389.41 ft (hydrography) ft, respectively. We could derive that the distance difference is increased if there are many changes in an area and map generalization is heavily conducted for the area. For example, in the downtown, the roads and buildings have not changed since they were mapped. In the residential area, new houses and roads are rapidly built in this case study area, Delaware County, OH. Therefore, Census maps might not contain these changes whereas reference maps from Delaware County, OH are updated continually whenever and wherever there are developments. Finally, in the hydrography area, map generalization must be done heavily in this area because river lines are so simplified in Census maps. Other factors that cause differences are erosion and meandering of the river lines. River lines in the real world keep changing while they remain fixed in static maps.
• Map match rates between Census maps and reference maps

The results of the types of map match show the exact match cases overall are 81.5% (residential), 89.6% (hydrography), and 91.1% (downtown), respectively. The partial match cases result from missing reference maps available, and topological surgery may be needed, depending on the irregularity of the problem domain.

Figure 7.17 An Example of Partial Match in the Case of Reference Maps Is Not Available.
Figure 7.17 shows an example of what occurs when corresponding features on reference maps are not available for the automated matching process. A Candidate control node is indicated by the yellow arrow in Figure 7.17. The Red line that passes through horizontally represent a small streamline and the purple line that passes through vertically represents a road in the Census map, but their corresponding layers are not available in the reference maps. Only the road line in the blue color is available. Therefore, applying the matching procedure produces a partial match.

Figure 7.18 An Example of Partial Match in the Case of Topology Surgery Is Needed.
Figure 7.18 shows how topological surgery may help. The candidate control node has four degrees of intersection with approximately 90-degree angle. However, there does not exist any exact match map features nearby because the candidate control node was removed through map generalization.

One particular 0-cell conflation test is verified in this case as follows.

$$\text{graphic limit} \leq k(\text{distance} \cdot \frac{\text{Scale in Source}}{\text{Scale in Reference}})$$

Where Scale in Source = 1:1,200,
   Scale in Reference = 1:100,000,
   distance in 1:1 scale = 21 ft as shown in Figure 7.17
   and set k = 1.

Therefore, it produces 0.000064 mm after scale reduction. This is below graphic and visual threshold that is 0.6 mm. For this candidate control node, one-to-many matching is necessary. In order words, the bottom Census polygons are matched to node B and the upper Census polygons are matched to node A.

- The real and virtual map transformation

Moellering’s real and virtual transformations (Moellering, 1984, 2000b) are used for the verification and modification of the matched pairs. There are 16 transformations between real and virtual map (Moellering, 1984, 2000). The verification and modification of the matched pairs is one of transformations from virtual 3 to virtual 1 and from virtual 1 to virtual 3. The deep structures of matched pairs are converted to surface structures to be incorporated with an existing surface map, the digital orthophoto.
• Map transformation between Census maps and reference maps

The replacement is implemented for the matched strings. Therefore, it provides the exact same geometry as the reference maps. Rubber-sheeting is implemented whenever an old boundary is updated to a new one, and it is applied to map objects that are inside the old boundary. The rubber-sheeting results of our rubber-sheeting produce a distance difference that is reduced by at least the half of the original distance before the rubber-sheeting in the residential and hydrography area. However, in the downtown area, the rubber-sheeting less not helpful because the displacement and distribution of control nodes does not correspond well with the displacement of map objects inside the boundary. As Figure 7.15 has been shown, there are systematic shifts in the x and y directions. Therefore, simple translation will improve the distance difference. The proof is shown in Table 7.5. For example, by simple translation, RMSE$_x$ is reduced to 13.63 from 51.76 and RMSE$_y$ is reduced to 20.82 from 50.95 in downtown in Table 7.5. Table 7.5 also shows the area that contains lot of changes and heavy map generalization does not get benefit from the simple translation but from rubber-sheeting. The transformation results that decrease the distance to the ground truths are used as the inputs for next level conflation. Therefore, it will help to find matched features in the next level because map features will be more closely moved to the ground truths by transformation.

• The importance of reference maps
With reference maps, map conflation can be done in a way that quantifies feature movement needed to achieve proper alignment. This dissertation explains not only how well our map conflation algorithm works compared to others but also how important good reference maps are. Currently, reference maps from local governments are better sources for Census map conflation rather than point features surveyed by the Census Bureau. The reasons are as follows.

A. The local data contain up-to-date information, and also contain complete and consistent geometry and topology.

B. Duplicate effort in creating reference data (point and linear features) by the Census Bureau can be avoided.

C. The staffs at local governments’ GIS departments are very familiar with their data and area.

D. The local governments’ GIS bases usually already contain the desired data. (i.e., digital orthophoto, GPS surveyed point data, large scale maps, etc.). For example, there are 78 counties that have GIS activities among 88 counties in Ohio in 2001 (David, 2001).

• Hierarchical conflation approach

The hierarchical conflation approach provides many advantages. Localization of data for transformation and matching reduces complexity, and the overall computational load is reduced.
7.3 Summary

This chapter provides the result of the proposed map conflation and discusses findings. First, the results of the selection of control nodes are shown for the three different areas. Next, matched strings are found based on these control nodes. The difference between the standard rubber-sheeting map conflation and the proposed map conflation are shown in Figure 7.12 and 7.13. The advantage of the proposed map conflation procedure is that it produces a map that is geometrically and topographically consistent with reference maps. Thus, various spatial analyses are possible between them. Finally, three different map transformation methods are implemented to study their performance in hierarchically bounded map conflation. The study shows that if an area has physically changed a lot since a real map was made, or if a real map contains heavy map generalization, then our three methods reduce the transformation distance by the half. Moreover, if an area has not changed much, or if the map contains little map generalization, then even a simple translation greatly reduce the distance. The results of map transformation may be used as inputs for the next level of conflation. Therefore, it will be easier to find matched features in the next level because map features will be more closely moved to the ground truths by transformation.

In the remaining final chapter, the dissertation summaries and conclusions are provided.
CHAPTER 8

SUMMARY AND CONCLUSIONS

8.1 Summary of the Research

In this dissertation a new hierarchically bounded map conflation is presented to increase the positional accuracy of Census Maps using the reference spatial data from local governments that have large-scale, complete, and consistent spatial data. This method guarantees producing a conflated result that is topologically consistent with reference maps.

Two map models are provided for map conflation. The first model is based on the cell model of a map in which a map is a cell complex consisting of 0-cells, 1-cells, and 2-cells. The other model is based on enforcing that a homeomorphism condition is maintained among all maps. This theory helps to localize matching feature in map conflation by assuming one-to-one relation is existed everywhere. Matching operations in map conflation are defined in two regions, that is, homeomorphic and non-homeomorphic cases. If subregions are homeomorphic, conventional matching strategy is applied. And if they are not homeomorphic, topological surgery may be implemented.
Map conflation principles to conflate Census maps consists of attribute embedding, finding meaningful nodes, cartographic 0-cell match, cartographic 1-cell match, and map transformation. An attribute-embedding component was developed to assign a spatial key to each line segment, which links to source database. A meaningful node is defined not according to geometry properties, but according to attribute properties. The categories to distinguish meaningful nodes are (1) same type and (2) same junction. Cartographic 0-cell matching uses a 0-cell conflation test to see if the neighborhood of the 0-cell has a legitimate topology, in other words, if the two subregions are homeomorphic. Geometry, topology, and attribute match tests are performed if the regions are homeomorphic. Using two matched 0-cell pairs and its correspondent 1-cell in the reference map, cartographic 1-cell match is implemented. Finally, map transformations are applied to transform the features (Census block groups and blocks) inside Census tract boundaries. The results of transformation are used as the inputs for the next level conflation.

Three map transformation methods are discussed to find a satisfactory method when a polygon is used as the boundary of the transformation. If there is a non-convex boundary and control points are picked on that boundary, narrow and thin triangles are intermediate results that are used for rubber-sheeting in Delaunay triangulation. In order to compensate for these narrow and thin triangles, we used the second order of Voronoi diagram to find triangles that correspond to an influence polygon for a transformed point (called Weighted Delaunay Triangulation). Our last method uses the linear features instead of point features to produce a transformation. Depending on how corresponding linear features are scaled, translated, and rotated, geometric relations are calculated and
applied to obtain the transformation. These transformations strongly depend on the movement and distribution of control points. The results of transformations are used as the inputs of the next level of conflation. Therefore, there is a reason for reducing the distance difference to the ground truths so map features will be found in closer location to the ground truths in the next level of conflation.

This dissertation also suggests how national maps may be updated effectively by federal and local governments by cooperating and exchanging spatial information. Federal governments such as Census Bureau and USGS have always produced small-scale maps, and these maps have not been regularly updated since they were created. However, local governments make large-scale maps to assist them in managing land title records, appraising real estate for taxation purposes, planning and subdivision of land, and more. These maps are maintained and updated on a daily basis. In our case study area, Delaware County, OH, we studied how spatial data are created and maintained, how GIS is operated, and what its role is. Those findings are provided in Appendix A. In the proposed map conflation, the spatial data from local government are used as reference maps since they are the source of up-to-date, positionally accurate, and geometrically and topologically complete spatial data. This also eliminates duplicate efforts by Census Bureau in rebuilding reference data.

The implementation issues are discussed in terms of the theoretical design of map conflation system and presented as results for the case study area, Delaware County, OH. Three different types of areas, downtown, residential, and high-density water-featured areas, were selected for their distinct geometry and topology. The downtown area
exhibits a typical regular grid pattern. Map generalization was not done much in this area because all mapped features were recognizable at the given scale (1:100,000). Also there have not been many changes since the area was mapped. The residential area exhibits much new development and construction. The areas with many water features depict natural shapes that are irregular and are subject to considerable map generalization. The mean distance between some selected points and their ground truths is 70.51ft (downtown), 261.45ft (residential), and 389.41ft (hydrography), respectively. A selected hierarchical boundary in each of those three areas was replaced with reference data. The results of the different transformations are presented and compared with other standard rubber-sheeting methods. Map transformations were implemented by Delaunay Triangulation (DT), Weighted Delaunay Triangulation (WDT), and Linear based feature morphing (LineMorp). The transformation results show that the distance difference before transformation is reduced by at least the half of the original distance in residential and water-feature-dense areas. The data also show that WDT performed better than DT in every area of the case study. LineMorp showed the best performance in the residential and water-feature-dense area and the worst result in the downtown area where control points do not corroborate the boundary’s offset from ground truth. LineMorp is more sensitive to the movement and distribution of control points because it is a globally computed transformation. These varied results illustrate one difficulty of map conflation: the non-uniformity of map conflation domains. Our case study results suggest that if an area changes a lot, or if the map contains substantial generalization, then all three methods reduce the error by at least half. On the other hand, if an area has not changed a
lot or if the map contains little generalization, then even a simple translation will greatly reduce the offset error.

8.2 Conclusions

The conclusions of this dissertation range from theoretical nature of a new model of conflation to the empirical results and findings of implementing the proposed map conflation system. The primary contributions of this research are as follows:

- Introduced the homeomorphism map conflation model that allows us to correct topology differences in maps to be conflated. To detect topology differences resulting from resolution-accommodating generalization, a 0-cell map conflation test is introduced.
- Introduced the new hierarchical-based conflation that can be incorporated in map conflation approach; anytime that loops are matched, candidate match searches can automatically be constrained to loop interiors or to loop exteriors. Therefore, complexity and computation load can be reduced.
- Suggested the combined use of real and virtual transformations to successfully verify and modify matching uncertainty in map conflation system
- Provided two new map transformation methods for map conflations and evaluated them. The test study showed that two new methods may move map features closer to actual location than traditional Delaunay triangulation.
- Through the test study, provided some insight into relationships between map transformation and map generalization or map revision.
• Discussed sources of reference data and other sources of conflation methodology. The details of reference data for this research are provided in the Appendix A.

• Finally, produced a topologically consistent conflated result with reference maps so that various spatial analyses could be possible between them.

The goals of this research were to formulate a conceptual framework and mathematical model of map conflation, develop a new match method, and investigate, explore, and analyze map conflation as a tool of transformation and integration of spatial data in GIS. The following paragraphs discuss how these goals are achieved through this research.

Traditional map conflation employed rubber-sheeting based on point features. The mathematics of rubber-sheeting has each triangle with control points as vertices transformed into its counterpart triangle. The limit of this model is that it cannot correct topology differences and cannot add more linear features to represent true topology and geometry. We introduced a new mathematical map conflation model that is based on a homeomorphism guarantee, i.e., there always exists a bijection between maps. When a map scale is reduced, many-to-one relations are drawn on virtual or real maps because of the graphic or visual limits. This process is called map generalization in cartography. In order to restore these many-to-one relations in map conflation, two approaches were examined. First, find inverse functions of map generalizations. Second, simulate map generalization using reference data, which is a larger scale. The first solution is hard to provide explicitly because map generalization itself is not accomplished systematically in all cases. Fortunately, the second approach is possible because we have good reference
maps from local governments to which we may try to conflate Census maps. For example, 78 of 88 Ohio counties had GIS activities in 2001 (David, 2001). Therefore, finding and using highly accurate reference data is very important in map conflation because the positional and topological accuracy are bounded to the reference data. Using one new conflation model, we may apply our 0-cell conflation test to modify topology. The combination of attribute, geometry, and topology matching criteria discussed in the Chapter 3 may also be applied to match features. After finding matched 0-cells, the string that connects those 0-cells may be found using the reference string from Census maps as a guide.

Map transformations may be used to replace exactly matched strings or transform map features inside a boundary. The transformed map may be used as input for the next stage of conflation. Transformations should reduce positional error. The traditional Delaunay (DT) triangulation does this in a limited way if control points are picked on a non-convex boundary because narrow and thin triangles are inevitably created. In that case, nearby control points have a lesser effect in the transformation. Two new methods have been explored and evaluated for map conflation in which a polygon boundary is used as a constraint. They are Weighted Delaunay Triangulation (WDT) and Linear feature based transformation (LineMorp). In order to provide more control points, neighboring triangles, which we selected by the second-order Voronoi diagram, were added in transformation in WDT. This method performed better than the DT in our test cases because it uses more nearby control points and has fewer elongated triangles. In other words, it produced a more regular distribution of triangle vertices as discussed in the
Chapter 4 and this was further reinforced by the case study results in Chapter 7. RMSE and mean distance for control points were provided in the Chapter 7 to substantiate this statement. For the subroutine LineMorp, all control points have some effect in transformation. However, nearer control points for a transformed point have greater effect. Weights based on relative constants control the magnitude of the effect. The weight constants used were $a = 0.5$, $b = 2$, and $p = 0.2$. These constants may be determined experimentally.

Map transformation depends on the movement and distribution of control point. The results of transformation for three different areas such as downtown, residential, and regions with dense water features also suggest a relationship between transformations and map generalization or map revision. If a mapped area has been changed a lot or if the map is highly generalized, then conflation map transformations such as DT, WDT, and LineMorp may reduce the offset error by half or more. On the other hand, if a mapped area has not changed a lot or if the map contains only a little generalization, then a simple translation by the mean offset may greatly reduce the individual offset errors. There may be uniform systematic errors when a real map is created or a real map is converted to a virtual map (real $\rightarrow$ virtual type 3 transformation) (Moellering, 1984). This situation corresponds to downtown areas where roads and buildings have not changed since they were mapped.

Finally, it is worthwhile emphasizing the importance of reference maps in Census map conflation. Map conflation operations may easily quantify the feature movement needed to achieve proper alignment. Currently, reference maps from local governments are the
most accurate sources for Census map conflation. They are even more accurate than the point features surveyed by the Census Bureau itself. The reasons for this greater accuracy are discussed in the Chapter 7, and the case study of a local government’s GIS is provided in the Appendix A.

8.3 Future Research

This research has provided new fundamental frameworks of map conflation, but this is only a first step. A number of future research activities could grow out of this dissertation. Further investigation is needed to understand the differences of virtual maps and real maps. Digital versions of maps have been created from real maps by digitizing or scanning and transformed into virtual maps (real → virtual) transformation (Moellerling, 1984, 2000b). With the development of technologies, a direct creation of digital version of maps is possible (virtual → virtual) transformation. The future of digital map creation will be based on (virtual → virtual) transformation. In this research, the cell and homeomorphism map model were provided to illustrate some fundamental differences of real and virtual maps and to establish a location correspondence between them. A 0-cell conflation test was proposed to account for resolution-dependent topology collapse. Only distance and displacement generalization transformations were considered in this research. Other generalization transformations such as exaggeration, refinement, and aggregation still need to be investigated.

Current map transformations are based on point-to-point or line-to-line relationships. Therefore, these relations are explicitly linked in transformation. Further research is
needed how map transformation can be done with line-to-string or coarse-string-to-detailed-string without explicitly restoring vertex-to-vertex relation on those strings. Cartographers and Geographic Information scientists have mainly focused on the results of map generalization in such ways that how the generalized map features deliver information correctly and how they are displayed effectively. However, map generalization theory has not yet explicitly considered the impediments to building inverse functions that could link original map features. This is a rich area of interest for future research in developing such an information-loss theory for map generalization.

This research only examined Census map conflation. There are other spatial datasets in computer mapping and GIS that need a sound conflation theory. They include map revision or update, old orthophotos replaced by new ones, maintaining map consistency of overlays, maintaining logical content-placement consistency (e.g., a river should flow under a bridge), and spatial data integration. Further investigations are needed to extend conflation theory to these fields.
APPENDIX A

GIS in Local Government, Delaware County Ohio

A.1 Introduction

In the United States, a wide range of federal and local government agencies are involved in creating geographic information. For example, USGS (the United States Geological Survey), NIMA (the National Imagery and Mapping Agency), and the U.S. Forest Service are typical federal agencies. The U.S. Census Bureau has become a mapmaker in order to manage its geographic data efficiently even though its mission is not to make maps. At the state and local government level, entities involved can be states, counties, cities, townships, and villages. According to a study of the relation between federal government and local government geoinformation (Masser, 1998), while the federal government has mainly produced small-scale maps and statistical data for the entire nation, state government has worked for some land titles registration and small – and large-scale mapping. Finally, local governments have focused on making land title registration and creating large-scale datasets. Large-scale maps play a critical role in local governments’ daily functions. It assists them in managing land title recordation, appraisal of real estate for taxation purposes, planning and subdivision of land, and more. Some
pioneer state and local governments have developed tools that are similar to geographic information systems since 1960s (Masser, 1998; Foresman, 1998). Those tools include interpreting and digitizing aerial photography into land use categories. More state and local governments participated and experimented with GIS programs and expanded their facilities and staff during 1970s and 1980s (Foresman, 1998). However, GIS was still not present in every state and local government office mainly due to the high cost associated with the creation of digital geographic data, hardware, software programs, and staff training. Since the 1990’s, mainly due to the rapid change in the computer industry and the birth of small but powerful workstations, GIS has been implemented on inexpensive platforms and therefore, widely used by the majority of state and local government entities (Warnecke, 1995). GIS software programs also underwent a revolutionary change. They became more user friendly and at the same time more powerful and yet less costly. Now, local governments can implement a GIS project with a small staff and inexpensive equipment and concentrate on generating more customized applications to provide higher quality service to their constituents. Internet invention is another important trend to GIS as well as to other industries because Internet GIS has allowed wider access to the general public than traditional GIS.

In terms of creating geographic data, the 1990’s were also an innovative decade for local governments because they were able to create their own highly accurate large-scale maps by the means of GPS field surveys, GPSvan (Center for Mapping, 1993), and conventional aerial photo or digital Orthophotos. This helped them manage their land/parcel information system. The Census Bureau has recognized that exchanging very
accurate digital data through public and private partnership will play an important role in improving the positional accuracy of its TIGER files (Liadis, 2000). In this chapter, the process by which a local government’s spatial data is created and maintained will be thoroughly examined. The case study is Delaware County, OH.

A.2 Delaware County and Its Administrative Organization

Delaware County was established and organized in 1808. The name Delaware is derived from the Delaware Indians who came from the Delaware River area near Philadelphia (Yost, 2000). The County encompasses 19 townships and 10 municipalities, of which the city of Delaware is the largest and its area is 459 square miles. Delaware County is located approximately 30 miles north of Columbus (Figure A.1), which is the capital of Ohio and it has been and remains the fastest growing county in the state of Ohio and the 15th fastest growing county in the nation (according to Census Bureau’s 2001 population estimates). For example, the population has increased from 66,929 in the 1990 census to 109,989 in the 2000 census. The number of real estate parcels has increased from 37,926 in 1990 to more than 59,000 as of now. The number of transactions in the County Recorder’s Office has jumped from 35,111 in 1997 to 44,037 in 2000, reflecting the pace of real estate transactions (Yost, 2000). The land use pattern has been dramatically changed as a result of this growth and while the majority of the landmass in the north is still predominantly agriculture, in the southern part of the county, the agriculture use is disappearing at a phenomenal pace. Table A.1 provides the land use of parcels and their
areas. The biggest portion of the number of parcels is residential but farmland has the biggest area. This represents the characteristics of fast growing area, which is close to metropolitan area.

<table>
<thead>
<tr>
<th>Land Use</th>
<th>Number of Parcel (%)</th>
<th>Area (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agricultural</td>
<td>8.74 %</td>
<td>55.75 %</td>
</tr>
<tr>
<td>Industrial</td>
<td>0.62 %</td>
<td>1.04 %</td>
</tr>
<tr>
<td>Commercial</td>
<td>4.09 %</td>
<td>4.03 %</td>
</tr>
<tr>
<td>Residential</td>
<td>82.39 %</td>
<td>20.65 %</td>
</tr>
<tr>
<td>Other</td>
<td>4.13 %</td>
<td>18.49 %</td>
</tr>
</tbody>
</table>

Table A.1 Land Uses of Parcel and Their Area, 2002 Year (1st Quarter)

The County provides various services to its people, including general government, courts of law, public safety, public works, human services, health services, community development, and conservation/recreation services. A three-member board of County Commissioners is elected at-large in even-numbered years for overlapping four-year terms. The commissioners serve as the taxing authority, contracting body, and administrators of public services for the County. The county Auditor serves as the chief fiscal officer for the County and tax assessor for all political subdivisions within the County. It is one of the duties of the Auditor's office to see that every parcel of land and the buildings thereon are fairly and uniformly appraised and then assessed for tax
purposes. Personal property is every tangible thing that is owned, except real property while real property is defined as land, growing crops, all buildings, structures, improvements and fixtures on the land.

The Treasurer is the custodian of all County funds and is responsible for the collection of all tax monies due to the County, as well as investing all idle funds of the County as specified by law. The duties of County Engineer are to plan, construct, and maintain road systems, survey properties, and assist the planning commission, building regulations, zoning commissions, and so on. Other elected officials include the Recorder, Prosecutor, Clerk of Courts, Sheriff, Coroner, and the Court of Common Pleas Judges. Figure A.2 shows the Delaware County’s administrative organization. The organizations of other counties are very similar to Delaware County.

As Figure A.2 shows, while several other offices have GIS Divisions, the Auditor’s GIS Division, Delaware Appraisal Land Information System or DALIS Project, is the one that creates the foundation of all base coverages for the entire county. The Engineer’s Office’s GIS Division maintains the bridge layer while the Regional Planning Commission (RPC)’s GIS Division maintains Zoning and comprehensive Plan layers. The RPC is also responsible for drafting the Subdivision Regulations and along with the County Engineer, also responsible for reviewing every plat in the county. Some of the fundamental coverages created by DALIS Project are parcels, road centerlines, road right of ways, digital orthophoto, hydrology, voting precincts, municipal and township boundaries, GPS points, and school district boundaries.
Figure A.1 Delaware County with Municipalities and Townships Boundaries, Ohio Location
Figure A.2 Delaware County’s Administrative Organization
A.3 The History of GIS in Delaware County, Ohio

The history of GIS in Delaware County can be traced to 1989 when the Delaware County Regional Planning Commission (DCRPC) hired a GIS Coordinator. The Coordinator with the help of student interns created a comprehensive dataset that was utilized in every planning related project. It is important to point out that prior to creation of RPC’s GIS, paper maps of USGS and county’s property maps at 1’ = 660” scale were utilized in the review of all subdivision plats. The process was very cumbersome and inaccurate.

Between 1990 and 1994, DCRPC’s GIS developed over 60 layers of information. The layers were primarily mined data from Ohio Department of Natural Resources (ODNR)’s OCAP data, USGS, the Census Bureau’s TIGER files, and the Environmental Protection Agency (EPA) datasets. The RPC staff also created zoning layers for the majority of townships as well as layers from existing township Master Plans. Other important layers included: Future Rezoning proposals, Active Subdivisions, Land in Speculation, waterlines, major sewer lines, flood plains, and more. This comprehensive dataset was then utilized in the creation of a County-wide Master Plan.

In 1994, the Delaware County Auditor established the Delaware Appraisal Land Information System (DALIS) Project to create a spatial database to primarily assist in the appraisal process. Between 1995 and 1997, DALIS Team created one of the most detailed and comprehensive GIS database in the region. The project has been the recipient of numerous awards including the Best GIS Practices in Ohio in 1998 and ESRI's Special Achievement Award in 2000. Table A.2 provides a chronology of DALIS Project’s GIS:
<table>
<thead>
<tr>
<th>Years</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990-1991</td>
<td>110 points of GPS monuments were established. First aerial fly-over provided conventional Orthophotos Using conventional Orthophotos, contractors conducted deed research &amp; manuscripted the cadastral base on the mylar Orthophotos</td>
</tr>
<tr>
<td>1992-1994</td>
<td>The contractor started delivering products to the county (tile based paper mylars).</td>
</tr>
<tr>
<td>1994-1996</td>
<td>GIS Director was hired in 1994 and DALIS Project was established 3 full time staffs were hired and Quality control (Q/C) for spatial database began. Consultants delivered Composite digital layers which were then Q/C'ed and separated by DALIS staff (in Arc/Edit module of Arc/Info software)</td>
</tr>
<tr>
<td>April 1997</td>
<td>Additional 228 GPS monuments were established The County was re-flown and digital Orthophotos were created. Contours and other planimetric layers such as road centerline and edge of pavement were also created.</td>
</tr>
<tr>
<td>1997-1998</td>
<td>DALIS Project database was published on CD ROMs and on the Internet. Delaware County was the second county in the state with parcels based data on the web. Started working on Census related projects</td>
</tr>
<tr>
<td>1998-1999</td>
<td>Other products such as topos, Orthophotos, flood plains and soils were published. Additional tools were added to the Internet application. Hired a company to create a GPS based point coverage DALIS Project received the Best GIS Practices Award from Ohio Geographically Referenced Program (OGRIP)</td>
</tr>
<tr>
<td>1999-2000</td>
<td>DALIS Project received national recognition for its efforts in the Census 2000 process DALIS Project created a voting precinct layer for the Board of elections and installed a customized GIS application for that office DALIS project became the recipient of ESRI's Special Achievement Award in 2000</td>
</tr>
<tr>
<td>2000-2001</td>
<td>DALIS Project assisted the County's E-911 system in adding a GIS</td>
</tr>
</tbody>
</table>
component to their CAD system

DALIS project’s CAUV and Census projects were presented at national and international conferences.

Implemented a Mobile based GIS to capture records of new construction’s addresses and photos.

| 2001-2002 | Planning to fly the county in April 2002 to make new Orthopics. |

Table A.2 Activities of DALIS Project in Delaware County, Ohio

A.4 Spatial Data Creation, Update, and Management in Delaware County, Ohio

As the previous section discussed, Delaware County had realized that they had to create accurate and qualified digital maps rather than use paper or converted digital maps that are from other sources such as USGS, EPS, and Census Bureau. The spatial data creation started in 1991 when the first conventional aerial photo was made. This section introduces the cost of spatial data creation for Delaware County, spatial data available up to now, and the maintenance process of spatial data.
A.4.1 The Cost of Spatial Data Creation

Local governments like Delaware County, usually make contracts with professional companies to create maps rather than doing the work by themselves. This offers many benefits including specialized service and experience, and is more cost effective than short-term self-work. The Table A.3 provides the cost of each spatial data, cost per parcel, and cost per square km.

<table>
<thead>
<tr>
<th>Spatial Data Type</th>
<th>Cost</th>
<th>Per Parcel</th>
<th>Per km²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1991 Conventional Orthophoto</td>
<td>$200,000</td>
<td>$4.16</td>
<td>$263.34</td>
</tr>
<tr>
<td>1991 Deed Research &amp; Manuscription of property line</td>
<td>$300,000</td>
<td>$6.25</td>
<td>$395.01</td>
</tr>
<tr>
<td>1991 Manual Digitizing</td>
<td>$400,000</td>
<td>$8.33</td>
<td>$526.68</td>
</tr>
<tr>
<td>1997 GPS Monuments</td>
<td>$83,000</td>
<td>$1.63</td>
<td>$109.28</td>
</tr>
<tr>
<td>1997 Digital Orthophoto &amp; Road Centerline</td>
<td>$193,000</td>
<td>$3.78</td>
<td>$254.12</td>
</tr>
<tr>
<td>1997 Contours &amp; DTM</td>
<td>$135,000</td>
<td>$2.65</td>
<td>$177.75</td>
</tr>
<tr>
<td>2002 Digital Orthophotos</td>
<td>$198,000</td>
<td>$3.31</td>
<td>$260.71</td>
</tr>
</tbody>
</table>

Table A.3 Spatial Data Creation Costs

To get started, the most important piece of spatial data is base map, that is, digital orthophoto of Delaware County. Based on it, other layers such as road centerline, parcel boundary, hydrology, etc. are created. Therefore, the positional accuracy of these data
depends on the digital orthophoto. DALIS chose three different scales for the base map to reflect the relationship between the cost and the density of the area (more development area or more rapidly changing terrain needs a larger scale. And larger scale means higher cost). The accuracy of digital orthophoto provided by the contracted company and the suggested accuracy of National Map Accuracy Standards (NMAS) are shown in Table A.4. NMAS requires that at least 90% of well-defined points on the map fall within 1/30th of an inch of their true position on the ground in the case of large scale maps (Korte, 1994).

<table>
<thead>
<tr>
<th>Scale</th>
<th>County</th>
<th>NMAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 inch = 100 feet</td>
<td>2.5 – 5.0 feet</td>
<td>3.3 feet</td>
</tr>
<tr>
<td>1 inch = 200 feet</td>
<td>5.00 – 7.5 feet</td>
<td>6.7 feet</td>
</tr>
<tr>
<td>1 inch = 400 feet</td>
<td>7.5 – 10.00 feet</td>
<td>13.3 feet</td>
</tr>
</tbody>
</table>

Table A.4 Positional Accuracy of Base Map, Digital Orthophoto

A.4.2 Available Spatial Data

DALIS project has several types of spatial information available. All spatial information used in DALIS project is in digital format. Paper maps are not used. Table A.5 provides each type of spatial data and its description.
<table>
<thead>
<tr>
<th>Spatial Data</th>
<th>Source</th>
<th>Scale</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital Orthophoto</td>
<td>DALIS, 1991 &amp; 1997, 2002</td>
<td>1/2 foot pixel resolution 1 foot pixel resolution 2 foot pixel resolution</td>
<td>3 different scales are for different areas, such as city, resident, and farm (higher density area, larger scale). Image format is TIFF and MrSid. Used as the base map for Delaware County</td>
</tr>
<tr>
<td>Parcels</td>
<td>DALIS, 1992-present</td>
<td>Same as the base map</td>
<td>Created by heads-up digitizing on conventional orthophotos as well as COGOing subdivision</td>
</tr>
<tr>
<td>Road Right of Ways</td>
<td>DALIS, 1992-present</td>
<td>Same as the base map</td>
<td>Created by heads-up digitizing on conventional orthophotos as well as COGOing subdivision</td>
</tr>
<tr>
<td>Road Centerline</td>
<td>DALIS, 1997</td>
<td>Same as the base map</td>
<td>Created from 1997 digital orthophoto</td>
</tr>
<tr>
<td>Township</td>
<td>DALIS, 1992-present</td>
<td>Same as the base map</td>
<td>Created by heads-up digitizing on conventional orthophotos as well as COGOing subdivision</td>
</tr>
<tr>
<td>Municipalities</td>
<td>DALIS, 1992-present</td>
<td>Same as the base map</td>
<td>Created by heads-up digitizing on conventional orthophotos as well as COGOing subdivision</td>
</tr>
<tr>
<td>Subdivisions</td>
<td>DALIS, 1992-present</td>
<td>Same as the base map</td>
<td>Created by heads-up digitizing on conventional orthophotos as well as COGOing subdivision</td>
</tr>
<tr>
<td>Schools</td>
<td>DALIS, 1992-present</td>
<td>Same as the base map</td>
<td>Created by heads-up digitizing on conventional orthophotos as well as COGOing subdivision</td>
</tr>
<tr>
<td>Hydrography</td>
<td>DALIS, 1992</td>
<td>Same as the base map</td>
<td>Created by heads-up digitizing on conventional orthophotos as well as COGOing subdivision</td>
</tr>
<tr>
<td>Soils</td>
<td>DALIS, 1998</td>
<td>1:12,000</td>
<td>Converted from quarterquad mylars to digital format</td>
</tr>
<tr>
<td>Flood Plains</td>
<td>FEMA*, 1999</td>
<td>1:24,000</td>
<td>Contains 100-year and 500-year flood plains, base flood elevation, cross section lines, and floodways for Delaware County, Ohio.</td>
</tr>
<tr>
<td>GPS Points</td>
<td>DALIS, 1991 &amp; 1997</td>
<td>None</td>
<td>Created 110 points in 1991 survey Created additional 228 points in 1997 to support aerial photogrammetric operations and density permanent control monumentation.</td>
</tr>
<tr>
<td>Topography (Contour)</td>
<td>DALIS, 1997</td>
<td>Same as the base map</td>
<td>Contours are generated in 2, 5, and 10 foot intervals. Created by photogrammetric method</td>
</tr>
<tr>
<td>Precincts</td>
<td>DCBE*</td>
<td>1:125,000</td>
<td>Conflated with DALIS database</td>
</tr>
<tr>
<td>Census 2000 files</td>
<td>Census Bureau, 2000</td>
<td>1:100,000</td>
<td>Contains census information with geographic entities.</td>
</tr>
<tr>
<td>Point Coverage</td>
<td>DALIS, 1999</td>
<td>None</td>
<td>Stored every structure’s X, Y, Z using GPS Van. Used for census project and county’s E-911 system</td>
</tr>
</tbody>
</table>

FEMA*: Federal Emergency Management Agency
DCBE*: Delaware County Board of Election

Table A.5 Available Spatial Data in Delaware County, Ohio

193
Table A.6 and A.7 provide attributes associated with parcels and road centerlines, respectively.

<table>
<thead>
<tr>
<th><strong>Attribute of Parcels</strong></th>
<th><strong>Attribute of Parcels (Continue)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parcel Identification Number</td>
<td>Number of living units</td>
</tr>
<tr>
<td>Tax District</td>
<td>Story Height</td>
</tr>
<tr>
<td>Old Tax ID</td>
<td>Basement</td>
</tr>
<tr>
<td>Owner's last &amp; first name</td>
<td>Crawl Space</td>
</tr>
<tr>
<td>Co-Owner's first name</td>
<td>Year Built</td>
</tr>
<tr>
<td>Street number and name</td>
<td>Year Remodeled</td>
</tr>
<tr>
<td>City, State and Zip code</td>
<td>Total Finished Living Area</td>
</tr>
<tr>
<td>Legal Description</td>
<td>Total Rooms</td>
</tr>
<tr>
<td>Property Class</td>
<td>Number of Bed-Rooms</td>
</tr>
<tr>
<td>Property Acreage</td>
<td>Family-Rooms</td>
</tr>
<tr>
<td>Land's Market Value</td>
<td>Dining-Rooms</td>
</tr>
<tr>
<td>Improvement's Market Value</td>
<td>Half Bath-Room</td>
</tr>
<tr>
<td>Total Market Value</td>
<td>Full Bath-Room</td>
</tr>
<tr>
<td>Land's Taxable Value</td>
<td>Attic</td>
</tr>
<tr>
<td>Improvement's Taxable Value</td>
<td>Garage Type</td>
</tr>
<tr>
<td>Total Taxable Value</td>
<td>Number of Garages</td>
</tr>
<tr>
<td>Current Agriculture Use Valuation</td>
<td>Recreation Room</td>
</tr>
<tr>
<td>Most recent sales date</td>
<td>Air-condition</td>
</tr>
<tr>
<td>Second recent sales date</td>
<td>Number of Fire Places</td>
</tr>
<tr>
<td>Most recent sales value</td>
<td>Heating Type</td>
</tr>
<tr>
<td>Second recent sales value</td>
<td>Grade</td>
</tr>
<tr>
<td>Most recent sales number</td>
<td>Grade Adjustment</td>
</tr>
<tr>
<td>Second recent sales number</td>
<td>Grade Adjustment Factor</td>
</tr>
<tr>
<td>Most recent conveyance number</td>
<td></td>
</tr>
<tr>
<td>Second recent conveyance number</td>
<td></td>
</tr>
</tbody>
</table>

Table A.6 The Attributes of Parcel Coverages


**A.4.3 GIS staff and GIS Software and Hardware**

A GIS in any setting including that of local government agencies consists of hardware, software, spatial data, and people. The DALIS project during the past 8 years has provided a comprehensive GIS dataset and excellent support to the Auditor and other offices within the County. The department has also enjoyed wonderful support from the county and has been continuously praised by the public and private sectors. The DALIS project created the foundation spatial data layers for the county and continues to update various data layers in this rapidly growing County. Another impressive aspect of the project is pursuing cutting-edge GIS applications and developing easy-to-use tools to further enhance the system. None of these tasks were feasible without a team of GIS professionals. The DALIS team is currently comprised of, a GIS Director, 5 full-time GIS Specialists, and 2 part-time GIS interns.
The Director is responsible for the following tasks:

- Prioritizing all projects, tasks and responsibilities and delegating those to appropriate team members.
- Designing and modifying various data models and cutting-edge solutions for different purposes (from web-based to desktop solutions & applications).
- Designing, implementing, & overseeing Quality Control protocols for the Cadastral Database.
- Frequently interviewing other department’s employees and the public to determine the need for enhancements and further training.
- Researching and Planning all existing and new projects and overseeing their implementations.
- Re-engineering work flows and procedures to enhance the overall operation (that frequently involves 2 other offices’ related work flows).
- Configuring and recommending hardware/software purchases and upgrades.
- Composing technical reports such as Request For Proposals and negotiating with consultants when outsourcing a service is required.
- Presenting and publishing papers at international and national GIS conferences.
- Communicating with and assist other government entities to utilize the database to the fullest and to maximize their benefits.
- Training Team members and other county and city employees in the GIS software program several times a year.
- Frequently giving public presentations to professional and community audiences, representing the Auditor.

Other team members are responsible for a variety of tasks including:

- Daily update of the cadastral database by entering/digitizing every recorded land transaction into the database. The update includes several data layers such as parcels, roads, right of ways, municipalities, townships, subdivisions, schools, & etc. This task also includes Quality Control (Q/C) of other colleagues’ work.
- Preparing the above database to be published on the Internet on a weekly basis, which involves conversion of files to a format acceptable by the Internet software program.
- Daily Q/C of all Lot-Splits entered/digitized by the Map Department into the database and correction of all errors.
- Daily system maintenance, which involves monitoring the nightly backup as well as monitoring the software license manager.
- Assigning Parcel Identification Numbers to newly recorded subdivisions.
- Weekly Check of Auditor’s appraisal database to ensure the currency and accuracy of the database.
• Assisting other colleagues with problems related to complicated deeds and legal descriptions.
• Creating or Q/C-ing new layers to enhance the quality of the database. This task can involve creation of spatial data (such as creating a new layer for mobile homes or condos) as well as adding or Q/C-ing attribute data (such as adding a registration code for mobile homes). These projects are assigned on a case by case basis.
• Assisting the Director in developing agency policies.
• Representing the Auditor’s office to the public and at various agency meetings.
• Creation and Q/C of final data sets to be published on CD-ROMS on a Quarterly basis.
• Training/assisting Interns and other full time employees.
• Assist other colleagues with problems related to complicated deeds and legal descriptions.
• Q/C the database entered on the Appraisal Internet web site.
• Creating maps for special projects (randomly).
• Programming in Arc Macro Language (AML); a procedural programming language to facilitate internal routines or to streamline tasks such as map-making.
• Hardware/Software installation, upgrades, and trouble shooting. Adding and/or modifying several extensive Documentations (for weekly as well as quarterly release process)

Generally, GIS consultants are hired on a case by case basis and for applications that the current staff is not trained to undertake. Finally, Table A.8 shows the hardware and software used by DALIS.
<table>
<thead>
<tr>
<th>Hardware</th>
<th>GIS Software</th>
<th>Other Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM Compatible PCs (8)</td>
<td>ArcGIS (Desktop and Workstation)</td>
<td>Windows NT</td>
</tr>
<tr>
<td>Internal Network</td>
<td>ArcView : Spatial Analyst, Image Analysis and 3D Analyst Extensions</td>
<td>Microsoft Office</td>
</tr>
<tr>
<td>Plotter (2)</td>
<td>ArcIMS</td>
<td>MS Access</td>
</tr>
<tr>
<td>Printer (2)</td>
<td>ArcInfo COGO Extensions</td>
<td>Web customized software</td>
</tr>
<tr>
<td>Web Server (3)</td>
<td>ArcExplorer</td>
<td>Microsoft Visual Studio</td>
</tr>
<tr>
<td></td>
<td>MapObjects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Erdas Imagine</td>
<td></td>
</tr>
</tbody>
</table>

Table A.8 Used Hardware and Software

A.4.4 Spatial Data Management and Update

Maintenance of existing spatial datasets is especially critical in Delaware County because the County is the fastest growing county in the state and 15th fastest growing county in the nation (according to the Census Bureau’s 2001 population estimates). The upkeep and delivery of an error-free dataset in such a fast growing entity provides for many challenges. Three components of maintenance consist of, system enhancements, database expansion, and routine system maintenance (updates and backups) (Montgomery, 1993). System enhancement includes upgrades to hardware and software. Upgrades to software are generally provided free of charge since the county is under a maintenance agreement with its GIS contractor. However, hardware upgrade is more dependent upon the
availability of fund, the application’s demand, and the age of the exiting hardware. Thankfully, the cost of hardware has been constantly dropping and that has played an important role in the County’s ability to upgrade more often. The County also acquires new software applications on a regular basis to keep up with technology and that strategy has played a critical role in keeping the DALIS Project’s GIS on the cutting edge. For example, while in 1995, the initial cost of the hardware was at $106,000, in 2000, the upgrade to a new set of much more powerful workstations and a server only cost the county about half of that amount.

Database expansion has also been an important and constant activity for the DALIS Project team. Some of the additional layers that have been added to the initial set of data layers include: road centerlines, GPS Master point coverage, flood plains, soils, and voting precincts.

The final critical factor is the routine system maintenance including the daily upgrade, Quality Control (Q/C), and the documentation of all related processes and procedures. The DALIS Project during the past 8 years has produced a comprehensive set of documents describing every process and method involved in the maintenance of its datasets. These sets of documents are regularly updated by the DALIS Project staff whenever changes occur or new features and applications are added to the datasets. The entire dataset is also backed up on a nightly basis and the staff monitors the process routinely. The combination of all the above activities has created an extremely efficient process, one that staff turnover or changes in technology cannot negatively impact.
A.4.4.1 Map Sheet and Parcel Identification Number

The DALIS project divides parcel maps for the entire area into tiles (map sheets) in order to reduce the data size manageable in the GIS software and to avoid update problems, i.e., if someone update a map while it is being hold, it creates conflicts between them. Figure A.3 shows the tile partition of parcel maps.

![Figure A.3 Tile Partition For Parcel Maps](image)

The partition follows the township boundary as shown in Figure A.3. The first partition is one quarter of the township and the next level is always $\frac{1}{4}$ of the previous-level partition. Finally, the third and last level partition is therefore $\frac{1}{4}$ of $\frac{1}{4}$ of $\frac{1}{4}$ of the township (Figure
A.4). These tiles also represent the scales at which the base map (digital orthophoto) was created. They are 1:4800, 1:2400, and 1:1200. The total numbers of tiles are 368.

![Figure A.4 Map Sheet Partition](image)

Each map sheet is assigned a unique six digit number as follows.

1 : *Township Number*, derived from indexing of the United States Military District, townships are numbered south to north (3, 4, 5, 6)

2 & 3 : *Range Number*, derived from indexing of the United States Military District, ranges are numbered east to west (16, 17, 19, 20)

4 : ¼ *Township Number*, each township in the United States Military District was divided into quarter township, and numbered counter clockwise. (1) being the NE quarter, (2) being the NW quarter, (3) being the SW quarter, and (4) begin the SE quarter.

5 & 6: If the two spaces are 00, the parcel is mapped at a scale of 1:4800.
If space (5) has a 1 through 4 and space (6) has a 0, the parcel is mapped at a scale of 1:2400 also a number 1 tells you it falls into the NE NE ¼, a number 2, the NW NE ¼, a number 3, the SW NE ¼, and a number 4, the SE NE ¼, of said ¼ ¼ township.

If space (6) has a 1 through 4, the parcel is mapped at a scale of 1:1200 and is located as follows: (1) NE NE NE, (2) NW NE NE, (3) SW NE NE, and (4) SE NE NE, of said ¼ ¼ ¼ township.

The map sheet number is the part of the Parcel Identification Number (PIN) that consists of 14 digits as follows.

```
1 2 3 4 5 6 7 8 9 10 11 12 13 14
   [Map Sheet]     [Block]     [Parcel]     [Unit]
```

1 – 6: Map Sheet Number, provide geographic location and scale.

7 & 8: Block Number, utilized to group parcels by physical boundaries, for the ease of locating parcels on the map.

9, 10, & 11: Parcel Number, used in inventorying all parcels of land in the block, and is done in a systematic fashion to facilitate phases of revaluation programs.

12, 13, & 14: Unit Number, used for splits and special interest. When space (12) is a “0”, this denotes a split of the parcel. And the numbers (13) and (14) become a parcel split from the parcel number. If space (12) is “1” through “9” this denotes special interest as follows:
A.4.4.2 Subdivision

The biggest chore in updating databases comes from land subdivision because Delaware County is a fast growing area. As Table A.1 shows, the biggest part by area (55.75%) is farmland in Delaware County. These areas are converted into residential, industrial, or commercial use when the area is developed. The subdivision of land within Delaware County is reviewed by the DCRPC as well as many other county agencies in order to ensure that the public health, safety and welfare are protected. The review consists of external factors such as water supply and sewer capacity, as well as the internal design (safe street patterns and proper drainage). When subdividing land, standards for improvements and platting are prescribed by the Delaware County Subdivision Regulations. From small property divisions to large tract divisions, the regulations provide guidance to the public for accomplishing their goals while meeting DCRPC’s standards.

The subdivision is surveyed by a surveyor and is converted into digital map. Engineering CAD drawing software is used and DXF (Digital Exchange Format) is used when the file
is delivered to the DALIS Project. This DXF form is inserted into the proper map sheet, which includes those subdivisions. Transformation and adjustment are needed in this process to insert into the GIS base. Edge match type of map conflation is required here because most of the time, the DXF file containing the subdivision does not fit within the existing boundary in the GIS base. The rubber-sheeting function and manual editing provided by ArcInfo is used to match unfitted edges. Visual checks for quality are performed with the orthophotos as background. Other quality processes such as topology and deleting extraneous lines are done by ArcInfo functions, AML, and edit operations. Finally, annotation and labels are placed in the appropriate positions.

A.4.4.3 Quality Control (Q/C)

The measurements of spatial data quality consist of positional accuracy, attribute accuracy, logical consistency, completeness, and lineage. The main goal of Q/C is to satisfy these components as much as possible. The positional accuracy defines the closeness of location information to the ground truth. In section 3.4, the relative and absolute positional accuracy between Census maps and the DALIS GIS base was discussed. When spatial data is added to the existing database, Q/C-ing for the positional accuracy in DALIS strongly depends on the base map and the digital orthophoto. Therefore, if there is a conflict between new added features and the existing GIS base, the forced fit to the existing GIS base is implemented. The method does not provide the absolute solution for Q/C-ing of positional accuracy because the orthophoto itself
contains inherent errors depending on the altitude and the mapping scale as shown in Table A.4. However, this level of accuracy is deemed adequate for appraisal purposes and the DALIS Project staff relies on these photos and perceives them as “accurate”. Additionally, even if a higher level of accuracy were desired, the time and resources necessary to conduct field checks to increase the accuracy would have been cost prohibitive. In other words, the DALIS Team decided that as long as the consistency was maintained within the based map, the positional accuracy was satisfied.

The attribute accuracy is defined as the correctness of the attribute values compared to their true object. The DALIS Project’s datasets include several attributes such as lot number, subdivision names, Parcel Identification Number, deed book and page number, and etc., all of which are embedded in its various spatial layers. Q/C-ing those attributes are conducted manually and programmatically whenever a new land transaction occurs. New land transactions may include lot splits, combinations and subdivision plats. Visual cross-check based on the deed or plat by different staff is implemented when digital data is processed. Then multiple AML programs are run to provide errors such as missing attributes, incorrect attributes, and inconsistencies between DALIS and Appraisal database. Finally, those errors are carefully reviewed and corrected.

Completeness is defined as the degree to which the data exhausts the universe of possible items (Chrisman). For example, were all the lots in a subdivision inputted into the county’s dataset? This information is checked programmatically and then corrected manually. Lineage is defined as the record of the data sources and of the operations,
which created the database (Chrisman). For example, who, when, and how the subdivision data were created.

Finally, logical consistency mainly defines the topological consistency. For example, check to see if a parcel polygon is closed, if nodes are formed at the intersection, if road centerlines are connected to each other, if the neighbor’s information is accurate, if any duplications exist, or if edges are matched. These topological integrities are checked and reported by several AML routines.

A.4.4.4 Spatial Data Backup

The entire cadastral dataset is backed up on digital tapes on a nightly basis (+/- 15 Gig). The server running the back up software and containing the tape drive resides on DALIS’s server and the accuracy of the back up is monitored on a daily basis. This server is located in the County’s Data Center and is secure and protected, not in DALIS’s office. The DALIS team maintains a daily, weekly, and quarterly backup and quarterly back up tapes are regularly over written while the quarterly tapes are placed in a fire-proof safe on a quarterly basis.

A.5 The role of GIS in Local Government

GIS has been used in numerous states and local government for various types of applications. Foresman (Foresman, 1998) divides GIS use in state and local government into environmental/natural resources, general state government (planning, administration,
finance, revenue, asset management), cultural resources, infrastructure, human services, and other uses. This study reveals that the most popular application is related to environmental/natural resources in all states. General state government applications are second most popular. However, the purposes or reasons in starting a GIS base are different to state and local governments. Those purposes or reasons include automating manual work in mapping, improvement of the planning and management process, advanced analysis on geographic related information, and modernized land/parcel information system (Foresman, 1998).

The initial purpose of the GIS base in Delaware County was to improve land/parcel information system. This initial purpose is now integrated with other objectives in managing and helping the County because they find that the role of GIS in managing local government is nearly infinite. For example, accurate and quality tax mapping, planning and subdivision, precinct and school districting, transportation, utilities, engineering and surveys, environment, emergency response, and decision-making are some of those roles in local governments.

The best way to describe some of projects conducted in the DALIS project is to understand how GIS helps local governments. Some of projects conducted by DALIS, the Census project, CAUV, and Internet GIS, are discussed in the next sections.

**A.5.1 Census Project**

In the 1990 Census survey, Delaware County’s housing units were undercounted because of the rapid growth of the area. However, during the post-census appeal process, they
could not correct the figures because of the lack of tools, adequate staff, and given time. The local elected officials and specifically the County Auditor were fully aware of the significance of an accurate count for the county. They knew that an accurate count would mean not only receiving more federal and state funds and grants in the next decade, but also achieving more precise reapportionment and representation (Elhami, 2000). Therefore, in 1997, Delaware County participated in the Census Bureau’s LUCA (Local Update of Census Addresses), which is the opportunity for local government to correct undercount figures. The clear mission of Delaware County is to review the Census Bureau’s MAF (Master Address File) and correct it in Census 2000. Using conflated Census maps and spatial data created by the DALIS project, it was able to identify more than 11,000 homes in Delaware County that were not part of the U.S. Census Bureau records. Many of the people living in these homes would undoubtedly not have been counted. As a result, untold additional state and federal dollars will flow into Delaware County during the next decade, because population is one of the primary factors used in the distribution of funds. More details of the Census project will be found in (Elhami, 2000).

A.5.2 Ohio's Current Agricultural Use Valuation (CAUV)

Delaware County has more than 730 active farms with an average size of 230 acres. Approximately 56% of the County’s area is still dedicated to agricultural use – and most of it is family-owned. Corn, wheat, and soybeans are the leading crops. The purpose of
Ohio's Current Agricultural Use Valuation (CAUV) is to provide tax breaks to farmers who keep their land in farming. The CAUV value of a parcel is determined by a formula that involves yield information, crop patterns, non-land production costs, and capitalization rates. It is also critical to identify soil types, land uses, and the area for each soil type in the parcel to calculate the final CAUV value. It is an impossible task if there are no GIS tools to support those calculations. However, one piece of GIS customized software, Avenue, is used to create the CAUV application for the Delaware County Auditor's office. Figure A.5 shows the Avenue based Graphic User Interface (GUI) for this application. The main dialog consists of "Call Find Table", "Search by Pin", "Manual Label", "Auto Label", "Show Map", and "Print Map".

Figure A.5 CAUV Application System
The "Call Find Table" assists in finding a Parcel Identification Number (PIN), cross-referenced with the owner name, by providing a search table that is sorted by PIN. After inserting the PIN and selecting the "Search by Pin" button, the search engine finds the proper parcel and intersects it with the soils/land use layer and calculates each soils/land use combinations' acreage. This application requires that the parcel's deeded acreage will be maintained and therefore, the ratio of the total calculated acreage and total deeded acreage is calculated and then it is applied to each combination to then be inserted into the final table on the layout. The operator only sees the results, which are a parcel highlighted by a blue outline and centered on the window. At this point, there are two options to label a combination of soil types and land use texts on each intersected soil/land use type polygons. The manual option guides the operator by highlighting each polygon selected from the table. The auto option uses Arc/View's labeling options. The "Show Map" button, displays the layout including the map of the parcel overlaid with the soils and land uses, a table including each soils/land use combinations' acreages, and finally the "Print Map" button sends the customized layout to the printer.

With the help of CAUV application, county officials can save time and effort from manual work and farmers receive a substantial reduction in the valuation of land.

A.5.3 Internet GIS

Internet has become a place of information dissemination because it allows people to access GIS anytime and anywhere. Various services are available by Internet GIS. For
example, showing location and change, performing spatial analysis and geo-processing, and providing spatial data are some examples of Internet GIS. These applications can be distinguished by interactive and non-interactive capabilities. The early stage of Internet GIS provides mainly non-interactive maps and their associated information. But the current trend is to generate maps on the fly according to specific queries requested by the user. Therefore, more and more diverse and rich services to the public can be implemented.

The DALIS internet site provides current property tax information, pictures of structures on real estate, appraisal information, and various map services in an easy-to-use format.

Figure A.6 Delaware County Public Print Module
Figure A.6 shows one of services of Delaware County’s Internet GIS. In this service, county resident can download their parcel maps in PDF format. Search queries by owner name, parcel ID, and address are provided to assist users in finding their parcel. For example, Smith Brian is queried by owner name and the highlighted area is shown in red in Figure A.6. An animated plane is guide to the highlighted area. Finally, the parcel map is shown and downloadable if a user clicks the highlight area.

The county’s Internet applications have had a dramatic impact on the quality of delivered data to the public and private sector in Delaware County. Many members of the public and private sectors are now conducting their business online and the web site’s increasing hit rate is a proof of this fact. With the help of the Internet-based GIS, County officials are also saving a great deal of time since an updated set of spatial and attribute data is readily and accurately available to them at all times.

Figure A.7 shows that how many users access the DALIS Internet service. The number of hits is stable for all year and shows an increase from January 2002 after further enhancements were applied to the web site.
Figure A.7 Average Number of Hits Per Day By Month From June 2001 to March 2002
APPENDIX B

LIST OF AVENUE CODES

B.1 Avenue Script that Builds a Polyline Shape File

The input polyline data format is as follows.

```avenue
1 // the number of polyline
1 // polyline ID
1 // the number of parts
7 // number of x and y points
10 10 // x, y coordinate
10 64.06218
10 110
79.18258 110
110 110
115 45
110 10
```

```avenue
'=============================================================================
',
' Make polyline shape from polyline data
',
',
' Created by Hoseok Kang.
',
'=============================================================================

inFile = LineFile.Make( "c:\script\test.txt".AsFileName, #FILE_PERM_READ )
'=============================================================================

theView = av.Project.FindDoc( "View1" )
class = polyline

def = av.Project.MakeFileName("theme", "shp")
def = FileDialog.Put(def, ".shp", "New Theme")

if (def <> nil) then
    theTable = FTab.MakeNew(def, class)
    if (theTable.HasError) then
```

214
if (theTable.HasLockError) then
    MsgBox.Error("Unable to acquire Write Lock for file " +
    def.GetBaseName, ")
else
    MsgBox.Error("Unable to create " + def.GetBaseName, ")
end
return nil
end

fld = Field.Make("ID", #FIELD_DECIMAL, 8, 0)
fld.SetVisible( TRUE )
theTable.AddFields({fld})
'
read data
totallineNum=inFile.ReadElt
totallineNum=totallineNum.AsNumber
for each j in 1..totallineNum
    idNum=inFile.ReadElt
    partNum=inFile.ReadElt
    partCnt=partNum.AsNumber
    theListofPointList=List.Make
    for each k in 1..partCnt
        lineNumber=inFile.ReadElt
        lineCnt=lineNumber.AsNumber
        thePointList=List.Make
        for each i in 1..lineCnt
            XPoints=inFile.ReadElt
            xpnt=XPoints.extract(0)
ypnt=XPoints.extract(1)
            thePoint=Point.Make(xpnt.AsNumber, ypnt.AsNumber)
            thePointList.Add(thePoint)
        end
        theListofPointList.Add(thePointList)
    end
ShapeField=theTable.FindField("Shape")
thePolygon=Polygon.Make(theListofPointList)
if (theTable.CanAddRecord) then
    re=theTable.AddRecord
    theTable.SetValue(ShapeField, re, thePolygon)
    theTable.SetValue(fld, re, idNum)
end
end

theTheme = FTheme.Make(theTable)
theView.AddTheme(theTheme)
theTheme.SetActive(TRUE)
theTheme.SetVisible(TRUE)
av.GetProject.SetModified(true)
end
inFile.Close
B.2 Avenue Script that Builds a Polygon Shape File

The input polygon data format is as follows.

```plaintext
1  // the number of polygon
1  // polygon ID
1  // the number of parts
7  // number of x and y points
10 10  // x, y coordinate
10 64.06218
10 110
79.18258 110
110 110
115 45
110 10
```

'================================================================='
'
Make polygon shape from polygon data
'
'
Created by Ho-seok Kang.
'
'=================================================================

```plaintext
inFile = LineFile.Make( "c:\script\test.txt".AsFileName, #FILE_PERM_READ )
'=================================================================

theView = av.GetProject.FindDoc( "View1" )
class = Polygon

def = av.GetProject.MakeFileName("theme", "shp")
def = FileDialog.Put(def, "*.shp", "New Theme")

if (def <> nil) then
  theTable = FTab.MakeNew(def, class)
  if (theTable.HasLockError) then
    MsgBox.Error("Unable to acquire Write Lock for file " +
def.GetBaseName, "")
  else
    MsgBox.Error("Unable to create " + def.GetBaseName, "")
  end
  return nil
end

fld = Field.Make("ID", #FIELD_DECIMAL, 8, 0)
fld.SetVisible( TRUE )
theTable.AddFields({fld})

'read data
totallineNum=inFile.ReadElt
```

216
MsgBox.Info(totallineNum,"info")
totallineNum=totallineNum.AsNumber
for each j in 1..totallineNum
  idNum=inFile.ReadElt
  partNum=inFile.ReadElt
  partCnt=partNum.AsNumber
  theListofPointList=List.Make
  for each k in 1..partCnt
    lineNum=inFile.ReadElt
    lineCnt=lineNum.AsNumber
    thePointList=List.Make
    for each i in 1..lineCnt
      Xpoints=inFile.ReadElt
      xpnt=Xpoints.extract(0)
      ypnt=Xpoints.extract(1)
      thePoint=Point.Make(xpnt.AsNumber, ypnt.AsNumber)
      thePointList.Add(thePoint)
    end
    theListofPointList.Add(thePointList)
  end
ShapeField=theTable.FindField("Shape")
thePolygon=Polygon.Make(theListofPointList)
if theTable.CanAddRecord then
  re=theTable.AddRecord
  theTable.SetValue(ShapeField, re, thePolygon)
  theTable.SetValue(fld, re, idNum)
end
end

theTheme = FTheme.Make(theTable)
theView.AddTheme(theTheme)
theTheme.SetActive(TRUE)
theTheme.SetVisible(TRUE)
av.GetProject.SetModified(true)
end

inFile.Close
LIST OF REFERENCES


Beard, M. Kate, 1991, Constraints on Rule Formation, Map Generalization, pp. 121-135.

Beier, Thaddeus, 1992, Feature-Based Image Metamorphosis, Siggraph '92


Car, Adrijna, 1994b, General Principles of Hierarchical Spatial Reasoning-The Case of Wayfinding, Proceeding of Sixth Int. Symposium on Spatial Data Handling, pp. 646-664.


Callahan, George and Broome, Frederick, The Joint Development of a National 1:100,000-Scale Digital Cartographic Data Base.


Center for Mapping, 1993, Clark County Mapping Project, Center for Mapping Report, The Ohio State University


Cobb, Maria A, etc., 1998a, A Rule-based Approach for the Conflation of Attributes Vector Data, GeoInformatica 2:1, pp. 7-35


Cuenin, Par. R., 1972, Cartographie Generale, Eyrolle


Elhami, Shoreh and Kang, Hoseok, 2000, Lessons Learned: Addressing and LUCA in Delaware County, OHIO, URISA 2000 conference


221

Farin, Gerald, 1990, Surfaces over Dirichlet Tessellation, Computer Aided Geometric Design, pp. 281-292


Filin, S., Doytsher, Y., 2000a, The detection of corresponding objects in a linear-based map conflation, Surveying and Land Information Systems, Vol. 60, no. 2, pp 117-128

Filin, S., Doytsher, Y., 2000b, A Linear Conflation Approach For The Integration of Photogrammetric Information And GIS Data, IAPRS, Vol. XXXIII, Amsterdam, pp 282-288

Foresman, Timothy W., 1998, The History of Geographic Information Systems; Perspectives from the Pioneers, Prentice Hall


Gans, David, 1969, Transformations And Geometries, Appleton-Century-Crofts

Garzon-Rojas, Tomas, 1996, Study On the Use of Spatial Information for Decision Making Processes at the Local Government Level in Clark County, OHIO

Gillman, Daniel, 1985, Triangulation For Rubber-Sheeting, AutoCarto 7, pp 191-199


Green, W. B., Jepsen, P.L. etc.,1975, Removal of instrument signature from Mariner 9 television images of Mars. Applied Optics 14:105

223


Hake, G., 1975, Kartographie, Sammlung Goschen Band


Korte, George B., 1994, GIS Book, Onword Press


Marx, Robert W., 1984, Developing An Integrated Cartographic/Geographic Data Base For The United States Bureau Of The Census, Bureau of the Census


Masser, Ian, 1998, Governments and Geographic Information, Taylor & Francis


Moellering , Harold , 2000b, The Scope and Conceptual Content of Analytical
Cartography, Cartography And Geographic Information Science, Vol. 27, No.3, pp 205-223

Mohar, Bojan and Thomassen, Carsten, 2001, Graphs On Surfaces, Johns Hopkins University Press


Nyerges, Timothy L., 1981, Cartographic Information Modeling As A Theoretical Basis For Cartographic Data Base Structures, Second International HBDS Seminar, Richmond, Virginia


Rosen, Barbara and Saalfeld, Alan, 1985 , Match Criteria for Automatic Alignment, AutoCarto 7, pp. 456-462


Saalfeld, Alan, 1986, Shape Representation for Linear Features in Automated Cartography, Technical Papers of the 1986 ACSM-ASPRS Annual Convention, V.1, pp 143-152.


Saalfeld, Alan, 1993, Conflation : Automated Map Compilation, Dissertation, University of Maryland

Saalfeld, Alan, 1998, Topologically Consistent Line Simplification with the Douglas-Peucker Algorithm, the presented paper from Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University.

Saalfeld, Alan, 1999, GS786 Class Note, Geodetic Science, The Ohio State University.

Saalfeld, Alan, 2000, Complexity and Intractability: Limitations to Implementation in Analytical Cartography, Cartography and Geographic Information Science, 27(3)


Sibson, Robin, 1981, A Brief Description of Natural Neighbor Interpolation, Ch. 2 Interpreting Multivariate Data, Wiley series in probability and mathematical statistics


Thompson, D’Arcy, 1947, On Growth and Form, Cambridge University Press.


230


Warnecke, L., 1995, Geographic Information/GIS Institutionalization in the 50 states: Users and Coordinators, Santa Barbara, CA: National Center for Geographic information and Analysis, University of California


Yost, David A., 2000, Comprehensive Annual Financial Report For the Year Ended December 31, 2000, Delaware County, Ohio